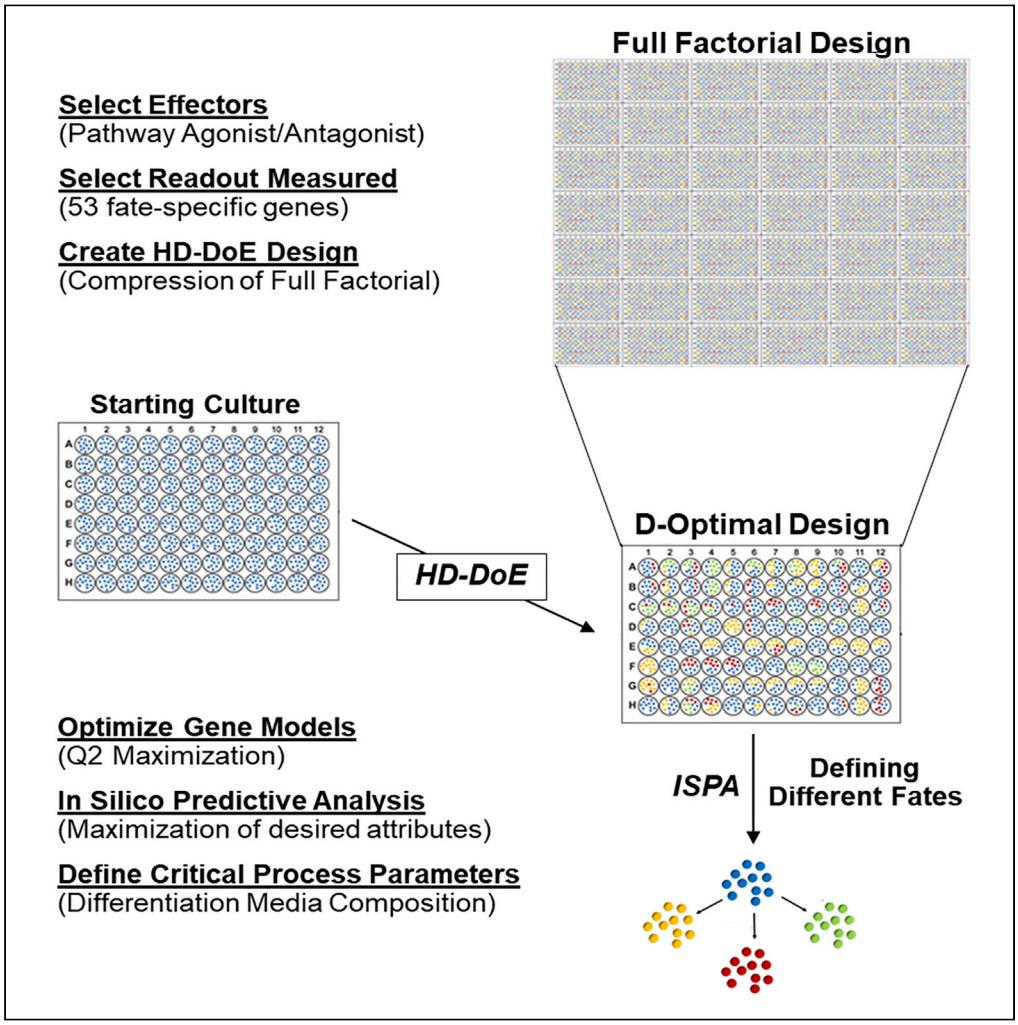
iScience

CellPress
OPEN ACCESS

Article

High-Dimensional Design-Of-Experiments Extracts Small-Molecule-Only Induction Conditions for Dorsal Pancreatic Endoderm from Pluripotency



Michael A. Bukys, Alexander Mihas, Krystal Finney, ..., Yong Wang, Jose Oberholzer, Jan Jensen

jensenj2@ccf.org

HIGHLIGHTS

Method development for addressing multifactorial problems in directed differentiation

Generation of endodermal populations without the use of TGF-β agonism

Small-molecule-based pancreatic differentiation protocol

Bukys et al., iScience *23*, 101346
August 21, 2020 © 2020
https://doi.org/10.1016/j.isci.2020.101346

iScience

CellPress
OPEN ACCESS

Article

High-Dimensional Design-Of-Experiments Extracts Small-Molecule-Only Induction Conditions for Dorsal Pancreatic Endoderm from Pluripotency

Michael A. Bukys,[1] Alexander Mihas,[1] Krystal Finney,[2,4] Katie Sears,[1] Divya Trivedi,[2,4] Yong Wang,[3] Jose Oberholzer,[3] and Jan Jensen[1,2,4,5,*]

SUMMARY

The derivation of endoderm and descendant organs, such as pancreas, liver, and intestine, impacts disease modeling and regenerative medicine. Use of TGF-β signaling agonism is a common method for induction of definitive endoderm from pluripotency. By using a data-driven, High-Dimensional Design of Experiments (HD-DoE)-based methodology to address multifactorial problems in directed differentiation, we found instead that optimal conditions demanded BMP antagonism and retinoid input leading to induction of dorsal foregut endoderm (DFE). We demonstrate that pancreatic identity can be rapidly, and robustly, induced from DFE and that such cells are of dorsal pancreatic identity. The DFE population was highly competent to differentiate into both stomach organoids and pancreatic tissue types and able to generate fetal-type β cells through two subsequent differentiation steps using only small molecules. This alternative, rapid, and low-cost basis for generating pancreatic insulin-producing cells may have impact for the development of cell-based therapies for diabetes.

