nature cell biology

TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors

Inês Cebola[1,12], Santiago A. Rodríguez-Seguí[2,3,4,12], Candy H-H. Cho[5,12,13], José Bessa[6,7,12], Meritxell Rovira[2,3,12], Mario Luengo[8], Mariya Chhatriwala[9], Andrew Berry[10], Joan Ponsa-Cobas[1], Miguel Angel Maestro[2,3], Rachel E. Jennings[10], Lorenzo Pasquali[2,3,13], Ignasi Morán[1], Natalia Castro[2,3], Neil A. Hanley[10,11,14], Jose Luis Gomez-Skarmeta[8,14], Ludovic Vallier[5,9,14] and Jorge Ferrer[1,2,3,14]

The genomic regulatory programmes that underlie human organogenesis are poorly understood. Pancreas development, in particular, has pivotal implications for pancreatic regeneration, cancer and diabetes. We have now characterized the regulatory landscape of embryonic multipotent progenitor cells that give rise to all pancreatic epithelial lineages. Using human embryonic pancreas and embryonic-stem-cell-derived progenitors we identify stage-specific transcripts and associated enhancers, many of which are co-occupied by transcription factors that are essential for pancreas development. We further show that TEAD1, a Hippo signalling effector, is an integral component of the transcription factor combinatorial code of pancreatic progenitor enhancers. TEAD and its coactivator YAP activate key pancreatic signalling mediators and transcription factors, and regulate the expansion of pancreatic progenitors. This work therefore uncovers a central role for TEAD and YAP as signal-responsive regulators of multipotent pancreatic progenitors, and provides a resource for the study of embryonic development of the human pancreas.

The human genome sequence contains instructions to generate a vast number of developmental programmes. This is possible because each developmental cellular state uses a distinct set of regulatory regions. The specific genomic programmes that underlie human organogenesis, however, are still largely unknown[1,2]. Knowledge of such programmes could be exploited for regenerative therapies, or to decipher developmental defects underlying human disease.

The pancreas hosts some of the most debilitating and deadly diseases, including pancreatic ductal adenocarcinoma and diabetes mellitus. Classic mouse knockout models and human genetics have uncovered multiple transcription factors (TFs) that regulate embryonic formation of the pancreas[3,4]. For example, GATA6 (GATA binding protein 6; refs 5–7), PDX1 (pancreatic and duodenal homeobox 1; refs 8,9), HNF1B (HNF1 homeobox B; ref. 10), ONECUT1 (one cut homeobox 1; ref. 11), FOXA1/FOXA2 (forkhead box A1/forkhead box A2; ref. 12), SOX9 (SRY (sex determining region Y)-box 9; refs 13,14) and PTF1A (pancreas specific transcription factor, 1a; ref. 15) are essential for the specification of pancreatic multipotent progenitor cells (MPCs) that arise from the embryonic gut endoderm, or for their subsequent outgrowth and branching morphogenesis. However, little is known concerning how these pancreatic TFs are deployed as regulatory networks, or which genomic sequences are required to activate pancreatic developmental programmes.

One obvious limitation to studying the genomic regulation of human organogenesis lies in the restricted access and the difficulties of manipulating human embryonic tissue. Theoretically, this can be circumvented by using human embryonic stem cells (hESCs) to derive cellular populations that express organ-specific progenitor markers,

[1]Department of Medicine, Imperial College London, London W12 0NN, UK. [2]Genomic Programming of Beta-cells Laboratory, Institut d'Investigacions August Pi i Sunyer (IDIBAPS), 08036 Barcelona, Spain. [3]CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), 08036 Barcelona, Spain. [4]Laboratorio de Fisiología y Biología Molecular, Departamento de Fisiología, Biología Molecular y Celular, IFIBYNE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EGA Buenos Aires, Argentina. [5]Wellcome Trust and MRC Stem Cells Centre, Anne McLaren Laboratory for Regenerative Medicine, Department of Surgery and Wellcome Trust—Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge CB2 0SZ, UK. [6]Instituto de Biologia Molecular e Celular (IBMC), 4150-180 Porto, Portugal. [7]Instituto de Investigação e Inovação em Saúde, Universidade do Porto, 4200-135 Porto, Portugal. [8]Centro Andaluz de Biología del Desarrollo, Consejo Superior de Investigaciones Científicas/Universidad Pablo de Olavide, 41013 Sevilla, Spain. [9]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [10]Centre for Endocrinology and Diabetes, Institute of Human Development, Faculty of Medical & Human Sciences, Manchester Academic Health Sciences Centre, University of Manchester, Manchester M13 9PT, UK. [11]Endocrinology Department, Central Manchester University Hospitals NHS Foundation Trust, Manchester M13 9WU, UK. [12]These authors contributed equally to this work. [13]Present addresses: Genomics Research Center, Academia Sinica, Taipei 115, Taiwan (C.H-H.C.); Division of Endocrinology, Germans Trias i Pujol University Hospital and Research Institute and Josep Carreras Leukaemia Research Institute, 08916 Badalona, Spain (L.P.).
[14]Correspondence should be addressed to N.A.H., J.L.G-S., L.V. or J.F.
(e-mail: Neil.Hanley@manchester.ac.uk or jlgomska@upo.es or lv225@cam.ac.uk or j.ferrer@imperial.ac.uk)

Received 30 August 2014; accepted 13 March 2015; published online 27 April 2015; DOI: 10.1038/ncb3160

although it is unclear if such cells can truly recapitulate broad genomic regulatory programmes of genuine progenitors.

In the current study, we dissected pancreatic buds from human embryos and used hESCs to create stage-matched pancreatic progenitor cells. We processed the two cellular sources in parallel and validated *in vitro* MPCs as a model to study gene regulation in early pancreas development. We created an atlas of active transcripts and enhancers in human pancreatic MPCs, and mapped the genomic binding sites of key pancreatic progenitor TFs. Using this resource, we show that TEA domain (TEAD) factors are integral components of the combination of TFs that activates stage- and lineage-specific pancreatic MPC enhancers.

RESULTS

Regulatory landscape of *in vivo* and *in vitro* MPCs

To study the genomic regulatory programmes of the nascent embryonic pancreas, we dissected pancreatic buds from Carnegie stage (CS) 16–18 human embryos. At this stage, the pancreas has a simple epithelial structure formed by cells expressing markers of pancreatic MPCs (including PDX1, HNF1B, FOXA2, NKX6-1 (NK6 homeobox 1) and SOX9), without obvious signs of endocrine or acinar differentiation, and is surrounded by mesenchymal cells (Supplementary Fig. 1a)[16]. For simplicity, we refer to this pancreatic MPC-enriched tissue as *in vivo* MPCs. Because human embryonic tissue is extremely limited and less amenable to perturbation studies, in parallel we used hESCs for *in vitro* differentiation of cells that express the same constellation of markers as *in vivo* MPCs (Supplementary Fig. 1a)[17]. We refer to these cells as *in vitro* MPCs. We carried out RNA sequencing (RNA-seq) and chromatin immunoprecipitation sequencing (ChIP-seq) analysis of *in vivo* and *in vitro* MPCs to profile polyadenylated transcripts, genomic sites bound by FOXA2 (a developmental TF that is specific to epithelial cells within the pancreas), and genomic regions enriched in the enhancer mark histone H3 monomethylated at Lys 4 (H3K4me1) (Fig. 1a and Supplementary Tables 1 and 2).

Earlier studies have shown that hESC-derived pancreatic progenitors express appropriate markers[17–20]. However, the extent to which they provide a suitable model to study global genome regulation of genuine pancreatic MPCs has not been tested. Several observations validated our artificial progenitors for this purpose, namely that *in vitro* MPCs recapitulated expression of known pancreatic MPC TFs (Fig. 1b and Supplementary Fig. 1b), that *in vitro* and *in vivo* MPCs showed a high correlation of transcript levels (Spearman's $\rho = 0.5876$, $P < 2.2 \times 10^{-16}$, Supplementary Fig. 1c) and of transcript enrichment relative to other human tissues (Spearman's $\rho = 0.5881$, $P < 2.2 \times 10^{-16}$, Fig. 1b,c and Supplementary Fig. 1d), and that the transcripts that are selectively enriched in either *in vitro* or *in vivo* MPCs relative to 22 non-pancreatic tissues (Fig. 1b) share common functional annotations, including pancreas development, chordate embryonic development and WNT signalling (Fig. 1d and Supplementary Table 3). The enrichment of WNT signalling genes included numerous non-canonical WNT regulators, including *FZD2*, *SFRP5*, *CELSR2* and *VANGL2* (Fig. 1d and Supplementary Table 3), whose orthologues have also been listed as selectively expressed in mouse embryonic pancreatic buds (Supplementary Table 4)[21,22], suggesting an evolutionarily conserved signalling mechanism operating in early pancreas development. This indicates that, despite the artificial origin of *in vitro* MPCs, and the presence of non-epithelial cell types in dissected embryonic pancreas, there are meaningful similarities in their transcriptomes. Integration of these data sets enabled us to define a core set of 500 genes that showed enriched expression in both sources of pancreatic MPCs (Supplementary Table 5), providing a resource to study genes important for early human embryonic pancreas development.

We next compared FOXA2 binding sites in the *in vivo* and *in vitro* pancreatic MPCs with other tissues where this TF is expressed (embryonic liver, adult liver and adult pancreatic islets; Fig. 1e,f). FOXA2 largely bound to the same genomic regions in both sources of MPCs, yet bound to different genomic sites in other tissues, despite the fact that a similar sequence motif was recognized in all cases (Fig. 1f and Supplementary Fig. 1e). Furthermore, *in vivo* and *in vitro* MPCs shared cell-specific H3K4me1 enrichment at *in vivo* FOXA2-bound sites (Fig. 1g and Supplementary Fig. 1f). Finally, genes with two or more nearby H3K4me1-enriched FOXA2-bound regions in the *in vivo* MPCs showed enriched messenger RNA expression in both *in vivo* and *in vitro* MPCs relative to 23 control tissues (Fig. 1h). Thus, *in vitro* and *in vivo* MPCs showed common FOXA2 and H3K4me1 occupancy patterns near pancreatic MPC-enriched genes. Taken together, our analyses suggest that artificial pancreatic MPCs recapitulate significant transcriptional and epigenomic features of genuine embryonic MPCs, and can thus be exploited as a tool to study genome regulation of human pancreas development.

An atlas of human pancreatic MPC enhancers

To map active *cis*-regulatory elements in human pancreatic MPCs, we employed *in vitro* MPCs to profile histone H3 acetylated at Lys 27 (H3K27ac), which marks active transcriptional enhancers[23,24]. We then selected all genomic regions that showed H3K27ac and H3K4me1 enrichment in chromatin from *in vitro* MPCs, and that were also enriched in H3K4me1 in human CS16–18 pancreas (*in vivo* MPCs). After exclusion of annotated promoters, this disclosed 9,669 regions that carried an active enhancer chromatin signature in pancreatic MPCs (Fig. 2a and Supplementary Fig. 2a,b and Supplementary Table 6).

The *cis*-regulatory map included known pancreatic MPC enhancers (Fig. 2a). As expected, predicted MPC enhancer sequences showed strong evolutionary conservation (Fig. 2b), they were preferentially located near genes with increased expression in CS16–18 pancreas (Fig. 2c) and they were often associated with core MPC-specific genes (hypergeometric test, $P < 10^{-15}$). In keeping with the cellular and temporal specificity of enhancers, 35% of pancreatic MPC enhancers showed no overlap with active enhancers from at least six of seven non-pancreatic tissues, and were thus defined as MPC-selective enhancers (Fig. 2d and Supplementary Fig. 2c, Supplementary Tables 7 and 8). Notably, 47% showed no overlap with enhancers from adult human islets[25] (Fig. 2d). As expected from this cell-specific and stage-specific profile, genes near MPC-selective enhancers are enriched in functions relevant for pancreas development (Supplementary Fig. 2d and Supplementary Table 9). This analysis therefore uncovered a large collection of candidate active enhancers of the nascent human embryonic pancreas.

A combinatorial code for pancreatic MPC enhancers

To understand the regulatory sequence code that drives early human pancreas development, we examined this collection of MPC enhancers

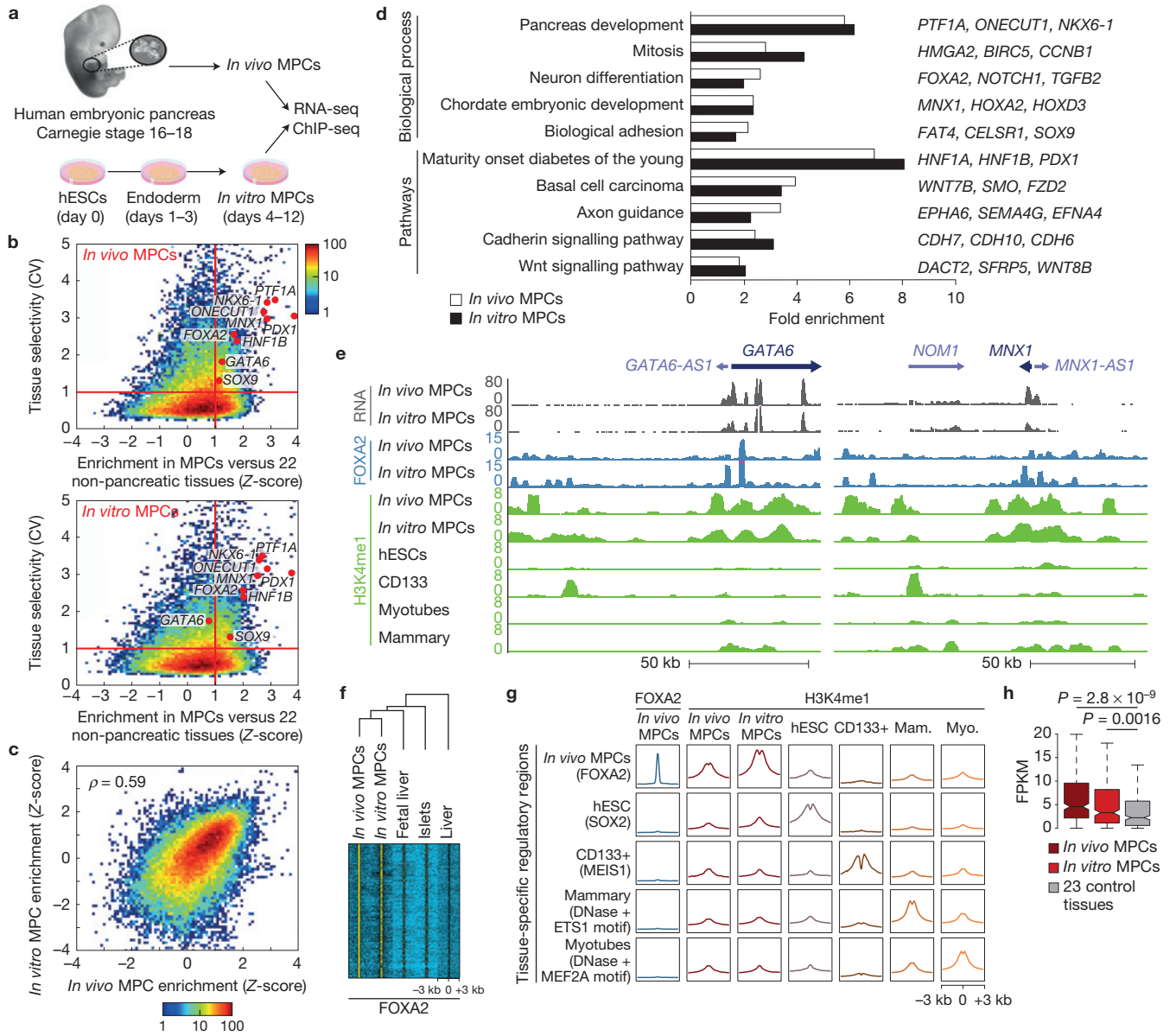


Figure 1 Human *in vitro* MPCs recapitulate transcriptional and epigenomic features of *in vivo* MPCs. (**a**) Experimental set-up. Pancreases were dissected from human CS16–18 embryos (*in vivo* MPCs). *In vitro* MPCs were derived from hESCs. (**b**) *In vitro* and *in vivo* MPCs share tissue-selective genes. Tissue selectivity of RNAs was determined by the coefficient of variation (CV) across 25 embryonic and adult tissues or cell types. Enrichment of RNAs in MPCs relative to non-pancreatic tissues was quantified as a *Z*-score. Red lines define genes that are both tissue selective and enriched in MPCs ($CV > 1$, $Z > 1$). Most known pancreatic regulatory TFs are in this quadrant in both sources of MPCs. The colour scale depicts the number of transcripts. (**c**) *Z*-scores of genes expressed in at least one source of MPCs were highly correlated for *in vitro* versus *in vivo* MPCs (see also Supplementary Fig. 1d for a comparison of unrelated tissues). The Spearman's coefficient value is shown. The colour scale depicts the number of transcripts. (**d**) *In vivo* and *in vitro* MPC-enriched genes have common functional annotations. Shown are the most significant terms for *in vivo* MPC-enriched genes, and their fold enrichment in both sources of MPCs. Representative genes from each category that are enriched in both MPCs are shown on the right. More extensive annotations are shown in Supplementary Table 3. (**e**) RNA, FOXA2 and H3K4me1 profiles of indicated samples in the *GATA6* and *MNX1* loci. (**f**) *In vivo* MPC FOXA2 occupancy is largely recapitulated by *in vitro* MPCs, but not by other tissues expressing FOXA2. Hierarchical clustering was carried out on the normalized FOXA2 ChIP-seq signal centred on all 5,760 *in vivo* MPC FOXA2 peaks. (**g**) *In vitro* MPCs recapitulate cell-specific H3K4me1 enrichment observed in chromatin from *in vivo* MPCs. Aggregation plots show H3K4me1 enrichment at occupancy sites of tissue-specific TFs. Mam., mammary; Myo., myotubes. (**h**) Genes with at least three regions enriched in FOXA2 and H3K4me1 at *in vivo* MPCs are preferentially expressed in both *in vivo* and *in vitro* MPCs. Boxes show RNA interquartile range (IQR) and notches indicate median 95% confidence intervals ($n = 327$ genes). *P*-values were calculated with the Wilcoxon rank-sum test. FPKM, fragments per kilobase of exon per million fragments mapped.

and found that the most enriched sequence motifs match binding sites of known pancreatic regulators, including FOXA, HNF1, SOX, PDX1, GATA and ONECUT (Fig. 3a and Supplementary Fig. 3a and Supplementary Tables 10 and 11). The single most enriched recognition motif, however, matched that of TEAD TFs, which have not been previously implicated in pancreas development (Fig. 3a).

