

Published in final edited form as:

*Nat Cell Biol.* 2023 March ; 25(3): 481–492. doi:10.1038/s41556-022-01075-8.

## Expansion of Ventral Foregut is linked to changes in the Enhancer Landscape for Organ Specific Differentiation

Yan Fung Wong<sup>#1</sup>, Yatendra Kumar<sup>#2</sup>, Martin Proks<sup>1</sup>, Jose Alejandro Romero Herrera<sup>1,3</sup>, Michaela Mrugala Rothová<sup>1</sup>, Rita S. Monteiro<sup>1</sup>, Sara Pozzi<sup>1</sup>, Rachel E. Jennings<sup>4</sup>, Neil A. Hanley<sup>4</sup>, Wendy A. Bickmore<sup>2,\*</sup>, Joshua M. Brickman<sup>1,\*</sup>

<sup>1</sup>Novo Nordisk Foundation Center for Stem Cell Medicine (reNEW), University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen N, Denmark

<sup>2</sup>MRC Human Genetics Unit, Institute of Genetics and Cancer, University of Edinburgh, Crewe Road, Edinburgh EH4 2XU, UK

<sup>3</sup>Center for Health Data Science, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen N, Denmark

<sup>4</sup>Faculty of Biology, Medicine & Health, Manchester Academic Health Science Centre, University of Manchester, Oxford Road, Manchester, M13 9PT, UK

# These authors contributed equally to this work.

### Abstract

Cell proliferation is fundamental for almost all stages of development and differentiation that require an increase cell number. Although cell cycle phase has been associated with differentiation, the actual process of proliferation has not been considered as having a specific role. Here we exploit human embryonic stem cell derived endodermal progenitors that we find are an *in vitro* model for the ventral foregut. These cells exhibit expansion dependent increases in differentiation efficiency to pancreatic progenitors that are linked to organ-specific enhancer priming at the level of chromatin accessibility and the decommissioning of lineage inappropriate enhancers. Our findings suggest that cell proliferation in embryonic development is about more than tissue expansion, it is required to ensure equilibration of gene regulatory networks allowing cells to become primed for future differentiation. Expansion of lineage specific intermediates may therefore be an important step in achieving high fidelity *in vitro* differentiation.

---

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: <https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms>

\*Correspondence: joshua.brickman@sund.ku.dk and wendy.bickmore@ed.ac.uk.

#### Author contributions

Y.F.W., W.A.B., and J.M.B. conceived of the project. Y.F.W. and Y.K. conducted experiments. W.A.B. and J.M.B. designed the experiments and obtained funding for the study. Y.F.W., Y.K., M.P., J.A.R.H., M.M.R., R.S.M., and S.P. performed data analysis. M.M.R. conducted the single-cell sequencing experiment. N.A.H. and R.E.J. provided insight into organ specific enhancer regulation and H3K27ac ChIP-seq data from human embryos. Y.F.W., Y.K., W.A.B., and J.M.B. wrote the manuscript.

#### Competing Interests Statement

The authors declare no competing interests.

## Introduction

The regulation of gene expression during differentiation is considered a linear process involving the action of signalling and transcription factors (TFs). Cell proliferation is regarded as peripheral to differentiation, although it has a clear function in the selection of specific cell types. While cell cycle phase has been linked to differentiation<sup>1,2</sup>, here we explore the notion that differentiation requires progenitor proliferation itself to enhance the processing of lineage promoting information.

The visceral organs are formed during embryonic development from the endoderm germ layer<sup>3</sup>. These cells are initially specified during gastrulation and undergo extensive proliferation as they prepare to differentiate into distinct organ primordia<sup>4</sup>. In particular, the liver and pancreas are derived from the anterior definitive endoderm (ADE). ADE is formed as a result of the anterior migration of cells from the anterior region of the primitive streak at the beginning of gastrulation. The anterior-most DE will then migrate ventrally to form the ventral foregut, containing a bipotent precursor of liver and ventral pancreas<sup>5,6</sup>, a population that has recently been shown to expand and retain potency for both lineages *in vivo* over a period of several days of mouse development<sup>7</sup>.

Pluripotent embryonic stem cells (ESCs) can be differentiated *in vitro* to form all embryonic germ layers including endoderm<sup>8,9</sup>. As a result, directed linear ESC differentiation is used to produce organ-specific cell types such as pancreatic beta cells<sup>10–12</sup> and hepatocytes<sup>13,14</sup>. An alternative to directed differentiation is the use of ESC derived expandable endodermal progenitors (EPs) as a staging platform for further differentiation<sup>15–17</sup> and the expansion of endodermal cells from human ESCs (hESCs) promotes the generation of more mature pancreatic beta cells<sup>15</sup>.

Here we find that *in vivo* identity of human EP cells is ventral foregut and that continued proliferation of these cells results in lineage priming that is correlated to organ-specific enhancer accessibility. Lineage priming is not accompanied by large changes in transcription of organ-specific genes, but instead prepares appropriate enhancers for their activation and decommissions enhancers normally present in other lineages. Our findings suggest that the extensive cell proliferation that characterizes normal embryonic development is not merely required for tissue expansion, but ensures equilibration of gene regulatory networks for future high-fidelity differentiation.

## Results

### Expanding endoderm progenitors mimic ventral foregut *in vitro*

To characterize the impact of expansion on endodermal differentiation, we focused on 3D human EP culture<sup>15</sup>. This protocol expands endoderm in the presence of FGF2, BMP4, VEGF, EGF<sup>15,17</sup>, cytokines known to act in the ventral foregut region. We quantitated gene expression during expansion by single cell RNA-seq and found transient ADE cells comprised two sub-populations (ADE.1 and ADE.2) while EP culture was homogenous (Extended Data Fig. 1a, left). In human development, ventral foregut endoderm has been described at Carnegie Stages 8-9<sup>18</sup> and we compared hESC derived endoderm to single

cell RNA from these stages of human embryos with our Cluster Alignment Tool (CAT)<sup>19</sup>. For this analysis we used a recently published dataset containing human embryonic foregut (hFG.1-4), the lip formed from ventral foregut - referred to as the lip of the anterior intestinal portal (hAL), midgut (hMG.1-3), and hindgut (hHG.1-2)<sup>20</sup> (Extended Data Fig. 1a, right). We found that ADE aligns to the foregut hFG.2 and midgut hMG1 clusters (Fig. 1a). In contrast, EP cells align with hAL and hMG1 - a population of MG cells located adjacent to the hAL<sup>20</sup>. As EP cells align to both these clusters, we assessed gene expression specifically enriched in RNA-seq from H9-derived EP cells (Extended Data Fig. 1b), and asked if this set contained genes that differentially expressed when the hAL and hMG.1 clusters were compared. With a few exceptions, genes expressed at higher levels in hAL were also elevated in EP cells (Fig. 1b). The hAL or ventral foregut identity of EP cells was confirmed by immunohistochemistry of the hAL markers HHEX<sup>20</sup> and TBX3<sup>15</sup> (Extended Data Fig. 1c).

As murine ventral foregut endoderm is actively cycling<sup>21</sup>, we measured the proliferation rate of hEP cultures and found it increased with time in culture (p6, p8, p12, and p15) (Fig. 1c). In mouse, HHEX is known to support ventral foregut expansion and morphogenesis<sup>21</sup>. To further confirm the identity of hEP we knocked down HHEX by shRNA and observed a reduction in growth without induction of apoptosis (Extended Data Fig. 1d, e). We measured actively proliferating cells in ADE, EP cells and HHEX knock down EP cells by 5-ethynyl-2'-deoxyuridine (EdU) labelling followed by cell-cycle analysis based on DAPI staining (Extended Data Fig. 1f). The percentage of S-phase cells increased with expansion in a HHEX dependent fashion, while the fraction in G2M was reduced (Fig. 1d, e). Based on expression of ventral foregut markers, the cytokines used in these cultures, and the function of HHEX in proliferation, we conclude hEP cells are an *in vitro* model for human ventral foregut and refer to them hereafter as ventral foregut progenitor cells (VFGs).

To probe VFG differentiation efficiency, we established VFG cultures from a hESC line containing a pancreatic reporter (PDX1-eGFP)<sup>22</sup> and determined the minimal cytokine set required to transform VFG spheres into proliferating pancreatic spheroids or hepatic organoids (Extended Data Fig. 2a). Removal of BMP4 from VFG culture resulted in negligible PDX1 reporter expression (< 2% GFP<sup>+</sup>), no PDX1 protein, and no dramatic transcriptional change at single-cell level (Extended Data Fig. 2, b-d). Subsequent addition of FGF7 and 10, and to a lesser extent FGF2, significantly stimulated PDX1-eGFP expression and induced robust transcriptional change (Extended Data Fig. 2, d-f). In response to initial cytokine treatment, we could separate PDX1<sup>+</sup> and PDX1<sup>-</sup> cells, and expand PDX1<sup>+</sup> cells as pancreatic spheroids, or PDX1<sup>-</sup> cells as hepatic organoids (Fig. 1f and Extended Data Fig. 2, g-i) in defined media<sup>23,24</sup>. These observations indicate that human VFG culture is poised to generate expanding hepatic and pancreatic endoderm.

### Expansion enhances pancreatic differentiation of VFG cells

To compare the differentiation efficiency of expanding VFGs to standard differentiation we employed aspects of three established protocols for the derivation pancreatic endoderm (PE) from ESCs<sup>10,12,22</sup> (Extended Data Fig. 3a). In two of these protocols<sup>12,22</sup> we observed relatively inefficient differentiation (< 20% PDX1<sup>+</sup>) (Fig. 2a and Extended Data Fig. 3b).

However, a protocol coupling BMP inhibition, FGF and WNT activation<sup>10</sup> resulted in > 80% PDX1<sup>+</sup> induction, suggesting that VFG cultures are adapted to protocols harnessing signals regulating ventral pancreatic specification. VFG-derived PE expressed pancreatic markers including *PDX1* and *NKX6-2*, Glycoprotein 2 (GP2)<sup>22,25</sup> and the ventral pancreatic marker Roundabout2 (ROBO2)<sup>26</sup> (Extended Data Fig. 2c). Consistent with the observation that ventral pancreatic bud expands more than the dorsal bud<sup>18</sup>, cells differentiated via this third protocol and not the other two, proliferate (Fig. 2b, c).

The efficiency of pancreatic differentiation increased with time in expansion and was maintained at a similar level following six passages (Fig. 2d and Extended Data Fig. 3d). Later passage VFG cells reintroduced into differentiation also produced more insulin-positive (INS<sup>+</sup>) endocrine cells (Fig. 2e, f). Similarly, extended VFG expansion produced enhanced hepatic, but not intestinal, differentiation (Extended Data Fig. 3e, f). Expression of primitive-streak and early endoderm genes, *GSC*, *GATA6* and *CER1*, decreased upon expansion (Fig. 2g). General endoderm markers expressed in the ventral foregut, such as *FOXA2*, *HHEX*, and *SOX17*, were expressed throughout expansion at levels comparable to those in transient ADE cells. Expression of foregut marker, *HNF1B*<sup>27</sup> and ventral foregut markers, *TBX3*, *ID2*, and *GATA3*<sup>15,28,29</sup>, were elevated in early passaged (p3 and p4) VFG cells and maintained during expansion. The pancreatic progenitor marker *PDX1* was never detected during VFG expansion.

### Chromatin accessibility is finely tuned in VFG expansion

Principle component analysis (PCA) of VFG RNA-seq data at multiple passages showed VFG cells form a cluster separated from ADE, PE and ESC (Fig. 3a). Different passages of VFGs, cultured with and without BMP4, cluster together and separate in the first principle component from PE. Comparison between VFGp3 and p6 cells shows a small set of genes (21 up- and 102 down-regulated) with significant changes in expression ( $\log_2FC > 2$ ,  $P < 0.05$ ), including down regulated primitive-streak markers (*GSC*, *CER1*, and *LEFTY1*) (Extended Data Fig. 4a and Supplementary Table 1a). The only GO terms for gene set enrichment with expansion were associated with chromatin modification and cell-cycle transition (Extended Data. 4b). We used ATAC-seq to map chromatin accessibility during the progression of hESC to pancreatic progenitors, at five defined stages of differentiation and expansion: hESC, ADE, VFGp3, VFGp6, and PE. Unlike the transcriptome of different passage VFG cultures that cluster together by PCA, we observed considerable change in the ATAC-seq profile as a function of time in culture, with the higher passage VFGs moving toward PE (Fig. 3b).

We used general linear modelling<sup>30</sup> to define the dynamic changes in chromatin accessibility at promoter-distal ATAC-seq peaks (putative enhancers) across these five stages of differentiation. This resulted in a dynamic set of 57803 sites (Extended Data Fig. 4c) showing significant chromatin opening or closing in at least one stage of differentiation. Temporal patterns of chromatin accessibility were defined using c-means clustering, producing 8 clusters corresponding to six distinct groups of putative enhancers (Fig. 3c and Supplementary Table 1b). The largest group of sites are where chromatin accessibility is reduced at the start of differentiation and remains closed for the duration through to PE. The

VFG<sup>OFF</sup> cluster contains sites that become accessible during ESC to ADE differentiation, but that then lose accessibility during VFG differentiation or expansion, so that they are inaccessible in PE. The PE<sup>OFF</sup> cluster also appears at ADE and then loses accessibility but only after VFG expansion. The PE<sup>ON</sup> cluster encompasses regions that only open up during differentiation to PE. We defined two VFG clusters, VFG<sup>TR</sup> and PE<sup>PRIMED</sup> clusters.

Chromatin accessibility for the PE<sup>PRIMED</sup> cluster increases gradually during VFG expansion and is most accessible in PE. An element located ~5 Kb upstream of the *GLIS3* transcriptional start site (TSS) (Fig. 3d, left, Extended Data Fig. 4d) is an example of this. *In vivo* *Glis3* is expressed in pancreatic endocrine progenitors and then beta cells<sup>31</sup>. RNA-seq shows that *GLIS3* is not expressed until PE differentiation from expanded VFGs (Fig. 3d, right). We also observed increases in accessibility in the conserved enhancer regions (Area IV) of *PDX1*<sup>32,33</sup> (Extended Data Fig. 4e). The VFG<sup>TR</sup> cluster contains regions where chromatin accessibility increases during VFG expansion and is then shut down during differentiation to PE. The putative enhancers located ~7 Kb upstream of the *TBX3* TSS (Fig. 3e, left, Extended Data Fig. 4f) are an example of this. *TBX3* is expressed in the developing human posterior foregut, and liver bud progenitors<sup>20,34</sup> and is expressed specifically in VFGs, but then silenced during the differentiation to PE (Fig. 3e, right).

To link these enhancer clusters to changes in gene expression we defined significantly changing genes in the transition from expansion into further differentiation ( $\log_2\text{FC} > 1.5$ ,  $P < 0.05$ ) (Supplementary Table 1c). To pair enhancers with specific genes, we considered enhancers located either within 25 or 200 Kb of the single nearest gene's TSS and we excluded low level changes in basal gene expression (Supplementary Table 1d, e). While filtering out gene expression noise that occurs with passaging reduces the size of the gene set, we were able to define enhancers located within either 25kb or 200kb of up-regulated PE genes. Regardless of which enhancer set used, we observed significant enrichment of both PE<sup>PRIMED</sup> and PE<sup>ON</sup> enhancer classes with up-regulated PE genes (Fig. 3f and Extended Data Fig. 4g, h), although the enrichment is greater for enhancers located closest to the genes they regulate. We also identified enhancers at the same distance from genes down-regulated in differentiation (Supplementary Table 1f, g). These down-regulated PE gene sets were associated with the PE<sup>OFF</sup> and VFG<sup>TR</sup> enhancer categories (Fig. 3g; Extended Data Fig. 4i, j). Taken together, this suggests that VFG expansion primes some pancreatic enhancers for later target gene induction while decommissioning enhancers driving gene expression inappropriate for the PE lineage.

### Differentiation imperfectly realizes the VFG enhancer landscape

To understand the extent the enhancer network induced during expansion is normally exploited in directed differentiation we compared our data to a previous study that profiled chromatin accessibility by ATAC-seq during the differentiation of hESC through definitive endoderm (DE), and posterior foregut (FG) stages to pancreatic progenitors (PP1)<sup>35</sup>. Based on this analysis we could define a common set of putative enhancers activated either in VFGs or FG<sup>35</sup> and that then remain accessible in later differentiation (PE or PP1), respectively (PE-PP1 common); and a class of element that is not induced in the absence

of expansion (PE-not-PP1) (Extended Data Fig. 5a, b; Supplementary Table 2a). Many of the peaks that closed down during or after VFG expansion (VFG<sup>OFF</sup>-in-DE-PP1 and VFG<sup>TR</sup>-in-PP1) remains accessible in the FG or PP1 stages (Extended Data Fig. 5a, c). Together, these data suggest that VFG expansion allows for the commissioning of enhancers relevant to pancreatic differentiation and the decommissioning of enhancers for alternative lineages. This process appears bypassed in directed differentiation.

Mapping of these enhancer elements to potential target loci (located with 200 Kb) (Supplementary Table 2b, c), reveals an enrichment for the two pancreatic endoderm enhancer clusters, PE-PP1 common and PE-not-PP1, in the vicinity of genes up-regulated in VFG-derived PE (the same gene set used for Fig. 3f, g) (Extended Data Fig. 5d). However, elements induced in directed differentiation, but not active in VFG-derived PE (VFG<sup>OFF</sup>-in-DE-PP1 or VFG<sup>TR</sup>-in-PP1), do not correlate with our PE up-regulated gene set. Moreover, the PE down-regulated gene set correlates with VFG<sup>TR</sup>-in-PP1 elements. These observations suggest expansion is required for appropriate enhancer decommissioning.

In embryogenesis, the pancreas is derived from two buds that originate in different regions of the posterior foregut, dorsal and ventral<sup>6</sup>. As the ventral pancreas is derived from the ventral foregut, we assessed the expression of markers thought to distinguish the dorsal pancreatic lineages<sup>36</sup>. Extended Data Fig. 6a shows the increase in expression of these markers in directed differentiation as foregut-like cells gives rise to PP1, and suggests that directed differentiation has more of a dorsal identity.

To explore global correlations between genes differentially regulated in pancreatic endoderm derived from VFGs and directed differentiation, we plotted gene expression from both protocols (Extended Data Fig. 6b) and focused on the two classes of expansion dependent elements, PE-not-PP1 and VFG<sup>TR</sup>-in-PP1 (Extended Data Fig. 6c, d, left). Genes in the vicinity of PE-not-PP1 elements are better induced in VFG-derived PE than directed differentiation, whereas genes mapped to elements decommissioned as a result of expansion - VFG<sup>TR</sup>-in-PP1 - are more extensively down regulated when PE is differentiated from expanding VFGs. Examples of expansion dependent up-upregulation include *FRMD6* and *FGFR2*; and for those ectopically expressed in directed differentiation, *IHH* and *EPHA4* (Extended Data Fig. 6c, d, right). These analyses suggest that there are significant differences in mRNA expression related to expansion dependent changes in enhancer accessibility.

### VFG expansion captures human fetal organ specific enhancers

To determine how the enhancer landscape captured during VFG expansion and PE differentiation *in vitro* corresponds with pancreatic development *in vivo*, we compared our ATAC-seq data with H3K27ac data obtained from micro-dissected endodermal (pancreatic, liver, lung, and stomach), mesodermal (adrenal and heart), and ectodermal (RPE and brain) tissues collected from Carnegie stages 15-22 human embryos<sup>37</sup> (Supplementary Table 3). Consistent with the VFG identity of our cultures, the PE<sup>PRIMED</sup> class of element is enriched for both liver and pancreatic enhancers, while the PE<sup>ON</sup> class overlaps more extensively with pancreatic elements (Extended Data Fig. 7a and Supplementary Table 4a). Enhancer clusters that shut down as expanded VFGs differentiate to PE (VFG<sup>TR</sup> and PE<sup>OFF</sup>) are



most enriched for enhancers active in the developing liver, consistent with their role in non-pancreatic VFG differentiation. Elements decommissioned in early differentiation or expansion (ADEF<sup>OFF</sup> or VFG<sup>OFF</sup>) are non-VFG enhancers, including elements spanning the ectodermal and mesodermal lineages (Extended Data Fig. 7b; Supplementary Table 4a, b).

We assessed how the enhancers classes that differ between *in vitro* VFG expansion and direct differentiation from pluripotent cells compare to human organogenesis. Not surprisingly, the PE-PP1 common class of element was enriched in enhancers accessible in the ventral foregut derived pancreas and liver, while expansion dependent PE-not-PP1 enhancers were more enriched in pancreatic elements (Extended Data Fig. 7c, d). Moreover, the set of enhancers accessible in directed differentiation, but decommissioned as consequence of expansion (VFG<sup>OFF</sup>-in-DE-PP1) or VFG differentiation to PE (VFG<sup>TR</sup>-in-PP1), did not contain significant numbers of pancreatic elements.