INTRODUCTION

Endoderm is the germ layer that creates the majority of cells within most of the internal organ systems, such as lung, stomach, pancreas, liver, and gut. The ability to robustly generate endodermal descendant tissues will impact the studies and therapy modalities of multiple human diseases. Until present, almost all efforts on inducing endoderm from pluripotent cells have relied on using a TGF-β pathway agonist, most commonly Activin A (AA), to push pluripotent cells through an *in vitro* gastrulation event (D'amour et al., 2005; Gadue et al., 2006). This results in an endodermal population that can be successfully used for generating multiple descendant fates including intestinal (Spence et al., 2011), pancreatic (Kroon et al., 2008; Rezania et al., 2014; Pagliuca et al., 2014), and liver (Sampaziotis et al., 2015). Generation of more anterior endodermal fates, such as lung, has been achieved by providing patterning inputs at a subsequent stage (Green et al., 2011). However, recent studies argue that initial patterning of definitive endoderm may occur during its generation (Matsuno et al., 2016; Loh et al., 2014).

The pancreas is of particular interest for cell-based therapy in diabetes, which is characterized by defects in, or loss of, insulin-producing cells. The pancreas is formed from two spatially distinct primordia arising on the dorsal and ventral sides of the primitive gut tube, which subsequently fuse. Although both pancreatic buds are capable of generating all lineages of the adult pancreas (Matsuura et al., 2009), distinct transcriptional programs control the initial induction of the pancreatic domains on opposing sides of the gut tube. In mice, the ventral pancreatic bud forms first at approximately embryonic day 8.5 (E8.5) from a region of endoderm possessing bipotential competence for pancreas and liver (Angelo et al., 2012; Deutsch et al., 2001; Tremblay and Zaret, 2005). This early ventral endoderm field consists of a progenitor population that co-expresses *Pdx1/Sox17* transiently, which by E9.5 splits to form the ventral pancreas and the extra-hepatobiliary system, respectively (Spence et al., 2009). Specification of the ventral pancreas relies on *HHex* expression. Gene ablation models have demonstrated complete ventral agenesis without affecting dorsal pancreatic bud formation (Bort et al., 2004). In contrast, the dorsal pancreatic bud in mice emerges at approximately embryonic day 9.0 and forms from an outgrowth caudal to the antral stomach region. Studies in mice have also identified factors involved in dorsal pancreatic specification with no effect on ventral organogenesis. *Mnx1* (Hlxb9) knockout models have shown dorsal agenesis occurs without a ventral

[1]Department of Biomedical Engineering, Lerner Research Institute, Cleveland Clinic Foundation, 9500 Euclid Avenue, Cleveland, OH 44195, USA

[2]Trailhead Biosystems Inc, 10000 Cedar Avenue, Cleveland, OH, USA

[3]Division of Transplantation, University of Virginia, Charlottesville, VA 22903, USA

[4]Cleveland Clinic, Cleveland, OH 44195, USA

[5]Lead Contact

*Correspondence: jensenj2@ccf.org

https://doi.org/10.1016/j.isci.2020.101346

- **Quality by Design (QbD)** – The concept that building quality into a manufactory process is a prerequisite for success. This is achieved through defining the critical quality attributes and the critical process parameters needed to attain the quality target product profile. This is a systemic approach that focuses on process understanding and process control to obtain a defined goal.
- **Critical Quality Attributes (CQA)** – Aspects of the product that have the greatest impact on quality and are therefore crucial to study and control.
- **Critical Process Parameters (CPP)** – The variables within a production process that are critical to control to ensure the products quality.
- **Critical Material Attribute** – The variables within starting material or the product that are critical to control to ensure the products quality.
- **Design of Experiments (DoE)** – DoE is a QbD compliant methodology for optimization within a design space which offers a cost effective method for attaining a greater process understanding than traditional methodologies.
- **Design space** – Interrogates the relationship of the process inputs to the CQAs.
- **D-Optimal** – An algorithm used for choosing compressed experimental designs that seeks to interrogate all interactions within the an experimental space. Though this method does not guarantee that a global maximum will be contained within the experimental space.

Box 1. Definition of Key Terms Used within This Study.

phenotype (Li et al., 1999). *Mnx1* expression is observed in the ventral field but only following *Pdx1* expression, whereas in the dorsal field, *Mnx1* precedes *Pdx1* expression. *Raldh2* knockout models resulted in a dorsal-specific agenesis attributed to the loss of *Pdx1* and *Prox1* expression in the dorsal bud (Martin et al., 2005; Molotkov et al., 2005). Furthermore, studies in chick have shown that the initial budding of the dorsal pancreas is dependent on the selective inhibition of SHH within the dorsal midgut (Hebrok et al., 1998). Although it is unclear if the murine system is conserved between species, a recent study using laser capture followed by deep sequencing analysis described some fundamental differences between the ventral and dorsal pancreas during human development (Jennings et al., 2017).

Despite differential pathway utilization and distinct cell intrinsic factors the dorsal and ventral pancreatic programs have much in common. HNF1β (*Tcf2*) is required for pancreas specification in both pancreatic buds and is critical through pancreatic development. *Tcf2* knockout mice fail to generate a ventral pancreas and have a greatly reduced dorsal bud incapable of differentiating or proliferating (Haumaitre et al., 2005). HNF1β is expressed in the pre-pancreatic foregut and functions at the apex of a sequential transcriptional cascade resulting in the activation of Hnf6 (*Oc1*) followed by *Pdx1* (Poll et al., 2006). Conditional inactivation of HNF1β results in a loss of *Glis3* and *Ngn3* expression and results in a pancreas characterized with cystic ducts and a loss of the pro-endocrine field (de Vas et al., 2015). In human development, the importance of HNF1β is highlighted by the occurrence of "maturity-onset diabetes of the young type 5" (MODY5) syndrome, a condition attributed to mutations in the HNF1β gene. Although a heterozygous mutation in HNF1β does not display a phenotype in mouse studies, in humans heterozygous mutation of HNF1β have been shown to be associated with MODY5 or complete pancreatic agenesis suggesting a more important role for HNF1β in human pancreatic development than in mouse (Body-Bechou et al., 2014).