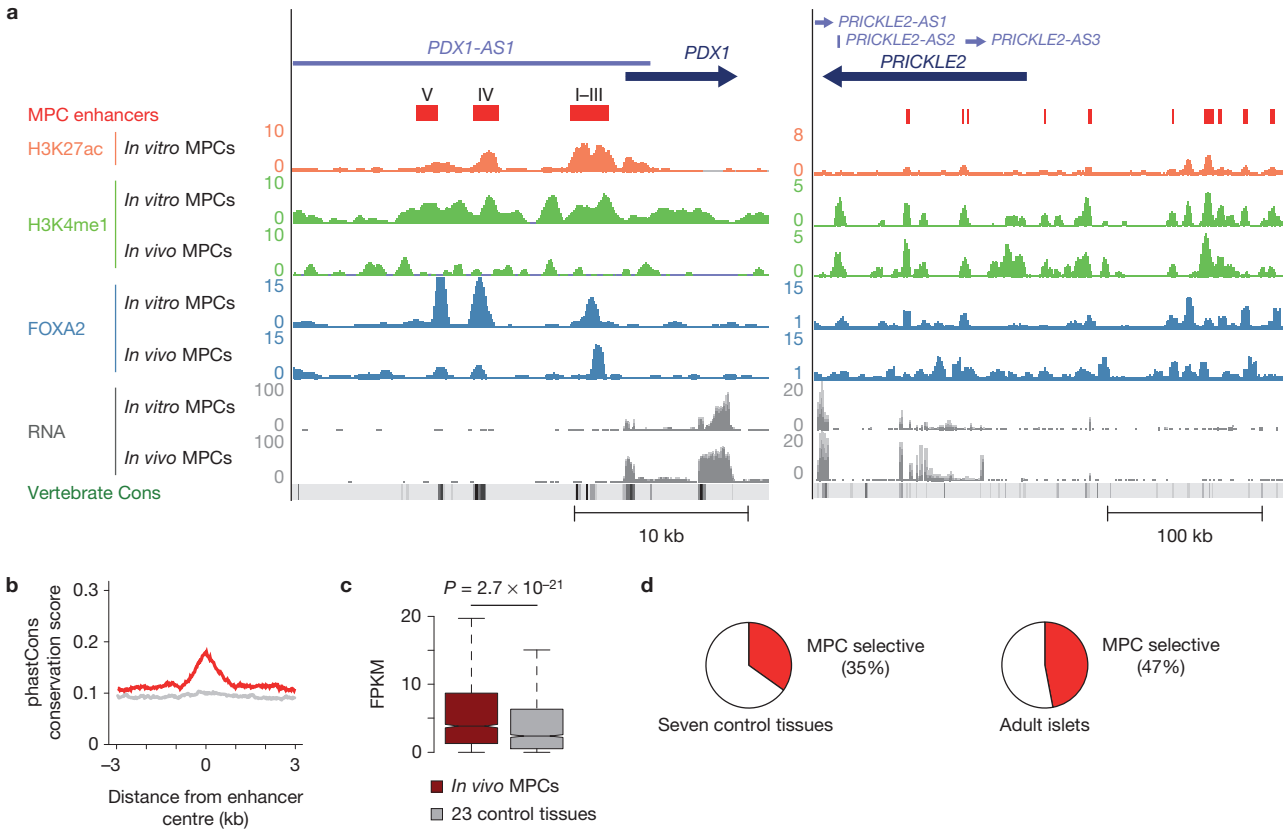


Figure 2 A compendium of active enhancers in human pancreatic MPCs. (**a**) Predicted enhancers were defined by enrichment in H3K27ac and H3K4me1 (see schematic representation in Supplementary Fig. 2b). Shown are examples in the vicinity of *PDX1*, including a previously unannotated enhancer which we coin area V, upstream of known enhancers (areas I–IV; refs 41,42), and several enhancers near *PRICKLE2*, a non-canonical WNT signalling component (Supplementary Table 4). (**b**) MPC enhancer sequences are evolutionary conserved (17 species vertebrate phastCons score). Conservation plots of random non-exonic sequences are shown as a light grey line. (**c**) Genes that are associated with three or more MPC enhancers show enriched expression in dissected *in vivo* MPCs relative to 23 other tissues. The boxes show the IQR of RNA levels, whiskers extend to 1.5 times the IQR or extreme values and notches indicate 95% confidence intervals of the median. The *P*-value was calculated with the Wilcoxon rank-sum test ($n = 2{,}093$ genes). (**d**) Many MPC enhancers are tissue and stage selective. We defined enhancers of eight control tissues using criteria identical to those in MPCs (Supplementary Fig. 2c and Supplementary Table 8) and show the proportion of enhancers that are inactive in at least six out of seven non-pancreatic tissues (left) or inactive in adult pancreatic islets (right).

TEAD motifs were similarly enriched in regions bound by FOXA2 in CS16–18 pancreas as well as *in vitro* MPCs, but not in regions bound by FOXA2 in adult pancreatic islets or liver (Fig. 3b and Supplementary Fig. 3b). Because TFs are thought to function in a combinatorial manner, we identified combinations of multiple motifs that were most enriched at pancreatic MPC enhancers relative to non-pancreatic enhancers (Fig. 3c and Supplementary Table 12). This showed that the most enriched combinations contained TEAD motifs adjacent to known pancreatic TF recognition sequences (Fig. 3c). These results therefore revealed that pancreatic MPC enhancers contain combinations of motifs that match known as well as previously unrecognized pancreatic regulatory TFs.

TEAD1 is a core component of pancreatic progenitor *cis*-regulatory modules

Mouse and human genetics have revealed numerous TFs that are essential for the specification, growth and morphogenesis of pancreatic MPCs (refs 3,26), yet very little is known about how such factors promote these processes. The availability of large numbers of *in vitro* MPCs enabled us to carry out ChIP-seq analysis to profile the occupancy sites of several TFs that are essential for early pancreas development, namely HNF1B (ref. 10), ONECUT1 (ref. 11), PDX1 (refs 8,9) and GATA6 (refs 5–7), in addition to FOXA2 (ref. 12; Supplementary Table 2). On the basis of our computational predictions we also profiled TEAD1, a TEAD homologue that is highly expressed in MPCs from human embryonic pancreas (Supplementary Fig. 4a), defining binding sites for a total of six TFs in human MPCs (Fig. 4a).

All six TFs preferentially bound to known cognate recognition sequences that were widely distributed throughout the genome (Supplementary Fig. 4b), although there was marked preference for binding to MPC enhancers and annotated promoters (Fig. 4a,b and Supplementary Fig. 4c–d). Furthermore, the six TFs very frequently co-occupied the same regions, predominantly at MPC enhancers (Fig. 4a,b and Supplementary Fig. 4c–e). For example, enhancers bound by PDX1 and GATA6, the TFs with the lowest total number of binding sites, showed co-binding by at least one of the other five TFs in 94.5% and 95.3% of instances, respectively (Supplementary Fig. 4e). Remarkably, TEAD1 showed a similar co-binding pattern as

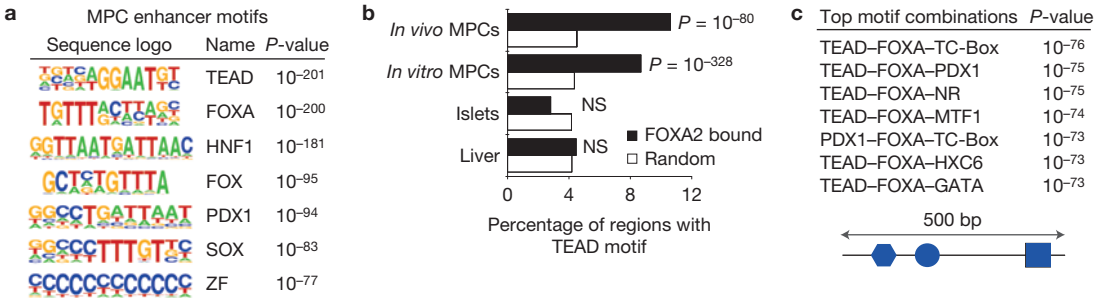


Figure 3 MPC enhancers are enriched in DNA binding motifs for TEAD and known pancreatic TFs. (**a**) TEAD recognition motifs were strongly enriched in a *de novo* motif search in MPC enhancers. Other enriched matrices match binding sites of known pancreatic regulators. See Supplementary Tables 10 and 11 for a complete list of motifs enriched in MPC and MPC-selective enhancers, respectively. ZF, zinc-finger transcription factor. (**b**) TEAD motifs are highly enriched in genomic regions bound by FOXA2 in both *in vivo* and *in vitro* MPCs, but not in regions bound by FOXA2 in islets or liver. Binomial distribution *P*-values were obtained using HOMER (ref. 57). NS: nonsignificant. (**c**) Combinations of recognition motifs for TEAD and other pancreatic regulators are specifically enriched in pancreatic MPC enhancers. We searched for combinations of three sequence motifs that were contained within 500 base pairs (bp) and were most enriched in pancreatic MPC enhancers relative to eight other tissue enhancers. The top 50 most enriched motif combinations are shown in Supplementary Table 12. TC-Box, TC-rich motif.

the five known pancreatic regulators analysed in this study (Fig. 4c and Supplementary Fig. 4d,e). Consistently, strong TEAD1 occupancy was observed not only at known targets from other cell types, such as *CTGF* or *CYR61* (ref. 27; Supplementary Fig. 4f and Supplementary Table 13), but also in 27% of all pancreatic MPC enhancers. Furthermore, 45% of enhancer-associated genes had at least one TEAD1-bound enhancer (Fig. 4d and Supplementary Table 14). In support, we confirmed TEAD1 binding to 10/12 enhancers in CS16–18 embryonic pancreas (Fig. 4e and Supplementary Table 15), and observed that TEAD1 binding was enriched in enhancers bound by FOXA2 *in vivo* (Fig. 4f). Altogether, computational and ChIP-seq analyses indicate that known pancreatic regulatory TFs show widespread co-binding at MPC enhancers, and that TEAD1 is an unexpected component of this combinatorial TF code.

Given the high degree of TF co-occupancy in MPC enhancers, we defined 2,945 regions within enhancers that are bound by two or more TFs, and coined these *cis*-regulatory modules (CRMs; Fig. 4a and Supplementary Fig. 4c). CRMs provided greater spatial resolution of *cis*-regulatory sequences than H3K27ac/H3K4me1-enriched regions alone, which often appear to merge several adjacent evolutionary conserved sequences bound by multiple TFs.

A large number of CRMs mapped near known pancreatic regulatory genes, including *HNF1B*, *FGFR2*, *HHEX*, *FOXA2*, *NKX6-1* and *SOX9* (Fig. 4a and Supplementary Fig. 4c and Supplementary Table 16). More generally, CRMs mapped near core MPC-enriched genes ($P = 3.32 \times 10^{-12}$). Notably, spatial clusters of CRMs were associated with genes that were highly enriched in gene functions relevant for early pancreas development, including epithelial cell proliferation and WNT signalling (Fig. 4g and Supplementary Tables 17 and 19). Notably, non-canonical WNT regulatory genes were enriched near clusters of CRMs ($P = 1.18 \times 10^{-9}$; Supplementary Table 19), in agreement with our transcriptome analysis of pancreatic MPCs (Fig. 1c and Supplementary Table 4) and transcriptome analysis of mouse pancreas development[21,22].

Interestingly, CRMs bound by any of the six TFs were associated with the same functional annotations (Fig. 4g). This included TEAD1-bound CRMs, despite the fact that this TF is widely expressed across multiple tissues and developmental stages (Fig. 4g). TEAD1-bound CRMs thus mapped to known or plausible pancreatic regulatory genes, including *FGFR2*, *RBPJ*, *FZD5/7/8*, *FRZB*, *JAG1*, *CDC42EP1*, *MAP3K1*, *NKX6-1*, *HHEX*, *GATA4*, *GATA6*, *FOXA2*, *HES1* and *SOX9* (Fig. 4a and Supplementary Fig. 4c and Supplementary Table 20). This is consistent with a broad combinatorial function of regulatory TFs in the establishment of the MPC-specific transcriptional programme.

To functionally validate these human embryonic pancreas CRMs, 32 sequences were transfected into *in vitro* MPCs, and 20 (62.5%) yielded significant enhancer activity (Mann–Whitney test for CRMs versus control regions, $P = 0.0144$; Fig. 5a and Supplementary Fig. 5a). To directly test the function of TEAD1 binding to CRMs, we mutated TEAD recognition sequences in three CRMs that were bound by TEAD and other pancreatic TFs, which disrupted enhancer activity in all cases (Fig. 5b).

We selected ten CRMs for validation using zebrafish transgenesis, and in eight cases we demonstrated enhancer activity in Pdx1$^+$/Nkx6.1$^+$ pancreatic endoderm MPCs (Fig. 5c–e, Supplementary Fig. 5b and Supplementary Table 21). Amongst these, we examined a CRM in the locus encoding *SOX9*, an essential regulator of the self-renewal of mouse pancreatic MPCs that is mutated in humans with pancreas hypoplasia[13,14] (Fig. 5c,d). This CRM showed pancreas-specific enhancer activity in zebrafish transgenics, whereas mutation of the TEAD recognition sequence abolished enhancer activity, providing further confirmation that TEAD1 binding is required for the *in vivo* function of pancreatic MPC enhancers (Fig. 5c).

Taken together, this analysis provided a rich source of *cis*-regulatory elements in human embryonic pancreatic progenitors. It also revealed widespread co-occupancy of pancreatic developmental TFs at MPC enhancers, and uncovered TEAD as a hitherto unrecognized core component of this combination of TFs.

TEAD and YAP regulate a pancreatic developmental programme

We next examined TEAD-dependent gene regulation during pancreas development. TEAD proteins interact with the active nuclear form of the coactivator Yes-associated protein (YAP). YAP is negatively regulated by Hippo signalling, which triggers YAP phosphorylation