### Enhancers explicitly correlating with VFG proliferation

Although differentiation efficiency increased with time in VFG culture, we wished to exclude alterations to enhancer accessibility that could result from the shift to VFG culture and variations in pancreatic differentiation arising between the dorsal and ventral lineages. We therefore defined a restricted set of enhancers specifically regulated between passages 3 and 6, correlating with enhanced pancreatic and hepatic, but not intestinal differentiation. We segregated defined enhancers activated or inactivated for the first time at passage 3 (VFGp3<sup>OPEN</sup> and VFGp3<sup>CLOSE</sup>), and those responding to increased passaging (VFGp6<sup>OPEN</sup> and VFGp6<sup>CLOSE</sup>) (Fig. 4a). While the chromatin accessibility of VFGp3<sup>OPEN</sup> and VFGp3<sup>CLOSE</sup> enhancer elements also respond to expansion, the influence of passaging is difficult to resolve from an initial response to the change in culture medium.

To investigate whether there was a change in chromatin state of enhancers specifically responding to expansion, we performed ChIP-qPCR for H3K27 acetylation (H3K27ac) and H3K4 mono methylation (H3K4me1) for multiple expansion-regulated elements (Fig. 4b, c). There were robust changes in H3K27ac deposition at these elements between VFG p3 and 6, while changes in H3K4me1 were more subtle. We also paired these explicitly expansion dependent enhancers to specific genes (within 200 Kb of the single nearest gene's TSS) (Supplementary Table 5a). We identified 480 enhancers explicitly correlated to expansion and located them within 200kb of PE up-regulated genes (the same gene set being used in Fig. 3f) (Supplementary Table 5b). Chromatin accessibility at both VFGp3<sup>OPEN</sup> and VFGp6<sup>OPEN</sup> enhancers correlated with gene expression (Fig. 4d). Similarly, for genes downregulated during VFG expansion ( $\log_2FC < -1.5$ ,  $P < 0.05$ ), we observed good correlation with decommissioning (Supplementary Table 5, c-e), where in this instance, only the expansion specific VFGp6<sup>OFF</sup> correlates well with gene expression (Fig. 4e) here.

To ask whether enhancers correlating directly with expansion are also related to ventral foregut specific differentiation, we compared these enhancers to the *in vivo* regulatory landscape in fetal organ development (Supplementary Table 6). Consistent with the interpretation that extended VFG culture lays the groundwork for further differentiation, both the VFGp3<sup>OPEN</sup> and the expansion-specific VFGp6<sup>OPEN</sup> clusters overlap with

active enhancer sets from the fetal pancreas and liver, but not stomach, lung or other non-endodermal organs (Fig. 5a, b). Both sets of VFG<sup>OPEN</sup> enhancers are enriched in the endoderm lineage, while the VFG<sup>CLOSED</sup> enhancers contain more mesodermal and ectodermal elements (Fig. 5c). Finally, we compared expansion clusters to directed-differentiation clusters (Fig. 5d). Both VFG<sup>OPEN</sup> enhancer clusters that are not regulated in directed differentiation (VFGp3<sup>OPEN</sup> and VFGp6<sup>OPEN</sup>-not-PP1) overlap with fetal pancreas and liver enhancers sets, while VFG decommissioned enhancers that remain accessible in directed differentiation (VFGp3<sup>CLOSED</sup> and VFGp6<sup>CLOSED</sup>-in-PP1) have little in common with pancreatic and hepatic elements.

### Transcription factors FOXA and HHEX in pancreatic priming

To determine factors responsible for VFG enhancer priming we assessed TF motifs in different enhancer classes (Fig. 6a and Supplementary Table 7), focusing on those linked directly to expansion, and regulated between P3 and P6. TF motifs in VFGp6<sup>OPEN</sup> enhancers included FOXA factors and to a lesser extent, a number of unrelated endodermal/hepatic factors broadly classed as hepatic nuclear factors (HNFs)<sup>38</sup> and TEAD1. In contrast, motifs in VFGp6<sup>CLOSED</sup> elements included early endoderm and mesendoderm factors included GATA and EOMES. To further refine the association of specific TFs with these enhancer classes, we used *k*-means clustering to define patterns of mRNA expression associated with enhancers that are up regulated or down regulated during VFG expansion (Extended Data Fig. 8a, b) and selected clusters that correlated with differentiation. In those enhancers related to clusters of up-regulated gene expression in pancreatic differentiation, we identified motifs for TF classes relevant to human pancreatic and liver development<sup>36,39</sup>, such as FOXA, HNF1B, TEAD, and the architectural factor CTCF. For those enhancers mapping to down-regulated clusters, we observed no motifs linked pancreatic differentiation or function (Extended Data Fig. 8c).

Of TFs known to recognize FOX DNA binding motifs, embryonic expression patterns and phenotypes in mouse development suggest FOXA1 and 2 could be relevant to VFG mediated enhancer priming<sup>33,40</sup>. FOXA2 is required for pancreas development and differentiation in both mouse<sup>33</sup> and human ESCs<sup>35</sup>, and a requirement for FOXA1 in pancreas development is observed in the context of FOXA1/2 double mutants. FOXA factors are known ‘pioneer TFs’ that access regulatory regions and prepare them for later activation<sup>41</sup>. However, FOXA1 mutant ESCs undergo apparently normal directed pancreatic endoderm differentiation<sup>35</sup>. To assess their function in pancreatic priming during human VFG expansion we knocked down FOXA1 and FOXA2 by shRNA during VFG expansion (Fig. 6b and Extended Data Fig. 9a). Neither factor produced a significant reduction in VFG marker expression (Extended Data Fig. 9b, c), although FOXA2, but not FOXA1, KD impaired VFG expansion. When VFG cells knocked-down for either FOXA1 or FOXA2 were challenged in pancreatic differentiation, expression of pancreatic markers was significantly reduced (Fig. 6c). We confirmed FOXA1 binding, using ChIP-qPCR, at the *PDX1* enhancer area IV, the PE<sup>PRIMED</sup> enhancers of *GLIS3*, and VFGp6<sup>OPEN</sup> enhancer of *SFRP5*, but not in the VFG<sup>TR</sup> element of *TBX3*. Binding was reduced in the stable FOXA1 KD VFG lines (Fig. 6d). Knock down of FOXA1 led to a significant reduction in H3K4me1



and H3K27ac at primed enhancers associated with *PDX1* and *SFRP5* (Fig. 6e), but not the enhancer associated with *TBX3*.

HHEX is suggested to be an essential transcriptional regulator of directed ESC differentiation to pancreatic endoderm<sup>42</sup>. We therefore asked whether HHEX was required for expansion linked pancreatic enhancer regulation. Knock down of FOXA1 or HHEX produced similar defects in pancreatic differentiation and the double knock-down had a combinatorial effect on PDX1 induction (Extended Data Fig. 10a, b) consistent with the specific influence they have on each other's binding at the PDX1 enhancer (Extended data Fig. 10c, d). At VFG linked enhancer elements, HHEX has a particularly pronounced effect on H3K27ac (Extended data Fig. 10e, f). To gain insight into the relation of HHEX binding to our enhancer dataset, we aligned HHEX ChIP-seq data from directed differentiation<sup>42</sup> to the enhancer regions from the different classes defined here (Extended data Fig. 10g). HHEX binding at both direct differentiation stages (FG and PP1) was detected at PE-PP1 common enhancers, but was depleted at the expansion dependent PE-not-PP1 class of elements. Moreover, the VFG decommissioned enhancers VFG<sup>OFF</sup>-in-DE-PP1 and VFG<sup>TR</sup>-in-PP1 elements, which are incompletely silenced in directed differentiation, were still occupied by HHEX during these stages of directed differentiation. As a result, it appears that all enhancer classes defined here and represented in the directed differentiation dataset are occupied by HHEX, including those normally decommissioned during VFG expansion. Perhaps, the binding of HHEX at these elements in directed differentiation prevents their decommissioning during rapid directed differentiation.

## Discussion

A portion of the pancreas comprising the uncinate process, in addition to the liver and gall bladder, are all derived from the VFG region of the developing embryo beginning at E8.5 in mouse or at Carnegie stage 10 (25-27 days post coitum) in human<sup>18,43</sup>. Based on gene expression and differentiation competence, hESC-derived EP cells were found to recapitulate ventral foregut. While prior studies have shown that VFG expansion can produce functional pancreatic endocrine cells<sup>15</sup>, here we demonstrate that this is a direct consequence of time in VFG culture. *In vivo*, pancreas development begins from two locations, the dorsal and ventral foregut, promoting organ development via distinct signalling. Dorsal pancreas is induced by factors derived from the notochord and dorsal aorta (retinoic acid, activin, and FGF2)<sup>44</sup>, while the ventral pancreas differentiates in the absence of signals driving hepatic specification (FGF2 produced by the cardiac mesoderm and BMP4 originating in the septum transversum)<sup>45</sup>. Ventral foregut embryonic explants therefore default to pancreatic differentiation in the absence of exogenous signalling<sup>5</sup>. However, *in vivo*, VFG progenitors and their descendants retain multi-potency up to E11.5 in mouse where the cell cycle time has been estimated to be between 17.3 and 26.6 hours<sup>7</sup>. As these progenitor cells are located close to both the cardiac mesoderm (FGF source) and septum transversum mesenchyme (BMP source), both components in VFG culture media, these founder populations may persist via self-renewing cell division *in vivo* exploiting their proliferation to insure efficient onward differentiation.

An increasing set of TFs have the ability to bind DNA in chromatin and to destabilise nucleosomes. These pioneer factors include the FOXA proteins identified here as important for VFG priming. FOXA proteins are associated with enhancer priming during foregut development<sup>45</sup> and associate with mitotic chromatin<sup>46</sup>. Yet, FOXA1 is not required for directed differentiation to pancreatic endoderm *in vitro*. While we have not shown a direct relationship between the cell cycle and enhancer priming by FOXA proteins, the major variable in our experiments is the amount of time in VFG culture and we cannot formerly exclude the influence of prolonged culture in these conditions on the enhancer network. However, it is possible that FOXA1 pioneer activity in VFG culture depends on proliferation, leading to a progressive equilibration of the enhancer network, involving both commissioning and decommissioning. Although pioneer factors are known to recognize their sites in chromatin, they may have an enhanced ability to bind their sites prior to the full restoration of heterochromatic marks following replication and then remain at these positions through mitosis. The human ESC directed differentiation protocol that comes closest to reproducing the proliferative nature of early ventral foregut is the one instance where a role for FOXA1 was previously suggested<sup>47</sup>. HHEX is also associated with enhancer priming in VFGs and can influence the stability of FOXA1 binding at the PDX1 enhancer. As HHEX physically interacts with FOXA1 in both gut tube and pancreatic progenitor stages of directed human ESC differentiation<sup>42</sup>, it is possible that HHEX could act together with FOXA1 to enhance the stability of binding to targets in mitotic chromatin.

While we are not aware of many progenitor culture systems where the impact of proliferation on differentiation has been explored, the transition into expanded primed pluripotent cells alters the type of endoderm induced by the same cytokines<sup>48</sup>. It is intriguing to hypothesize that the reconfiguration of the enhancer network during the transition from naïve to primed pluripotency<sup>49</sup>, may also involve proliferation as cells at gastrulation stages proliferate rapidly with a cell cycle as short as five hours<sup>50,51</sup>. Moreover, in both naïve and primed pluripotency the binding of pluripotency TFs to differentiation specific genes determines how these enhancers will respond to signalling and whether differentiating cells retain plasticity<sup>52,53</sup> suggesting that TFs function to set the enhancer network for lineage specific progenitors to respond to signaling. In addition to preparing enhancers for later activation, we also found that enhancer decommissioning exploits expansion, perhaps as a result of going through multiple rounds of replication in the absence of specific TFs that protect these enhancers from nucleosome occlusion following replication. In VFGs, these decommissioned elements contain motifs for GATA factors, with GATA4 and 6 being downregulated in the early stages of VFG culture. While FOXA1 can bind mitotic chromatin, GATA factors are only partially retained<sup>54</sup>, suggesting that expansion could provide FOXA proteins with a competitive advantage. In this way, expansion not only primes differentiation, but shields the later developing endoderm from the lingering action of early endoderm enhancers.

We observe that the proliferation or expansion of lineage-restricted progenitors may be essential for high efficiency later differentiation. Proliferation is therefore not just about producing sufficient numbers of cells, but fine-tuning the response of these cells to upcoming differentiation cues. Progenitor cell expansion can also equalize the differentiation efficiency of poorly performing hESCs<sup>16,55,56</sup> suggesting that the lineage

potential of different pluripotent cell lines may be determined by the extent they proliferate in differentiation. Moreover, as proliferation and growth are a hallmark of later fetal development, additional expansion steps could enhance the efficiency with which more mature organ specific cell types can be obtained from human pluripotent cells.

## Materials and Methods

### Experimental Design

**Maintenance of human ESC**—Undifferentiated human ESCs H9 (WA09, WiCell Madison, WI) were maintained on tissue culture plates pre-coated with 0.1% gelatine with irradiated C57BL6 mouse embryonic fibroblast feeder cells (25,000 cells/cm<sup>2</sup>) in H9 ESC media: DMEM/F12 GlutaMAX medium (Thermo Fisher Scientific, 10565018) supplemented KnockOut Serum Replacement (Thermo Fisher Scientific, 10828010), NEAA, beta-mercaptoethanol (Thermo Fisher Scientific, 21985023), and 10 ng/mL FGF2 (Peprotech, 100-18B). Cells were passaged as clusters with collagenase IV (Thermo Fisher Scientific, 17104019) when reaching approximately 70% confluence, and maintained in 20% O<sub>2</sub>/5% CO<sub>2</sub>/37°C. Undifferentiated ESC HUES4 WT and PDXeG clone 170-3<sup>22</sup> were adapted and maintained in DEF-CS (Takara, Y30017). When reaching approximately 80% confluence, cells were dissociated with TrypLE (Thermo Fisher Scientific, 12604013) and counted with the automated NucleoCounter NC-200 cell counter (Chemometec). Cells were re-plated at a density of 40,000 cells/cm<sup>2</sup> and maintained in 20% O<sub>2</sub>/5% CO<sub>2</sub>/37 °C. All hESC lines were routinely screened for mycoplasma, and all were negative. All cell lines were approved for use in this project by De Videnskabetiske Komiteer, Region Hovedstaden under number H-4-2013-057.

**Transient differentiation of ADE cells**—Transient ADE cells were generated from wild type H9 and HUES4 ESCs, as well as HUES4 PDX1-eGFP reporter (PDXeG clone 170-3) ESC cell line<sup>22</sup> as described in Cheng et al.<sup>15</sup>. In brief, ESC cells at 70-80% confluence were collected with Accutase (Thermo Fisher Scientific, 00455556), re-plated at a density of 50,000 cells/cm<sup>2</sup> on polystyrene cell culture plates (Corning, 353047) pre-coated with un-diluted growth factor reduced (GFR) Matrigel (Corning, 354230), cultured in either H9 ESC or DEF-CS media for 48 hours with 10µM ROCK inhibitor Y-27632 (STEMCELL Technologies, 72302) for the first 24 hours and maintained in 20% O<sub>2</sub> / 5% CO<sub>2</sub> / 37 °C. The ESC clusters were used to generate transient ADE cells in three-dimensional differentiation under hypoxic conditions (5% O<sub>2</sub>/5% CO<sub>2</sub>/37°C) for 5 days. On day 1, the cell clusters were cultured in RPMI 1640 GlutaMAX (Thermo Fisher Scientific, 61870036) with 10% SFD media<sup>57</sup> supplemented with Activin A [100 ng/mL] (Peprotech, 120-14P), CHIR99021 [3 µM] (Tocris, 4423) and 4.5x10<sup>-4</sup> M Monothioglycerol (Sigma-Aldrich, M6145). On day 2, the medium was changed to RPMI 1640 GlutaMAX supplemented with Activin A [100 ng/mL], BMP4 [0.5 ng/mL] (Peprotech, 120-05ET), FGF2 [10 ng/mL], VEGF [10 ng/mL] (Peprotech, 100-20), 0.5 mM ascorbic acid (Sigma-Aldrich, A92902), and 4.5x10<sup>-4</sup> M monothioglycerol. The same media was applied at day 3. At day 4, differentiation media was changed to SFD media supplemented with Activin A [100 ng/mL], BMP4 [0.5 ng/mL], FGF2 [10 ng/mL], VEGF [10 ng/mL], 0.5 mM ascorbic acid, and 4.5x10<sup>-4</sup> M monothioglycerol.

**Generation and expansion of VFG**—EP/VFG expansion was performed as described<sup>15</sup> with minor modifications. In brief, day 5 transient ADE clusters were dissociated with 1 volume of trypsin-EDTA [0.25%] (Thermo Fisher Scientific, 25200056) for 5 mins at 37 °C, and the enzyme then inactivated with 0.5 volume of fetal bovine serum (FBS) (Sigma-Aldrich, F4135). Single cells suspensions were obtained by repeatedly washing with 10 volumes of ice-cold washing buffer, which contains 3% FBS in Phosphate Buffered Saline without Calcium and Magnesium (PBS-/-) (Thermo Fisher Scientific, 10010023). Single cells were incubated with 1:100 CD184-PEcy7 (BD Biosciences, 560669) and CD117-APC (BD Biosciences, 561118) for 45 mins at 4 °C and stained with 4',6-diamidino-2-phenylindole (DAPI) (Thermo Fisher Scientific, D3571) to exclude dead cells. CD184-CD117 double-positive cells were sorted into SFD media with 1:100 Penicillin-Streptomycin (Thermo Fisher Scientific, 15140122) by Fluorescence-Activated Cell Sorting (FACS) on a SH800 (SONY SH800 Software). Sorted cells were re-plated at a density of 20,000-30,000 cells/cm<sup>2</sup> on polystyrene cell culture plates pre-coated with GFR-Matrigel and pre-seeded with low density (8,000 cells/cm<sup>2</sup>) irradiated DR4 mouse embryonic fibroblast feeder cells (MEFs) (ATCC, SCRC-1045). Cells were cultured in complete EP/VFG media (SFD medium supplemented with BMP4 [50 ng/mL], FGF2 [10 ng/mL], VEGF [10 ng/mL], EGF [10 ng/mL] (Peprotech, AF-100-15), 0.5 mM ascorbic acid, and 4.5x10<sup>-4</sup> M monothioglycerol) and maintained under hypoxic conditions (5% O<sub>2</sub>/5% CO<sub>2</sub>/37 °C).

Media was changed every other day until cells reached confluence, at 80,000-120,000 cells/cm<sup>2</sup>. When VFG cells reached approximately 100 µm in diameter, they were passaged by dissociation using 1 volume of trypsin-EDTA [0.25%] for 5 mins at 37°C, detached from the plate using a cell scraper and then supplemented with 0.5 volume of FBS for enzyme inactivation. Single cells suspension was obtained by repeatedly washing with 10 volume of ice-cold washing buffer. VFG single cells were re-plated on the pre-coated GFR-Matrigel with feeders at 15,000-20,000 cells/cm<sup>2</sup>. Antibody information is listed on Supplementary Table 1.

**Single cell preparation for RNA-seq and index sorting**—Dissociated ADE and VFG single cells with treatments (mock, BMP4 withdrawal, and BMP4 withdrawal plus FGF2 stimulation) were incubated with 1:100 CD184-PEcy7 and CD117-APC for 45 mins at 4°C and cells were stained with DAPI to exclude dead cells. The single cells from BMP4 withdrawal plus FGF2 stimulated VFG culture were incubated only with 1:100 CD117-APC in a similar condition to that described above. Cells were sorted using a BD FACS Aria III (FACSDiva™) with a 100 µm nozzle and 20 psi sheath pressure. Forward scatter (FSC) and side scatter (SSC) were used to define a homogeneous population. FSC-H/FSC-W gates were used to exclude doublets and dead cells were excluded based on DAPI inclusion. The boundary between positive and negative populations were set based on a negative population of unstained cells. Sorting speed was kept at 100-300 events/s to eliminate sorting two or more cells into one well. Single cell sorting was verified colorimetrically based on a previously described protocol<sup>58</sup>. Cells were sorted directly into lysis buffer containing the first RT primer and RNase inhibitor, immediately frozen and later processed by the MARS-seq1 protocol as described previously<sup>59</sup>. All single-cell RNA-seq libraries were sequenced

using Illumina NextSeq 500 at a median sequencing depth of 225,000 reads per single cell. Antibody information is listed on Supplementary Table 1.

**Immuno-histochemical analysis**—Media was removed completely and matrigel dome containing 3D clusters were gently mixed with fresh undiluted matrigel 1:1 and transferred to 8-well  $\mu$ -slides (Ibidi, 80826) wells (20  $\mu$ L/cm<sup>2</sup> well) for whole mount immunostaining. When the matrigel was solidified at 37°C, room-temperature 4% paraformaldehyde (PFA) (Sigma-Aldrich, 158127) was added and cultures were fixed at room temperature for 10 mins, blocked and permeabilized with 2% donkey serum (Jackson Immuno Research, 017-000-121), 0.3 % Triton X-100 (Sigma-Aldrich, X100) and 0.1 % BSA (Sigma-Aldrich, A7906) in PBS-/- for 1 hour at RT. Primary antibodies were incubated with 3 % FBS in PBS-/- overnight at 4°C, subsequently incubated with the appropriate secondary antibody (AlexaFluor, Molecular probes) and DAPI at room temperature for 1 hour at RT. Antibody information is listed on Supplementary Table 1. Brightfield and fluorescent imaging were done using a Leica SP8 confocal microscope with Las X software (3.5.7.23225) and processed in Imaris 9.6.