Most current laboratory efforts at directing the differentiation of pluripotent stem cells rely on emulating developmental signaling event(s) leading to the generation of desired cell type. This is generally accomplished by assaying a single pro-differentiation factor at time. The limitation of this methodology is that it relies on studying each pro-differentiation factor individually, hindering the detection of synergistic or systemic influences. A way to address this problem is using a systems biology approach (Kitano, 2002; Carinhas et al., 2012) capable of assaying multiple factors simultaneously in a manner capable of elucidating individual and synergistic effects. Because experimental design size increases exponentially as additional factors are incorporated, this greatly limits traditional methods from approaching a systems biology level of interrogation. To overcome this limitation, we have developed a novel approach focusing on key aspects of a manufacturing process (key concepts are defined within Box 1). Using Design-of-Experiments (DoE) mathematics (Chakrabarty et al., 2013), we are able to greatly increase the dimensionality of our differentiation experiments by relying on a compression of the design space (Gerin et al., 2014; Mendes et al., 2016). This systematic approach minimizes the number of experimental runs needed to interrogate multiple parameters simultaneously within a single experimental design (Rathore et al., 2014). Combined with a deep set of lineage-informative transcript level measurements a better understanding of the cell culture

behavior is obtained within the design space (Mercier et al., 2013). Such an approach is integral to a Quality-by-Design (QbD) process (Juran, 1993; McConnell et al., 2010; Swain et al., 2018). It provides process understanding, allowing for consistent product manufacturing (Kumar et al., 2014; Lipsitz et al., 2016). In the present study, we use DoE-based optimization to produce a pancreatic directed differentiation protocol defining the critical process parameters relevant to the differentiation process and future manufacturing. We monitor gene expression throughout the differentiation process as a critical material attribute that defines the cellular phenotype. We then identify which pathway control elements are the critical process parameters that must be controlled to ensure proper differentiation.

Contrary to most methods of endoderm induction, we demonstrate that effective and regionalized patterned endoderm can be robustly differentiated directly from pluripotency without the use of TGF-β agonism. Exploring for optimal endodermal fate conversion conditions, we used HNF1β expression as an initial waypoint for pancreas and other endodermal derivatives. This led to a novel and highly robust protocol for inducing specialized human endoderm representative of the dorsal foregut region of the gut tube. We demonstrate that this population can be effectively converted into dorsal pancreatic progenitors that subsequently are able to adopt endocrine fates, including the generation of fetal-like beta cells. By inspecting the critical process parameters we created a three-stage protocol that converts PSCs into fetal-like beta cells using a series of small molecules.

RESULTS

Application of Systems Biology to Understand Developmental Models

We combined a high-dimensional application of Design of Experiments (HD-DoE) methodology with deep response set measurements to generate predictive models for pluripotent culture forward differentiation. This method allowed interrogation of pathway interactions, providing a more comprehensive understanding of the biological inputs impacting endodermal differentiation. By challenging pluripotent cultures with a perturbation matrix composed of multiple morphogen pathway agonists and antagonists simultaneously, and measuring multiple fate determining genes, we extracted a systems-level effector/response dynamic. The application of DoE substantially compressed the number of experimental runs compared with a full factorial design. Yet it remained possible to statistically determine any first-order pathway/pathway interactions. It also revealed the system behavior covered by the tested dimensions within the concentration ranges used (i.e., "known space").

The morphogens (effectors) used in initial experimental designs were AA, BMP4, FGF2, WNT3a, SHH, and RA. We also included respective small-molecule antagonists for each pathway except RA. These factors were chosen because previous publications within the stem-cell field have shown these inputs to elicit forward differentiation from pluripotency into various descendant fates. We generated D-optimal DoE designs to test all of these effectors in a single experiment. Designs specified the combinations of effectors to be included in each well of a 96-well plate. The designs were constrained to prevent unproductive combinations of agonist/antagonist of the same pathway being used together in the individual reactions. Robotic assembly of experimental media (i.e., the Perturbation Matrix) eliminated human errors while ensuring accuracy. Response measurements were custom-chosen early lineage-determining genes. Following experiment execution, a multivariable regression model was generated for each response gene as a function of effector contribution. All response data were mathematically fitted to maximize predictive power (Q^2 maximization). As a result, we obtained an *in silico* representation of the behavior of each response gene as related to each effector input. These models allowed us to predict conditions that would achieve desirable induction, or suppression, of any of the genes monitored. We refer to these interrogations as *in silico predictive analysis* (ISPA), noting that such predictions were calculated on the basis of statistical models resting on the entire set of the DoE design.

For ISPA, various tools were needed to achieve specific desirable outcomes and identify critical process parameters. Coefficient Plots may be generated for each individual response gene. These plots display the coefficient for each effector term in the regression model for the respective gene. coefficient plots are scaled and centered and thus also provide graphical representations of model term significance. We used coefficient plots to inspect individual effectors' contributions to the activation of individual genes. The "optimizer" function used the regression models to identify which media compositions contributed to a desired differentiation event. The optimizer provides the relative Factor Contribution (FC) for each effector. FC is proportional to how important the individual effector is to the differentiation event. FC thereby helps