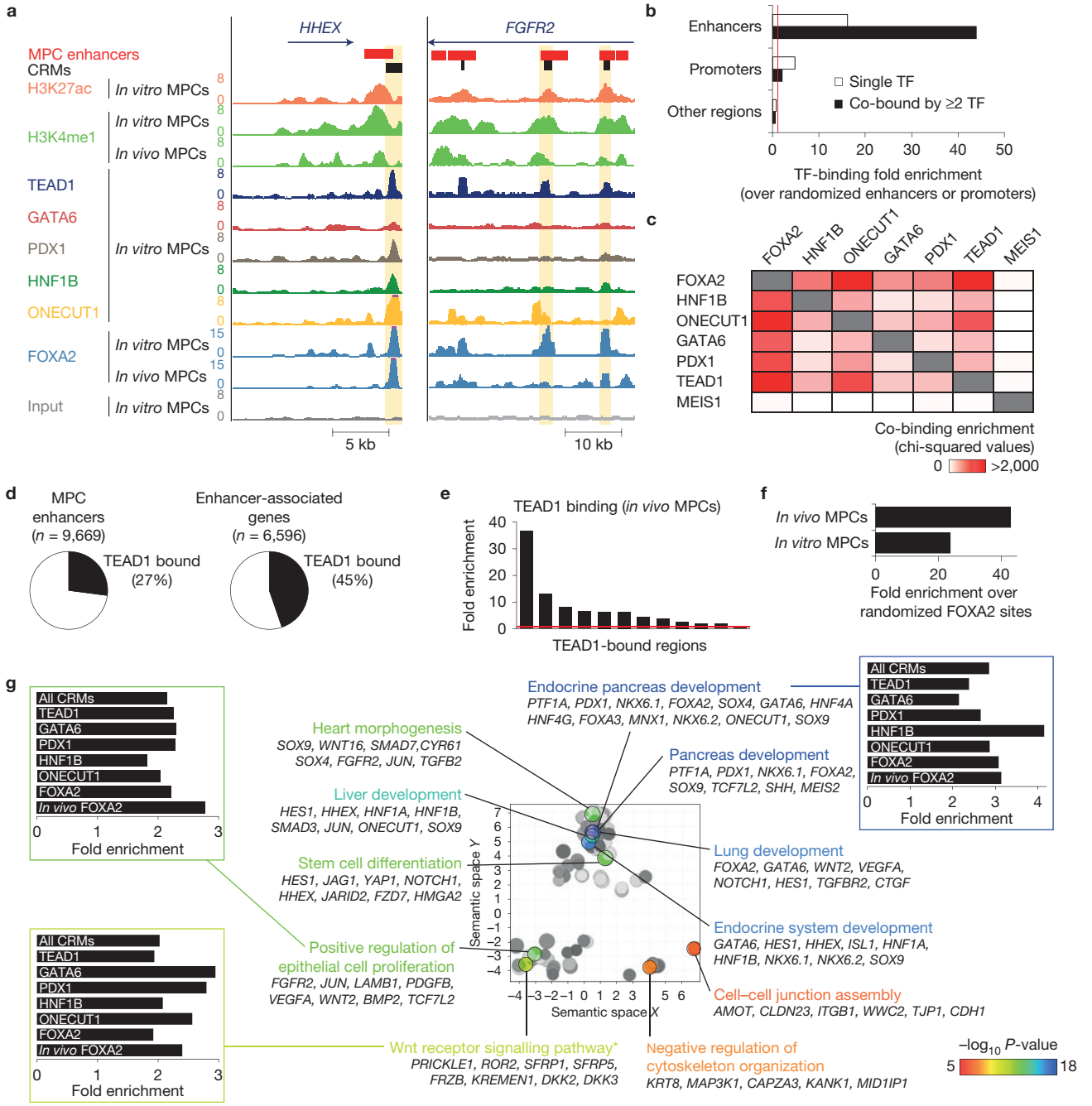


Figure 4 TEAD1 is a core component of human pancreatic MPC CRMs. (**a**) ChIP-seq was used to locate binding sites of six TFs in MPCs, as illustrated in two loci encoding pancreatic TFs. CRMs were defined as enhancer regions with at least two overlapping TF-bound sites. Examples are highlighted in yellow. (**b**) TFs preferentially occupy MPC enhancers, and this is most pronounced for regions bound by at least two TFs. Binding enrichment was calculated over 1,000 permutations of enhancer or promoter genomic positions in the mappable genome. For comparison, we analysed all other genomic regions after exclusion of MPC enhancers or promoters. The red line indicates a fold enrichment of unity. (**c**) Pancreatic TFs co-occupy genomic regions, and TEAD1 shows a similar co-occupancy pattern to other known pancreatic TFs. Binding sites of MEIS1 in a non-pancreatic cell type were used as control. The heat map depicts chi-squared values for all pairwise comparisons of observed versus expected co-binding. The latter was estimated by permuting each set of TF peaks independently 1,000 times. (**d**) Over one-quarter of MPC enhancers are bound by TEAD1, whereas 45% of genes associated with MPC enhancers include at least one TEAD1-bound enhancer. (**e**) ChIP-qPCR (quantitative PCR) with *in vivo* MPCs confirms TEAD1 binding at *in vitro* MPC TEAD1-bound regions (regions and associated genes in Supplementary Table 15). (**f**) TEAD1 binding is enriched in regions bound by FOXA2 in either *in vitro* or *in vivo* MPCs. We calculated TEAD1–FOXA2 co-binding over the median expected value after generating 1,000 permutations of *in vitro* or *in vivo* FOXA2 peak positions. (**g**) CRMs underlie a pancreas developmental regulatory network. The 2,956 genes associated with CRMs were functionally annotated using GREAT (ref. 53), and REVIGO (ref. 54) was used to visualize annotation clusters. The most significant terms from each cluster are highlighted according to the *P*-value colour scale. Bar graphs show that Gene Ontology (GO) terms are similarly enriched in CRMs bound by different TFs. *Several WNT-pathway-related terms were enriched, although manual annotation in this category revealed that most genes were either non-canonical WNT signalling mediators or antagonists of canonical WNT signalling (full annotations in Supplementary Table 17).

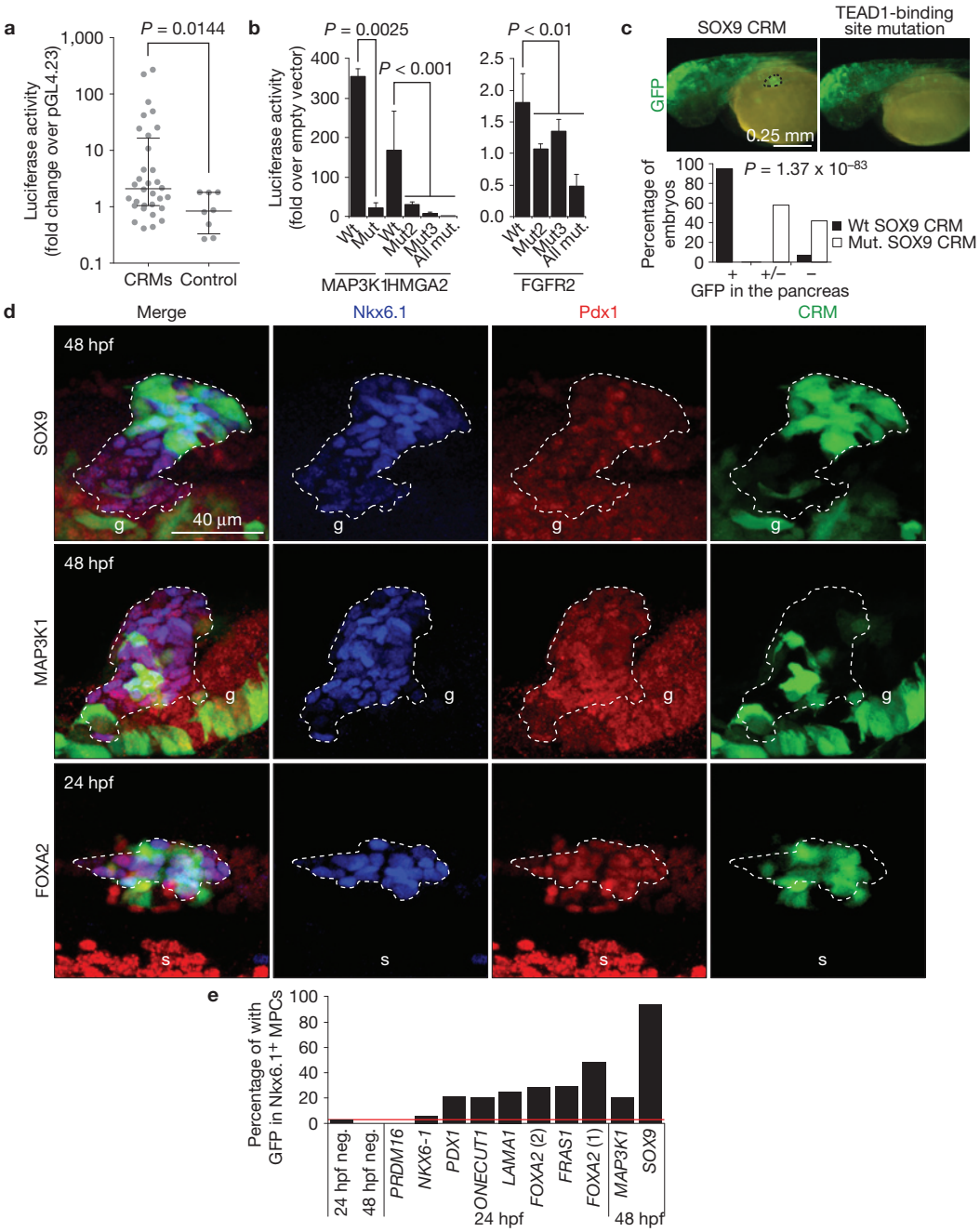


Figure 5 Functional validation of CRMs as transcriptional enhancers. (**a**) Thirty-two CRMs were cloned into the pGL4.23 vector and tested in reporter assays, where 20 (62.5%) yielded significant activation of a minimal promoter driving luciferase in human pancreatic MPCs. Lines represent median with IQR. Two-tailed Mann–Whitney test *P*-value is shown ($n=4$ replicate wells). (See also Supplementary Fig. 5a.) (**b**) TEAD-binding sites are essential for MPC enhancer activity. Mutation of one or more canonical TEAD-binding sites in three CRMs abolished their activity in luciferase reporter assays in *in vitro* MPCs. Locations of the *FGFR2* and *MAP3K1* CRMs are highlighted in Fig. 4a and Supplementary Fig. 4c, respectively. Two-tailed *t*-test *P*-values are listed in Supplementary Table 22 ($n=3$–4 transfections per construct, in one or two independent experiments). Error bars represent s.e.m. Wt, wild type; mut., mutant. (**c**,**d**) A TEAD1-bound CRM near *SOX9* (Fig. 7e) was fused to a minimal promoter and green fluorescent protein (GFP), and injected into zebrafish embryos. In (**c**), a *SOX9* CRM drove strong GFP expression in the pancreatic domain of 48 hpf zebrafish embryos (dotted circle, left panel), which was disrupted by a mutation in the TEAD recognition sequence (right). A midbrain-specific enhancer was used as an internal control for transgenesis. Note that this experiment assessed the activity of a single *SOX9* CRM, which does not necessarily fully recapitulate the expression of endogenous *sox9b*. In the graph, +, +/− and − represent strong, weak and absent GFP expression in the pancreatic domain, respectively ($n=110$–140 embryos per condition, chi-squared test $P=1.37\times10^{-83}$). (**d**) Immunofluorescence analysis of pancreatic MPCs in zebrafish embryos injected at one- to two-cell stage with constructs containing *SOX9*, *MAP3K1* and *FOXA2* CRMs driving GFP. Images show GFP in Pdx1$^+$/Nkx6.1$^+$ cells at 24/48 hpf, as indicated. In total, 8/10 CRMs yielded activity in Pdx1$^+$/Nkx6.1$^+$ progenitors (see also Supplementary Fig. 5b). The pancreatic progenitor domain is revealed by co-expression of Pdx1$^+$ and Nkx6.1$^+$ (dashed lines). Note that in zebrafish Nkx6.1 is specific to MPCs within embryonic pancreas[68]. g, Pdx1$^+$ gut cells; s, somites showing cross-reactivity with anti-Pdx1 serum. (**e**) Percentage of transgenic embryos with CRM-driven GFP expression in MPCs, or in negative controls (neg.) (quantifications shown in Supplementary Table 21).

and nuclear exclusion[27]. We examined nuclear localization of YAP throughout differentiation, and found that YAP was highly expressed in the nucleus of hESCs, and subsequently showed low yet detectable immunoreactivity throughout intermediary stages of the *in vitro* pancreatic differentiation protocol (Supplementary Fig. 6a), as well as in the nucleus of dorsal foregut epithelial cells of CS10 human embryos (Supplementary Fig. 6b). Strong YAP expression was subsequently observed in the nucleus of *in vitro*-derived pancreatic MPCs, as well as human and mouse *in vivo* pancreatic MPCs (CS18 and embryonic day 10.5 (E10.5)–E14.5 embryos, respectively; Fig. 6a and Supplementary Fig. 6c–f,h), in keeping with recent descriptions in mice[28]. By contrast, YAP immunoreactivity was undetectable or delocalized to the cytoplasm in NGN3$^+$ endocrine-committed progenitors, differentiated acinar cells or endocrine cells (Fig. 6b,c and Supplementary Fig. 6c–g,i), although nuclear expression was maintained in ductal cells (Supplementary Fig. 6f). Furthermore, in pancreatic MPCs YAP bound to most tested TEAD1-bound regions (Fig. 6e), similar to what has been observed in other cell types that exhibit nuclear YAP expression[27]. Thus, during embryonic pancreas development the coactivator YAP shows stage-specific nuclear localization in MPCs. This suggests a YAP-dependent function of TEAD1 during early pancreas development that is confined to MPCs, and is then inactivated on differentiation of pancreatic lineages.

To study YAP-dependent TEAD function in pancreatic MPCs, we first used verteporfin (VP), a chemical compound that disrupts the TEAD–YAP complex[29]. VP treatment of human *in vitro* MPCs and pancreatic bud explants dissected from E11.5 mouse embryos and grown *ex vivo* caused decreased expression of a subset of genes associated with TEAD1-bound enhancers, including genes that are established critical regulators of progenitor cell growth in the embryonic pancreas, such as *FGFR2* (ref. 30) and *SOX9* (refs 14,31), as well as mediators of growth regulatory pathways, such as *NOTCH1* and the known Hippo target *CCND1* (encoding cyclin D1; Fig. 7a,b and Supplementary Fig. 4f). Consistently, exposure of mouse explants to VP for 24 h significantly reduced epithelial cell proliferation by 39% ($P = 0.006$; Fig. 7c) and limited the growth of pancreatic buds to 27% of control organs after 3 days in culture ($P = 0.038$; Fig. 7d). These results suggest that the TEAD–YAP complex has direct effects on several known regulators of pancreatic progenitors, and is required for the proliferation and growth of early embryonic pancreas epithelium.

To further test the *in vivo* function of YAP and TEAD in pancreas development, we carried out genetic perturbations in zebrafish. In keeping with our chemical inhibition studies, morpholino inhibition of *yap1* caused a reduction in the pancreas size at 48 hours post fertilization (hpf), with hypoplasia in 65% of embryos ($n = 46$; Supplementary Fig. 7a), and a marked reduction of *sox9b*-expressing pancreatic MPCs (Fig. 7f,g). This effect was partially rescued by co-injection with *yap1* mRNA, confirming the morpholino specificity (Supplementary Fig. 7a). In agreement, zebrafish embryos expressing a TEAD protein fused to the transcriptional repressor domain of Engrailed (TEAD–EnR) (ref. 32) phenocopied the morpholino inhibition of *yap1* (Fig. 7g and Supplementary Fig. 7a). In summary, inhibition of Yap1 and TEAD proteins in zebrafish suppressed pancreatic *sox9b* expression and cell growth, in agreement with our mouse and human *in vitro* studies. Given that TEAD directly regulates a *SOX9* enhancer (Fig. 5c), and that *SOX9* regulates mouse and human pancreatic MPC growth[13,14,31], we hypothesize that the effects of TEAD and YAP on pancreatic progenitors are partially mediated through *SOX9*. Taken together, genetic and chemical inhibitor experiments support a model whereby YAP co-activation of TEAD1-bound MPC enhancers regulates a genomic regulatory programme that is required for the expression of stage-specific genes and for the outgrowth of pancreatic progenitors.

DISCUSSION

We have created and validated a map of active enhancers in human embryonic pancreatic progenitors. This effort expands the current list of known active enhancers in the embryonic pancreas from a handful of examples to thousands of stage-specific *cis*-regulatory elements. This included clustered enhancers, which were linked to a core cell-specific transcriptional programme, in analogy to earlier studies in diverse cellular lineages[25,33]. Our studies also show that pancreatic embryonic progenitor cells derived from hESCs mimic salient transcriptional and epigenomic features of pancreatic progenitors from human embryos, illustrating the power of pluripotent stem cell biology to dissect regulatory mechanisms underlying human embryogenesis.

This atlas of pancreatic MPC enhancers should facilitate the discovery of non-coding mutations that cause human diseases linked to abnormal pancreas development. In support for this claim, H3K4me1-, PDX1- and FOXA2-binding data from *in vitro* MPCs enabled the identification of recessive mutations that map to a previously unannotated enhancer of *PTF1A* and cause isolated pancreas agenesis[34]. Sequence variation in MPC enhancers could hypothetically increase the susceptibility to type 2 diabetes mellitus by impacting pancreas development and thereby affecting the pancreatic beta cell mass. Finally, germ-line or somatic variants in MPC enhancers could also influence the development of pancreatic adenocarcinoma, which has been associated with dedifferentiation of adult exocrine cells[35,36] and with YAP activation[37,38].

Our study identifies binding sites of several TFs that are known to be essential for early pancreas development, and show that they co-occupy pancreatic MPC enhancers, consistent with a combinatorial TF code. Unexpectedly, our results revealed that TEAD proteins—exemplified by TEAD1—and the coactivator YAP are central components of this combinatorial code, activating key regulatory genes and promoting the outgrowth of pancreatic MPCs.

The TEAD-dependent transcriptional mechanism provides a means for signal-responsive dynamic regulation of MPC enhancers during pancreas development. The coactivator YAP is a component of the Hippo signalling cascade, which phosphorylates YAP, leading to its retention in the cytoplasm or to its degradation[39]. Our data shows that, as human pancreatic MPCs transition to endocrine and acinar lineages, YAP undergoes immediate nuclear exclusion and downregulation. On the basis of our chemical and genetic experiments, this dynamic change is expected to lead to inhibition of MPC enhancers during pancreatic differentiation.