**EdU labelling and apoptosis assay**—Cells were incubated with 10  $\mu$ M EdU (Click-iT EdU) (Thermo Fisher Scientific, C10634) in medium for 4 hours at 5% O<sub>2</sub>/5% CO<sub>2</sub>/37°C. The 3D clusters were prepared for whole mount immunostaining as described above. Dissociated cells were collected for flow cytometry as described above. Permeabilization, blocking and Click-iT reaction for EdU detection were performed according to the manufacturer's instructions. Immunostaining of EdU-labeled 3D clusters were performed with antibodies supplied with the kit and with DAPI (1  $\mu$ g/mL) for nuclear staining. Flow cytometry of EdU-labeled dissociated cells was performed with DAPI (10  $\mu$ g/mL) staining cells for DNA content. Cell apoptosis was measured by Annexin V Conjugates for Apoptosis Detection kit (Thermo Fisher Scientific, A13202) according to the manufacturer's instructions.

**Flow cytometry**—For surface marker staining, dissociated cells were incubated with conjugated antibodies for 1 hour at 4°C and were stained with DAPI (1  $\mu$ g/mL) to exclude dead cells. For intracellular staining, cells were stained with Ghost Dye 450 (TONBO biosciences, 13-0868) prior to 4% PFA fixation to stain dead cells. Fixed cells were permeabilized in PBS with 5% donkey serum and 0.3% Triton X-100 for 30 mins at room temperature. Cells were incubated with primary antibodies in 1x PBS-/- with 5% donkey serum and 0.1% Triton X100 overnight at 4°C. The following day, cells were washed twice in 1x PBS and unconjugated antibodies were further incubated with secondary antibodies (Alexa Fluor conjugates) for 2 hours. Antibody sources and concentrations are indicated in Supplementary Table 1. Cells were analysed using an LSR Fortessa (BD Bioscience) or FACS sorted by SH800 (SONY SH800 Software). All data were analysed with FCS Express 6 software (BD Biosciences). Antibody information is listed on Supplementary Table 1.

**Generation of PDX1-eGFP positive and negative cells with minimal cytokine sets for pancreatic spheroid and hepatic organoid expansion**—PDX1-eGFP reporter VFG cells passage 6 was plated at 25,000 cells/cm<sup>2</sup> on polystyrene cell culture



plates pre-coated with un-diluted GFR-Matrigel and pre-seeded with  $8 \times 10^3$  cells/cm<sup>2</sup> MEF. The cells were cultured in BMP4 withdrawal media (SFD medium supplemented with FGF2 [10 ng/mL], VEGF [10 ng/mL], EGF [10 ng/mL], 0.5 mM ascorbic acid, and  $4.5 \times 10^{-4}$  M Monothioglycerol) and maintained under hypoxic conditions (5% O<sub>2</sub>/5% CO<sub>2</sub>/37°C) for 5 days with medium changing every other day. For generating PDX1-eGFP positive and negative fractions, cells were further differentiated in DMEM high glucose GlutaMAX Supplement (Thermo Fisher Scientific, 10566016) with 1% vol/vol B27 supplement (Thermo Fisher Scientific, 17504044), 50 ng/mL FGF2, FGF7 (Peprotech, 100-19), or FGF10 (Peprotech, 100-26) for 5 days with medium changed every day. Both BMP4 withdrawal and FGFs stimulation were performed under hypoxic conditions (5% O<sub>2</sub>/5% CO<sub>2</sub>/37°C).

The single PDX1-eGFP positive and negative cells generated from the BMP4 withdrawal and FGF10 stimulated VFG culture were sorted by FACS using a SH800. GFP<sup>+</sup> cells were expanded as pancreatic spheroids and GFP<sup>-</sup> cells as hepatic organoids according to the described protocols<sup>23,24</sup>, except that the cultures were maintained under hypoxic conditions (5% O<sub>2</sub>/5% CO<sub>2</sub>/37°C).

**Pancreatic differentiation**—VFG cells at passages 6-8 were plated at 25,000 cells/cm<sup>2</sup> on polystyrene cell culture plates pre-coated with un-diluted GFR-Matrigel and pre-seeded with 8000 cells per cm<sup>2</sup> MEF in the VFG medium. Day 5 expanding VFG cells were used for pancreatic differentiations under hypoxic conditions (5% O<sub>2</sub>/5% CO<sub>2</sub>/37°C) according to protocols described as below:

For the protocol adapted from Ameri et al.<sup>22</sup>, day 5 expanding VFG cells were treated with DMEM high glucose GlutaMAX Supplement with 1% vol/vol B27 supplement as basal media throughout the differentiation, and were supplemented with 2 μM retinoic acid (RA) (Sigma-Aldrich, R2625) for 3 days; then with 64 ng/mL FGF2 and 50 ng/mL hNOGGIN (R&D systems, 6057-NG-100/CF) for 3 days; and finally with 64 ng/mL FGF2, 50 ng/mL hNOGGIN, and 0.5 μM TPB (PKC activator) (Merck Millipore, 565740) for 3 days, with the media changed every day.

For the protocol adapted from Rezanian et al.<sup>12</sup>, day 5 expanding VFG cells were exposed to MCDB 131 basal medium (Thermo Fisher Scientific, 10372019) throughout differentiation, and supplemented with 1.5 g/L sodium bicarbonate (Thermo Fisher Scientific, 25080094), 1x GlutaMAX Supplement (Thermo Fisher Scientific, 35050061), 10 mM D-(+)-Glucose (Thermo Fisher Scientific, G8270) 0.5% BSA, 0.25 mM ascorbic acid and 50 ng/mL FGF7 for 2 days; and then with 2.5 g/L sodium bicarbonate, 1x GlutaMAX Supplement, 10 mM glucose, 2% BSA, 0.25 mM ascorbic acid, 1:200 Insulin-Transferrin-Selenium-Ethanolamine (ITS-X) (Thermo Fisher Scientific, 51500056), 50 ng/mL FGF7, 1 μM RA, 0.25 μM SANT-1 (Sigma-Aldrich, S4572), 100 nM LDN193189 (Tocris, 6053) and 80 nM TPB (EMD Millipore) for 2 days; and finally with 2.5 g/L sodium bicarbonate, 1x GlutaMAX, 10 mM glucose, 2% BSA, 0.25 mM ascorbic acid, 1:200 ITS-X, 2 ng/mL FGF7, 0.1 μM RA, 0.25 μM SANT-1, 200 nM LDN193189 and 40 nM TPB for 3 days.



For the protocol adapted from Nostro et al.<sup>10</sup>, day 5 expanding VFG cells were fed SFD media supplemented with 50 ng/mL of FGF10, 3 ng/mL mWnt3a (R&D systems, 1324-WN-010/CF) and 0.75  $\mu$ M dorsomorphin (Sigma-Aldrich, P5499) for 3 days with the medium changed every day. Media was then changed to DMEM high glucose GlutaMAX Supplement with 1% vol/vol B27 supplement, 50 ng/mL FGF10, 50 ng/mL hNOGGIN, 50  $\mu$ g/mL ascorbic acid, 2  $\mu$ M RA, with 0.25  $\mu$ M KAAD-cyclopamine (Sigma-Aldrich, 239804) for one day. Finally, media was changed to DMEM high glucose GlutaMAX Supplement with 1% vol/vol B27 supplement, 50 ng/mL hNOGGIN, 50 ng/mL EGF, 10 mM nicotinamide (Sigma-Aldrich, N0636), and 50  $\mu$ g/mL ascorbic acid for 4 days with the medium changed every day.

The protocol adapted from Nostro et al.<sup>10</sup> was used to assess efficiency of pancreatic differentiation in a directed protocol from ADE cells, VFG passages 3, 6, and 12 cells generated from the PDX-eGFP reporter. Day 5 transient ADE cells were generated as described previously and directly used for differentiation. Differentiation of wild type H9 and HUES4 VFG passages 3 and 6 cells to pancreatic beta-like cells were performed as reported<sup>15,55</sup> with modifications during endocrine differentiation. In brief, day-13 differentiating VFG cells were re-aggregated following treatment with 1 mL Corning<sup>®</sup> Cell Recovery Solution (Sigma-Aldrich, CLS354270) and cultured on the membrane surface of Millicell insert (Millipore, PICM03050) in the same media described in Tiya et al<sup>55</sup>.

**Hepatic and intestinal differentiations**—Hepatic and intestinal differentiations were started from day-5 expanding VFG cells according to the protocols described in Cheng et al<sup>15</sup>.

#### **Total mRNA purification, reverse transcription and quantitative PCR Analysis**

—Two hundred thousand cells were washed in 1x PBS twice, lysed in RLT buffer (RNeasy Micro kit) (Qiagen, 74004) containing 1%  $\beta$ -mercaptoethanol (Sigma-Aldrich, M6250) and stored at  $-80^{\circ}\text{C}$  until processing. Total mRNA was isolated using the RNeasy Micro kit according to the manufacturers' instructions and digested with RNase-free DNase I, (Qiagen, 79254) to remove genomic DNA. First strand cDNA synthesis was performed with SuperScript<sup>™</sup> III First-Strand Synthesis System (Thermo Fisher Scientific, 18080051) using random hexamers (Thermo Fisher Scientific, N8080127) and amplified using SYBR Green PCR Master Mix (Thermo Fisher Scientific, 4309155). PCR primers were designed using Primer3Plus<sup>60</sup> and validated for efficiency ranging between 95-100%. Primer sequences used in RT-qPCR are listed in Supplementary Table 2. StepOnePLUS Real-Time PCR System (Thermo Fisher Scientific) was used for RT-qPCR in 96 well plates format. Expression values for each gene were normalized against *ACTB*, using the delta-delta CT method.

**Sample preparation for Bulk RNA-seq**—Total mRNA amount and RNA integrity were assessed using a Fragment Analyzer (AATI). Ribosomal RNA was removed from samples using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, E7490L). Sequencing libraries were prepared from 100 ng of purified total mRNA using NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB, E7770L) according to the manufacturer's

instructions. RNA-seq libraries were sequenced for 75 cycles in single-end mode on NextSeq 500 platform (Illumina, FC-404-2005).

**Sample preparation for ATAC-seq**—Dissociated single cells were washed with ice-cold PBS<sup>-/-</sup> and pelleted at 500 x g for 10 mins at 4°C. Fifty thousand cells were taken from a diluted stock in PBS buffer to prepare ATAC-seq libraries as described in Buenrostro et al.<sup>61</sup> with slight modifications. Nuclei were prepared by resuspending the cells in 100 µL ice-cold ATAC lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub> and 0.1% NP40) followed by incubation on ice for 15 mins while mixing every 5 mins. Nuclei were then collected by centrifuging at 1000 x g for 10 mins at 4°C and pellet resuspended in 50 µL transposition buffer (10 mM Tris pH 8, 5 mM MgCl<sub>2</sub> and 10% dimethylformamide). Tagmentation was performed by adding 2.5 µL Tn5 transposase (Illumina, 20034197) and incubating at 37°C while shaking in a thermomixer set at 1000 rpm. Tagmentation reactions were stopped and purified with MinElute PCR Purification Kit (Qiagen, 28004) and tagmented DNA eluted in 10 µL elution buffer (10 mM Tris pH 8.0). A 50 µL PCR reaction was assembled containing 10 µL of tagmented DNA, 25 µL NEB-Next High-Fidelity PCR Mix (NEB, M0541S), 5 µL of SYBR Green (Invitrogen, S7563) and primers at 2 µM concentration. Ten microliters of each PCR reaction were used to decide the optimum number of PCR cycles required with following conditions: 5 mins at 72°C; 30 sec at 98°C; and 20 cycles of 10 sec at 98 °C, 30 sec at 63 °C and 60 sec at 72 °C. The reaction was monitored in a LightCycler-480 qPCR (Roche) and the number of cycles required was deduced from the amplification curve. The remaining PCR reaction was then subjected to this number of PCR cycles. PCR reaction was purified with an equal volume of AMPure XP beads (Beckman, A63880) following manufacturer's protocol and was eluted in 20 µL Tris pH 7.8. Libraries were quantified with Qubit dsDNA High-sensitivity Assay (Invitrogen, Q32851) and fragment profiles were checked using Bioanalyzer High Sensitivity assay (Agilent) or Fragment Analyzer (AATI). Samples that showed nucleosomal bands were sequenced for 75-150 cycles in paired-end mode on an Illumina HiSeq-2000 platform or NextSeq 500.

**Generation of shRNA knockdown VFG cell lines**—Short hairpin (shRNA) targeting *HHEX*, *FOXA1*, and *FOXA2* transcripts were designed using RNAi consortium (TRC) GPP Web Portal (Broad Institute) (<https://portals.broadinstitute.org/gpp/public>) (*HHEX*, *FOXA1*, and *FOXA2* shRNA sequences, see Supplementary Table 2). A vector delivering a scrambled sequence was used as control (scrambled shRNA sequence, see Supplementary Table 2). All shRNA sequences were cloned into a lentiviral vector (pL-U6-sgRNA-SFFV-Puro-P2A-EGFP), a gift from Kristian Helin (Addgene, 12247)<sup>62</sup>, using *BsmBI* sites. HEK293FT packaging cells were co-transfected with the pL-U6-sgRNA-SFFV-Puro-P2A-EGFP carrying individual shRNAs and pAX8 and pCMV-VSV using lipofectamine 2000 supplemented with Polyethylenimine (PEI) (Sigma-Aldrich, 408727) according to standard protocols. SFD medium carrying lentivirus produced from HEK293FT cells (48 hrs post-transduction) was applied 1:1 with fresh VFG expansion media to one 12-well of day 2 VFG cell culture (passed at 25,000 cells/cm<sup>2</sup> at day 0). Transduction was performed in presence of 1:1000 polybrene infection/transfection reagent (Merck Millipore, TR-1003-G) at 8 µg/mL. Forty-eight hours after transduction with the sgRNA-encoding lentiviral

plasmids, the VFG cells were selected and maintained at 0.25  $\mu\text{g}/\text{mL}$  puromycin in standard VFG condition.

**ChIP-qPCR**—Chromatin immunoprecipitation (ChIP) was carried out using the True MicroChIP kit (Diagenode, C01010132) with modifications. One hundred thousand sorted CD184-CD117 double-positive cells ADE, VFGp3, and VFGp6 cells; or shRNAs (scrambled, FOXA1, FOXA2, or HHEX) knockdowns VFGp6 cells were fixed in 1% formaldehyde (Thermo Scientific, 28906) in ADE or VFG media for 10 mins at room temperature followed by a 5-minute quench with glycine (in True MicroChIP kit, Diagenode) at room temperature. Cells were lysed and immunoprecipitation performed using the True MicroChIP kit (Diagenode, AB-002-0016) with the following modifications. Up to 100,000 cells were sonicated in one lysate and split into 50,000 equivalents after sonication. Samples were lysed using 50  $\mu\text{L}$  of buffer tL1 and incubated for 5 mins on ice. One hundred fifty microliters of Hank's buffered salt solution with 1x protease inhibitor cocktail (in True MicroChIP kit, Diagenode) was added, and the lysate was sonicated in 0.65 mL Bioruptor Pico Microtubes (Diagenode, C30010020). Chromatin was sheared using a Bioruptor Pico (Diagenode) with 10 cycles (30 sec on, 30 sec off). Sonicate was aliquoted in 100  $\mu\text{L}$  (for 50,000 cells), and an equivalent volume of complete ChIP buffer tC1 was added. For immunoprecipitation, the following antibodies and amounts of antibody were used for the 50,000-cell ChIP: 2  $\mu\text{g}$  of FOXA1 (1:50) (Abcam, ab170933), 2  $\mu\text{g}$  of H3K4me1 (1:50) (Abcam, ab8895), 2  $\mu\text{g}$  of H3K27ac (1:50) (Abcam, ab4279), and 2  $\mu\text{g}$  of HHEX (1:100) (R&D, MAB83771). Immunoprecipitation and washes were as described in the True MicroChIP protocol, then purified by phenol chloroform extraction and ethanol precipitation. The pull-down DNA was eluted in 100  $\mu\text{L}$  elution buffer and qPCR was performed as described in the True MicroChIP protocol for different genomic loci. Enrichment was calculated as percentage of input. Antibody information is listed on Supplementary Table 1. The primer sequences used in ChIP-PCR are listed in Supplementary Table 2.

**in vitro scRNA-seq analysis**—Sequences were mapped to the hg38 assembly of the human genome, de-multiplexed and filtered as previously described<sup>59,63</sup> extracting a set of UMIs that define distinct transcripts in single cells for further processing. We estimated the level of spurious UMIs in the data using statistics on empty MARS-seq wells as previously described<sup>64</sup>. Mapping of reads was done using HISAT (version 0.1.6)<sup>65</sup>. Reads with multiple mapping positions were excluded. Reads were associated with genes if they mapped to an exon. Raw counts were further analyzed using Seurat (4.0.1)<sup>66</sup> (<https://satijalab.org/seurat/>). Cells were filtered with the following thresholds (lower bound: 2,000 UMIs; 550 genes and upper bound: 35,000 UMIs; 4,950 genes). Additionally, cells with more than 20% of mitochondria content were removed. In Extended Data Fig. 1a, we subset ADE and VFG cells (505 cells). Raw counts were further normalized, log-transformed and scaled using *NormalizeData* and *ScaleData* respectively. PCA was computed on 2,000 highly variable genes without cell cycle regression. The dataset was clustered using Louvain with 0.7 resolution followed by UMAP dimension reduction on top 20 PCs. In Extended Data Fig. 2d, we subset for treated and withdrawal cells (562 cells). We follow the same steps above

adjusting only clustering resolution set to 0.5. Detailed analyses can be found in <https://github.com/brickmanlab/wong-et-al-2022/>.

***In vivo* scRNA-seq re-analysis**—The Li et al.<sup>20</sup> dataset HRA000280 was downloaded from Genome Sequence Archive. Cells with low quality and mitochondrial content higher than 20% were filtered out (lower bound: 3,000 genes and upper bound: 9,000 genes; 400,000 UMIs). Additionally, cells labelled as “Poor quality” were also discarded. We followed the same preprocessing steps as mentioned above without clustering. We subsetted the final dataset for hMG, hHG, hFG and hAL population.

**Cluster alignment tool (CAT)**—We used CAT to determine similarity between clusters from *in vivo* and *in vitro* studies. CAT calculates mean gene expression of randomly sampled cells with replacement for each cluster 1,000 times. Euclidian distance is measured between all pairs of clusters. A small distance represents high similarity. A detailed explanation of the method can be found in Rothova et al<sup>19</sup>.

**Analysis of bulk RNA-seq data**—Fastq files from bulk RNA-seq samples were aligned to the hg38/GRCh38 genome using STAR v2.5.3a<sup>67</sup>. Transcript expression levels were estimated with the `--quantMode GeneCounts` option and GRCh38p10.v27 annotations. FastQC v0.11.7 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) was used for QC metrics, and multiqc v1.7<sup>68</sup> for reporting. Data analysis was then performed with R/Bioconductor<sup>69</sup> (<https://www.R-project.org>). Normalization was performed with DESeq2 (v1.24.0)<sup>70</sup>. The Lee et al.<sup>35</sup> dataset was retrieved from NCBI GEO (GSE114102) and analyzed as above. Differential Gene Expression was performed using DESeq2 (R package version 1.32.0). Z-scoring was calculated as previously described for each dataset separately. Gene Set Enrichment Analysis (GSEA) was performed by Webgestalt (<http://www.webgestalt.org>) (log2 fold change between VFG passage 3 and 6) for Gene Ontology Biological Process (GO-BP) with False Discovery Rate (FDR) < 0.05.

**Processing of ATAC-seq datasets**—The quality of the sequencing reads was assessed with FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) followed by trimming of poor-quality base calls and adaptor sequences with cutadapt<sup>71</sup>. Read-pairs were then aligned to the hg19 reference genome using bowtie2<sup>72</sup> with the following parameters: `bowtie2 --no-discordant --no-mixed --no-unal --very-sensitive -X 2000`. Samtools<sup>73</sup> was used for sorting alignments and format conversions. Alignments from PCR duplicates were removed using Picard (<http://broadinstitute.github.io/picard/>). Alignments were then converted into BED format using bedtools<sup>74</sup>. The 5' ends of the reads were offset by +4 bases for the reads on Watson strand and by -5 bases for the reads on Crick strand, to reflect the exact location of Tn5 insertion site. Single-base genome-wide coverage was computed using a 30 bp fragment centred at the Tn5 insertion site in BigWig format. We called peaks using Macs2<sup>75</sup> with following parameters: `macs2 callpeak --nomodel --extsize 150 --shift -75 -g 'hs' -p 0.01`. For each condition, data from two biological replicates was used to create a set of highly reproducible peaks using Irreproducible Discovery Rate (IDR <= 0.05, ref<sup>76</sup>). Deeptools<sup>77</sup> was employed to compute Pearson's correlation among the conditions/replicates and for PCA plots. Bedtools `intersect` command was used to find overlapping or

unique (with parameter “-v”) enhancer positions (bed format) between two conditions in question (Fig. 3b).