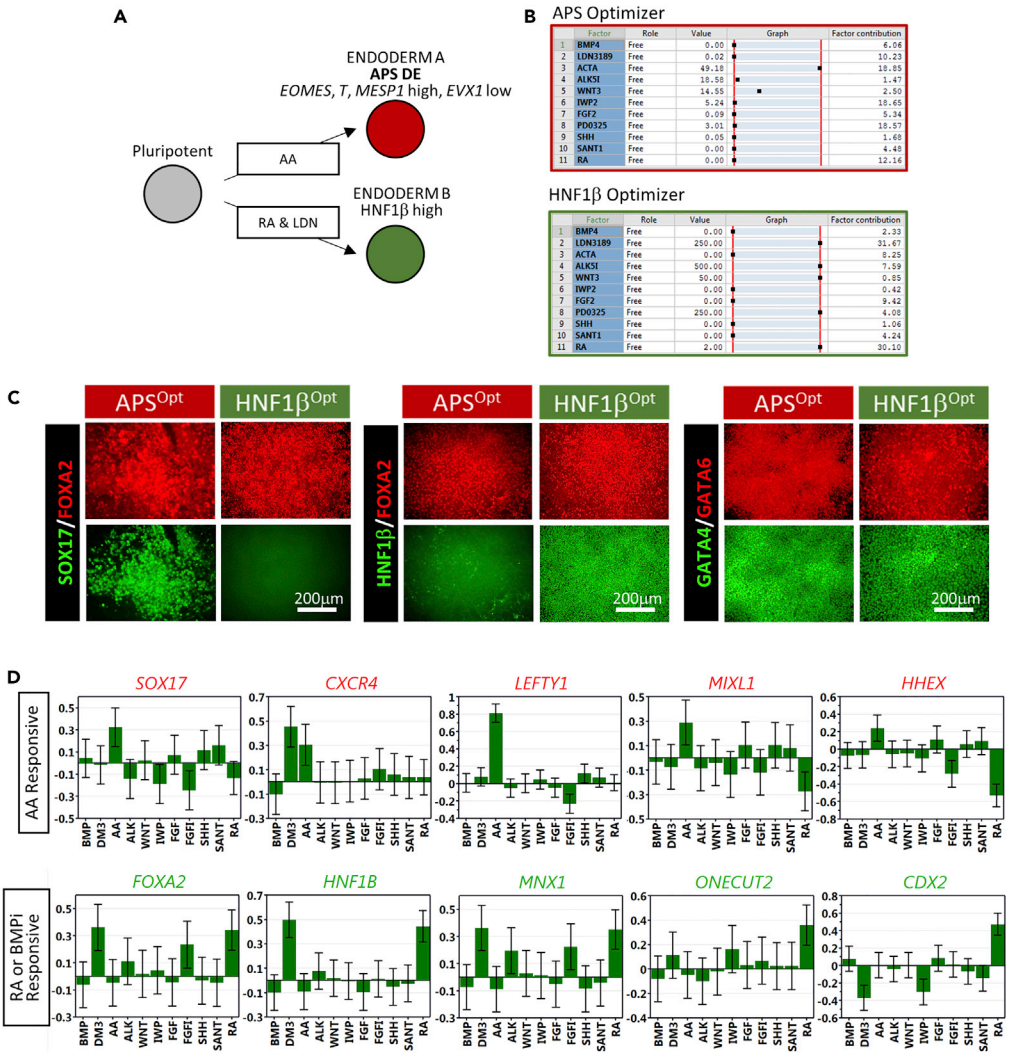


Figure 1. Effective Endoderm Induction in Absence of AA/WNT

(A) Schematic of endodermal generation through AA induction of Anterior Primitive Streak (APS) versus endoderm induction optimizing for HNF1β induction.

(B) Predicted conditions satisfying APS gene induction (red) versus HNF1β induction (green).

(C) Validation of endodermal marker induction using the two separate protocols.

(D) Coefficient plots (primary effectors only) for select endodermal genes. Error bars within the Coefficients Plots represent 95% confidence intervals. Significant terms are identified as terms with confidence intervals that do not overlap the y axis.

identify criticality of a process parameter. For practical proposes, we considered FC < 10 as low relevance, 10–20 as relevant, and >20 as highly relevant. Particularly valuable, we used the "Dynamic Profiling" to visualize expression behavior for multiple genes simultaneously at any given input condition (set point).

Two Separate and Distinct Pathways Exist for Endoderm Induction

We initially set out to predict conditions needed to define a definitive endoderm population as suggested by the literature. This was accomplished by defining an anterior primitive streak (APS) population by modeling for the maximal expression of *MESP1*, *EOMES*, and *BRACHYURY/T* while minimizing *EVX1* (posterior primitive streak marker) (Loh et al., 2014). Through ISPA, the conditions predicted to generate this differentiation event consisted of low tolerance to Wnt inhibition (FC = 18.65) and high levels of AA (FC = 18.85) (Figures 1A and 1B). These conditions agree with current protocols for generating definitive endoderm (DE) (D'amour et al., 2005). Indeed, when using this condition to differentiate pluripotent

cultures, a FOXA2+/SOX17 + population was obtained within a 3-day period (Figure 1C). Through ISPA, a number of other genes were predicted to be highly expressed under these conditions including *COL6A1*, *HHEX*, *MESP2*, *SOX17* (Figure 1D and data not shown). These genes are all known to be elevated in definitive endoderm. However, through ISPA, we noted that not all known early endoderm-expressed genes were uniquely maximized through the APS conditions. Inspecting the known space from the aforementioned experiment, a quite different solution set could be obtained focusing on HNF1β (*TCF2*) expression, also known to be expressed in definitive endoderm. Maximizing for HNF1β expression also led to expression of accompanying genes such as *FOXA2*, *HNF4A*, *MNX1*, *CXCR4*, *MTF1*, all known markers of endoderm, whereas expression of *HHEX* and *SOX17* remained low (ISPA results not shown). We went on to characterize these distinct states and the requirements for their induction.