Two recent reports showed that pancreas-specific disruption of the upstream Hippo kinases Mst1/2 leads to increased proliferation of *adult* acinar pancreatic cells, which acquire a duct-like morphology, exhibit increased nuclear localization of Yap and show ectopic

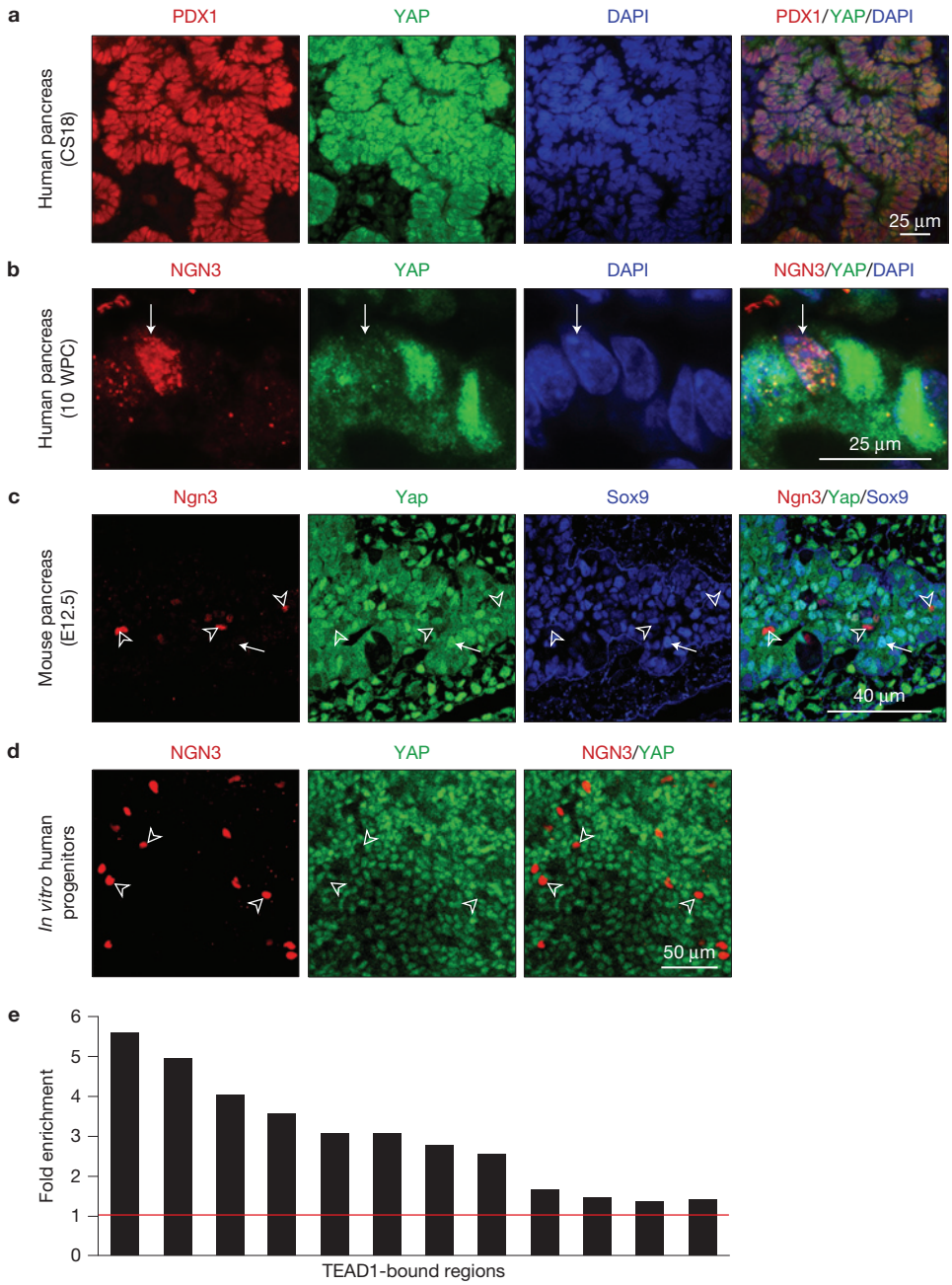


Figure 6 YAP is expressed in the nucleus of pancreatic MPCs, and shows co-occupancy with TEAD1 at MPC enhancers. (**a**) YAP is detected in the nucleus of PDX1$^+$ *in vivo* MPCs from human CS18 pancreas. DAPI, 4′,6-diamidino-2-phenylindole. (**b**) In 10 weeks post conception (WPC) human pancreas YAP expression is strong in nuclei of PDX1$^+$ progenitors, but shows markedly diminished signal intensity in NGN3$^+$ progenitors (white arrow). The image depicts five cells in human embryonic pancreas 10 WPC. (**c**) Yap is detected in the nucleus of Sox9$^+$ MPCs from mouse E12.5 embryonic pancreas (white arrow), whereas Yap is diffuse in or absent from Ngn3$^+$ endocrine progenitor cells (hollow arrowheads). (**d**) YAP is excluded from the nucleus in hESC-derived pancreatic NGN3$^+$ progenitor cells (hollow arrowheads). (**e**) ChIP-qPCR analysis of YAP occupancy in chromatin from *in vitro* MPCs shows that TEAD1-bound regions are often co-bound by YAP.

expression of the TEAD target Sox9 (refs 28,40). These observations do not address whether Hippo signalling or TEAD are important for pancreatic progenitors, but they are consistent with failed suppression of a progenitor programme in adult cells, and therefore support the predictions from our findings. Collectively, existing data suggest a model whereby TEAD proteins provide a regulatory switch that activates a stage-specific transcriptional programme in pancreatic MPCs, and facilitates signal-responsive inactivation of this programme during pancreatic cell differentiation (Fig. 8).

Further studies should explore this regulatory mechanism in human disease. The reactivation of the YAP–TEAD-dependent MPC enhancer programme in adult acinar cells could conceivably activate a progenitor-like cellular programme during early stages of pancreatic carcinogenesis[35,36], and/or contribute to YAP-dependent cancer

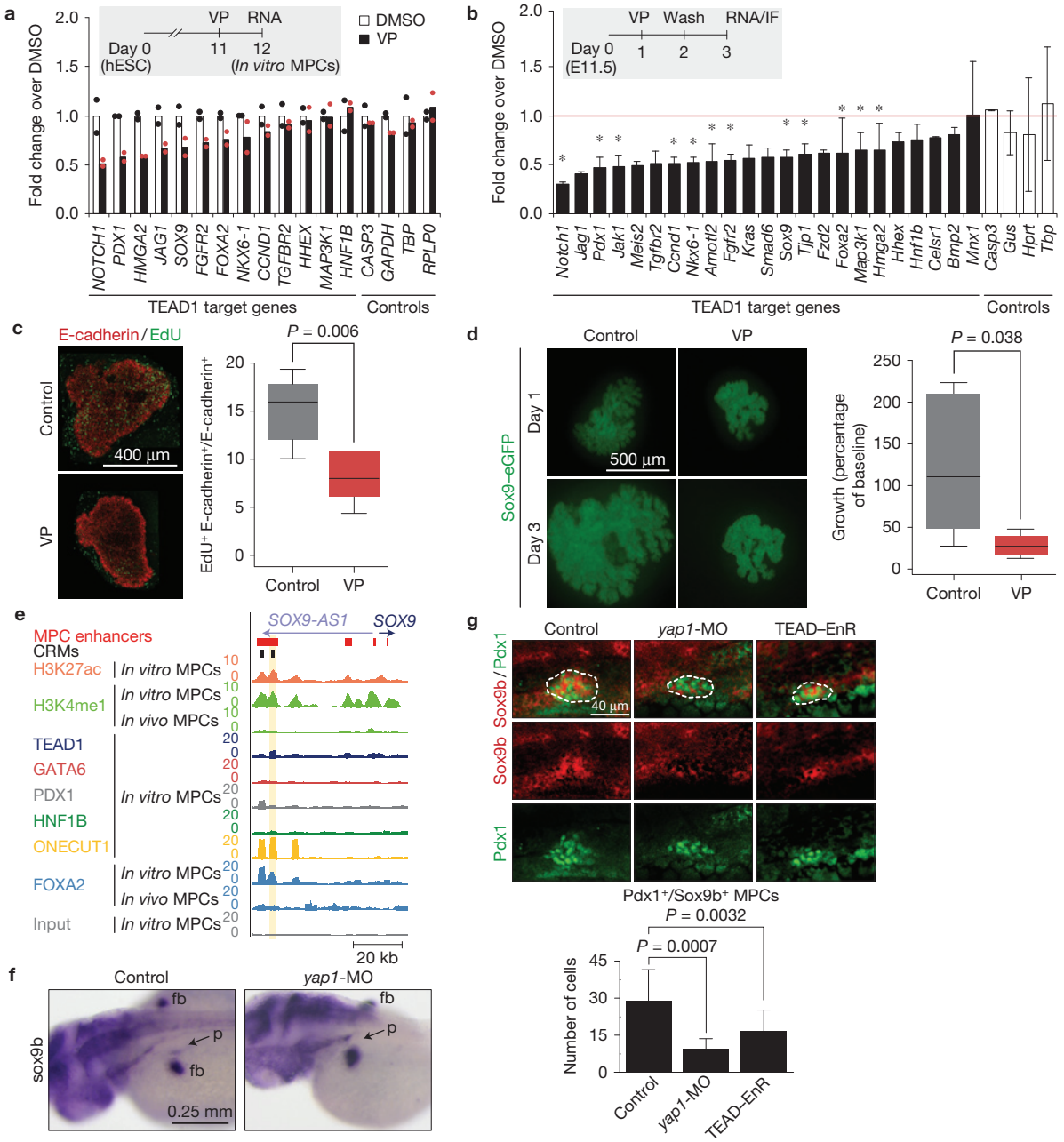


Figure 7 TEAD and YAP regulation of pancreas development. (**a**) Human *in vitro* MPCs were incubated with VP 24 h to disrupt TEAD–YAP interactions, causing downregulation of genes associated with TEAD1-bound enhancers. Data were normalized by *PBGD*. Bars show mean values from two independent experiments, and points represent mean of two technical replicates. DMSO, dimethylsulphoxide. (**b**–**d**) VP treatment of E11.5 mouse pancreatic explants downregulated orthologues of TEAD1-bound genes, inhibited proliferation and reduced growth of pancreatic epithelial cells. Explants were treated with VP for 24 h, washed and incubated for 24 h before analysis. Data were normalized to *Gapdh*. *Two-tailed *t*-test, $P < 0.05$ (individual values listed in Supplementary Table 22). Error bars represent s.d. from three independent experiments (each with $n = 2$–4 embryos per condition). IF, immunofluorescence. (**c**) The percentage of proliferating epithelial cells was quantified with E-cadherin and 5-ethynyl-2′-deoxyuridine (EdU) immunolocalization. The two-tailed Mann–Whitney *P*-value is shown for three experiments (each with $n = 2$–3 pancreases per condition). (**d**) GFP$^+$ area in Sox9–eGFP (enhanced GFP) transgenic embryo explants is shown at day 3, compared with day 1. Two-tailed Mann–Whitney test *P*-values are shown for three experiments (each with $n = 2$–4 buds per condition). In **c** and **d** boxes are IQR and median, whiskers are $1.5 \times$ IQR or extreme values. (**e**) Snapshot of the human *SOX9* locus, encoding a regulator of MPC growth[14]. The CRM tested in functional assays in Fig. 5c and Fig. 7f is highlighted. (**f**) *yap1* inhibition decreased pancreatic *sox9b* expression. Injection of *yap1* morpholino oligonucleotide (*yap1*-MO) caused a reduction or absence of *sox9b* mRNA in the pancreatic domain (p; arrow) in 50/102 48 hpf embryos. Control embryos showed pancreatic *sox9b* expression in 100/100 embryos (chi-squared $P = 2.61 \times 10^{-15}$). Note that control and morphant embryos always showed *sox9b* expression in fin buds (fb). (**g**) Injection of *yap1*-MO ($n = 10$ embryos) or the TEAD–EnR dominant negative ($n = 12$ embryos) caused a decreased number of *sox9b*$^+$/Pdx1$^+$ pancreatic progenitors (dotted lines) in 24 hpf embryos versus controls ($n = 9$ embryos). *Sox9b* was detected by *in situ* hybridization and Pdx1 by immunofluorescence. The graph reflects the total number of pancreatic progenitors in each embryo. *yap1*-MO also increased ectopic expression of pancreatic markers Supplementary Fig. 7b). Student's *t*-test *P*-values and s.d. are shown.

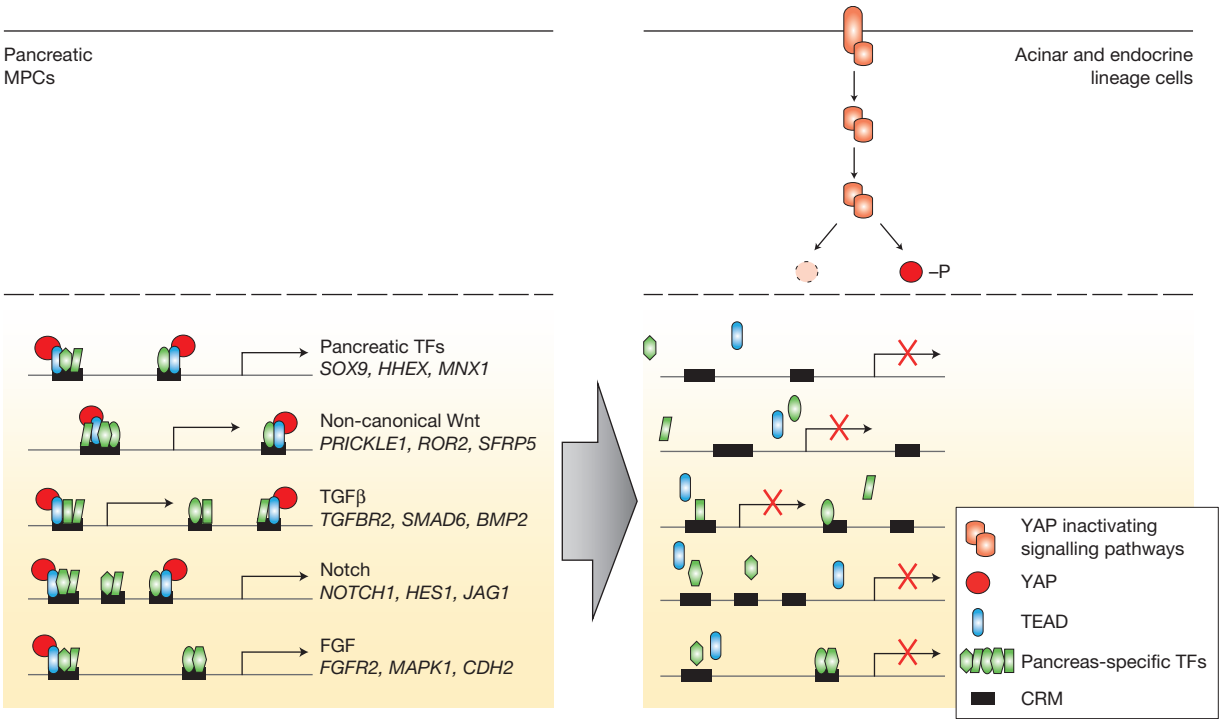


Figure 8 YAP–TEAD-dependent activation provides a regulatory switch for pancreatic MPC enhancers. A significant number of pancreatic MPC enhancers is co-bound by known stage-specific TFs along with TEAD and YAP. During pancreatic differentiation YAP is rapidly excluded from the nucleus and its expression is reduced, causing inactivation of MPC stage-specific enhancers. This simplified model depicts inhibition of YAP through Hippo kinase-induced phosphorylation or degradation, although further non-mutually exclusive mechanisms for dynamic inhibition of YAP signalling are plausible. The model is supported by evidence showing that chemical or genetic inhibition of YAP and TEAD function causes inhibition of MPC enhancers.

progression[37,38]. This same genetic programme could potentially be exploited to control growth and differentiation during the generation of artificial pancreatic cells. □

METHODS

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

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS
The research was supported by the National Institute for Health Research (NIHR) Imperial Biomedical Research Centre. Work was funded by grants from the Ministerio de Economía y Competitividad (CB07/08/0021, SAF2011-27086, PLE2009-0162 to J.F., BFU2013-41322-P to J.L.G-S.), the Andalusian Government (BIO-396 to J.L.G-S.), the Wellcome Trust (WT088566 and WT097820 to N.A.H., WT101033 to J.F.), the Manchester Biomedical Research Centre, ERC advanced starting grant IMDs (C.H-H.C. and L.V.) and the Cambridge Hospitals National Institute for Health Research Biomedical Research Centre (L.V.). R.E.J. is a Medical Research Council clinical training fellow. The authors are grateful to C. Wright (Vanderbilt University) for zebrafish Pdx1 antiserum, J. Postlethwait (Purdue University) for a Sox9b clone, H. Sasaki (Kumamoto University) for a TEAD–EnR clone, C. Vinod and L. Abi for research nurse assistance, and clinical colleagues at Central Manchester University Hospitals NHS Foundation Trust. The authors thank J. Garcia-Hurtado for technical assistance (IDIBAPS).

AUTHOR CONTRIBUTIONS
J.F. coordinated the overall project and supervised epigenomic analysis and mouse studies, N.A.H. supervised human embryo characterization, L.V. supervised hESC differentiation studies and J.L.G-S. supervised zebrafish studies. I.C., S.A.R-S., C.H-H.C., J.B., M.R., M.L., M.C., A.B., M.A.M. and R.E.J. designed, carried out and analysed experiments. N.C. carried out experiments. I.C., S.A.R-S., J.P-C., L.P. and I.M. carried out computational analysis. I.C., S.A.R-S. and J.F. wrote the manuscript with contributions from C.H-H.C., J.B., M.R., M.L., J.P-C., N.A.H., J.L.G-S. and L.V.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Published online at http://dx.doi.org/10.1038/ncb3160
Reprints and permissions information is available online at www.nature.com/reprints

1. Fang, H. *et al.* An organogenesis network-based comparative transcriptome analysis for understanding early human development *in vivo* and *in vitro*. *BMC Syst. Biol.* **5,** 108 (2011).
2. Fang, H. *et al.* Transcriptome analysis of early organogenesis in human embryos. *Dev. Cell* **19,** 174–184 (2010).
3. Pan, F. C. & Wright, C. Pancreas organogenesis: from bud to plexus to gland. *Dev. Dynam.* **240,** 530–565 (2011).
4. Zaret, K. S. & Grompe, M. Generation and regeneration of cells of the liver and pancreas. *Science* **322,** 1490–1494 (2008).
5. Lango Allen, H. *et al.* GATA6 haploinsufficiency causes pancreatic agenesis in humans. *Nat. Genet.* **44,** 20–22 (2012).
6. Xuan, S. *et al.* Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. *J. Clin. Invest.* **122,** 3516–3528 (2012).
7. Carrasco, M., Delgado, I., Soria, B., Martín, F. & Rojas, A. GATA4 and GATA6 control mouse pancreas organogenesis. *J. Clin. Invest.* **122,** 3504–3515 (2012).
8. Offield, M. F. *et al.* PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122,** 983–995 (1996).
9. 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).
10. Haumaitre, C. *et al.* Lack of TCF2/vHNF1 in mice leads to pancreas agenesis. *Proc. Natl Acad. Sci. USA* **102,** 1490–1495 (2005).
11. 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).
12. Gao, N. *et al.* Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev.* **22,** 3435–3448 (2008).