**Detection of differential chromatin accessibility and temporal dynamics of enhancers from ATAC-seq data**—A consensus set of ATAC-seq peaks was created using reproducible peaks from all five stages of differentiation. Next, we computed normalized read coverage (rpkm) for the consensus peak-set in all stages. General Linear Modelling (GLM) was applied to the normalized counts from the step above in order to detect changes in chromatin accessibility across the stages and in both directions. We used the following parameters for differential accessibility:  $\log_2\text{-fold} > 2$  or  $\log_2\text{-fold} < -2$  at adjusted  $p\text{-value} < 0.005$  (TC-seq, ref<sup>30</sup>). We then defined stage-specific peaks using c-means clustering of dynamic peak-set from the step above. We called 8 clusters that gave a functionally relevant pattern along the timeline of differentiation. Some clusters were merged, as they were too similar to be dealt with separately. This led to formation of the six groups of dynamic enhancers (Fig. 3c, right). RPKM normalized BigWig tracks from merged replicates were used to plot heatmaps in deepTools<sup>77</sup>. For locus-specific visualizations, we used UCSC Genome Browser (<http://genome.ucsc.edu>, ref<sup>78</sup>) to load BigWig tracks.

**Enrichment scoring of defined ATAC-clusters from the mapped gene sets that are up or down regulated at the PE stage compared to VFGp6**—ATAC-seq peaks were assigned to genes using GREAT<sup>79</sup> with the setting of single nearest gene within 25 or 200kb (Supplementary Table 1b). The enrichment of gene-annotated ATAC-clusters in differential expression gene sets was calculated by  $\log_2$  ratio between number of observed overlaps and number of expected overlaps from the dataset. We compared the impact of very low levels of background gene expression noise (those genes not reaching greater than 100 or 1000 reads in a particular sample, baseMean 100 or 1000) on these gene sets (Supplementary Table 1, d-g). While filtering out gene expression noise reduces the size of the gene set, it can be expanded by considering enhancers located within 200 Kb of a target gene.

**Motif analysis from ATAC-clusters**—Enrichment of known and de-novo transcription factor binding motifs was calculated with the HOMER v4.11.1 suite<sup>80</sup> using the *findMotifsGenome* function with default parameters.

**Hierarchical k-means clustering of expression patterns of genes annotated to ATAC-peaks clusters**—Bulk RNA-seq gene expression levels were normalized using DESeq2 R package version 1.32.0<sup>70</sup>. The mean of normalized expression was calculated for each condition and transformed into z-scores. Gene expression levels were then separated into the different annotated ATAC-peaks clusters. Finally, gene expression patterns were grouped using hierarchical clustering ( $k = 10$ ) based on Euclidian distances.

**Mapping and analysis of H3K27ac data from human embryo samples**—Preprocessing and alignment of ChIP-seq reads was as described in Gerrard et al.<sup>37</sup>. Single-end reads were aligned to hg19 genome assembly with bowtie 1.0.0 (parameters: -m1 -n 2 -1 28, uniquely mapped reads only). These alignments were received in compressed



BAM format from European Genome-Phenome Archive (<https://ega-archive.org/>) under accession no: EGAS00001003163 and EGAS0001004335. We converted the alignments to BED format and called peaks with HOMER (parameters: findPeaks -style histone) against a pooled input sample. We then used bedtools-2.30<sup>74</sup> to select the peaks present in both replicates (bedtools intersect -f 0.50 -r -u -a rep1.bed -b rep2.bed) of most tissue types except for stomach.

Lineage-specific sets of H3K27ac regions were generated by concatenating peaks from relevant tissues as follows: ectoderm (RPE; brain), endoderm (pancreas; liver; lung; stomach) and mesoderm (heart; adrenal). To identify unique regions for each germ layer, we use *bedtools intersect* command, followed by sorting regions using sort option and finally merging smaller regions which are subset of larger regions using *bedtools merge* command. This process ensures a unique count of peaks even if a given peak is part of a larger regulatory region. Similarly, we identified regions unique to tissue types. To map different ATAC clusters to the H3K27ac regions described above, we took regions in different enhancer classes and intersected these with different classes of H3K27ac regions from human fetal samples with *bedtools intersect* command. These overlaps were used in generating over-representation scores defined as  $\log_2$  (observed/expected).

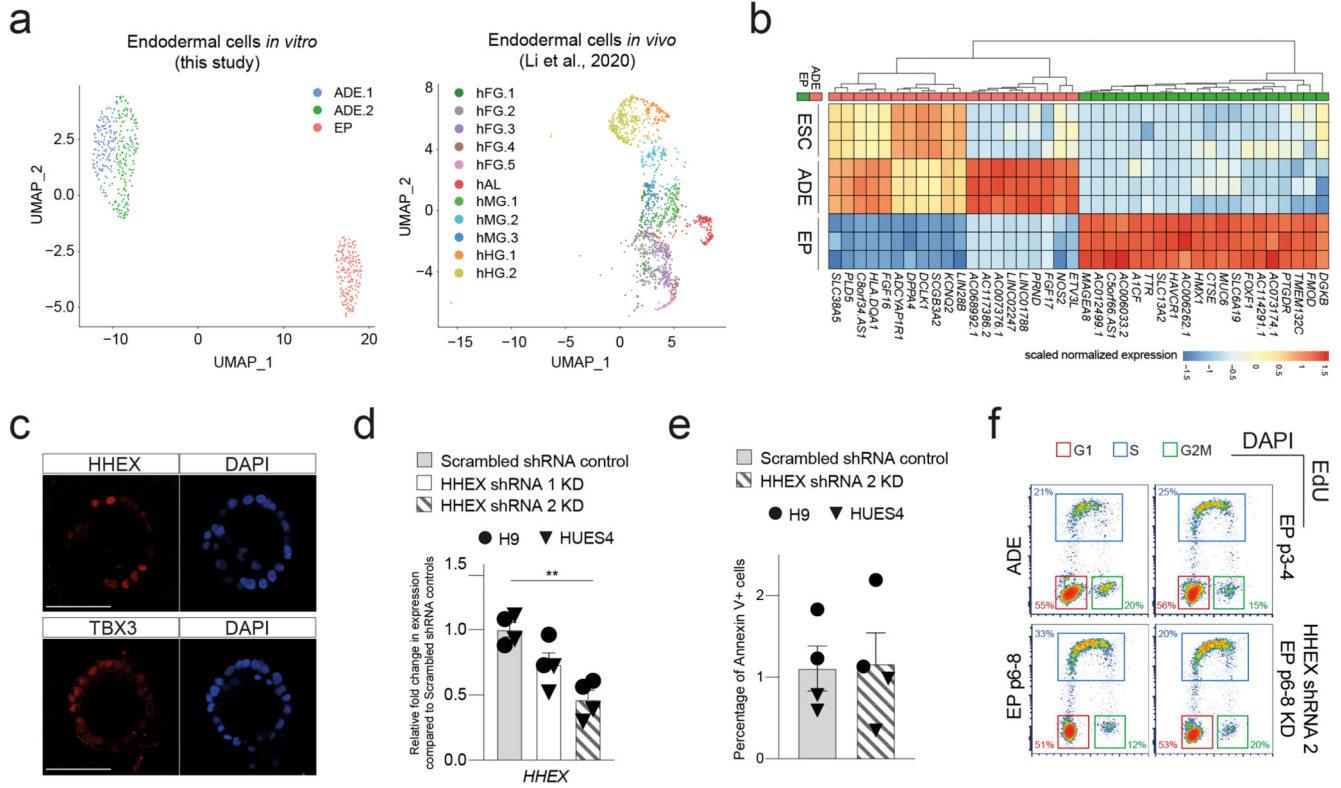
**Enrichment scoring of dynamic ATAC-seq clusters with H3K27Ac regions from human embryonic tissues**—The enrichment of ATAC-clusters in different lineage- and tissue-specific H3K27ac groups was calculated based on the ratio between number of observed overlapped regions (between ATAC and H3K27ac peaks) and number of expected overlapped regions from the datasets.

**Analysis of HHEX ChIP-seq dataset**—We aligned HHEX ChIP-seq data from Yang et al.<sup>42</sup> to hg19 assembly using bowtie-1.3.1<sup>81</sup> with default parameters and converted the alignments to HOMER tag-directory format. We created depth-normalized bigwig files using the HOMER<sup>80</sup> makeUCSCfile program. ComputeMatrix (deeptools suite<sup>77</sup>) was used to plot the coverage centered at the midpoint of enhancer regions in different classes (Extended Data Fig. 10f).

**Statistical Analyses and reproducibility**—No statistical methods were used to pre-determine sample size. Data distribution was assumed to be normal, but this was not formally tested. The experiments were not randomized. Data collection and analysis were not performed blind to the conditions of the experiments. No data points were excluded from the analyses. Data collection was performed using Microsoft Office Excel (16.16.2). Data representation and statistical analyses were performed using GraphPad Prism. Unless mentioned otherwise, data are shown as mean  $\pm$  SEM and *N* numbers refer to biologically independent replicates. Statistical significance ( $P < 0.05$ ) was determined as indicated in figure legends using one-way ANOVA Tukey's multiple comparison test (Figs. 1c, 1d, 2d, 4b, 4c; Extended data Figs 3d, 3e, 3f), one-way ANOVA Dunnett's multiple comparison test (Figs. 2a, 2b, 2g, 6b, 6c; Extended data Figs 2h, 2i, 9a, 9c, 10a, 10b), unpaired two-tailed t-test (Figs. 2d, 2e), unpaired one-tailed t-test (Figs. 6d, 6e, and 6f; Extended Data Figs. 10c, 10d, 10e, 10f), and Chi-squared test (Figs. 3f, 3g, 4d, 4e; Extended data Figs. 5d).



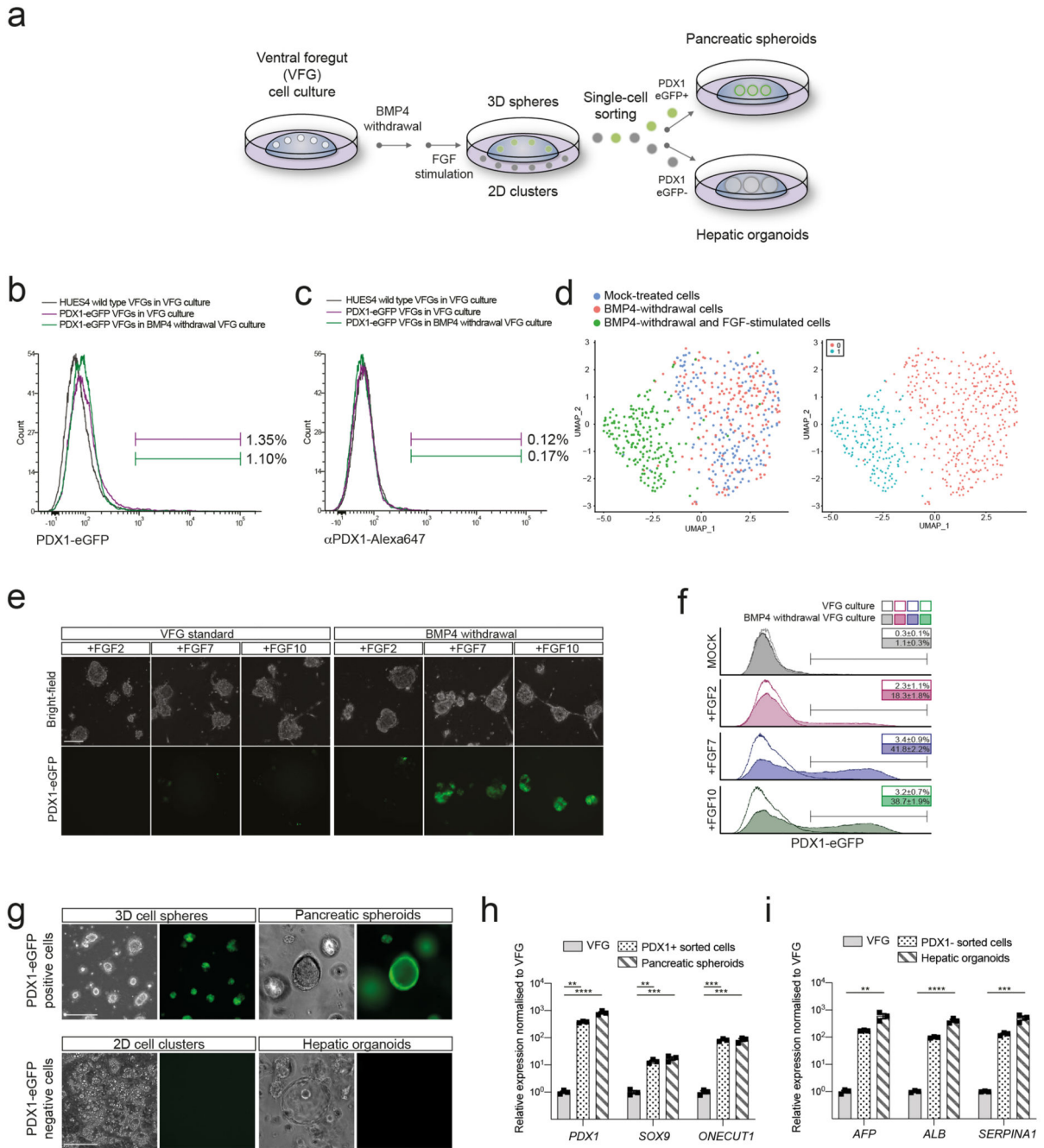
Extended Data



**Extended Data Fig. 1. Ventral foregut identity of expanding endodermal progenitors.**

**a**, Left: UMAP visualization of single cells from the transient ADE (ADE.1 and ADE.2) and EP passage 6 samples. Right: UMAP visualization of single cells from different endodermal populations from early human embryos reported in Li et al.<sup>20</sup>. **b**, Heatmap illustrating gene expression in H9-derived ESC, ADE, and EP cells ( $N=3$  independent experiments) from bulk RNA-seq dataset. Scaled normalized expression of the top 20 differentially expressed genes for each condition (ADE vs EP) is shown. **c**, Representative immunostaining of hAL markers, HHEX and TBX3, in EP passage 6 cells derived from H9 ESC cells. Images represent three independent experiments. Scale bar = 50  $\mu$ m. **d**, Expression analysis in HHEX shRNA KD cells (set 1 and set 2) and scrambled shRNA control by RT-qPCR. Relative fold change in mRNA of HHEX gene in KDs and control EP/VFG cells was assayed by RT-qPCR. Expression is normalized to *ACTB*. Circles and triangles mark cells derived from H9 and HUES4 WT ESCs respectively. Data are represented as mean  $\pm$  SEM ( $N=4$  independent experiments). Statistics analysis (\*\* $P<0.01$ , unpaired two-tailed t-test) was performed between KD and control EP/VFG cells. Comparisons without an indicated P value are not significant. **e**, Apoptosis assay in HHEX shRNA (set2) KD and scrambled shRNA control EP/VFG. Bar plot showing percentage of Annexin V+ cells for each assay. Circles and triangles mark cells derived from H9 and HUES4 WT ESCs respectively. Data are represented as mean  $\pm$  SEM ( $N=4$  independent experiments). No statistical difference (unpaired two-tailed t-test) was found between HHEX shRNA KD and scrambled shRNA control. **f**, Representative flow cytometry plots used to analyze the cell cycle in transient

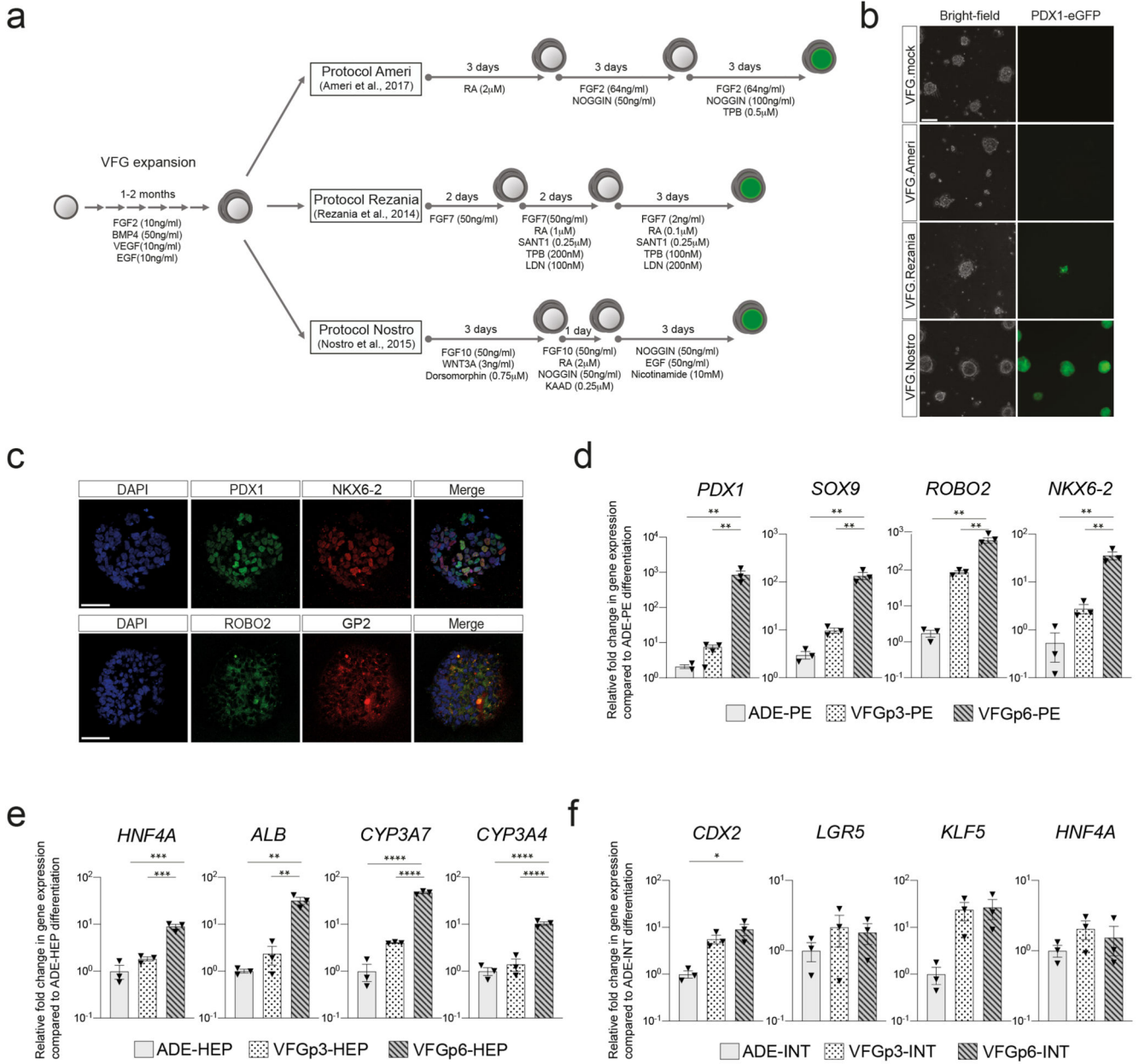
ADE, early EP/VFG (p3-4), expanding EP/VFG (p6-8) and HHEX depleted EP/VFG (p6-8). Cells were stained with EdU and DAPI. Cells in G1 (red), S (blue), and G2M (green) were gated and percentages of each fraction shown. Flow Cytometry plots represent three independent experiments.



**Extended Data Fig. 2. Human VFG cultures can be readily transformed to either pancreatic or hepatic lineages.**

**a.** Schematic representation showing the conversion of VFG culture to pancreatic and hepatic expansion. The figure illustrates the generation of PDX1-eGFP positive (PDX1<sup>+</sup>)

and negative (PDX1<sup>-</sup>) cells from VFG culture after BMP4 withdrawal and subsequent stimulation with FGF. **b-c**, Flow cytometry of eGFP expression (**b**) or intracellular PDX1 (**c**) for HUES4 wild type (grey), PDX1-eGFP reporter (purple) in VFG culture, and the reporter following BMP4 withdrawal (green). Fractions of eGFP<sup>+</sup> or PDX1<sup>+</sup> were gated and percentages are shown. Flow Cytometry plot represents three independent experiments. **d**, Left: UMAP visualization of 526 cells isolated from mock-treated VFG (blue), VFG cells grown in the absence of BMP4 (red), and transient pancreatic induction by FGF2 simulation (green). Right: UMAP visualization of Seurat clustering from the samples described on the left. **e**, Representative bright-field (top) and fluorescent (bottom) images for the PDX1-eGFP reporter VFGs (left) or following BMP4 withdrawal (right), and then treated with FGF2, FGF7, or FGF10. Images represent three independent experiments. Scale bar = 50  $\mu$ m. **f**, Flow cytometry of eGFP expression for the conditions described in (**e**), including mock-treated cells. Percentages of PDX1<sup>+</sup> cells were shown in the rectangle boxes of each histogram. Flow Cytometry plots represent three independent experiments. **g**, PDX1<sup>+</sup> cells form 3D spheres and expand as pancreatic spheroids (Top). PDX1<sup>-</sup> cells form 2D clusters and expand as hepatic organoids (bottom). Images represent three independent experiments. Scale bar = 50  $\mu$ m. **h-i**, Relative fold change in mRNA of pancreatic markers (*PDX1*, *SOX9*, and *ONTGUT1*) (**h**) and hepatic markers (*AFP*, *ALB*, and *SERPINA1*) (**i**) in the VFGs and VFG-derived cell types (as described in **a** and **g**). Expression is normalized with *ACTB*. Data are represented as mean  $\pm$  SEM ( $N=3$  independent experiments). \*\* $P<0.01$ , \*\*\* $P<0.001$ , \*\*\*\* $P<0.0001$  (one-way ANOVA Dunnett's multiple comparison test compared with VFG).