Retinoic Acid and Bmp Inhibition Synergistically Induce an Endodermal Program Mutually Exclusive to Activin-Induced Endoderm

Using ISPA, we inspected the fundamental logics governing endodermal gene induction. For the APS-derived DE, it was clear that many early endoderm genes were under the direct control of TGF-β signaling, displaying strong and positive coefficient terms from AA in their complex regulatory models. These genes included, but were not limited to, *SOX17*, *CXCR4*, *LEFTY1*, *MIXL1*, and *HHEX* (Figure 1D). This was not the case for multiple other known endoderm markers. A sub-group of endoderm genes did not respond to AA stimulation but were directly dependent on RA signaling and to a lesser extent required the inhibition of the BMP pathway. These genes included *FOXA2*, *EPCAM*, *ONECUT1*, *CDX2*, and *MNX1* (Figure 1D) and were predicted to be directly controlled through the synergistic effects of retinoic acid and BMP inhibition with high factor contributions including *FOXA2* (FC for RA = 24.8, BMPi = 25.3), *HNF1β* (FC for RA = 30.1, BMPi = 31.7), and *MNX1* (FC for RA = 22.7, BMPi = 30.5) (Figure 1B and data not shown). Of note, AA was predicted to have no contribution to activating these genes; rather, inhibition of the TGF-β pathway was predicted to benefit the expression of these genes. Factor contributions for Alk5i were 7.59 for HNF1β, 8.88 for FOXA2, and 18.98 for MNX1 (Figure 1B and data not shown). To assess the robustness of the methodology, the same perturbation matrix design was used on a second pluripotent cell line, H9 (female) (Figure S1). Both the BMPi (Factor Contribution of 11.0) and the provision of RA (Factor Contribution 29.9) were again shown to be critical factors for the activation of HNF1β. Although the factor contribution for BMPi decreased, there was a corresponding increased importance in the absence of AA (Factor Contribution 20.1 that AA is not included) (Figures S1A′–S1B′). Altogether, this suggests that retinoic acid input when provided in the absence of TGF-β signaling is a critical process parameter for induction of HNF1β. Comparative analysis using dynamic profiling at the HNF1β set point revealed that effectors controlling HNF1β, MNX1 (both retinoic acid and TGF-β inhibition responsive), and HHEX (AA responsive) were very similar between these cell lines, which differ in sex and prior culturing conditions (the male H1 in Essential 8 and the female H9 in mTesR media) (Figures S1C and 1C′). The predicted conditions for HNF1β optimization (HNF1β^{Opt}) were tested on differentiating pluripotent cells for validation. Based on protein expression, HNF1β, MNX1, and FOXA2 could all be activated as expected. The target gene for optimization, HNF1β, was present in 97.9 ± 0.7% of the cells (Figures S2A and 2C).

The two paths to endoderm activation were fundamentally distinct and rested on conflicting input logic suggesting that the pathways were mutually exclusive. We tested this by creating hybrid protocols assaying the effects of RA and BMPi in the presence of AA (Figure 2A). Inclusion of RA into the APS-based DE-generating protocols (D'amour et al., 2005) proved to only moderately increase gene expression for retinoic acid-responsive genes; only *CDX2* (data not shown) and *OSR1* (Figure 2B) were significantly up-regulated in this manner. Conversely, known TGF-β-responsive genes were shown to be significantly down-regulated when retinoic acid was included in APS-type DE-generating reactions including *HHEX*, *SOX17*, and *GSC* (Figure 2B and data not shown). This demonstrates that presence of either of the key protocol drivers (AA versus RA) will suppress the other. Furthermore, also as predicted by ISPA, the genes up-regulated in the presence of retinoic acid and LDN3189 were activated more efficiently when AA was excluded from these reactions. Importantly, the two key protocol inputs for the HNF1β^{OPT} conditions, RA and LDN3189, sufficed to initiate differentiation comparable with the full-input HNF1β^{Opt} conditions (Figure 2B) with 97.1 ± 1.8% of the cells within the culture expressing HNF1β (Figures S2A and S2C).

Retinoic Acid/TGF-β Inhibition-Induced Endoderm Is of a Dorsal Foregut Character

To gain a better understanding of the differing nature of the endodermal populations, we subjected cultures for RNA sequencing. Common endodermal genes were expressed at similar levels in both