13. Piper, K. *et al.* Novel SOX9 expression during human pancreas development correlates to abnormalities in Campomelic dysplasia. *Mech. Dev.* **116,** 223–226 (2002).
14. Seymour, P. A. *et al.* SOX9 is required for maintenance of the pancreatic progenitor cell pool. *Proc. Natl Acad. Sci. USA* **104,** 1865–1870 (2007).
15. Krapp, A. *et al.* The p48 DNA-binding subunit of transcription factor PTF1 is a new exocrine pancreas-specific basic helix-loop-helix protein. *EMBO J.* **15,** 4317–4329 (1996).
16. Jennings, R. E. *et al.* Development of the human pancreas from foregut to endocrine commitment. *Diabetes* **62,** 3514–3522 (2013).
17. Cho, C. H. H. *et al.* Inhibition of activin/nodal signalling is necessary for pancreatic differentiation of human pluripotent stem cells. *Diabetologia* **55,** 3284–3295 (2012).
18. Xie, R. *et al.* Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell* **12,** 224–237 (2013).
19. Kroon, E., Martinson, L. A., Kadoya, K. & Bang, A. G. Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nature* **26,** 443–452 (2008).
20. Borowiak, M., Maehr, R., Chen, S., Chen, A. E. & Tang, W. Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. *Cell Stem Cell* **4,** 348–358 (2009).
21. Rodríguez-Seguel, E. *et al.* Mutually exclusive signaling signatures define the hepatic and pancreatic progenitor cell lineage divergence. *Genes Dev.* **27,** 1932–1946 (2013).
22. Cortijo, C., Gouzi, M., Tissir, F. & Grapin-Botton, A. Planar cell polarity controls pancreatic β cell differentiation and glucose homeostasis. *Cell Rep.* **2,** 1593–1606 (2012).
23. Rada-Iglesias, A., Bajpai, R., Swigut, T. & Brugmann, S. A. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–283 (2011).
24. Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107,** 21931–21936 (2010).
25. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46,** 136–143 (2014).
26. Oliver-Krasinski, J. M. & Stoffers, D. A. On the origin of the β cell. *Genes Dev.* **22,** 1998–2021 (2008).
27. Zhao, B. *et al.* TEAD mediates YAP-dependent gene induction and growth control. *Genes Dev.* **22,** 1962–1971 (2008).
28. George, N. M., Day, C. E., Boerner, B. P., Johnson, R. L. & Sarvetnick, N. E. Hippo Signaling Regulates Pancreas Development through Inactivation oelf Yap. *Mol. Cell. Biol.* **32,** 5116–5128 (2012).
29. Liu-Chittenden, Y. *et al.* Genetic and pharmacological disruption of the TEAD-YAP complex suppresses the oncogenic activity of YAP. *Genes Dev.* **26,** 1300–1305 (2012).
30. Elghazi, L., Cras-Méneur, C., Czernichow, P. & Scharfmann, R. Role for FGFR2IIIb-mediated signals in controlling pancreatic endocrine progenitor cell proliferation. *Proc. Natl Acad. Sci. USA* **99,** 3884–3889 (2002).
31. Lynn, F. C. *et al.* Sox9 coordinates a transcriptional network in pancreatic progenitor cells. *Proc. Natl Acad. Sci. USA* **104,** 10500–10505 (2007).
32. Sawada, A. *et al.* Tead proteins activate the Foxa2 enhancer in the node in cooperation with a second factor. *Development* **132,** 4719–4729 (2005).
33. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
34. Weedon, M. N. *et al.* Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **46,** 61–64 (2014).
35. Hezel, A. F., Kimmelman, A. C., Stanger, B. Z., Bardeesy, N. & DePinho, R. A. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev.* **20,** 1218–1249 (2006).
36. Rooman, I. & Real, F. X. Pancreatic ductal adenocarcinoma and acinar cells: a matter of differentiation and development? *Gut* **61,** 449–458 (2012).
37. Kapoor, A. *et al.* Yap1 activation enables bypass of oncogenic Kras addiction in pancreatic cancer. *Cell* **158,** 185–197 (2014).
38. Zhang, W. *et al.* Downstream of mutant KRAS, the transcription regulator YAP is essential for neoplastic progression to pancreatic ductal adenocarcinoma. *Sci. Signal.* **7,** ra42 (2014).
39. Zhao, B., Tumaneng, K. & Guan, K. L. The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal. *Nat. Cell Biol.* **13,** 877–883 (2011).
40. Bardeesy, N. & Stanger, B. Z. Hippo signaling regulates differentiation and maintenance in the exocrine pancreas. *Gastroenterology* **144,** 1543–53–1553.e1 (2013).
41. Fujitani, Y. *et al.* Targeted deletion of a cis-regulatory region reveals differential gene dosage requirements for Pdx1 in foregut organ differentiation and pancreas formation. *Genes Dev.* **20,** 253–266 (2006).
42. Gannon, M., Gamer, L. W. & Wright, C. V. Regulatory regions driving developmental and tissue-specific expression of the essential pancreatic gene pdx1. *Dev. Biol.* **238,** 185–201 (2001).

METHODS

Human samples. Human embryos were collected with informed consent with approval from the North West Regional Ethics Committee (08/H1010/28) following termination of pregnancy and staged immediately by stereomicroscopy according to the Carnegie classification[43]. The collection, use and storage of material followed guidelines from the UK Polkinghorne Committee, legislation of the Human Tissue Act 2004 and the Codes of Practice of the Human Tissue Authority, UK. The analysis of human embryonic tissue was also approved by the Comitè Ètic d'Investigació Clínica del Centre de Medicina Regenerativa de Barcelona and Departament de Salut, Generalitat de Catalunya. Human embryonic pancreas and liver were dissected at CS16–18, which correlates to ∼37–45 days post-conception. These stages were the earliest at which pancreatic epithelial cells could be efficiently dissected away from surrounding mesenchyme with minimal contamination. After isolation tissues were rinsed with PBS, incubated for 10 min in 1% formaldehyde and 5 min in 125 mM glycine, rinsed in PBS containing protease inhibitor cocktail (Roche) at 4 °C, and snap-frozen and stored at −80 °C. RNA was extracted using TRIzol and DNase.

Human ESCs (H9, WiCell) were imported under guidelines from the UK Stem Cell Bank Steering Committee (SCSC10-44). Differentiation of pancreatic MPCs has been described[17]. Briefly, definitive endoderm was induced by growing hESCs in AFBLy: chemically defined medium with polyvinyl alcohol (CDM-PVA) + activin A (100 ng ml^{-1}), BMP4 (10 ng ml^{-1}), basic fibroblast growth factor (bFGF; 20 ng ml^{-1}) and LY294004 (10 μM). The CDM-PVA AFBLy cocktail was replenished daily, and daily media changes were made until differentiation day 10. After the definitive endoderm stage (days 1–3), cells were cultured in Advanced DMEM (Invitrogen) with SB-431542 (10 μM; Tocris), FGF10 (50 ng ml^{-1}; Autogen Bioclear), all-*trans* retinoic acid (2 μM; Sigma) and Noggin (150 ng ml^{-1}; R&D Systems) during days 4–6. For days 7–9, cells were supplemented with human FGF10 (50 ng ml^{-1}; Autogen Bioclear), all-*trans* retinoic acid (2 μM; Sigma), KAAD-cyclopamine (0.25 μM; Toronto Research Chemicals) and Noggin (150 ng ml^{-1}; R&D Systems). On days 10–12, cells were cultured in human FGF10 (50 ng ml^{-1}; Autogen Bioclear), all-*trans* retinoic acid (2 μM; Sigma), and KAAD-cyclopamine (0.25 μM; Toronto Research Chemicals).

For RNA-seq and ChIP-seq, cells from three independent differentiation experiments were pooled. For ChIP, cells were fixed as described above, snap-frozen and kept at −80 °C. Total RNAs were extracted from hESCs or differentiated progenitors using the RNeasy Mini Kit (Qiagen) and treated with RNase-free DNase (Qiagen).

Immunolocalization. Immunolocalization was carried out as described[16,17,44,45]. Antibodies are listed in Supplementary Table 23.

Pancreatic explants from E12.5 C57BL/6 mouse embryos and whole-mount stainings were carried out as described[46] with modifications. Briefly, pancreases were fixed for 20 min in 4% paraformaldehyde, blocked in 0.5% Triton X-100/10% fetal bovine serum (FBS)/PBS overnight at 4 °C and incubated for 24 h with primary antibody at 4 °C then overnight with secondary antibody at 4 °C, and finally DAPI stained. EdU staining was carried out using a Click-iT EdU Alexa Fluor 488 Imaging Kit (Invitrogen). All images presented show representative results obtained from at least three independent experiments.

ChIP. Either seven human CS16–18 pancreatic buds, four CS16–18 liver buds (as described above) or about 10 million cells from a pool of three pancreatic progenitor *in vitro* differentiation experiments were pooled in 1 ml of lysis buffer containing protease inhibitor cocktail (Roche) and sonicated for 10–15 cycles essentially as described[25,47]. We verified that a substantial portion of chromatin fragments were in the 200–600-bp range by gel electrophoresis.

ChIP was carried out with 50–300 μl of sonicated chromatin as described[25,48,49], with minor modifications. Briefly, sonicated chromatin was diluted with ChIP dilution buffer (0.75% Triton X-100, 140 mM NaCl, 0.1% sodium deoxycholate, 50 mM HEPES at pH 8.0, 1 mM EDTA, 1× protease inhibitor cocktail) to achieve a final SDS concentration of 0.2%, pre-cleared with A/G Sepharose beads (GE Healthcare) for 1 h, incubated overnight at 4 °C with 1–1.5 μg antibodies (Supplementary Table 23), rotated for 2 h at 4 °C with A/G Sepharose beads, and then sequentially washed and processed[25,48,49].

RNA-seq. All samples had RNA integrity number (RIN) > 9. RNA-seq was generated from DNase-treated polyA + RNA from a CS17 pancreatic bud or from a pool of three *in vitro* pancreatic MPC differentiation experiments, sequencing 90-nucleotide reads with an Illumina HiSeq 2000 instrument. RNA-seq data sets from 23 tissues and their sources are listed in Supplementary Table 1. Reads were aligned to the NCBI36/hg18 genome using TopHat v1.2.0 (ref. 50) with default parameters, allowing only one mismatch per read. For comparison of RNA levels, we processed and calculated FPKM values for each transcript as described[51]. For a global comparison of gene expression profiles (Supplementary Fig. 1c), we analysed 44,699 UCSC gene variants expressed at more than 5 FPKM in at least one sample. Expression values were median-centred and scaled by the root mean square. Spearman's correlation values were calculated for each pair.

RNA enrichment analysis. Tissue selectivity of each transcript was assessed by computing the FPKM CV in the 25 samples described in Supplementary Table 1. To obtain the enrichment of each transcript in each tissue, we calculated Z-scores as the difference between the $\log_2$-transformed expression level in the specific tissue and the mean of all tissues, divided by the standard deviation. For detection of MPC-specific transcripts, Z-score measurements were calculated without data from islets and either *in vitro* or *in vivo* MPCs. We defined tissue-specific genes as those with CV ≥ 1, expression ≥ 0.3 FPKM and Z-score ≥ 1 in any tissue.

Core MPC-specific genes were defined as UCSC-annotated genes that were tissue selective (CV ≥ 1) and enriched in *in vitro* MPCs (Z-score ≥ 1). We then sorted by *in vivo* MPC enrichment Z-score, and selected the top 500 (Supplementary Table 5).

Functional annotations. Transcript functional annotation was carried out with DAVID (ref. 52), using GO Biological Process (FAT), Pathways (KEGG, Panther) and annotation clustering. In Fig. 1c, we sorted terms by P-value and show the most significant term of each cluster. Annotations are provided in Supplementary Table 3.

Genes associated with enhancers and CRMs were analysed with GREAT v2.0.2 (ref. 53) applying default settings (basal plus extension; significant by both binomial and hypergeometric tests), and annotated with GO Biological Process plus all pathway annotations. Raw binomial P-value and binomial fold enrichment were used to present enrichments. Supplementary Tables 9, 17 and 19 list annotations associated with MPC-selective enhancers, CRMs and CRM clusters, respectively. GO Biological Process terms were further processed with REVIGO (ref. 54; 0.9 allowed similarity; term size database—whole UniProt; semantic similarity measure—normalized Resnik; cluster definition default parameters) taking the most significant term in each GO cluster.

ChIP-seq. Chromatin from replicate pools of *in vitro* MPCs was used for FOXA2 and H3K4me1 ChIP-seq experiments. Single libraries were prepared from chromatin pools for all other ChIP-seq experiments, except for FOXA2 *in vivo* MPC, in which libraries from two experiments were sequenced and reads were pooled for alignment. All libraries were prepared with 5–10 ng DNA, sequenced with the Illumina HiSeq 2000 platform and aligned to NCBI36/hg18 using Bowtie v0.12.7 (Supplementary Table 2), allowing unique alignment with at most one mismatch. Post-alignment processing included *in silico* extension, signal normalization based on the number of millions of mapped reads, extension to MACS fragment size estimation (v1.4.0beta) and retention of only unique reads. For signal normalization, the number of reads mapping to each base in the genome was counted with genomeCoverageBed (bedtools v2.17.0). TF enrichment sites were detected with MACS v1.4.0beta using default parameters and $P < 10^{-10}$. The background model was defined with the input DNA sequence. SICER v1.03 was used to call H3K4me1-enriched islands with window size 100 bp, gap size 800 bp and fragment size estimated by MACS v1.4.0beta. Enriched islands were called at false discovery rate (FDR) $< 10^{-3}$. For H3K27ac-enriched regions gap size was 200 bp. For replicate samples we retained overlapping peaks/islands in replicates. To compute FOXA2 and H3K4me1 signal correlations between duplicates we divided the genome into 1 or 5 kilobase (kb) bins, respectively, then counted unique reads in each bin and quantile-normalized results. Bins with values less than the fifth percentile in both samples were excluded from the analysis. Pearson correlation values were 0.8–0.9 in all biological replicates (Supplementary Fig. 1g). Public data sets were processed identically (listed in Supplementary Table 2).

TF and H3K4me1 aggregation plots. To compute aggregation plots (Fig. 1f), we first selected 'tissue-specific regulatory regions', defined by the intersections of H3K4me1 islands with TF 'peaks' in the same tissue. The resulting numbers of regions were as follows: FOXA2 in *in vivo* MPCs, 2,307; SOX2 in hESCs, 5,749; MEIS1 in CD133$^+$ cells, 2,210; DNase I peaks containing ETS1 motifs in mammary epithelial cells, 14,100, and DNase I peaks containing MEF2A motifs in myotube cells, 13,614. Next, regions spanning ±3 kb from the centre of TF peaks were divided into 100-bp bins. The coverage signal was obtained using coverageBed (bedtools v2.17.0). Data were quantile normalized after creating 100-bp bins in H3K4me1 islands from each tissue.

Clustering. To compare ChIP-seq signals between tissues (Fig. 1e and Supplementary Fig. 1f), we generated 6-kb windows centred on *in vivo* MPC FOXA2 peaks. Each window was divided into 100-bp bins and binned signal coverage was quantile normalized as described above. Hierarchical clustering was carried out with Cluster3 (ref. 55) with similarity metric set to Correlation (uncentred) and average linkage as the clustering method. Heat maps were visualized with TreeView[56].

DOI: 10.1038/ncb3160

Definition of enhancers and CRMs. Enhancers were defined as H3K27ac islands in the *in vitro* MPCs that overlapped H3K4me1 islands in both *in vitro* and *in vivo* MPCs. We discarded regions overlapping promoters (1 kb upstream and 2 kb downstream of RefSeq transcription start sites) or less than 50 bp. Enhancers in eight control tissues were defined with analogous criteria based on H3K27ac and H3K4me1 islands (Supplementary Table 8).

To define CRMs we merged all *in vitro* MPC TF peaks that were less than 500 bp apart, and retained 2,945 regions bound by at least two different TFs that overlapped MPC enhancers by at least 1 bp.

Clusters of CRMs were defined as described[25], essentially as any group of at least three CRMs in which all adjacent CRMs were separated by less than the 25th percentile of chromosome-specific randomized distances.

Enhancer selectivity. *MPC-selective* enhancers were defined as those that showed no overlap with an enhancer from at least six out of seven control tissues (hESCs, fetal muscle, fetal stomach, fetal thymus, mammary epithelial cells, myotubes and osteoblasts).

Conservation. Conservation was assessed in ±3-kb windows centred in enhancers, using the average vertebrate phastCons score from 17 species for 20-bp bins.

Motif analysis. *De novo* motif discovery was carried out with HOMER (ref. 57). For enhancers we searched for either short (length = 6, 8, 10, 12) or long (length = 14, 16, 18, 20) motifs as described previously[25], retaining non-redundant matrices (Pearson correlation < 0.65) with $P < 10^{-50}$. Motifs were annotated using HOMER (ref. 57), TOMTOM (ref. 58) and manual comparisons.

All possible combinations of three motifs from the 23 enriched motifs contained within 500-bp regions were computed in MPC enhancers versus enhancers from eight other tissues. We calculated eight MPC versus control tissue fold enrichment and P-values (chi-squared test), and then combined them in a unique P-value for each motif combination with a Z-weighted method[59]. Supplementary Table 12 shows the top 50 most enriched combinations.

For TF peaks, HOMER analysis was carried out in 200-bp windows centred on peak summits and motif lengths were set to 8, 10 and 12 bp. Co-enriched motifs were manually curated to exclude redundant motifs. Known DNA-binding motifs were associated with the *de novo* recovered matrix only if the HOMER score was more than 0.7.

Binding and co-binding enrichment analysis. To assess the enrichment of TF binding and co-binding in enhancers or promoters, the positions of the enhancers or promoters were randomized in all mappable hg18 coordinates using shuffleBed (bedtools v2.17.0). Mappable regions were defined as those not annotated as genome gaps and with a score of 1 in the CRG mappability 50-bp track of the UCSC browser[60]. Binding enrichment was calculated over the median of 1,000 permutations.