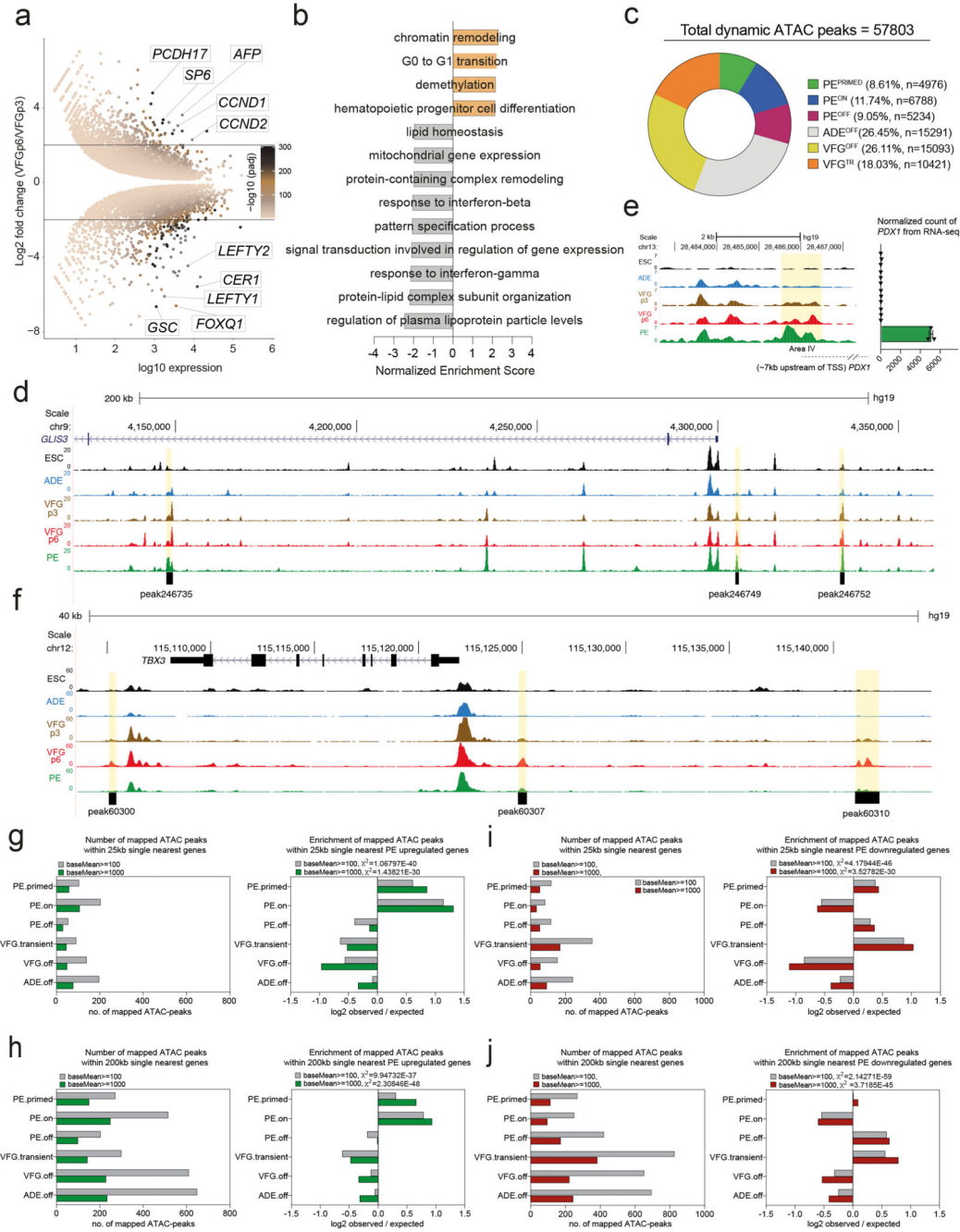


**Extended Data Fig. 3. *in vitro* differentiation of VFG culture towards pancreatic, hepatic, and intestinal endoderm.**

**a**, Schematic diagram for stepwise pancreatic differentiation with protocols from Ameri et al.<sup>22</sup>, Rezania et al.<sup>12</sup> and Nostro et al.<sup>10</sup> from established VFG culture. **b**, Representative bright-field (left) and fluorescent (right) images for the PDX1-eGFP reporter VFGs differentiated with protocols indicated. Images represent three independent experiments. Scale bar = 50  $\mu$ m. **c**, Representative immunostaining of PDX1 (green) and NKX6-2 (red) in the top row; ROBO2 (green) GP2 (red) in the bottom row, including DAPI (blue) for p6 VFG cells differentiated with protocol from Nostro et al.<sup>10</sup>. Images represent three independent experiments. Scale bar = 50  $\mu$ m. **d**, Bar plot showing relative fold change in mRNA of pancreatic markers *PDX1*, *SOX9*, *ROBO2*, and *NKX6-2* in pancreatic

differentiation from ADE and VFG at p3 and p6 cells. Data are represented as mean  $\pm$  SEM ( $N=3$  independent experiments). \*\* $P<0.01$ , (one-way ANOVA Tukey's multiple comparison test, only significant comparisons are shown). **e**, Bar plot showing relative fold change in mRNA of hepatic markers *HNF4A*, *ALB*, *CYP3A7*, and *CYP3A4* in hepatic differentiation from ADE and VFG at p3 and p6 cells. Data are represented as mean  $\pm$  SEM ( $N=3$  independent experiments). \*\* $P<0.01$ , \*\*\* $P<0.001$ , \*\*\*\* $P<0.0001$  (one-way ANOVA Tukey's multiple comparison test, only significant comparisons are shown). **f**, Bar plot showing relative fold change in mRNA of intestinal markers *CDX2*, *LGR5*, *KLF5*, and *HNF4A* in hepatic differentiation from ADE and VFG at p3 and p6 cells. Data are represented as mean  $\pm$  SEM ( $N=3$  independent experiments). \* $P<0.05$  (one-way ANOVA Tukey's multiple comparison test, only significant comparisons are shown). Comparisons without an indicated  $P$  value are not significant.



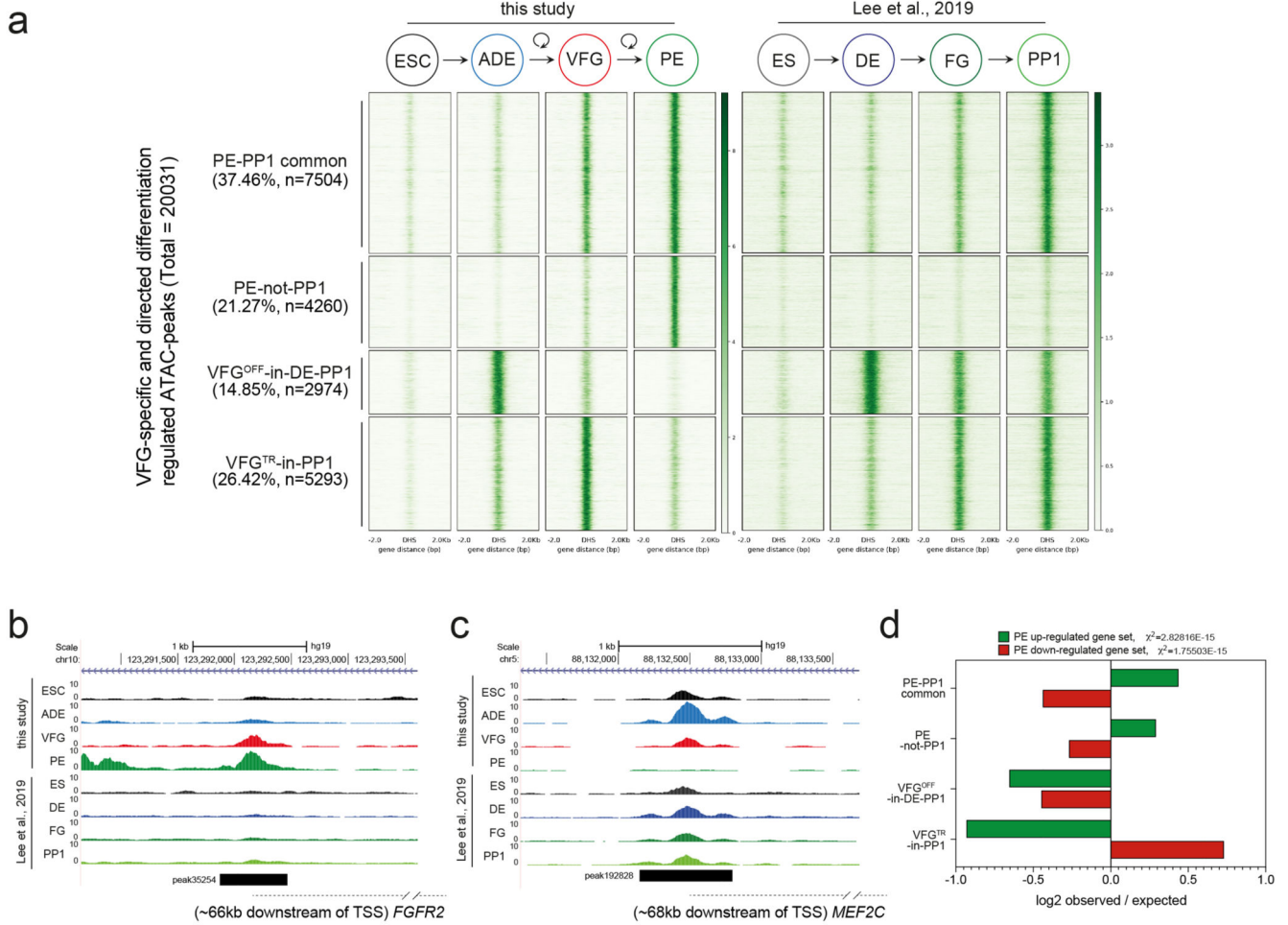


**Extended Data Fig. 4. The dynamic chromatin landscape and gene expression in VFG expansion and further differentiation.**

**a**, MA-plot representing differential expression in VFGp6 versus VFGp3 culture (Log2 fold change > 2,  $P < 0.05$ ) ( $N=3$  independent experiments). **b**, GSEA for GO-BP of VFGp6 compared to VFGp3 cells. Normalized Enrichment Score for significant terms for VFGp6 are shown as positive value, and VFGp3 as negative value (FDR < 0.05). **c**, Pie-chart showing distribution of dynamic ATAC-peaks (n=57803) with percentage and numbers of peak indicated per cluster in Fig. 3c. **d**, Representative UCSC Genome Browser



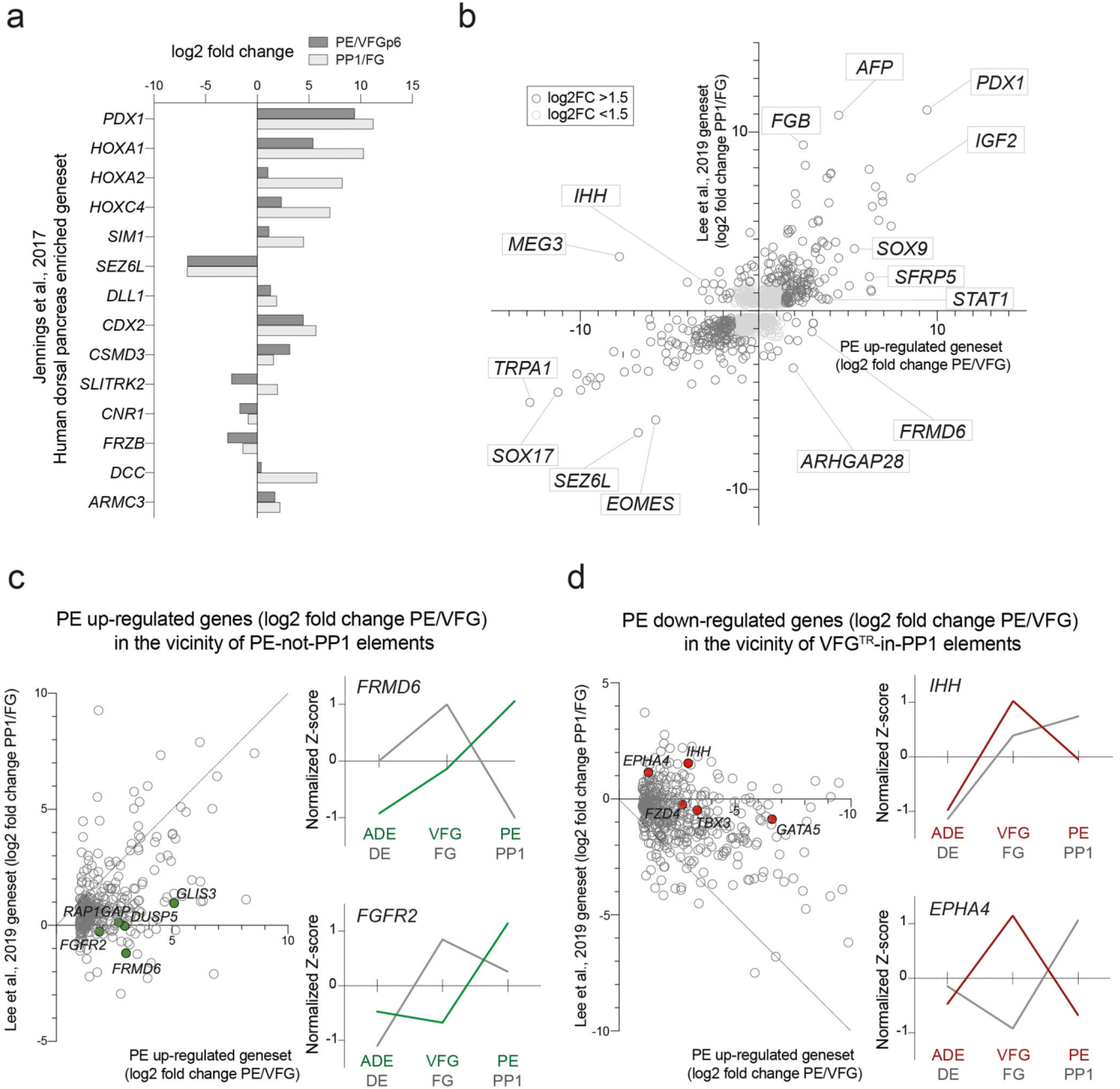
screen shot (from two independent experiments) at the *GLIS3* locus showing ATAC-seq data from ESC, ADE, VFGp3, VFGp6, and PE. Genome coordinates (bp) are from the hg19 assembly of the human genome. PE<sup>PRIMED</sup> elements (peaks 246735, 246749, and 246752) are shown at the bottom and the corresponding regions are highlighted in yellow. **e**, Left: Representative UCSC Genome Browser screen shot as in d. The region of the area IV enhancer is highlighted in yellow. Approximate distance between the region and *PDX1* TSS is indicated by a broken dashed line. Right: bar plots for expression (normalized RNA-seq counts,  $N=3$  independent experiments) for *PDX1* RNA across the same samples as ATAC-seq. **f**, Representative UCSC Genome Browser screen shot at the *TBX3* locus as in d. VFG<sup>TR</sup> elements (peaks 60300, 60307, and 60310) are shown at the bottom and the corresponding regions are highlighted in yellow. **g-h**, Mapping dynamic enhancer classes to gene expression (up-regulated genes). Left: Number of mapped ATAC peaks in each cluster defined in Fig. 3c located within 25 Kb (**g**) or 200 Kb (**h**) of the single nearest gene's TSS from the PE up-regulated gene set with baseMean >100 (grey) or >1000 (green). Right: Enrichment ( $\log_2$  observed/expected) of the PE up-regulated gene set with baseMean > 100 (grey) or >1000 (green), in proximity (within a 25 or 200 Kb window) to ATAC-clusters defined in Fig. 3c. **i-j**, Mapping dynamic enhancer classes to gene expression (down-regulated genes) for elements located within 25 Kb (**i**) or 200 Kb (**j**) of the single nearest gene's TSS from the PE down-regulated gene set, analysis and labels as in **g**. All data shown are significant by chi-square analysis.



**Extended Data Fig. 5. VFG expansion enables consolidation of an enhancer landscape that is imperfectly realized during directed differentiation.**

**a**, A comparison of chromatin accessibility of enhancers charted in this study (heatmap, left) with the Lee et al. dataset<sup>35</sup> (heatmap, right). Enhancers in the group “PE-PP1 common” are the pancreatic endoderm enhancers that are activated independent of VFG expansion (37.46%, n=7504). Enhancers in the group PE-not-PP1 are PE enhancers that are activated only if PE is differentiated from expanding VFGs (21.27%, n=4260). Enhancers in the “VFG<sup>OFF</sup>-in-DE-PP1” group, represent a subset of ADE enhancers that are inactivated during VFG expansion (14.85%, n=2974). The “VFG<sup>TR</sup>-in-PP1” enhancer group at the bottom of the heatmap (26.42%, n=5293) are inactivated in PE derived from expanding VFGs, but not in directed differentiation. **b-c**, Representative UCSC Genome Browser screen shot (from two independent experiments) showing examples of a PE-not-PP1 enhancer (peak35254), in an intron of the *FGFR2* locus (**b**) and a VFG<sup>OFF</sup>-in-DE-PP1 group (peak192828) contained within an intron of the *MEF2C* locus (**c**). Approximate distance between elements and TSS is indicated by a broken dashed line in each panel. **d**, Bar plot showing the prevalence (log<sub>2</sub> observed/expected) of ATAC peaks within a 200 Kb window from genes up-regulated (green) or down-regulated (red) between PE and VFGp6 across the

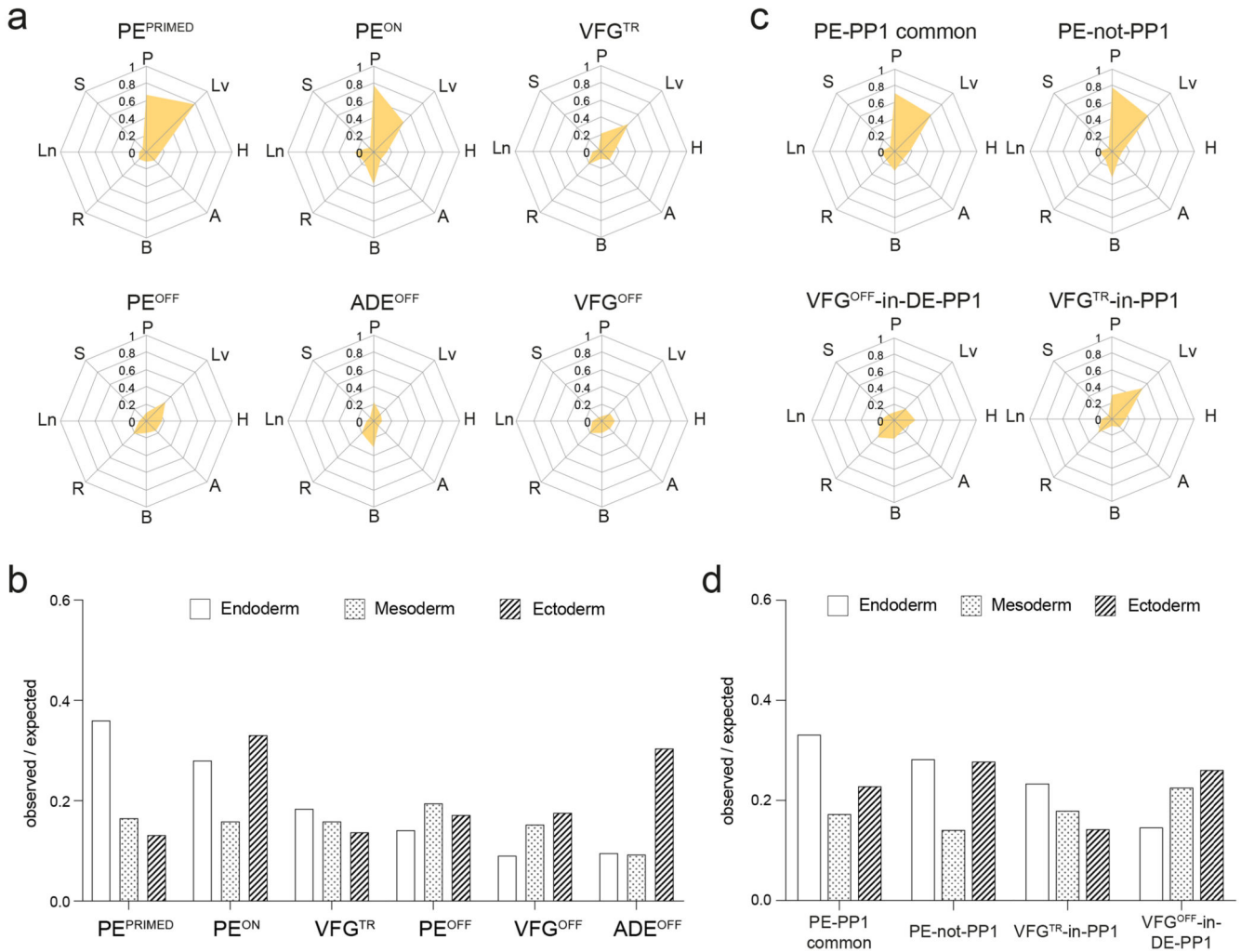
defined ATAC peak clusters. Genes considered have a base mean expression > 1000, log<sub>2</sub> fold change > 1.5 and *P* < 0.05. All data shown are significant in chi-square analysis.