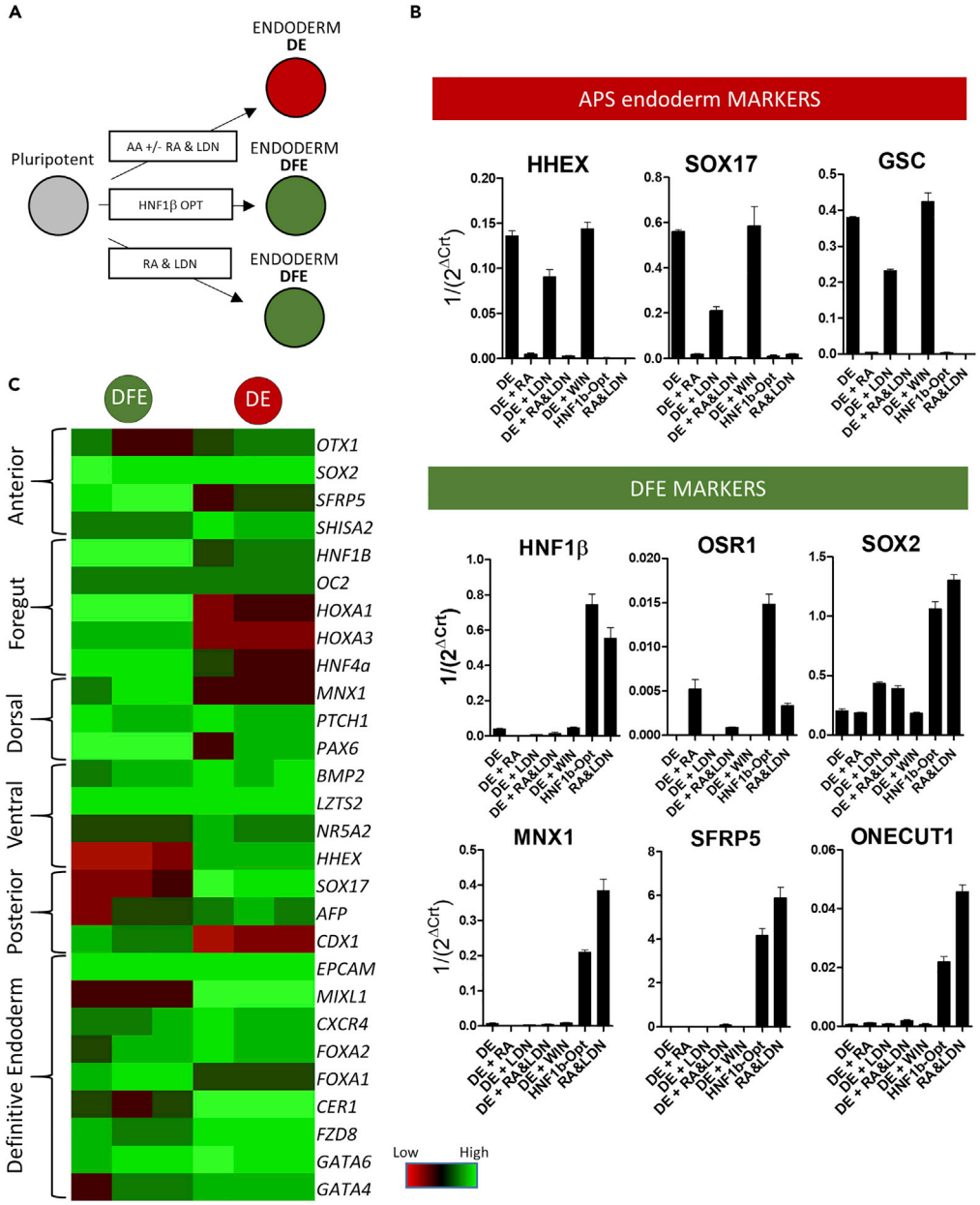


Figure 2. Non-APS-Derived Endoderm Is Critically Activated by Retinoic Acid and BMP Inhibition and Demonstrates a Dorsal Foregut Character

(A) Schematic of the experiments performed.

(B) Graphs showing the relative expression of several endodermal genes in response to retinoic acid and BMP inhibition. Experiments consist of quadruplicate biological replicates performed in parallel experiments within a single TC plate. Genes were normalized to the average expression of the endogenous levels of *YWHAZ*, *GAPDH*, and *TBP*. DE, definitive endoderm; RA, retinoic acid; LDN, LDN3189, a BMP inhibitor; WIN, Win 18446, an ALDH inhibitor. Error bars represent standard deviation/triplicate assays.

(C) Heatmap containing key endodermal genes from pluripotent cultures either subjected to classic definitive endoderm differentiation conditions or differentiated using conditions predicted in the HNF1β optimizer.

populations; these included *CXCR4*, *FOXA2*, *EPCAM*, *GATA4*, and *GATA6* (Figure 2C). However, significant differences were observed for genes associated with patterning revealing that HNF1β^{Opt} induced endoderm was enriched in genes characteristic of known dorsal (*MNX1* and *PAX6)* and foregut endoderm (*HOXA1*, *HOXA3*, *HNF4A*, and *HNF1β*) (Figure 2C). In contrast, APS-type DE showed an enrichment for

genes representative of *ventral* endoderm (*NR5A2*, *HHEX*, and *SOX17*) and more *posterior* endoderm (*SOX17* and *AFP*), although not *CDX1* (Figure 2C). Since the HNF1β^{Opt} culture appeared to have a stronger dorsal foregut endoderm (DFE) phenotype, we challenged it for differentiation competence toward stomach, pancreas, and liver using conditions previously shown to induce these fates from APS-derived DE (Figure S3A). These tissues are derived from posterior foregut, and liver is a ventral derivative only. We also differentiated APS-derived DE for comparison. Differences in competence were observed between the DFE and APS-DE populations. In all cases, APS-DE cultures activated liver genes to higher levels (*APOB*, *HHEX*, and *EVX1* (Figure S3B) than DFE, whereas stomach (*OSR1*) and pancreas (*PDX1*) genes were activated at higher levels in the DFE cultures. We also challenged the DFE culture to generate stomach organoids (Figures S3C and S3D [Mccracken et al., 2014]). OSR1 and PDX1 co-expression and SOX2 and PDX1 co-expression were both observed, suggesting that the stomach organoids preferentially converted into antral-type, posterior-most stomach (Figure S3E).

Organ-Field Specification Mechanisms from Dorsal Foregut Endoderm

Since DFE patterned cultures displayed a competence for PDX1 activation (Figure S3A), we next evaluated DFE pancreatic potential through a sequential DoE modeling experiment (as outlined Table S2). DoE designs included effectors previously shown to induce pancreas as well as effectors known to pattern along the mid-section of the developing gut tube. These included FGF2, FGF4, PD0325901, SHH, Sant1, BMP4, LDN3189, EGF, AA, A8301, and Wnt3a with agonist/antagonist constraints. Results from ISPA maximization of the expression of *SOX2*, *OSR1*, or *PDX1* were compared (Figures 3A and 3B). Dynamic profiling analysis (Figure 3C) revealed that both OSR1 (gastric) and PDX1 (pancreatic and wider) were under similar mechanisms of control. Both genes were strongly responsive to RA with FC = 30.68 and FC = 31.38, respectively (Figures 3B and 3C). However, differential responsiveness to FGF signaling were predicted. FGF4 was predicted to be important for *OSR1* (gastric) activation with an FC of 13.45, whereas MEK pathway inhibition with the small-molecule inhibitor PD0325901 was shown to strongly contribute to *PDX1* activation with a substantial FC of 29.62 (Figures 3B and 3C). Thus, the bipotentiality of gastric/pancreatic fates is resolved by FGF/FGFi inputs, respectively. Of note, SOX2 expression was strongly decreased by RA and highly increased by SHH with respective FC of −28.59 and 29.58 (Figures 3B and 3C). Our data argue that retinoic acid secures a posterior antral field and active inhibition of SHH contributes to the down-regulation of SOX2 expression, hereby allowing for a switch from a gastric to a pancreatic field.