Co-bound regions were defined with intersectBed (bedtools v2.17.0) as regions bound by at least two TFs. To calculate TF co-binding enrichment, we shuffled each TF individually in the mappable genome, and calculated the overlap with sites bound by the other TFs (median of 1,000 permutations generated by shuffleBed, bedtools v2.17.0). A chi-squared test was applied to assess the enrichment of each combination of two TFs over expected co-binding. For comparison, we applied the same pipeline to define 'co-binding' between MEIS1 in CD133$^+$ cells and the six MPC TFs (Supplementary Table 2).

Enhancer function assays in human cells. The pGL4.23[*luc2*/minP] vector (Promega) was modified by inserting a Gateway cassette upstream of the minimal promoter (pGL4.23-GW) for subsequent cloning of CRMs and control sequences. These 500–2,000-bp sequences were amplified from human genomic DNA with Phusion High-Fidelity DNA Polymerase (New England Biolabs), cloned into pENTR/D-TOPO (Invitrogen), shuttled into pGL4.23-GW and assessed by Sanger sequencing and restriction enzyme digestion. To mutate CRMs, we replaced a 3-bp sequence of the core of TEAD motifs, as this was previously shown to disrupt TEAD binding[27], and confirmed by Sanger sequencing. Oligonucleotides are available in Supplementary Table 15.

At day 10 of the differentiation protocol, cells were transfected in 24-well plates with pGL4.23-CRM plasmids (400 ng) and *Renilla* normalizer plasmid (4 ng) using Lipofectamine 2000 (Invitrogen). Luciferase was measured at day 13 with a Dual-Luciferase Reporter Assay System (Promega). The results shown represent the average and s.e.m. of three (HMGA2, GLIS3 and MAP3K10) or four (all other CRMs and all negative controls) independent transfections per construct. Eight of 32 plasmids in Fig. 5a and Supplementary Fig. 5a, and six of the nine CRMs in Fig. 5b, were retested in independent experiments that yielded comparable results. Statistical significance was assessed with a two-tailed Student t-test using all experiments (Supplementary Table 22).

Pancreatic explant experiments. Mouse experiments were approved by the Comitè Ètic d'Experimentació Animal (University of Barcelona) in accordance with national and European regulations. No statistical method was used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments or outcome assessment. Pancreatic explants were carried out as described[61] with minor modifications. Dorsal pancreatic buds from E11.5 CD-1 mouse embryos were cultured in RPMI medium with 10% FBS for 16 h (day 1) before VP 0.1 μM (Atomax) or DMSO (control) treatment 24 h in RPMI 3% FBS. After 24 h (day 2), the drug was washed out and buds were cultured for 1 day in RPMI 10% FBS (day 3).

For quantification of explant growth we used Sox9–eGFP transgenic embryos, which enabled visualization of pancreatic epithelial progenitors. We used ImageJ 1.46a to measure the area of eGFP-expressing cells on days 1 and 3. We carried out three independent experiments, and examined two to three pancreases per condition in each experiment. We expressed areas as percentage of the baseline in the same explant, and used the Mann–Whitney test to determine significance. Data failed to show normal distribution with the Kolmogorov–Smirnov test.

To study epithelial cell proliferation, explants were exposed to EdU (1 μM) after VP treatment for 30 min and analysed 24 h later. We examined two to four pancreases per condition in each of three independent experiments. The Mann–Whitney test was used for statistical significance.

We obtained RNA from pools of at least three pancreatic buds using the RNeasy Mini Kit (Qiagen), and collected three separate pools from independent experiments. qPCR with reverse transcription (qRT–PCR) was carried out using a 7300 Real Time PCR System (Applied Biosystems) and Power SYBR Green PCR Master Mix (Applied Biosystems). Each sample pool was amplified in duplicate using *Gapdh* for normalization. Oligonucleotides are shown in Supplementary Table 15. Statistical significance was assessed with a two-tailed Student t-test.

VP experiments in human progenitors. *In vitro* MPCs were subjected to VP (10 μM) or DMSO treatment for 16 h in duplicate on day 12 of the differentiation protocol. The drug was washed out with PBS and RNA was extracted with an RNeasy Mini Kit (Qiagen). Reverse transcription was carried out with 0.5 μg RNA using Superscript II (Invitrogen) and qPCR was carried out using SensiMix (Quantace). Oligonucleotide sequences are listed in Supplementary Table 15 and in ref. 17. qPCR reactions were normalized to *PBGD* and analysed with a two-tailed t-test.

Zebrafish experiments. Zebrafish embryos from the same cross were randomly selected for the control, morphant (*yap1*-MO), dominant negative (TEAD–EnR) and rescue (*yap1*-MO + *yap1* mRNA) conditions. Five nanolitres of 2 mM morpholino targeting a splice junction of *yap1* (*yap1*-MO, 5′-AGCAACATTAACAACTCAC TTTAGG-3′; previously reported[62]) were injected into the yolk of one- to two-cell stage zebrafish embryos. Morpholino activity was confirmed by qRT-PCR (oligonucleotides 5′-TGCCAGACTCATTCTTCACG-3′, 5′-TGGGAACCTTGC TTTACTGG-3′). For rescue experiments, *yap1* mRNA (50 pg) was co-injected with the morpholino. The mRNA of mouse *Tead2* fused with Engrailed repressor domain (TEAD2–EnR) was synthesized using an existing vector[52], and 200 pg was injected into the yolk of one- to two-cell stage zebrafish embryos. Embryos were fixed overnight at 4 °C in 4% paraformaldehyde in PBS. *In situ* hybridization for Sox9b (ref. 63) and insulin[64] was carried out as described[65] and revealed with NBT/BCIP substrate in 46–71 embryos per condition. After *in situ* hybridization, immunolocalization was carried out for some embryos using antibodies listed in Supplementary Table 23. The number of Pdx1$^+$/Sox9b$^+$ pancreatic progenitors was counted in each embryo using confocal microscopy, and differences between groups were analysed with a two-tailed Student t-test.

For transgenic analysis of wild-type and mutant CRMs, zebrafish embryos from the same cross were randomly selected. DNA fragments were recombined to an enhancer test vector that is sequentially composed of a Gateway cassette for insertion of CRMs, a *gata2* minimal promoter, an enhanced GFP reporter gene and a strong midbrain enhancer (z48) that works as an internal control for transgenesis. All these elements have been previously reported[66] and were assembled in a tol2 transposon[67]. Transgenesis was carried out as described[66] and embryos were grown to 24 and 48 hpf at 28 °C. GFP was documented using an epifluorescence stereomicroscope. Embryos positive for transposon integration were immunostained for simultaneous detection of Nkx6.1 plus either Pdx1 or insulin expression to identify pancreatic progenitors by confocal microscopy. Note that in zebrafish Nkx6.1 is expressed in pancreatic MPCs but not in endocrine cells, unlike mammalian embryos[68]. For each construct we counted embryos with GFP expression in Nkx6.1$^+$ pancreatic cells (Supplementary Table 21).

No statistical method was used to predetermine sample size. The investigators were not blinded to allocation during experiments or outcome assessment.

Reproducibility of experiments. Figure 5c shows representative data from one independent experiment with 110–140 zebrafish embryos per condition.

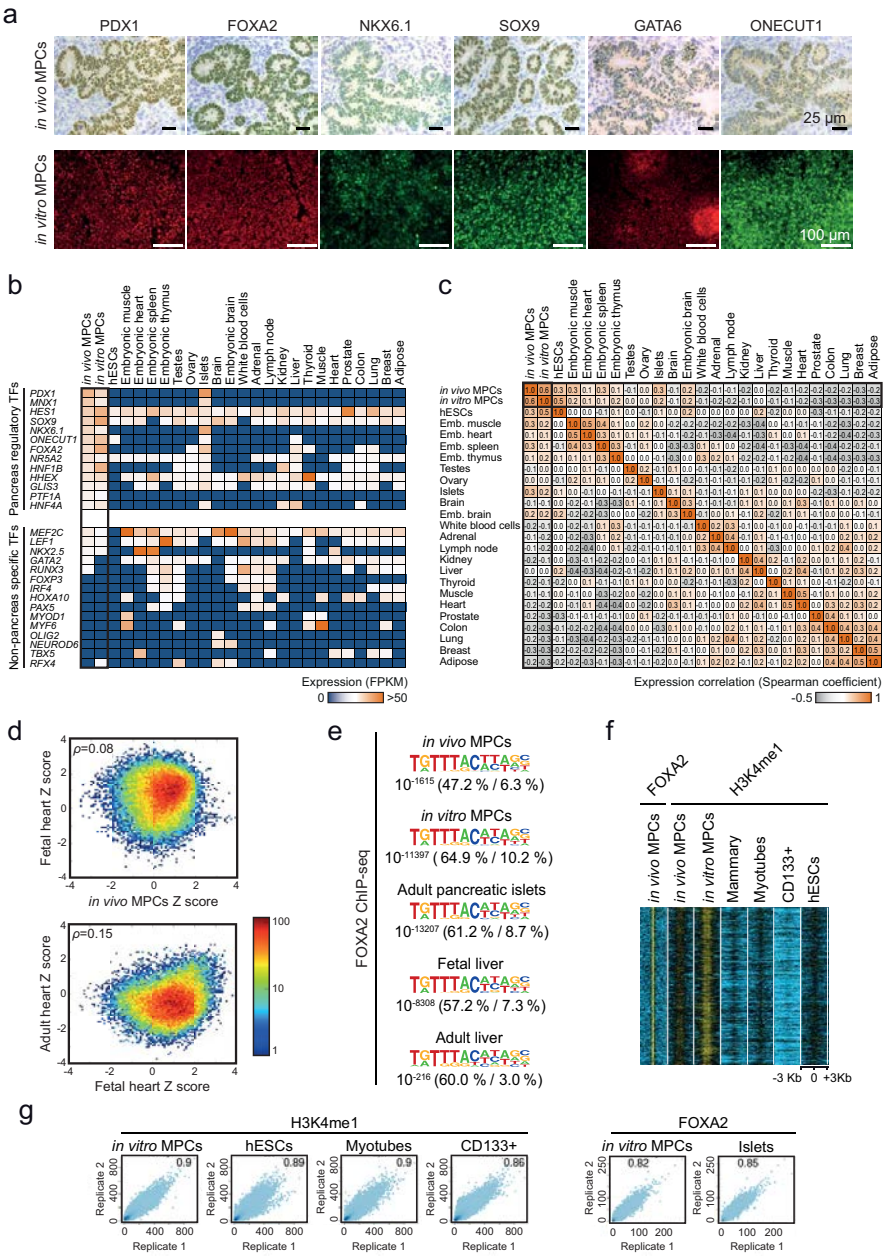
Figure 5d and Supplementary Fig. 5b show representative data from three or four independent experiments. Each independent experiment consisted of 50–120 injections. The exact number of zebrafish embryos analysed for each CRM is shown in Supplementary Table 21. Figures 6a–d and 7c,d,f and Supplementary Figs 1a, 6a–i and 7b show representative data from three independent experiments. Figure 7g shows representative data from one independent experiment with 9–12 zebrafish embryos per condition. Supplementary Fig. 4a shows representative data from six independent experiments. Three immunostainings were carried out independently for two human embryos (CS18 and CS19). Supplementary Fig. 7a shows representative data from one independent experiment with 46–71 zebrafish embryos per condition.

Accession numbers. Primary data sets generated here are available at ArrayExpress under accession numbers E-MTAB-1990 and E-MTAB-3061. Referenced data sets are listed in Supplementary Tables 1 and 2.

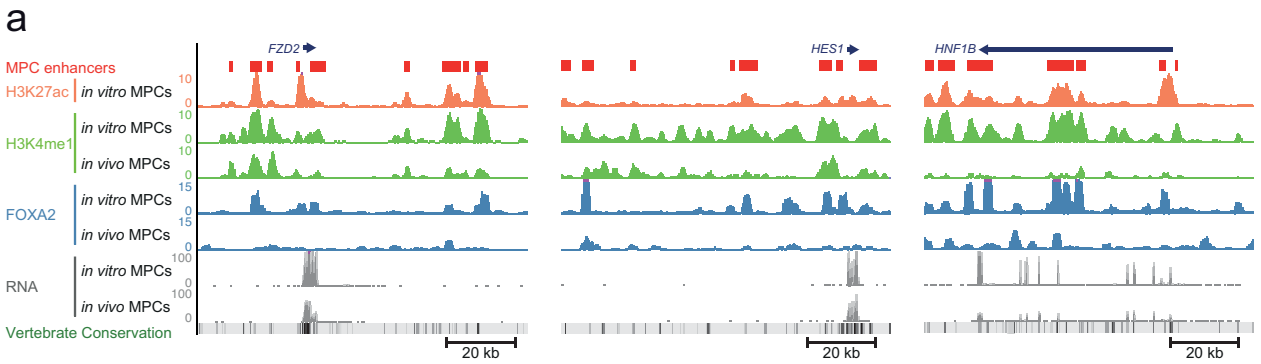
43. O'Rahilly, R., Müller, F., Hutchins, G. M. & Moore, G. W. Computer ranking of the sequence of appearance of 73 features of the brain and related structures in staged human embryos during the sixth week of development. *Am. J. Anat.* **180,** 69–86 (1987).
44. Maestro, M. A. *et al.* Hnf6 and Tcf2 (MODY5) are linked in a gene network operating in a precursor cell domain of the embryonic pancreas. *Hum. Mol. Genet.* **12,** 3307–3314 (2003).
45. Piper, K. *et al.* β cell differentiation during early human pancreas development. *J. Endocrinol.* **181,** 11–23 (2004).
46. Petzold, K. M. & Spagnoli, F. M. A system for *ex vivo* culturing of embryonic pancreas. *J. Vis. Exp.* **66,** e3979 (2012).
47. Gaulton, K. J. *et al.* A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42,** 255–259 (2010).
48. Luco, R. F., Maestro, M. A., Sadoni, N., Zink, D. & Ferrer, J. Targeted deficiency of the transcriptional activator Hnf1α alters subnuclear positioning of its genomic targets. *PLoS Genet.* **4,** e1000079 (2008).
49. 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).
50. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25,** 1105–1111 (2009).
51. 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).
52. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).
53. McLean, C. Y., Bristor, D., Hiller, M. & Clarke, S. L. GREAT improves functional interpretation of cis-regulatory regions. *Nature* **28,** 495–501 (2010).
54. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6,** e21800 (2011).
55. de Hoon, M. J. L., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics* **20,** 1453–1454 (2004).
56. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics* **20,** 3246–3248 (2004).
57. 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).
58. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol.* **8,** R24 (2007).
59. Whitlock, M. C. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J. Evol. Biol.* **18,** 1368–1373 (2005).
60. Derrien, T. *et al.* Fast computation and applications of genome mappability. *PLoS ONE* **7,** e30377 (2012).
61. Esni, F., Miyamoto, Y., Leach, S. D. & Ghosh, B. Primary explant cultures of adult and embryonic pancreas. *Methods Mol. Med.* **103,** 259–271 (2005).
62. Skouloudaki, K. *et al.* Scribble participates in Hippo signaling and is required for normal zebrafish pronephros development. *Proc. Natl Acad. Sci. USA* **106,** 8579–8584 (2009).
63. Chiang, E. F. *et al.* Two sox9 genes on duplicated zebrafish chromosomes: expression of similar transcription activators in distinct sites. *Dev. Biol.* **231,** 149–163 (2001).
64. Argenton, F., Zecchin, E. & Bortolussi, M. Early appearance of pancreatic hormone-expressing cells in the zebrafish embryo. *Mech. Dev.* **87,** 217–221 (1999).
65. Jowett, T. & Lettice, L. Whole-mount *in situ* hybridizations on zebrafish embryos using a mixture of digoxigenin- and fluorescein-labelled probes. *Trends Genet.* **10,** 73–74 (1994).
66. Bessa, J. *et al.* Zebrafish enhancer detection (ZED) vector: A new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev. Dynam.* **238,** 2409–2417 (2009).
67. Kawakami, K., Shima, A. & Kawakami, N. Identification of a functional transposase of the Tol2 element, an Ac-like element from the Japanese medaka fish, and its transposition in the zebrafish germ lineage. *Proc. Natl Acad. Sci. USA* **97,** 11403–11408 (2000).
68. Binot, A-C. *et al.* Nkx6.1 and nkx6.2 regulate α- and β-cell formation in zebrafish by acting on pancreatic endocrine progenitor cells. *Dev. Biol.* **340,** 397–407 (2010).