**Extended Data Fig. 6. Characterization of global transcriptional changes between PE/VFG and PP1/FG cells.**

**a**, Bar plot showing differential expression of 13 dorsal pancreas markers (log<sub>2</sub> fold change) in PE/VFGp6 and PP1/FG from RNA-seq dataset (*N*=3 independent experiments). **b**, Scatter plot of differential expression for gene regulated in PE/VFGp6 (horizontal axis) and in PP1/FG (vertical axis) (*N*=3 independent experiments). **c-d**, Left: Scatter plot of differential expression for genes up-regulated (**c**) or down-regulated (**d**) in PE vs VFG

(as defined in Extended Data Fig. 5d), and within 200Kb of minimum one ATAC peak in the PE-not-PP1 (c) or VFG<sup>TR</sup>-in-PP1 (d) clusters respectively, vs their expression after directed differentiation (PP1/FG). The diagonal line indicates where there is no difference in differential expression between two comparisons (datasets). Right: normalized z-score expression of representative candidates (c: *FRMD6* and *FGFR2* and d: *IHH* and *EPHA4*). Normalized z-score expression for each candidate was plotted for the ADE, VFG (p6), and PE conditions (green in c and red in d), and the DE, FG, and PP1 conditions (grey) ( $N=3$  independent experiments).

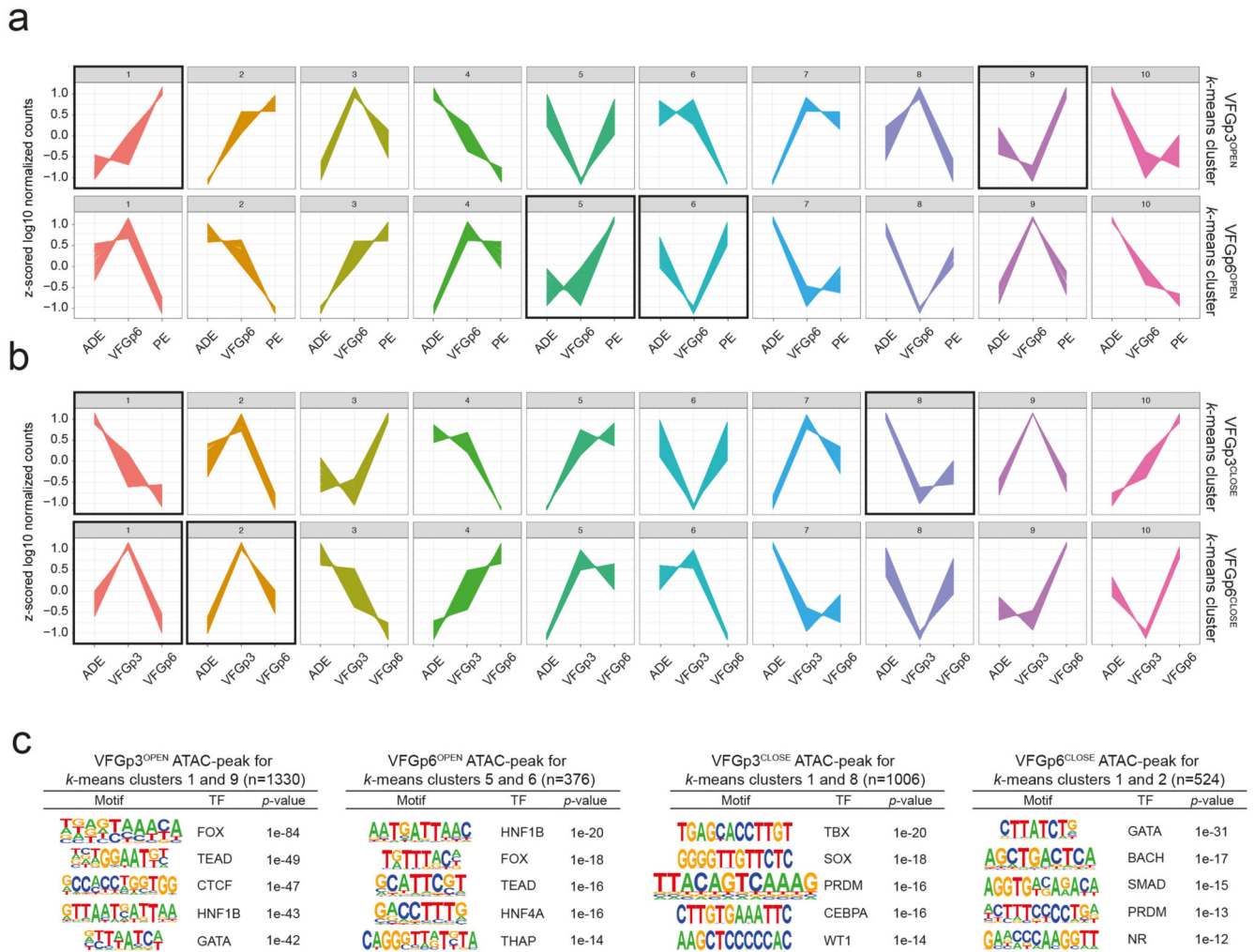


**Extended Data Fig. 7. VFG expansion insures higher fidelity regulation of enhancers normally exploited in fetal organogenesis**

**a-b**, Enrichment of tissue-specific (a) and lineage-specific (b) H3K27ac enhancers from human embryos (from two independent experiments for most tissue types, except for stomach where only one sample was available) in different ATAC clusters defined in Fig. 3c were displayed by enrichment score (observed/expected) in radar charts. **c-d**, Enrichment of tissue-specific (c) and lineage-specific (d) H3K27ac enhancers from human embryos (from two independent experiments for most tissue types, except for stomach where only

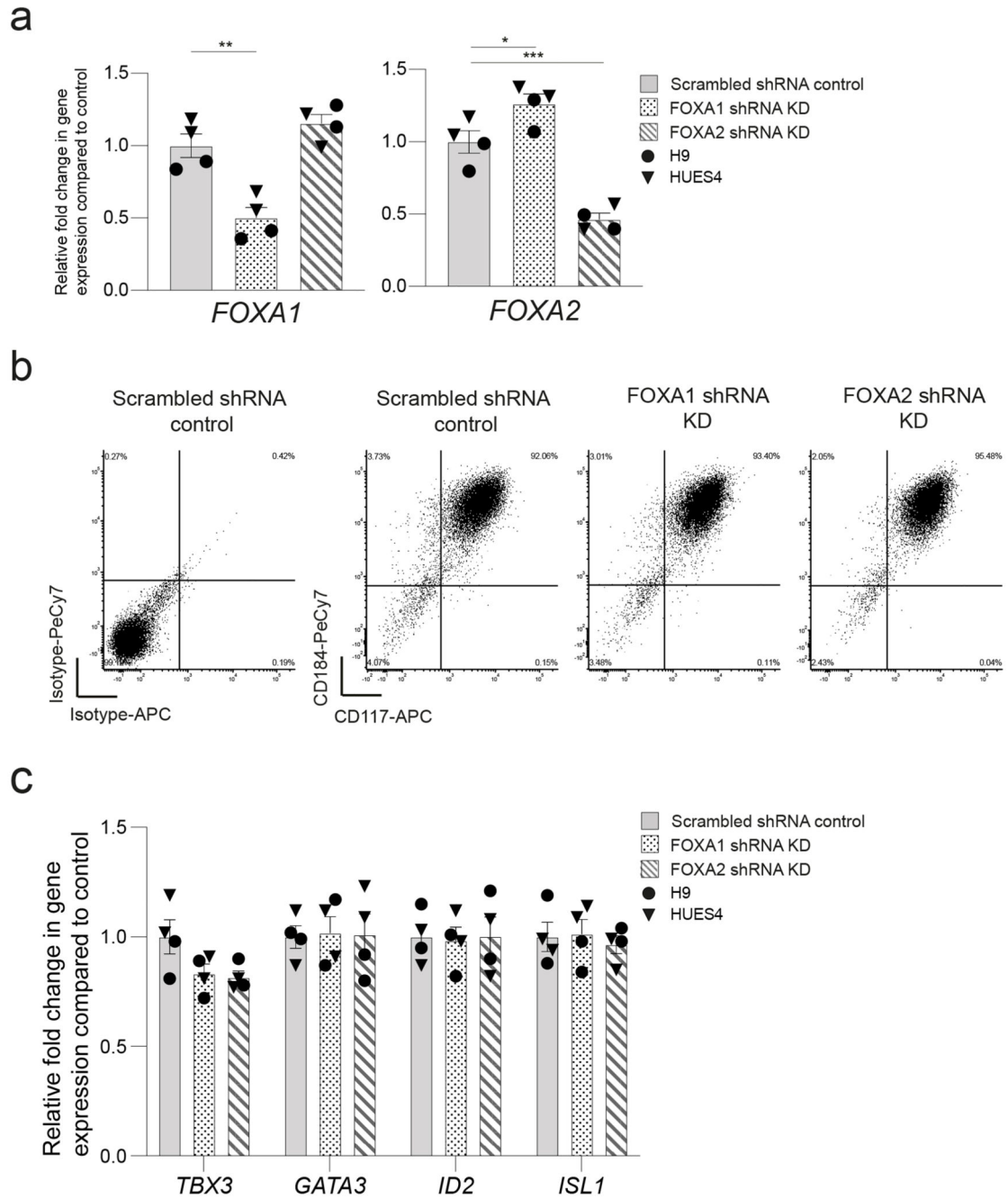


one sample was available) across different VFG-specific ATAC clusters defined in Extended Data Fig. 4a by enrichment score (observed/expected) in a radar chart. P: pancreas, Lv: liver, H: heart, A: adrenal, B: brain, R: RPE, Ln: lung, S: stomach.



**Extended Data Fig. 8. *k*-means clustering and motif analysis for VFG expansion dependent ATAC-clusters.**

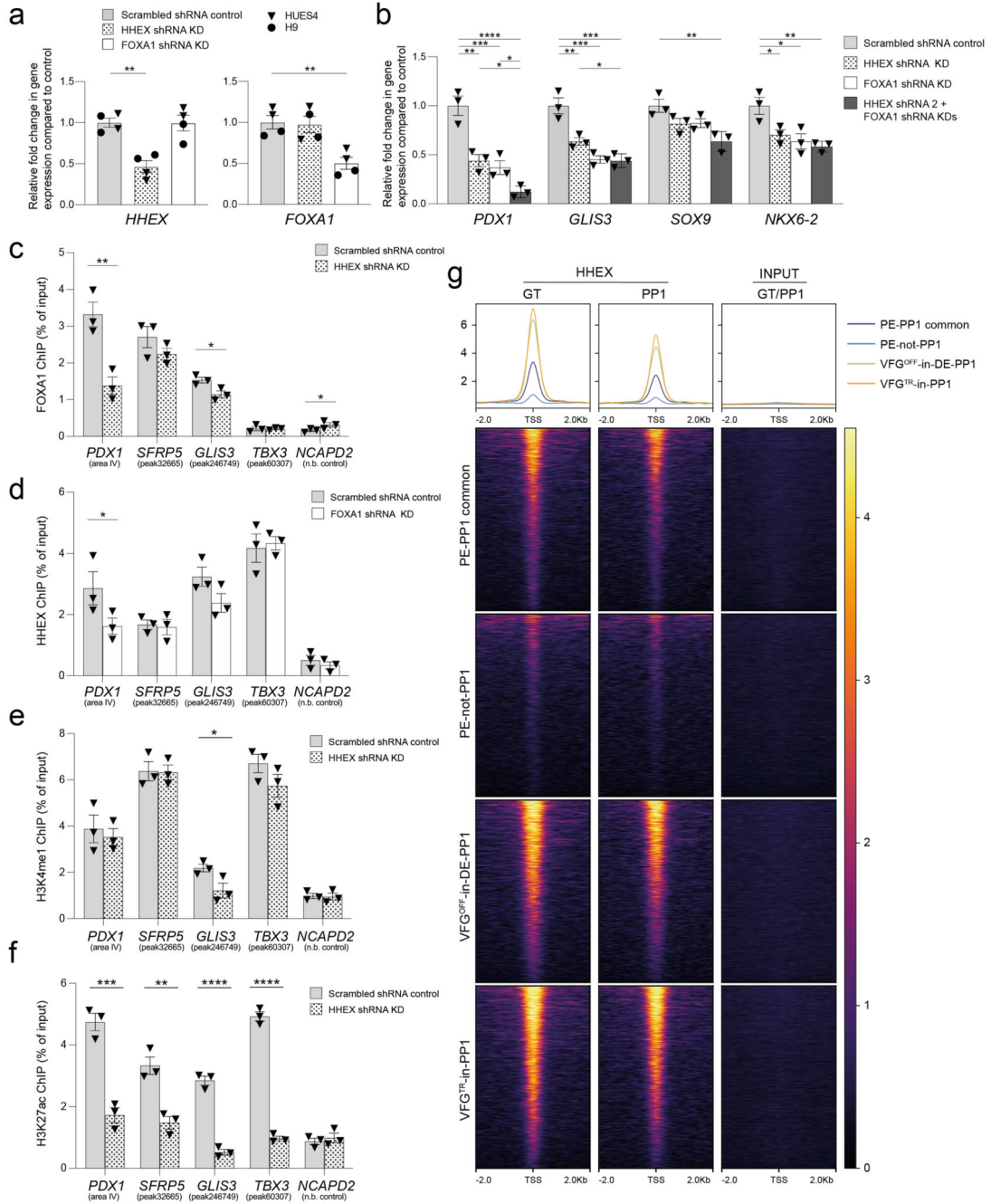
**a-b**, *k*-means clustering of genes within 200Kb of peaks in ATAC clusters as defined in Fig. 4a (VFGp3<sup>OPEN</sup>, VFGp6<sup>OPEN</sup>, VFGp3<sup>CLOSE</sup>, and VFGp6<sup>CLOSE</sup>). Z-scored log<sub>10</sub> normalized gene expression of ADE, VFGp6, and PE samples (**a**); and of ADE, VFGp3, and VFGp6 samples (**b**) were plotted for VFG<sup>OPEN</sup> and VFG<sup>CLOSE</sup> clustered genes respectively (n=10). **c**, De novo motif search was made using Homer findMotifsGenome and searched within ±200 bp of peak center for genes mapped to the vicinity of VFGp3<sup>OPEN</sup> (*k*-means clusters 1 and 9, n=1330), VFGp6<sup>OPEN</sup> (*k*-means clusters 5 and 6, n=376), VFGp3<sup>CLOSE</sup> (*k*-means clusters 1 and 8, n=1006), and VFGp6<sup>CLOSE</sup> (*k*-means clusters 1 and 2, n=524).



**Extended Data Fig. 9. Characterization of FOXA1 and FOXA2 shRNA KD VFG cells.**  
**a**, Expression analysis in FOXA1 and FOXA2 shRNA KD cells (described in Fig. 6b) by RT-qPCR. Expression of FOXA1 (left) and FOXA2 (right) in the KD cells was normalized relative to the expression in scrambled shRNA controls. Triangles and circles mark cells derived from HUES4 and H9 ESCs respectively. Data are represented as mean  $\pm$  SEM ( $N=4$  independent experiments). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$  (unpaired two-tailed t-test). Comparisons without an indicated  $P$  value are not significant. **b**, Representative flow cytometry density plots showing CD184 (CXCR4) and CD117 (KIT) expression



in scrambled shRNA control, FOXA1, and FOXA2 shRNA KD VFG cells. Bottom left quadrant indicates gating based on isotype staining controls in scrambled shRNA control VFG cells. Flow Cytometry plots represent three independent experiments. **c**, Expression analysis in FOXA1 and FOXA2 shRNA KD cells (described in Fig. 6b) by RT-qPCR. Expression of VFG markers *TBX3*, *GATA3*, *ID2*, and *ISL1* in the KD cells was normalized relative to that in scrambled shRNA controls. Triangles and circles mark cells derived from HUES4 and H9 ESCs respectively. Data are represented as mean  $\pm$  SEM ( $N=4$  independent experiments). No statistical difference (unpaired two-tailed t-test) was found in the comparisons.



**Extended Data Fig. 10. HHEX is also required alongside FOXA1 for enhancer priming in VFGs.**

**a**, RT-qPCR of HHEX (left) and FOXA1 (right) in the HHEX or FOXA1 KD cells. Expression was normalized relative to the scrambled shRNA controls. Triangles and circles mark cells derived from HUES4 and H9 ESCs respectively. Data are represented as mean  $\pm$  SEM ( $N=4$  independent experiments). \*\*  $P < 0.01$  (unpaired two-tailed t-test, only significant comparisons are shown). **b**, Differentiation of scrambled control, HHEX KD, FOXA1 KD and HHEX/FOXA1 double KDs VFG cells to PE. Relative fold change of pancreatic genes (*PDX1*, *GLIS3*, *SOX9* and *NKX6-2*) was assayed by RT-qPCR and normalized relative to

the scrambled shRNA controls. Data are represented as mean  $\pm$  SEM ( $N=3$  independent experiments).  $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$  (one-way ANOVA Dunnett's multiple comparison test compared with scramble control and HHEX/FOXA1 double KDs). **c-d**, FOXA1 (**c**) and HHEX (**d**) binding enrichment by ChIP-qPCR at enhancer regions of *PDX1* (area IV), *SFRP5* (peak32665), *GLIS3* (peak246749), and *TBX3* (peak60307) in HHEX and FOXA1 shRNA KD VFG and scrambled control cell lines. An intragenic region of *NCAPD2* served as non-bound control. Data are represented as mean  $\pm$  SEM ( $N=3$  independent experiments). Statistics analysis was performed between KD and control VFG cells ( $*P < 0.05$ ,  $**P < 0.01$ , unpaired one-tailed t-test, only significant comparisons are shown). **e-f**, H3K4me1 (**e**) and H3K27ac (**f**) enrichment by ChIP-qPCR at enhancer regions of *PDX1* (area IV), *SFRP5* (peak32665), *GLIS3* (peak246749), and *TBX3* (peak60307) in HHEX shRNA KD VFG and scrambled control cell lines. An intragenic region of *NCAPD2* served as a non-bound control. Data are represented as mean  $\pm$  SEM ( $N=3$  independent experiments). Statistics analysis was performed between the KD and control VFG cells ( $*P < 0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$ , unpaired one-tailed t-test, only significant comparisons are shown). **g**, HHEX signal plotted on VFG-specific enhancer classes (PE-PP1 common, PE-not-PP, VFG<sup>OFF</sup>-in-DE-PP1, and VFG<sup>TR</sup>-in-PP1) at the FG and PP1 stages of directed differentiation (ChIP-seq dataset<sup>42</sup>) (from two independent experiments).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

We thank Paul Gadue for sharing protocols for EP expansion; Henrik Semb for the HUES4 WT and PDXeG clone 170-3 cell lines; we thank the reNEW Genomics Platform, reNEW Flow Cytometry Platform, the reNEW Imaging Platform and the reNEW Stem Cell Culture Platform for training, technical expertise, support and the use of instruments. We also thank members of the Brickman and Bickmore labs for critical comments on this manuscript. We are grateful to Anne G. Botton for critical reading of the manuscript. This work was funded by HumEn under the European Union Seventh Framework Programme FP7/2007-2013 (HEALTH-F4-2013-602889). J.M.B. was supported by the Danmarks Frie Forskningsfond (DFF-6110-00009) and Lundbeckfonden (R198-2015-412). W.A.B. was supported by MRC University Unit grant (MC\_UU\_00007/2). N.A.H. was supported by MRC (MR/000638/1 and MR/S036121/1). R.E.J. is a Diabetes UK Harry Keen Clinician Scientist fellow. J.A.R.H. was also supported by the Novo Nordisk Foundation (grant number NNF200C0063268). R.S.M. was supported by a Lundbeckfonden post-doctoral fellowship (R303-2018-2939). The Novo Nordisk Foundation Center for Stem Cell Medicine is supported by Novo Nordisk Foundation (grant number NNF21CC0073729 and previously NNF17CC0027852).

## Data availability

Sequencing data generated in this study are available on NCBI GEO with accession GSE185670 (bulk RNA-seq), GSE188362 (single-cell RNA-seq), and GSE108623 (ATAC-seq). The Lee et al.<sup>35</sup> dataset is from NCBI GEO with accession GSE114102. H3K27ac ChIP-seq dataset of human embryos is from our previous study (Gerrard et al.<sup>37</sup>) and is available on European Genome Phenome repository (EGAS00001004335 and EGAS00001003163). Processed data and gene lists from various analysis are included as Supplementary Tables. Source data are provided with this study. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

Cell lines and reagents generated for this study are available from the corresponding author with a complete Materials Transfer Agreement.

## Code Availability

Code used to perform the analyses in this study is available at <https://github.com/brickmanlab/wong-et-al-2022/> or from the corresponding authors upon request.