DFE-Derived Pancreas Is of a Dorsal Identity

Using the ISPA-defined PDX1^{Opt} conditions (shown in Figure 3B), we next demonstrated that the PDX1-expressing DFE-derived cultures co-expressed several known pancreatic progenitor markers including FOXA2, NR5A2, GATA4, and SOX9 indicating that a true pancreatic endodermal (PE) state was rapidly induced (Figure 3D). This DFE-derived pancreatic induction was shown to be reproducible in the H9 female embryonic stem cell line, as well as in iPSC culture (Figure S4). An RNA-seq-based KeyGenes analysis (Roost et al., 2015) was used to verify that this DFE-derived PDX1-induction was truly a pancreatic fate (Figure 3E). We compared DFE-derived PE with previously published (Xie et al., 2013) APS-type DE-derived stage 4 PE (Figure 3F). Interestingly, both of the starting populations, DE and DFE cultures, initially displayed a similarity to brain, which was lost through the sequential stage in both protocols (Figure 3E). This loss of similarity to brain is likely attributable to emergent expression of neuronal fate suppressor genes (Jennings et al., 2017) activated during the sequential stage(s) in both populations (Figure 3F). KeyGenes analysis showed that the published APS-derived DE population displayed similarity to lung, whereas the DFE population did not (Figure 3E), presumably because the lung buds are exclusively derived from ventral endoderm. A hierarchical clustering comparison between DFE, DE, and their derived PE populations demonstrated that the DFE and DE populations were so similar that the DFE population clustered in-between the triplicate set of DE samples (Figure S5), attesting that both protocols attain a fundamental endodermal program. However, when comparing PE derived from either DFE or DE with genes highly enriched in the dorsal pancreas during human development (Jennings et al., 2017) we found that the DFE population already expressed several of these genes including *DLL1*, *CNR1 FRZB*, *HOXA1*, and *ARMC3*, whereas the APS-DE population displayed only low expression of *CNR1* and *FRZB*. Also, expression of *MNX1* (Hlxb9), a previously described dorsal marker, was expressed throughout the DFE culture, whereas detectable *MNX1* expression within the DE-derived PE only began at the PFG stage (Figure 3F). Subsequently, DFE-derived PE continued to express the vast majority of the dorsal-specific genes (Figure 3F).

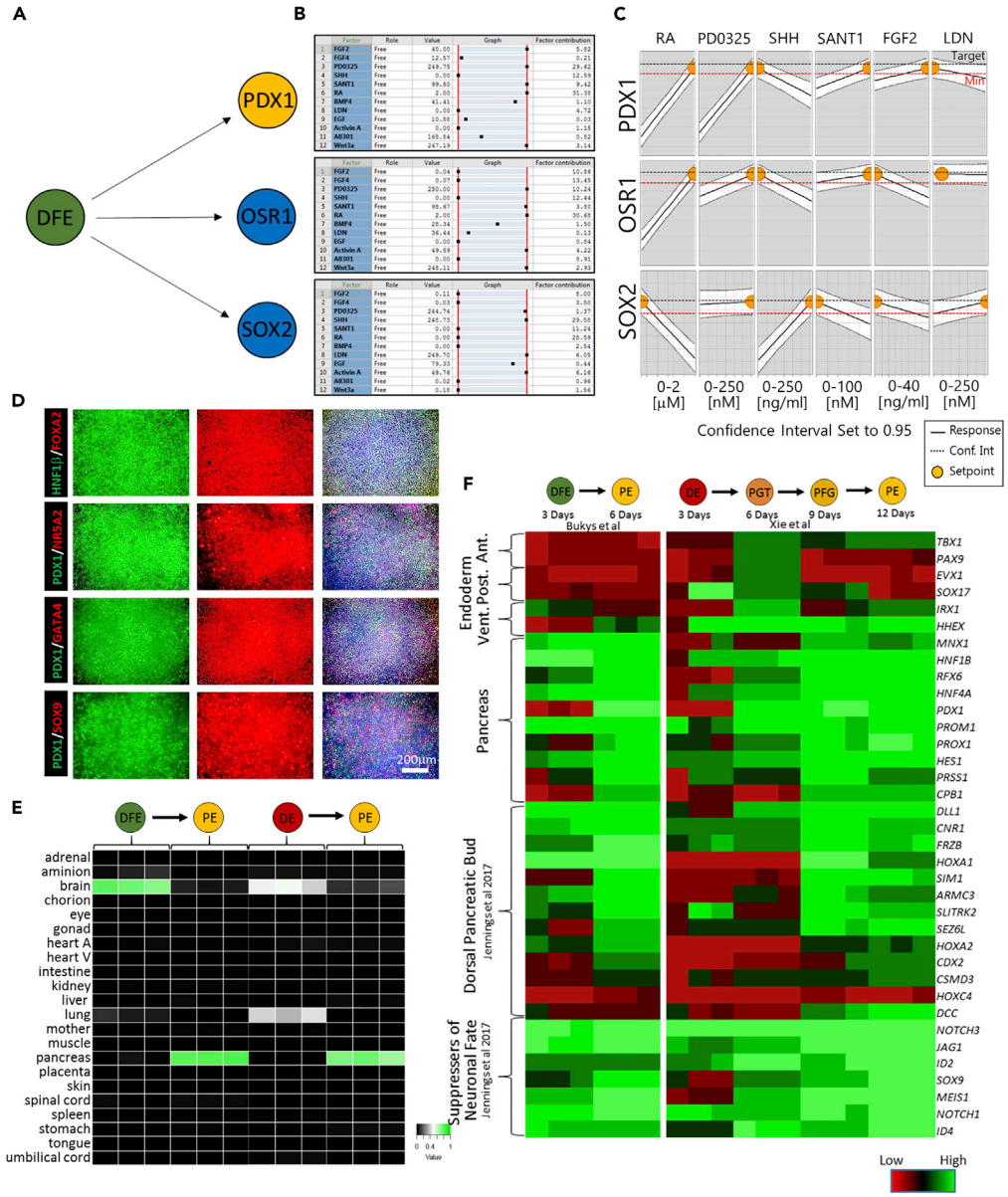


Figure 3. DFE Can Generate Pancreatic Endoderm with a Highly Dorsalized Nature

(A) Schematic showing modeling of the optimization for the stomach genes *SOX2*, *OSR1* and the pancreatic gene *PDX1*.