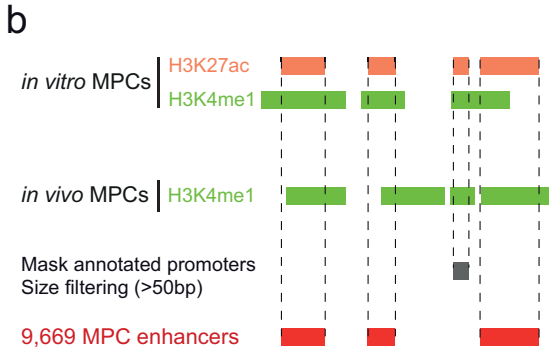
nature cell biology

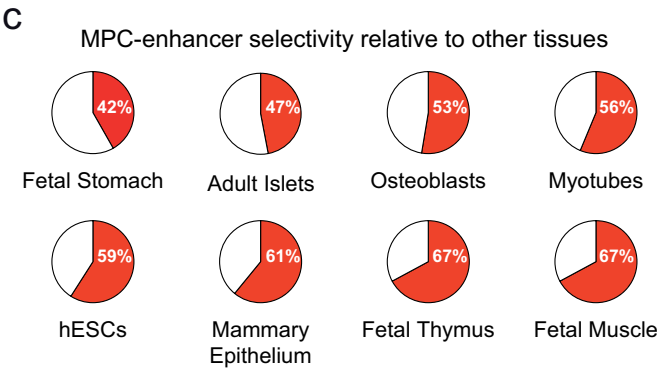
DOI: 10.1038/ncb3160

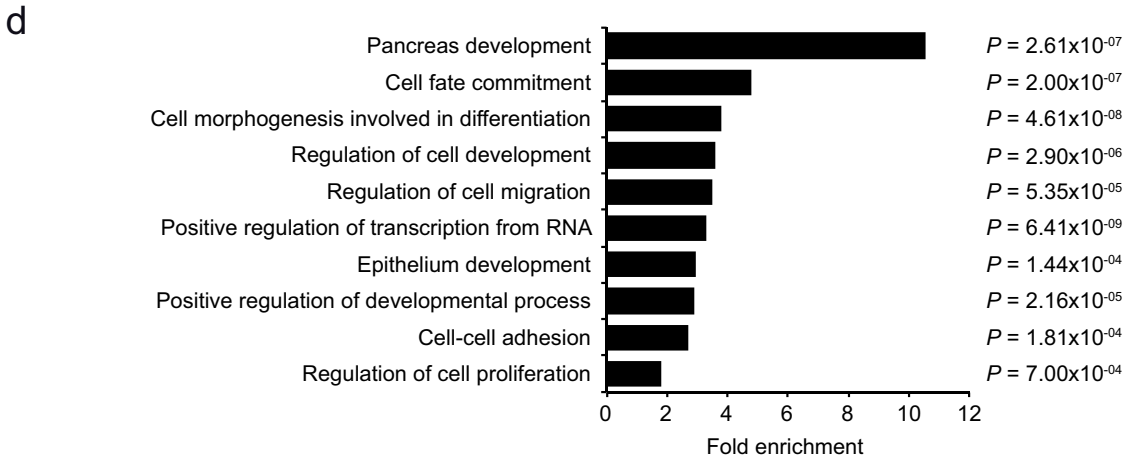


Supplementary Figure 1 Human *in vitro* MPCs recapitulate key features of *in vivo* MPCs. **(a)** Immunohistochemistry analysis of *in vivo* MPCs from Carnegie stage 16-18 human embryonic pancreas, and immunofluorescence analysis of *in vitro* MPCs show expression of stage-delimiting MPC TFs in both sources of MPCs. **(b)** Heatmap showing RNA-seq FPKM signal in MPCs and 23 control tissues for TFs that are enriched in pancreatic MPCs and for a similar number of known lineage-specific non-pancreatic TFs. **(c)** Expression correlation matrix showing Spearman coefficient values for transcript levels from *in vivo* and *in vitro* MPCs vs. 23 control tissues. **(d)** Z score correlation density plots. Comparisons of *in vivo* MPCs with an unrelated tissue (fetal heart, left panel), or between tissues from the same lineage, but different stages (adult and fetal heart, right panel), do not show high correlation, in contrast with data presented in **Figure 1c** that shows highly correlated Z-scores for *in vivo* and *in vitro* MPCs. Spearman coefficient values are shown for each comparison. Color scale depicts number of transcripts. **(e)** Motif discovery in different FOXA2 ChIP-seq datasets, shows a similar binding motif for this TF in all samples. *P* values and percentages of bound versus background regions are indicated below each motif logo. **(f)** Regions enriched in FOXA2 and H3K4me1 in chromatin from *in vivo* MPCs also show H3K4me1-enrichment in *in vitro* MPCs, but not in control samples (mammary epithelial cells, myotubes, CD133+ umbilical cord blood and hESCs). The heatmap shows FOXA2 and H3K4me1 signal centered on these regions (see Methods for details). Note that even though the regions were pre-selected from *in vivo* MPC data, H3K4me1-enrichment is stronger in chromatin from *in vitro* MPCs, reflecting the larger number of cells used for ChIP-seq. **(g)** H3K4me1 and FOXA2 signals in the whole genome were binned in 5 Kb for H3K4me1 and 1 Kb for FOXA2. These signals were highly correlated in biological replicates (R >0.8).

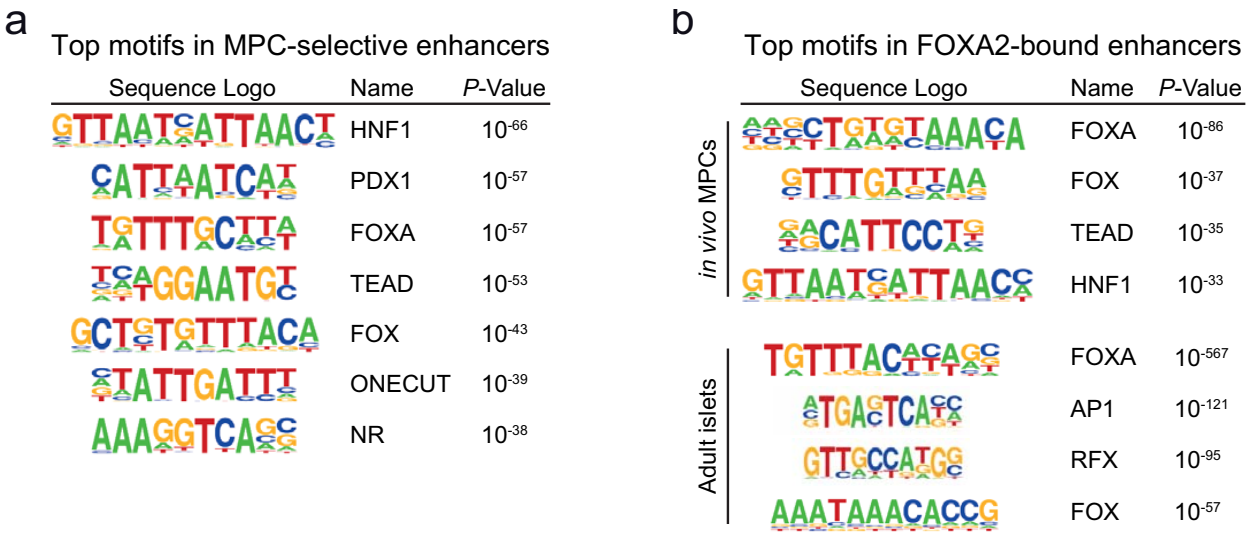




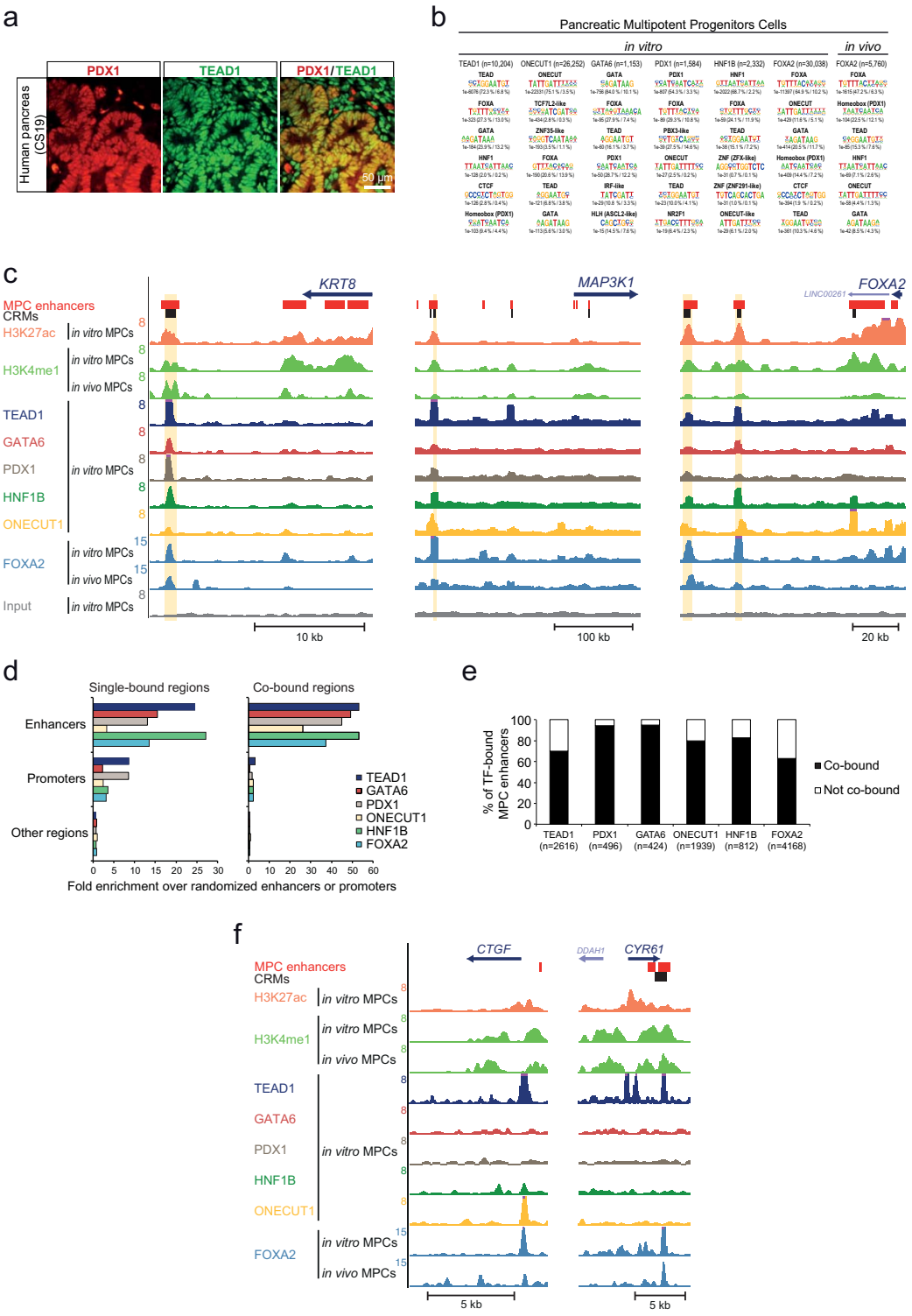




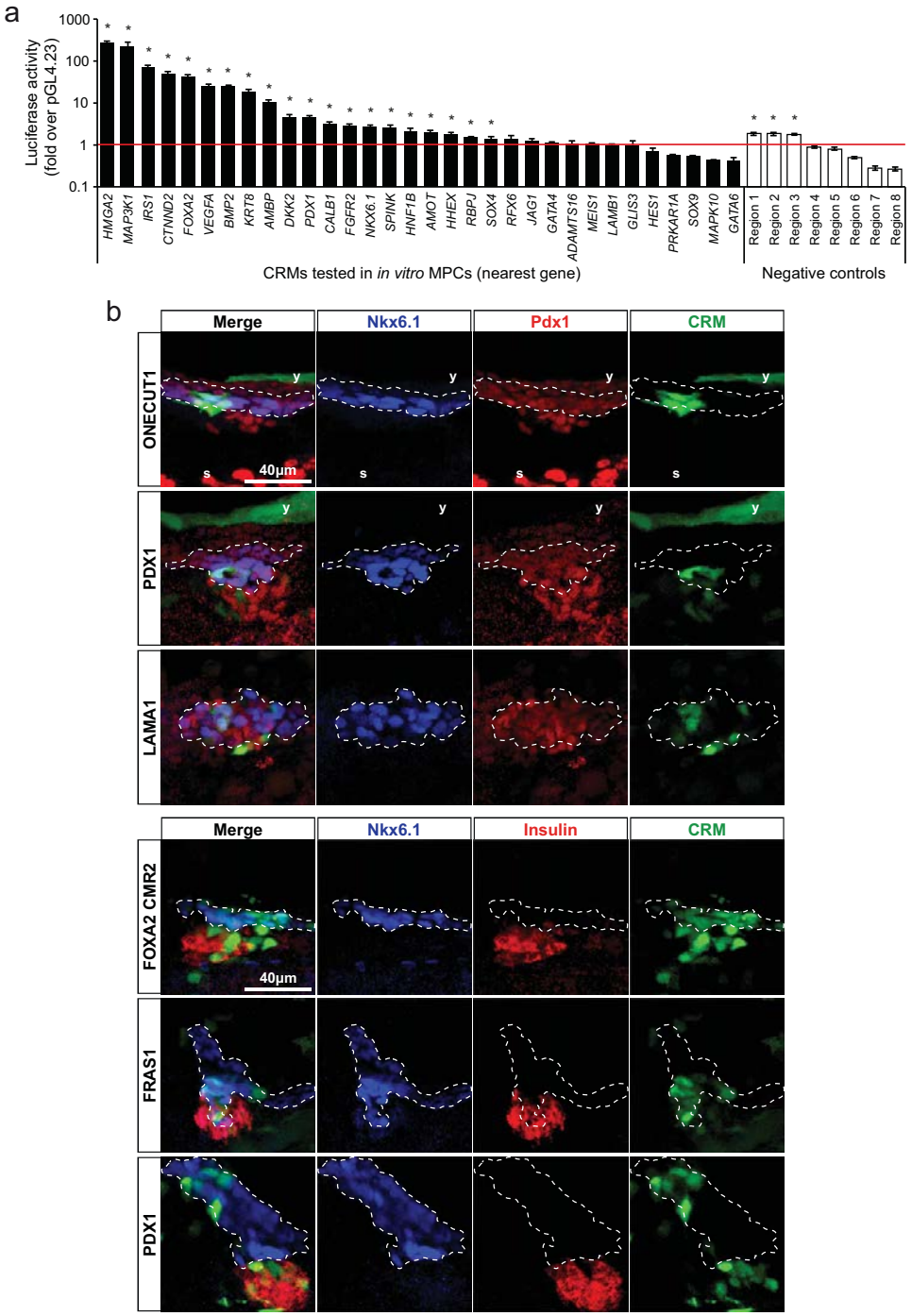
Supplementary Figure 2 Human pancreatic MPC enhancers. **(a)** Examples showing how *in vitro* MPCs recapitulate the epigenomic landscape of *in vivo* MPCs. *HNF1B* encodes a TF that important for pancreas development, *FZD2* is a non-canonical WNT signaling component, and *HES1* is a transcriptional repressor that controls growth and differentiation of pancreatic MPCs. **(b)** Enhancers were defined as H3K27ac islands in the *in vitro* MPCs that overlapped H3K4me1 islands in both *in vitro* and *in vivo* MPCs. We discarded regions overlapping promoters (1 Kb upstream and 2 Kb downstream of RefSeq TSS) and any regions smaller than 50 bp. This revealed 9,669 MPC enhancers. **(c)** MPC enhancers are tissue- and stage-selective. Enhancers were defined for 8 tissues in a similar manner to MPCs (Supplementary Table 8). Each pie chart shows in red the proportion of MPC enhancers that are inactive in each tissue. We defined MPC-selective enhancers as those that were inactive in at least 6 out of 7 non-pancreatic tissues. (d) Enriched annotated functions among genes that are associated with three or more MPC-selective enhancers. The graph shows fold enrichment values and P values calculated with GREAT[45].



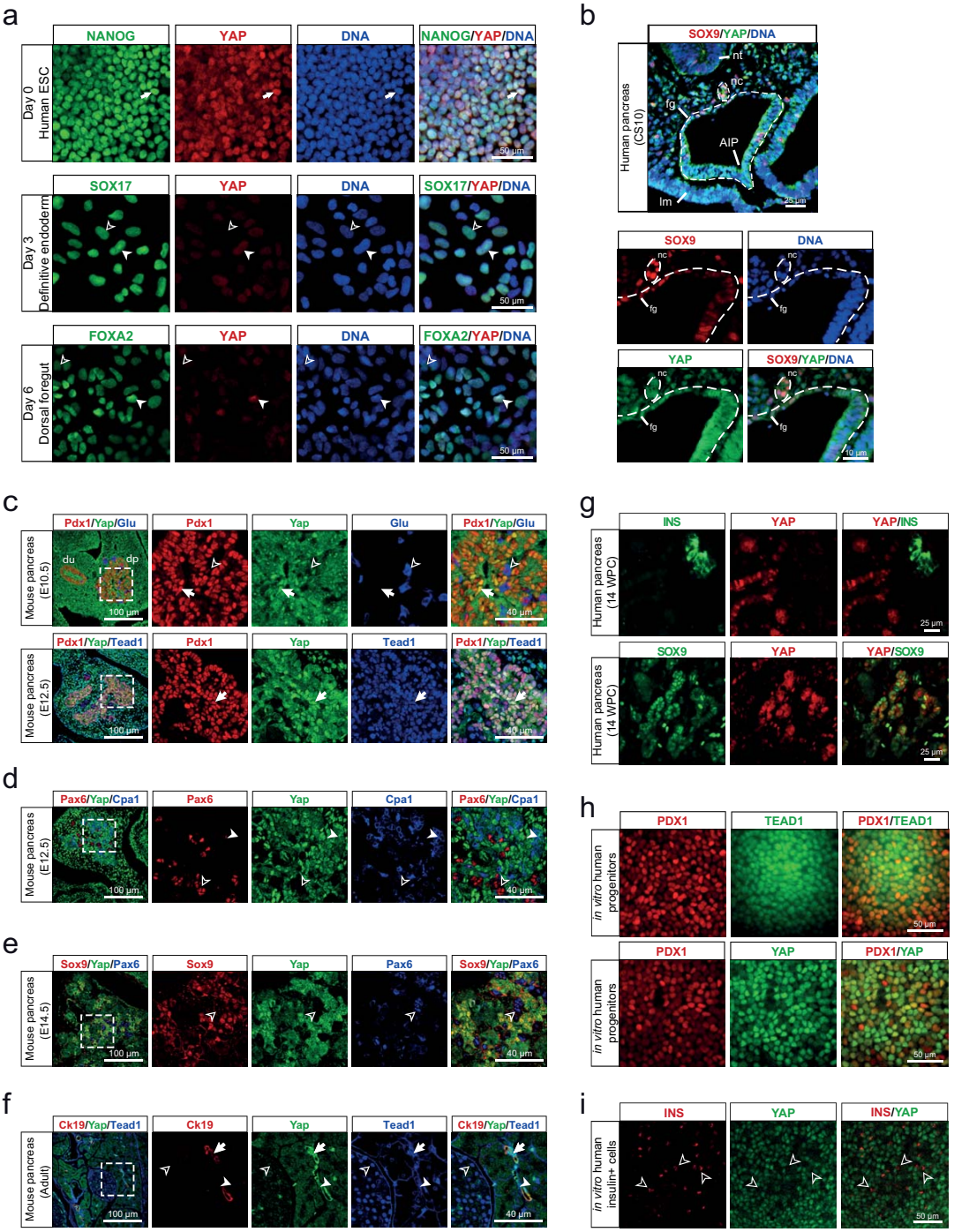
Supplementary Figure 3 MPC enhancers are enriched in TEAD motifs. (**a**) *De novo* motif search in MPC-selective enhancers revealed strong enrichment for TEAD recognition sequences, similarly to what we observed for the whole set of MPC enhancers. Other enriched matrices match binding sites of known pancreatic regulators. (**b**) TEAD motifs are highly enriched in enhancers bound by FOXA2 in both *in vivo* and *in vitro* MPCs, but not in enhancers bound by FOXA2 in adult pancreatic islets.