## References

1. Liu L, Michowski W, Kolodziejczyk A, Sicinski P. The cell cycle in stem cell proliferation, pluripotency and differentiation. *Nat Cell Biol.* 2019; 21: 1060–1067. [PubMed: 31481793]
2. Pauklin S, Vallier L. The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity. *Cell.* 2013; 155: 135–147. [PubMed: 24074866]
3. Wells JM, Melton DA. Vertebrate Endoderm Development. *Annu Rev Cell Dev Biol.* 1999; 15: 393–410. [PubMed: 10611967]
4. Miller SA, et al. Domains of differential cell proliferation suggest hinged folding in avian gut endoderm. *Dev Dyn.* 1999; 216: 398–410. [PubMed: 10633859]
5. Deutsch G, Jung J, Zheng M, Lora J, Zaret KS. A bipotential precursor population for pancreas and liver within the embryonic endoderm. *Development.* 2001; 128: 871–881. [PubMed: 11222142]
6. Tremblay KD, Zaret KS. Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. *Dev Biol.* 2005; 280: 87–99. [PubMed: 15766750]
7. Willnow D, et al. Quantitative lineage analysis identifies a hepato-pancreato-biliary progenitor niche. *Nature.* 2021; 597: 87–91. [PubMed: 34433966]
8. D'Amour KA, et al. Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol.* 2005; 23: 1534–1541. [PubMed: 16258519]
9. Yasunaga M, et al. Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nat Biotechnol.* 2005; 23: 1542–1550. [PubMed: 16311587]
10. Nostro MC, et al. Efficient Generation of NKX6-1+ Pancreatic Progenitors from Multiple Human Pluripotent Stem Cell Lines. *Stem Cell Reports.* 2015; 4: 591–604. [PubMed: 25843049]
11. Pagliuca FW, et al. Generation of functional human pancreatic  $\beta$  cells in vitro. *Cell.* 2014; 159: 428–439. [PubMed: 25303535]
12. Rezaia A, et al. Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. *Nature Biotechnology.* 2014; 32: 1121–1133.
13. Hay DC, et al. Efficient differentiation of hepatocytes from human embryonic stem cells exhibiting markers recapitulating liver development in vivo. *Stem Cells.* 2008; 26: 894–902. [PubMed: 18238852]
14. Hay DC, et al. Highly efficient differentiation of hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. *PNAS.* 2008; 105: 12301–12306. [PubMed: 18719101]
15. Cheng X, et al. Self-renewing endodermal progenitor lines generated from human pluripotent stem cells. *Cell Stem Cell.* 2012; 10: 371–384. [PubMed: 22482503]
16. Hannan NRF, et al. Generation of Multipotent Foregut Stem Cells from Human Pluripotent Stem Cells. *Stem Cell Reports.* 2013; 1: 293–306. [PubMed: 24319665]
17. Morrison GM, et al. Anterior Definitive Endoderm from ESCs Reveals a Role for FGF Signaling. *Cell Stem Cell.* 2008; 3: 402–415. [PubMed: 18940732]
18. Jennings RE, et al. Development of the human pancreas from foregut to endocrine commitment. *Diabetes.* 2013; 62: 3514–3522. [PubMed: 23630303]
19. Rothová MM, et al. Identification of the central intermediate in the extra-embryonic to embryonic endoderm transition through single-cell transcriptomics. *Nat Cell Biol.* 2022; 24: 833–844. [PubMed: 35681011]
20. Li L-C, et al. Single-cell patterning and axis characterization in the murine and human definitive endoderm. *Cell Research.* 2020; 1–19. DOI: 10.1038/s41422-020-00426-0 [PubMed: 31802008]

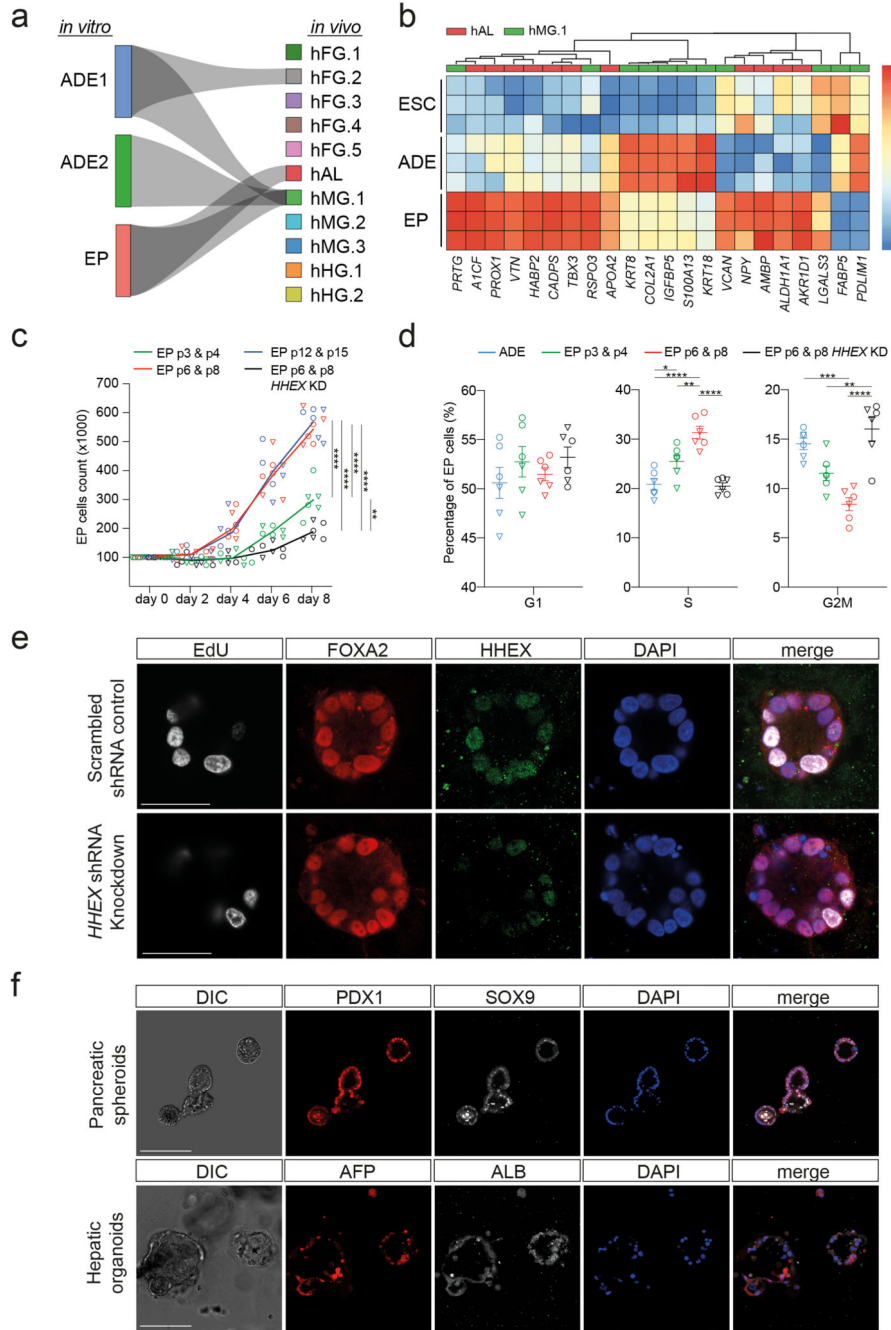
21. Bort R. Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. *Development*. 2004; 131: 797–806. [PubMed: 14736744]
22. Ameri J, et al. Efficient Generation of Glucose-Responsive Beta Cells from Isolated GP2 + Human Pancreatic Progenitors. *Cell Reports*. 2017; 19: 36–49. [PubMed: 28380361]
23. Akbari S, et al. Robust, Long-Term Culture of Endoderm-Derived Hepatic Organoids for Disease Modeling. *Stem Cell Reports*. 2019; 13: 627–641. [PubMed: 31522975]
24. Gonçalves CA, et al. A 3D system to model human pancreas development and its reference single-cell transcriptome atlas identify signaling pathways required for progenitor expansion. *Nat Commun*. 2021; 12 3144 [PubMed: 34035279]
25. Cogger KF, et al. Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. *Nature Communications*. 2017; 8 331
26. Escot S, Willnow D, Naumann H, Di Francescantonio S, Spagnoli FM. Robo signalling controls pancreatic progenitor identity by regulating Tead transcription factors. *Nature Communications*. 2018; 9 5082
27. El-Khairi R, et al. Modeling HNF1B-associated monogenic diabetes using human iPSCs reveals an early stage impairment of the pancreatic developmental program. *Stem Cell Reports*. 2021; 16: 2289–2304. [PubMed: 34450036]
28. Debacker C, Catala M, Labastie MC. Embryonic expression of the human GATA-3 gene. *Mech Dev*. 1999; 85: 183–187. [PubMed: 10415360]
29. Mukherjee S, French DL, Gadue P. Loss of TBX3 enhances pancreatic progenitor generation from human pluripotent stem cells. *Stem Cell Reports*. 2021; 16: 2617–2627. [PubMed: 34653400]
30. Wu, Mengjun; Gu, Lei. TCseq: Time course sequencing data analysis. 2021.
31. Scoville DW, Kang HS, Jetten AM. Transcription factor GLIS3: Critical roles in thyroid hormone biosynthesis, hypothyroidism, pancreatic beta cells and diabetes. *Pharmacology Therapeutics*. 2020; 215 107632 [PubMed: 32693112]
32. 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*. 2006; 20: 253–266. [PubMed: 16418487]
33. Gao N, et al. Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Development*. 2008; 22: 3435–3448. [PubMed: 19141476]
34. Ang LT, et al. A Roadmap for Human Liver Differentiation from Pluripotent Stem Cells. *Cell Reports*. 2018; 22: 2190–2205. [PubMed: 29466743]
35. Lee K, et al. FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation. *Cell Reports*. 2019; 28: 382–393. e7 [PubMed: 31291575]
36. Jennings RE, et al. Laser Capture and Deep Sequencing Reveals the Transcriptomic Programmes Regulating the Onset of Pancreas and Liver Differentiation in Human Embryos. *Stem Cell Reports*. 2017; 9: 1387–1394. [PubMed: 29056335]
37. Gerrard DT, et al. Dynamic changes in the epigenomic landscape regulate human organogenesis and link to developmental disorders. *Nat Commun*. 2020; 11 3920 [PubMed: 32764605]
38. Costa RH, Kalinichenko VV, Holterman A-XL, Wang X. Transcription factors in liver development, differentiation, and regeneration. *Hepatology*. 2003; 38: 1331–1347. [PubMed: 14647040]
39. Cebola I, et al. TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat Cell Biol*. 2015; 17: 615–626. [PubMed: 25915126]
40. Golson ML, Kaestner KH. Fox transcription factors: from development to disease. *Development*. 2016; 143: 4558–4570. [PubMed: 27965437]
41. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev*. 2011; 25: 2227–2241. [PubMed: 22056668]
42. Yang D, et al. CRISPR screening uncovers a central requirement for HHEX in pancreatic lineage commitment and plasticity restriction. *Nat Cell Biol*. 2022; 24: 1064–1076. [PubMed: 35787684]
43. Spence JR, et al. Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. *Dev Cell*. 2009; 17: 62–74. [PubMed: 19619492]



44. Schiesser JV, Wells JM. Generation of  $\beta$  cells from human pluripotent stem cells: Are we there yet? *Ann N Y AcadSci.* 2014; 1311: 124–137.
45. Zaret KS. Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. *Nat Rev Genet.* 2008; 9: 329–340. [PubMed: 18398419]
46. Iwafuchi M, et al. Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat Genet.* 2020; 52: 418–427. [PubMed: 32203463]
47. Wang A, et al. Epigenetic Priming of Enhancers Predicts Developmental Competence of hESC-Derived Endodermal Lineage Intermediates. *Cell Stem Cell.* 2015; 16: 386–399. [PubMed: 25842977]
48. Anderson KGV, et al. Insulin fine-tunes self-renewal pathways governing naive pluripotency and extra-embryonic endoderm. *Nat Cell Biol.* 2017; 19: 1164–1177. [PubMed: 28945231]
49. Tesar PJ, et al. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature.* 2007; 448: 196–199. [PubMed: 17597760]
50. Mac Auley A, Werb Z, Mirkes PE. Characterization of the unusually rapid cell cycles during rat gastrulation. *Development.* 1993; 117: 873–883. [PubMed: 8325243]
51. Snow MH, Bennett D. Gastrulation in the mouse: assessment of cell populations in the epiblast of tw18/tw18 embryos. *J Embryol Exp Morphol.* 1978; 47: 39–52. [PubMed: 722232]
52. Hamilton WB, et al. Dynamic lineage priming is driven via direct enhancer regulation by ERK. *Nature.* 2019; 575: 355–360. [PubMed: 31695196]
53. Mullen AC, et al. Master Transcription Factors Determine Cell-Type-Specific Responses to TGF- $\beta$  Signaling. *Cell.* 2011; 147: 565–576. [PubMed: 22036565]
54. Caravaca JM, et al. Bookmarking by specific and nonspecific binding of FoxA1 pioneer factor to mitotic chromosomes. *Genes Dev.* 2013; 27: 251–260. [PubMed: 23355396]
55. Tiyaonchai A, et al. GATA6 Plays an Important Role in the Induction of Human Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic  $\beta$  Cells. *Stem Cell Reports.* 2017; 8: 589–604. [PubMed: 28196690]
56. Yiangou L, Ross ADB, Goh KJ, Vallier L. Human Pluripotent Stem Cell-Derived Endoderm for Modeling Development and Clinical Applications. *Cell Stem Cell.* 2018; 22: 485–499. [PubMed: 29625066]
57. Gadue P, Huber TL, Paddison PJ, Keller GM. Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells. *Proc Natl Acad Sci U S A.* 2006; 103: 16806–16811. [PubMed: 17077151]
58. Rodrigues OR, Monard S. A rapid method to verify single-cell deposition setup for cell sorters. *Cytometry Part A.* 2016; 89: 594–600.
59. Jaitin DA, et al. Massively parallel single cell RNA-Seq for marker-free decomposition of tissues into cell types. *Science.* 2014; 343: 776–779. [PubMed: 24531970]
60. Untergasser A, et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 2012; 40: e115. [PubMed: 22730293]
61. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol.* 2015; 109: 21.29.1–21.29.9.
62. Müller I, et al. MPP8 is essential for sustaining self-renewal of ground-state pluripotent stem cells. *Nat Commun.* 2021; 12 3034 [PubMed: 34031396]
63. Keren-Shaul H, et al. MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. *Nat Protoc.* 2019; 14: 1841–1862. [PubMed: 31101904]
64. Jaitin DA, et al. Massively parallel single cell RNA-Seq for marker-free decomposition of tissues into cell types. *Science.* 2014; 343: 776–779. [PubMed: 24531970]
65. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods.* 2015; 12: 357–360. [PubMed: 25751142]
66. Hao Y, et al. Integrated analysis of multimodal single-cell data. *Cell.* 2021; 184: 3573–3587. e29 [PubMed: 34062119]

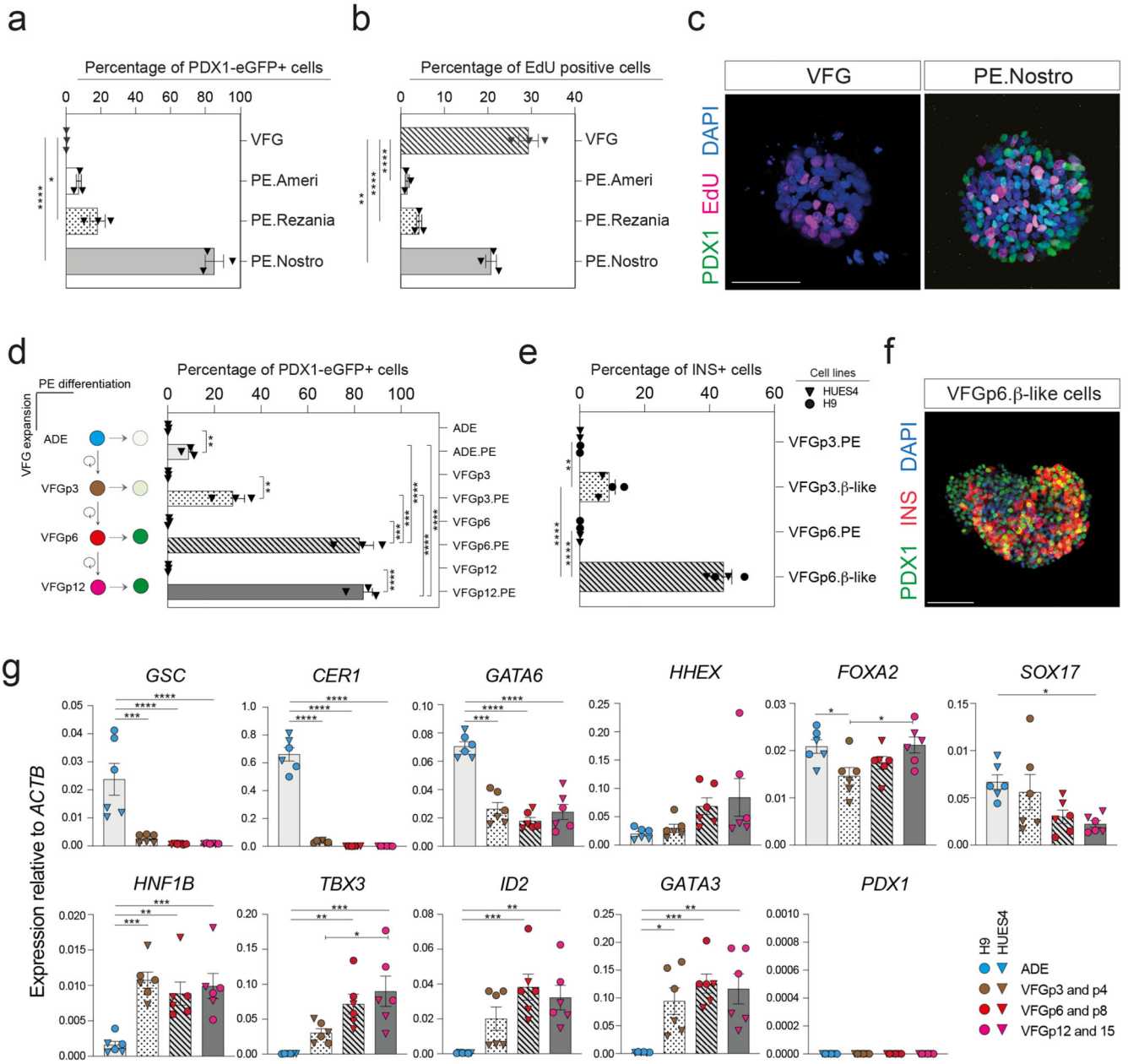


67. Dobin A, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013; 29: 15–21. [PubMed: 23104886]
68. Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016; 32: 3047–3048. [PubMed: 27312411]
69. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biology*. 2004; 5 R80 [PubMed: 15461798]
70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*. 2014; 15 550 [PubMed: 25516281]
71. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011; 17: 10–12.
72. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9: 357–359. [PubMed: 22388286]
73. Li H, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25: 2078–2079. [PubMed: 19505943]
74. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26: 841–842. [PubMed: 20110278]
75. Zhang Y, et al. Model-based Analysis of ChIP-Seq (MACS). *Genome Biology*. 2008; 9 R137 [PubMed: 18798982]
76. Li Q, Brown JB, Huang H, Bickel PJ. Measuring reproducibility of high-throughput experiments. *The Annals of Applied Statistics*. 2011; 5: 1752–1779.
77. Ramírez F, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016; 44: W160–165. [PubMed: 27079975]
78. Kent WJ, et al. The Human Genome Browser at UCSC. *Genome Res*. 2002; 12: 996–1006. [PubMed: 12045153]
79. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010; 28: 495–501. [PubMed: 20436461]
80. Heinz S, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010; 38: 576–589. [PubMed: 20513432]
81. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10 R25 [PubMed: 19261174]



**Fig. 1. Expanding endoderm progenitors as an *in vitro* model for ventral foregut.**  
**a**, Visualization of the CAT alignments between *in vitro* clusters (ADE.1, ADE.2, and EP) from this study and *in vivo* endodermal clusters from the Li et al. dataset<sup>20</sup>. Only significant CAT alignments between clusters are shown. **b**, Heatmap showing expression of hAL and hMG marker genes in ESC, ADE and EP cells (Bulk RNA-seq dataset, scaled normalized expression,  $N=3$  independent experiments). Only markers expressed significantly different between ADE and EP are shown ( $\log_2$  fold change > 1.5, adjusted  $P < 0.05$ ). **c**, Cumulative growth curves showing EP cell counts at different passages of expansion for control and

HHEX knockdown (EPs were derived from H9, circle, or HUES4, triangle, ESCs). Data are represented as mean  $\pm$  SEM ( $N=6$  independent experiments).  $**P<0.01$ ,  $****P<0.0001$  (one-way ANOVA Tukey's multiple comparison test was applied to analyze differences at day 8, only significant comparisons are shown). **d**, Dot plots showing percentage of G1, S, and G2M cycling cells assayed by flow cytometry with EdU and DAPI staining in control and HHEX knockdown EP expansion. Data are represented as mean  $\pm$  SEM ( $N=6$  independent experiments).  $*P<0.05$ ,  $**P<0.01$ ,  $***P<0.001$ ,  $****P<0.0001$  (oneway ANOVA Tukey's multiple comparison test, only significant comparisons are shown). **e**, Representative images (from three independent experiments) of control (top row) and HHEX shRNA (bottom row) EP cells stained with EdU, FOXA2, HHEX, DAPI. Scale bar = 50 $\mu$ m. **f**, Top row: representative immunostaining of PDX1 and SOX9, including DAPI, of VFG-derived pancreatic spheroids at passage 5. Bottom row: representative immunostaining of AFP and ALB, including DAPI, of VFG-derived hepatic organoids at passage 5. Images represent three independent experiments. Scale bar = 50 $\mu$ m.



**Fig. 2. Expansion enhances pancreatic differentiation of VFG cells.**

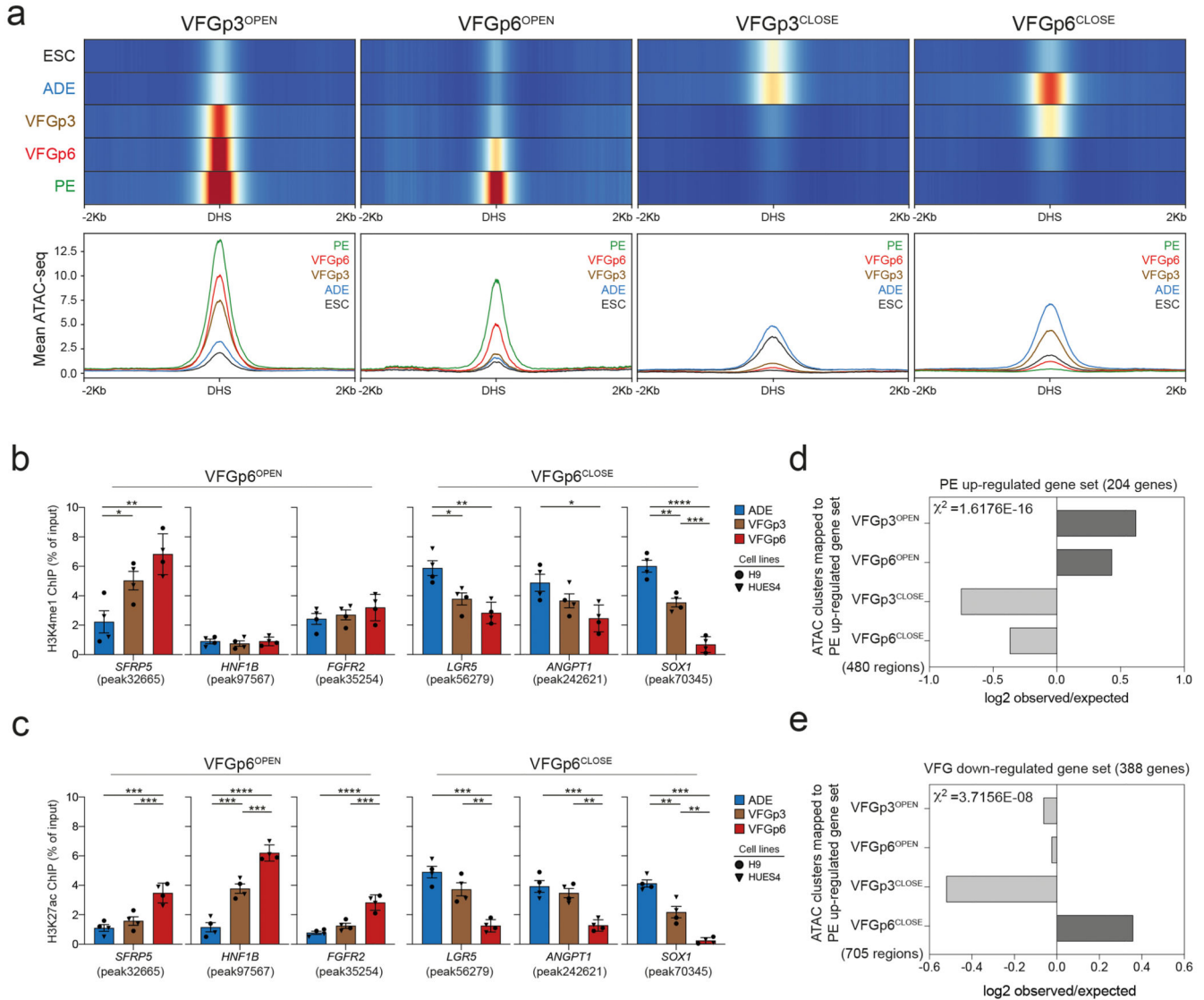
**a-b**, Bar plots showing percentage of (a) PDX1-eGFP<sup>+</sup> or (b) EdU<sup>+</sup> cells from flow cytometry analysis in VFG cells, and PE generated from VFG cells based on different differentiation protocols. Data are represented as mean ± SEM (N=3 independent experiments). \*P < 0.05, \*\*\*\*P < 0.0001 (one-way ANOVA Dunnett's multiple comparison test compared with VFG cells). **c**, Representative immunostaining (from three independent experiments) of VFG cells and PE, generated using conditions from Nostro et al.<sup>10</sup>, stained with PDX1, EdU, DAPI. Scale bar = 50 μm. **d**, Left: schematic of PE differentiation using conditions from Nostro et al.<sup>10</sup> from ADE, VFG at p3, p6, and p12. Right: Bar plot showing percentage GFP<sup>+</sup> positive cells generated for the indicated conditions. Data are represented

as mean  $\pm$  SEM ( $N=3$  independent experiments). Statistical analysis was performed for differentiation of each indicated cell type (\*\* $P<0.01$ , \*\*\* $P<0.001$ , \*\*\*\* $P<0.0001$ , unpaired two-tailed t-test), as well as comparisons between different differentiations (\*\*\* $P<0.001$ , \*\*\*\* $P<0.0001$ , one-way ANOVA Tukey's multiple comparison test, only significant comparisons are shown). **e**, Bar plots showing percentage of INS<sup>+</sup> cells generated from VFGp3 or p6 cultures derived from HUES4 (triangles) and H9 (circles) ESCs. Data are represented as mean  $\pm$  SEM ( $N=4$  independent experiments). Statistical analysis was performed for differentiation of each indicated cell type (\*\* $P<0.01$ , \*\*\*\* $P<0.0001$ , unpaired two-tailed t-test), as well comparisons between different differentiations (\*\*\*\* $P<0.0001$ , unpaired two-tailed t-test, only significant comparisons are shown). **f**, Representative immunostaining (from three independent experiments) of VFGp6-derived  $\beta$ -like cells, stained with PDX1, INS, DAPI. Scale bar = 50 $\mu$ m. **g**, Expression analysis of ESC-derived VFG cultures at different passages: RT-qPCR of the indicated genes in transient ADE and VFGs. Expression is normalized with *ACTB*. Data are represented as mean  $\pm$  SEM ( $N=6$  independent experiments). \* $P<0.05$ , \*\* $P<0.01$ , \*\*\* $P<0.001$ , \*\*\*\* $P<0.0001$  (one-way ANOVA Dunnett's multiple comparison test compared with ADE, only significant comparisons shown). \* $P<0.05$  (one-way ANOVA Dunnett's multiple comparison test compared with VFGp3-4, only significant comparisons shown).





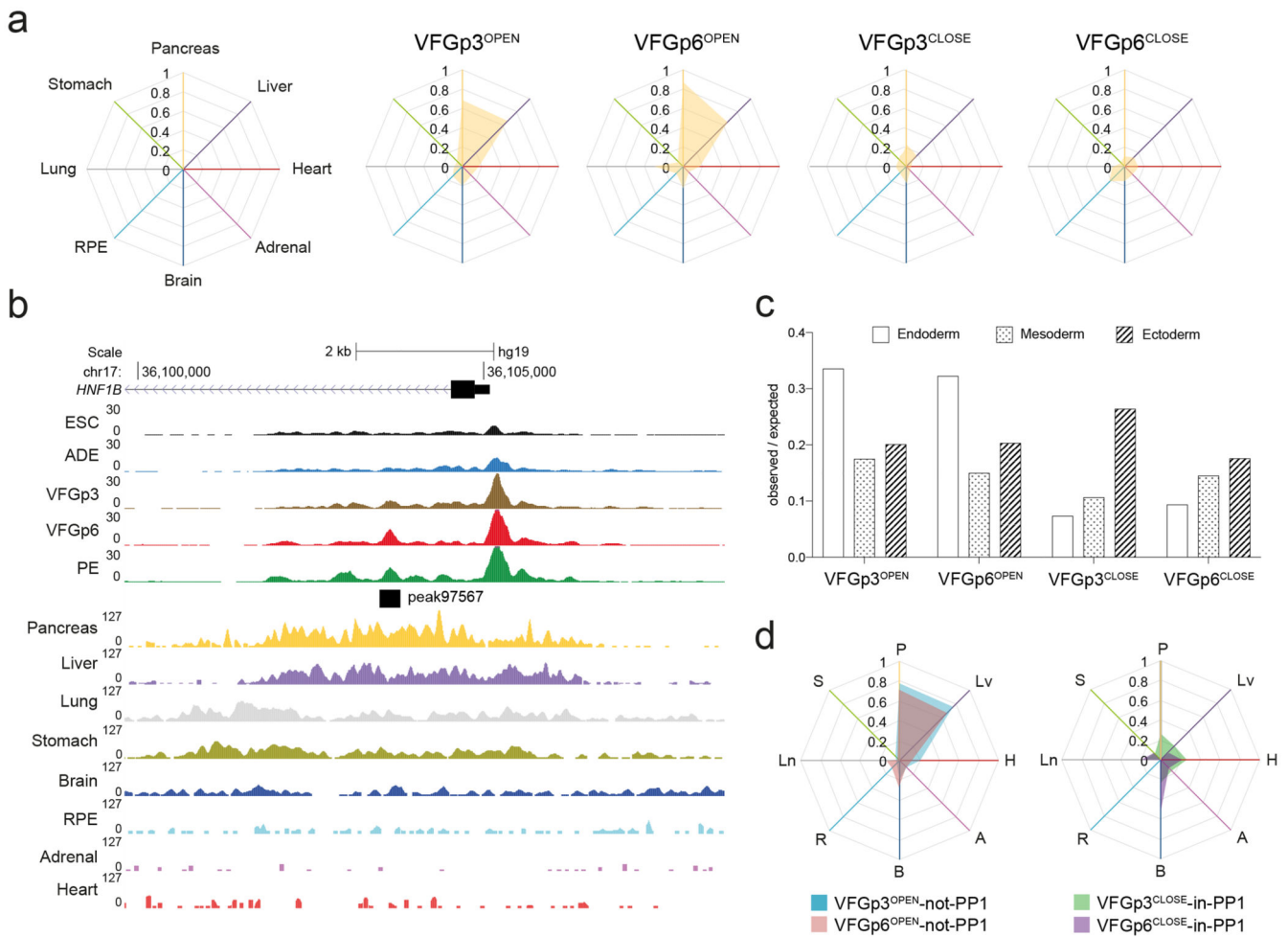
Right: Time course-sequencing (TCseq) trajectories for each cluster. Membership score reflects how well a given enhancer follows the pattern identified in time course analysis. **d-e**, Left: Representative UCSC Genome Browser screen shot (from two independent experiments) at the *GLIS3* (**d**) and *TBX3* (**e**) locus showing ATAC-seq data from ESC, ADE, VFGp3, VFGp6 and PE. Genome coordinates (bp) are from the hg19 assembly of the human genome. The PE<sup>PRIMED</sup> regulatory element (peak246749) (**d**) and VFG<sup>TR</sup> element (peak60307) (**e**) are shown with a black bar. Approximate distance between the element and the respective TSS is indicated by a broken dashed line in each panel. Right: RNA-seq data (normalized read count) for *GLIS3* (**d**) and *TBX3* (**e**) across the same conditions as the ATAC tracks. RNA-seq data are represented as mean  $\pm$  SEM ( $N=3$  independent experiments). **f-g**, Bar plot showing enrichment scores ( $\log_2$  observed/expected) of ATAC peak sets found within a 200 Kb window from genes upregulated (**f**) or downregulated (**g**) between PE and VFGp6 across the defined ATAC peak clusters. Genes considered here had a base mean expression  $> 1000$ ,  $\log_2$  fold change  $> 1.5$  and adjusted  $P < 0.05$ . For annotation see Supplementary Table 1, d-g. Analysis using lower base mean (100) or reduced genomic window sizes (25 Kb) are shown in Extended Data Fig. 3, g-j. All data shown are significant using chi-square analysis.



**Fig. 4. VFG proliferation dependent enhancers are associated active histone marks and correlate with later gene expression.**

**a**, Enhancer classification relative to VFG expansion time (from two independent experiments). Top panel shows heatmaps of normalized ATAC-seq signal in on enhancers that open (VFGp3<sup>OPEN</sup> and VFGp6<sup>OPEN</sup>) or close (VFGp3<sup>CLOSE</sup> and VFGp6<sup>CLOSE</sup>) at VFGp3 or p6. VFGp3<sup>CLOSE</sup> group comprises of ADE enhancers that are shutdown during VFG expansion at passage 3. Bottom panel shows average ATAC-seq signal in 10bp bins for these enhancers in same stages. **b-c**, H3K4me1 (**b**) and H3K27ac (**c**) enrichment by ChIP-qPCR for ADE, VFGp3, and p6 culture at VFGp6<sup>OPEN</sup> enhancer regions: *SFRP5* (peak32665), *HNF1B* (peak97567), and *FGFR2* (peak35254); and at VFGp6<sup>CLOSE</sup> enhancer regions of *LGR5* (peak56279), *ANGPT1* (peak242621), *SOX1* (peak70345). Circles and triangles mark cells derived from H9 and HUES4 WT ESCs respectively. Data are represented as mean  $\pm$  SEM ( $N=4$  independent experiments). \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$  (one-way ANOVA Tukey's multiple comparison test, only

significant comparisons shown). **d-e**, Bar plot showing the prevalence (log<sub>2</sub> observed/expected) of ATAC peaks within a 200 Kb window from genes up regulated between PE and VFGp6 (**d**); from genes down regulated between ADE and VFGp6 (**e**) across the ATAC peak clusters (defined in **a**). Genes considered here had a base mean expression > 1000, absolute log<sub>2</sub> fold change > 1.5 and adjusted  $P < 0.05$ . All data shown are significant using chi-square analysis.

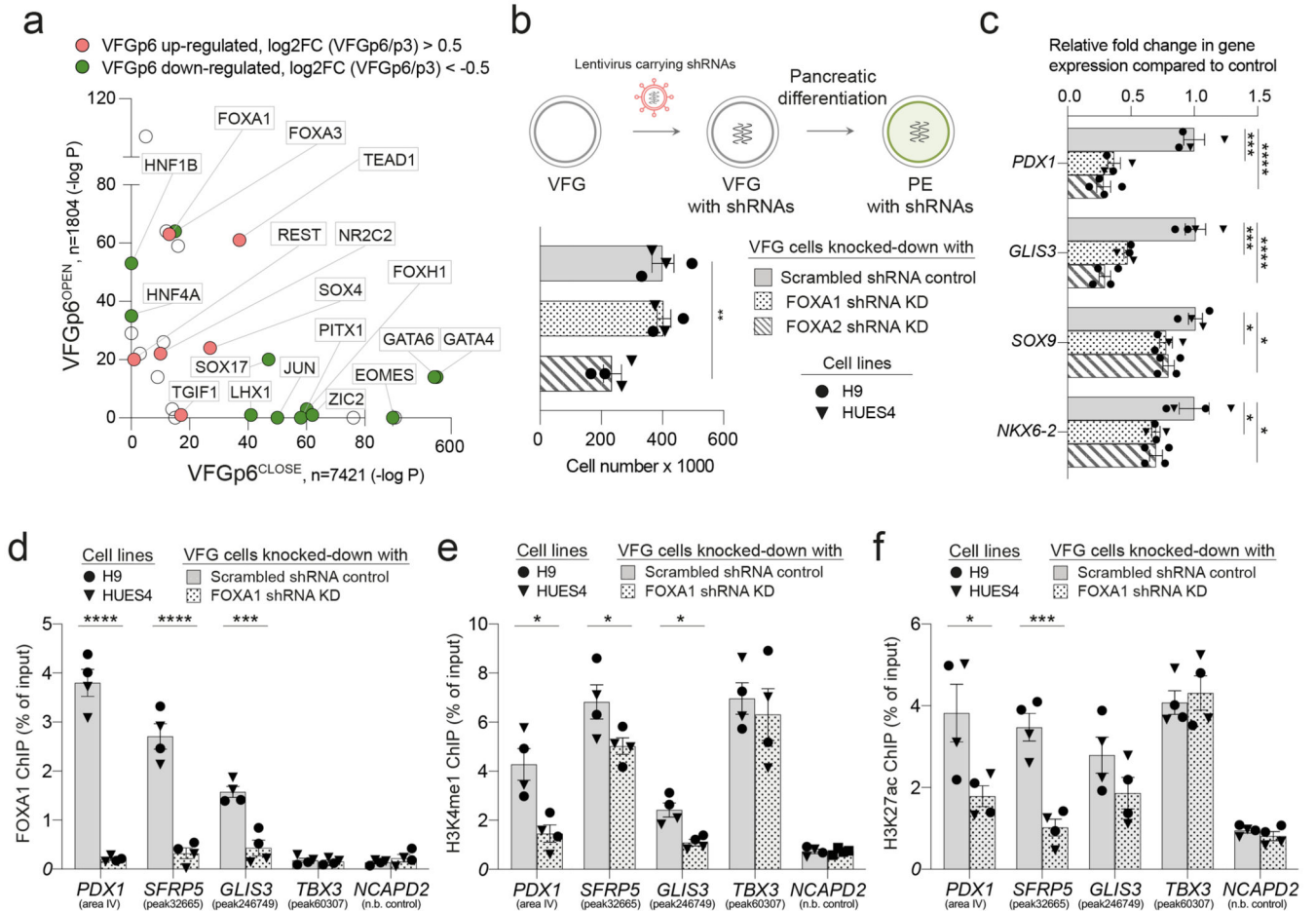


**Fig. 5. VFG expansion captures enhancers that are active during human ventral foregut derived organogenesis.**

**a**, Enrichment of tissue-specific H3K27ac enhancers from human embryos (from two independent experiments for most tissue types, except for stomach where only one sample was available) in different ATAC clusters defined in Fig. 4a displayed by enrichment score (observed/expected) in radar charts. **b**, Representative UCSC Genome Browser screen shot (from two independent experiments) at the *HNF1B* locus showing ATAC-seq data from this study (ESC, ADE, VFGp3, VFGp6, and PE) and H3K27ac ChIPseq data<sup>37</sup> from multiple human embryonic tissues (pancreas, liver, lung, stomach, brain, RPE, adrenal, and heart). Genome coordinates (bp) are from the hg19 assembly of the human genome. VFGp6<sup>OPEN</sup> (peak97567) element overlapping with pancreatic specific H3K27ac enhancer is shown at the bottom and the approximate distance between the elements and the *HNF1B* TSS is indicated. **c**, Enrichment of lineage-specific H3K27ac enhancers (endoderm, ectoderm, and mesoderm) from human embryos<sup>37</sup> in the different VFG expansion-specific ATAC clusters defined in Fig. 4a by enrichment score (observed/expected). **d**, Enrichment of tissue-specific H3K27ac enhancers from human embryos across different VFG<sup>OPEN</sup> and VFG<sup>CLOSE</sup> clusters (defined in Fig. 4a) that are not regulated in directed differentiation were

displayed by enrichment score (observed/expected) in radar charts. P: pancreas, Lv: liver, H: heart, A: adrenal, B: brain, R: RPE, Ln: lung, S: stomach.





**Fig. 6. FOXA proteins are required for VFG enhancer priming towards pancreatic differentiation.**

**a**, Transcription factor motif enrichment in VFGp6<sup>OPEN</sup> (n=1804) and VFGp6<sup>CLOSE</sup> (n=7421) ATAC clusters. n = peaks analysed. P values were derived from hypergeometric enrichment using HOMER default background. Candidate factors with P-value > 1e<sup>-10</sup> for both clusters was not included in the plot. Gene expression of the candidate factors that up regulated (red) or down regulated (green) from VFGp3 to p6 (Log2 fold change > 0.5, P<0.05) are labeled. **b**, Top: Schematic of FOXA1 and FOXA2 shRNA KD VFG cells and their PE differentiation. Bottom: Histogram for proliferation assay (cell counts) for FOXA1 and FOXA2 shRNA KD and scrambled shRNA control VFG cells. Data are represented as mean ± SEM (N=4 independent experiments). Statistics analysis was performed between KDs and control VFG cells (\*\*P<0.01, unpaired two-tailed t-test, only significant comparisons are shown). **c**, Differentiation of FOXA1 and FOXA2 shRNA KD and scrambled control VFG cells to PE, with legend shown in (b). Relative fold change in mRNA of pancreatic genes (PDX1, GLIS3, SOX9 and NKX6-2) was assayed by RT-qPCR. Expression is normalized to ACTB. Data are represented as mean ± SEM (N=4 independent experiments). \*P<0.05, \*\*\*P<0.001, \*\*\*\*P<0.0001 (one-way ANOVA Dunnett's multiple comparison test compared with control). **d-f**, FOXA1 binding (**d**), H3K4me1 (**e**) and H3K27ac (**f**) enrichment by ChIP-qPCR at enhancer regions of PDX1

(area IV), *GLIS3* (peak246749), and *TBX3* (peak60307) in FOXA1 shRNA KD VFG and scrambled control cell lines. An intragenic region of *NCAPD2* served as a non-bound control. Data are represented as mean  $\pm$  SEM ( $N=4$  independent experiments). Statistics analysis was performed between the KD and control VFG cells. \* $P<0.05$ , \*\*\* $P<0.001$ , \*\*\*\* $P<0.0001$ , unpaired one-tailed t-test, only significant comparisons are shown).