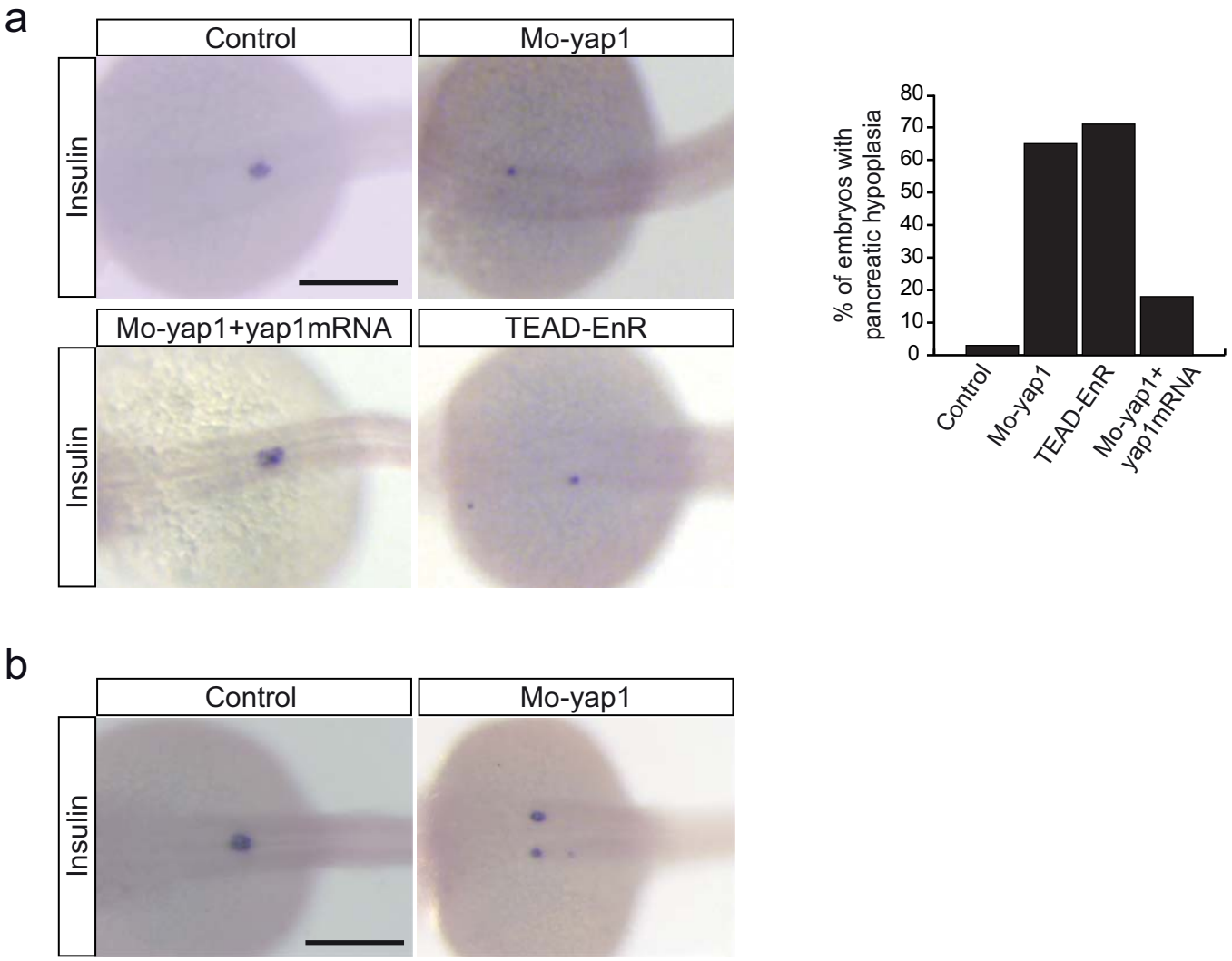
Supplementary Figure 4 TEAD1 is a core component of the combination of TFs that bind to MPC enhancers. **(a)** TEAD1 is expressed in PDX1$^+$ *in vivo* MPCs from human pancreas of Carnegie stage 19. **(b)** *De novo* analysis of over-represented sequence motifs for regions bound by each of the TFs examined in this study. As expected, each dataset showed a top scoring motif that coincided with the immunoprecipitated TF. A marked co-enrichment of many known pancreatic TF and TEAD1 motifs was observed. *P* values and percentages of bound vs. background regions are indicated below each motif logo. **(c)** Examples showing CRMs bound by multiple TFs (regions highlighted in yellow). **(d)** TF binding and co-binding preferentially occurs at MPC enhancers. Note that TEAD1 binding and co-binding enrichment is comparable to the enrichments found for other TFs. Binding fold enrichment was calculated over 1,000 permutations of enhancer or promoter genomic positions. **(e)** MPC enhancers bound by any of the pancreatic TFs or TEAD1 show a high degree of co-binding with other TFs. Total number of peaks for each TF is shown below the corresponding column. **(f)** Representative examples of known Hippo pathway targets showing TEAD1 binding at their promoter regions.



Supplementary Figure 5 Functional validation of CRMs as transcriptional enhancers in pancreatic MPCs. **(a)** Functional validation of CRMs as transcriptional enhancers in human progenitors. Thirty-two CRMs and 8 negative control regions were cloned into the pGL4.23 vector and tested in reporter assays. Reporter activity was compared to empty pGL4.23. * Two-tailed Student's t test $P<0.05$ (P values fully listed in Supplementary Table 22). n=3-4 independent transfections per enhancer, 8 of 32 constructs were tested in an independent experiment that yielded comparable results. **(b)** Functional validation of unannotated CRMs identified in the vicinity of MPC-enriched genes in 24 hpf zebrafish embryos. Eight out of 10 TEAD1-bound CRMs yielded activation of a minimal promoter driving GFP (see also Fig. 5c-e and Supplementary Table 21). Pancreatic progenitors were identified by co-staining Nkx6.1 and either Pdx1 or insulin. Note that in zebrafish Nkx6.1 is expressed in early pancreatic progenitors but not in endocrine cells, unlike in mammalian embryos, which show Nkx6.1 expression in both cellular compartments[47]. The percentage of transgenics showing activation of GFP in the pancreatic domain for each CRM and in control injections is presented in the bar plot in Figure 5e. Dashed lines demarcate the pancreatic progenitor domain (Nkx6.1$^+$ cells). y: yolk autofluorescence, s: somites showing crossreactivity with anti-Pdx1 serum.



Supplementary Figure 6 Developmental expression of YAP. **(a)** Immunofluorescence images of hESCs and different stages of differentiation show that YAP is strongly expressed in the nuclei of hESCs (white arrows), whereas a marked decrease in YAP immunoreactivity is observed in definitive endoderm and dorsal foregut stages (days 3 and 6)(white arrowheads). In days 3 and 6, YAP is not detected in a subset of SOX17^{+} and FOXA2^{+} cells, respectively (hollow arrowheads). **(b)** YAP is present in the nuclei of dorsal foregut endoderm cells of human Carnegie Stage 10 embryos. AIP: anterior intestinal portal, fg: foregut, lm: lateral mesoderm, nc: notochord, nt: neural tube. **(c)** Immunofluorescence images of mouse E10.5 and E12.5 embryonic pancreas show Tead1 and Yap expression in most nuclei of Pdx1+ MPCs (white arrows) and in the surrounding mesenchyme. Yap expression is absent in glucagon-expressing endocrine cells (hollow arrowheads). The squares in the leftmost panels depict areas shown at higher power in other panels. du: duodenum, dp: dorsal pancreas. **(d,e)** Yap is broadly expressed in the nuclei of pancreatic mesenchyme and epithelium from E12.5 and E14.5 mouse embryonic pancreas, yet shows cytoplasmic localization in Cpa1+ progenitor cells (white arrowheads) and is undetectable in early Pax6+ endocrine cells (hollow arrowheads). **(f)** In the adult mouse pancreas Yap is present in nuclei from ducts (white arrows), and not in endocrine (hollow arrowheads) or acinar cells (white arrowheads). (**g**) YAP is expressed in nuclei of SOX9+ epithelial cells but absent in insulin-expressing endocrine cells from 14 weeks post-conception (WPC) human pancreas. (**h**) PDX1 co-stains with YAP and TEAD1 in the nucleus of *in vitro* MPCs. **(i)** Nuclear YAP is not detected in differentiated insulin-expressing cells derived from hESCs (hollow arrowheads).



Supplementary Figure 7 Knockdown of Yap1 or dominant inhibition of Tead reduces pancreas size in zebrafish. **(a)** Injection of a morpholino targeting *yap1* (Mo-yap1) or mRNA encoding a TEAD protein fused to the transcriptional repressor domain of Engrailed (TEAD-EnR) decreased the number of insulin expressing cells detected by *in situ* hybridization. This phenotype was rescued by co-injection of Mo-yap1 with an *in vitro* synthesized *yap1* mRNA that is not sensitive to morpholino inhibition. The percentage of embryos from each condition showing reduced insulin-positive cells was quantified as an indication of pancreatic hypoplasia, and displayed in the graph shown on the side (n=46-71 embryos per condition). Scale bar = 0.25 mm. **(b)** Mo-yap1 increased ectopic expression of pancreatic markers. The panels show insulin *in situ* hybridization in control and Morpholino-treated 24 hpf zebrafish embryos. Scale bar = 0.25 mm.

Supplementary Tables

Supplementary Table 1 Alignment details for RNA-seq data used in this study. Read count, accession numbers and corresponding references are provided for all RNA-seq datasets used in this study.

Supplementary Table 2 Alignment and peak calling details for ChIP-seq data used in this study. Read count, MACS alignment details and accession numbers and corresponding references are provided for all ChIP-seq datasets used in this study.

Supplementary Table 3 Functional annotation of transcripts enriched in pancreatic MPCs. This table depicts the functional annotations of transcripts selectively enriched in MPCs (CV and Z score > 1) with DAVID[57], using Gene Ontology (GO) biological process (FAT), Pathways (KEGG, Panther) and annotation clustering. The analysis was carried out independently for *in vivo* and *in vitro* MPCs (red titles), which retrieved similar categories. To highlight this result we show the most enriched clusters in Figure 1c.

Supplementary Table 4 Curated list of non-canonical WNT pathway mouse genes expressed in pancreatic progenitors. Previous studies carried out in mouse embryos show an enrichment of non-canonical WNT signaling genes in pancreatic MPCs[21,22]. In this table we provide a list of non-canonical WNT mouse genes referenced to by Cortijo et al.[22] and Rodríguez-Seguel et al.[21] together with transcriptional analysis data obtained in our study for their human ortholog genes. Expression and enrichment Z scores are shown for both *in vivo* and *in vitro* MPCs. Coefficients of variation (CV) were calculated using RNA-seq datasets of 25 different cell/tissue samples, including *in vivo* and *in vitro* MPCs.

Supplementary Table 5 List of core MPC-specific genes. To define a core set of 500 MPC-specific genes, we first selected transcripts with CV and Z score > 1 and expression > 0.3 FPKM in both *in vitro* and *in vivo* MPCs. Transcripts were then ranked by Z score in *in vivo* MPCs and the top 500 genes were selected.

Supplementary Table 6 Genomic coordinates of the 9,669 MPC enhancers identified in this study and associated genes. Enhancers were defined as H3K27ac islands in the *in vitro* MPCs that overlapped H3K4me1 islands in both *in vitro* and *in vivo* MPCs. We discarded regions overlapping annotated promoters or <50 bp. The 9,669 MPC enhancers were then associated to genes using GREAT-v2.0.2[45]. Genomic coordinates shown are in hg18.

Supplementary Table 7 Genomic coordinates of all MPC-selective enhancers and associated genes. The MPC-selective enhancers (Fig 2d) were associated to genes using GREAT-v2.0.2[45]. Genomic coordinates shown are in hg18.

Supplementary Table 8 Number of pancreatic and non-pancreatic enhancers. Enhancers were defined in the same manner for MPCs and 8 control tissues based on H3K27ac and H3K4me1 enrichment. Annotated promoter regions were discarded.

Supplementary Table 9 Functional annotation of genes associated with 3 or more MPC-selective enhancers. Genes associated with 3 or more MPC-selective enhancers were annotated with GREAT-v2.0.2[45]. Given the extension of this list and the redundancy of some terms, we further processed the data by clustering similar functional annotation terms with REVIGO[46] (Supplementary Fig. 2d). Rows highlighted in light pink correspond to the most enriched terms for each REVIGO cluster.

Supplementary Table 10 Transcription factor motifs enriched in MPC enhancers. *De novo* motif discovery in MPC enhancers was performed with HOMER[44]. We searched for either short (length=6,8,10,12) or long (length=14,16,18,20) motifs as described previously[25], retaining non-redundant matrices (Pearson correlation <0.65) with $P<10^{-50}$. Motifs were annotated using HOMER[44], TOMTOM[60] and manual comparisons.

Supplementary Table 11 Transcription factor motifs enriched in MPC-selective enhancers. *De novo* motif discovery in MPC-selective enhancers was performed with HOMER[44]. We searched for either short (length=6,8,10,12) or long (length=14,16,18,20) motifs as described previously[25], retaining non-redundant matrices (Pearson correlation <0.65) with $P<10^{-50}$. Motifs were annotated using HOMER[44], TOMTOM[60] and manual comparisons.

Supplementary Table 12 Top 50 most enriched combinations of 3 motifs in MPC enhancers vs. other tissues. All possible combinations of 3 motifs from the 23 enriched motifs (Supplementary Table 10) contained within 500 bp regions were computed in MPC enhancers vs. enhancers from 8 other tissues. We calculated eight MPC vs. control tissue fold-enrichment and *P* values (Chi-squared test), and then combined them in a unique *P* value for each motif combination with a Z-weighted method[61].

Supplementary Table 13 Curated list of mediators of Hippo signaling along with TEAD1 occupancy and expression in MPCs. Known Hippo pathway or transcriptional target genes were selected from the literature. For this list only, we associated the TEAD1 ChIP-seq peaks, including peaks in promoters, with their nearest gene using GREAT-v2.0.2[45].

Supplementary Table 14 Genomic coordinates of TEAD1-bound MPC enhancers and associated genes. The TEAD1-bound MPC enhancers (Fig 4d) were associated to genes using GREAT-v2.0.2[45]. Genomic coordinates shown are in hg18.

Supplementary Table 15 Oligonucleotides used in this study. Oligonucleotide sequences are listed according to their application in the study (red titles).

Supplementary Table 16 Genomic coordinates of CRMs and associated genes. CRMs were associated to genes using GREAT-v2.0.2[45]. Genomic coordinates shown are in hg18.

Supplementary Table 17 Functional annotation of genes associated with at least one CRM. Genes associated with at least one CRM were annotated with GREAT-v2.0.2[45]. Given the extension of this list and the redundancy of some terms, we further processed the data by clustering similar functional annotation terms with REVIGO[46]. The REVIGO clustering results are shown in Figure 4g.

Supplementary Table 18 Genomic coordinates of CRM clusters and associated genes. Clusters of CRMs were defined as described[25], essentially as any group of ≥3 CRMs in which all adjacent CRMs were separated by less than the 25th-percentile of chromosome-specific randomized distances. CRMs clusters were then associated to genes using GREAT-v2.0.2[45]. Genomic coordinates shown are in hg18.

Supplementary Table 19 Functional annotation of genes associated with at least one cluster of CRMs. Genes associated with clusters of CRMs were annotated with GREAT-v2.0.2[45]. Note that noncanonical Wnt signaling pathway genes are highly enriched near clusters of CRMs (binomial raw *P* value = 1.18×10^{-9}, highlighted in red), which is not observed when all CRMs are analyzed together (Supplementary Table 17).

Supplementary Table 20 Genomic coordinates of TEAD1-bound CRMs and associated genes. The TEAD1-bound CRMs were associated to genes using GREAT-v2.0.2[45]. Genomic coordinates shown are in hg18.

Supplementary Table 21. Quantifications of GFP colocalization in NKX6.1^{+} pancreatic cells in zebrafish transgenics. This table contains the entire quantification data correspondent to Figure 5d,e and Supplementary Figure 5. In order to detect GFP co-localization in zebrafish pancreatic MPCs, the pancreatic progenitors domain was revealed by co-expression of Pdx1 and NKX6.1 or Insulin and Nkx6.1.

Supplementary Table 22 Statistical significance results for Supplementary Figure 5a, Figure 5b and Figure 7. This table contains all the *P* values calculated using two-tailed Student's t test with the data shown in Supplementary Figure 5a, Figure 5b and Figure 7b.

Supplementary Table 23 Details of the antibodies used for immunolocalization in this study. This table contains information per specimen on the species, dilution and supplier's details for all antibodies used in immunolocalization studies.