(B) The corresponding optimizers for the predicted maximal induction of *SOX2*, *OSR1*, and *PDX1*, respectively.

(C) Dynamic profiles for the effectors most responsible for the respective gene activation.

(D) Representative IHC of $PDX1^{Opt}$.

(E) KeyGenes prediction for the respective DE- and DFE-derived pancreatic endoderm.

(F) Heatmap assessing the differential expression of several pancreatic and dorsal-specific pancreatic genes between the two protocols.

Time Dependency of DFE-Derived Fetal-like Endocrine Cells

The importance of NOTCH pathway inhibition for terminal differentiation of endocrine cells is well known (Apelqvist et al., 1999; Jensen et al., 2000; Afelik et al., 2012). We examined the temporal effects of NOTCH inhibition toward endocrine commitment using three media inputs. All included a NOTCH pathway inhibitor (γ-secretase inhibitor XX), and we also evaluated single-SMAD (Alk5 inhibitor) and dual-SMAD inhibition (Alk5i/LDN) (as outlined in Figure S6A). Strikingly, prolonging the period of PE induction diminished endocrine competence (Figure S6B). Of note, increasing the duration at the PE stage led to an increase

in the competence toward acinar and, finally, ductal differentiation (Figure S6C) as assayed through the expression of CPB1, and MIST1 (BHLHA15) and F3, HNF1β and PROM1, respectively. Inspecting the resulting endocrine cultures, we found PDX1+/NKX6.1+ co-expression throughout the culture with patches of NKX2.2/insulin C-peptide co-expression throughout the culture (Figure S7B). Furthermore, mono-hormonal (INS, GCG, and SST) as well as polyhormonal expressions of INS+/GCG + or INS+/SST+ (Figure S7B) were observed. Maximum C-peptide levels occurred after 10 days of exposure to Notch/Alk5 inhibition (Figure S7C), and cellular aggregates demonstrated dithizone retention (Figure S7D). Classical GSIS and microfluidic analysis assays (Adewola et al., 2010; Wang et al., 2012) were consistent and demonstrated an immature βcell physiology capable of synthesizing and storing insulin but with limited functional profile when responding to glucose fluctuations (Figures S7E and S7F). This DFE-derived endocrine cells also display a fetal-like βcell state as previously observed from APS-DE-derived β cells (Hrvatin et al., 2014).

Development of a Small-Molecule Method for Pancreatic Induction

Inspection of the PDX1^{Opt} conditions identified that retinoic acid and PD0325901 had the highest factor contributions of 31.38 and 29.62, respectively (Figure 4A), whereas all other effectors tested had low factor contributions <10. Seeking to achieve a minimally complex method for pancreatic induction we combined the identified critical process parameters for HNF1β^{Opt} (RA and LDN [R/L]) with the critical process parameters of the PDX1^{Opt} (RA and PD0325901 [R/P]). PDX1 levels at the PE stage varied little between these different combinations of the induction methods. Inducing DFE using the full HNF1β^{Opt} and then switching between the PDX1^{Opt} and just the defined CPP components, RA and PD0325901, resulted in cultures that were either 85.6 ± 2.5% or 85.7 ± 2.5% positive for PDX1 expression, respectively (Figures S2B and S2C). However, when PDX1 induction was assayed on DFE induced using only RA and LDN3189, PDX1 levels varied between 81.6 ± 3.2% (for RA/LDN – PDX1^{Opt}) or 86.7 ± 3.6% (for RA/LDN – RA/PD0325901) positive (Figures S2B and S2C). Endocrine competence was then confirmed with an endocrine fate conversion using the aforementioned combination of NOTCH and ALK5 inhibition. DFE induction using HNF1β^{opt} followed by R/P significantly increased levels of the pancreatic markers assayed, and this increase was amplified when followed by an endocrine induction (Figure 4B). A further increase in pancreatic genes was observed when the HNF1β^{Opt} was replaced with the identified CPP (R/L). Using the R/L step, followed by PDX1^{opt} conditions, generated the highest levels of the pancreatic genes *INS*, *GCG*, *NKX6.1*, *CHGA*, *GLP1R*, and *PDX1* (Figure 4B). Yet, comparative changes were observed when both the HNF1β^{Opt} and PDX1^{Opt} were replaced with R/L and R/P in conjunction. Slight decreases in *INS*, *GCG*, *NKX6.1*, *CHGA*, *GLP1R*, and *PDX1* expression were observed under these conditions. However, moderate compensatory increases were observed in the expression of *NEUROD1*, *NGN3*, *PAX4*, *MAFA*, and *SST*. We conclude that the only critical process parameters examined within this study that need to be controlled for the directed differentiation of pluripotent cells to pancreatic endoderm are retinoic acid, BMP, and FGF pathways and that fetal β cells can be rapidly induced through three sequential stages using five small molecules: RA, LDN3189, PD0325901, γ-XX, and A8301.

DISCUSSION

Others have argued that biological complexity is a barrier to fulfilling the potential of biotechnology (Sadowski et al., 2016). A solution to this barrier was proposed to require large numbers of complex experiments, combined with sophisticated software and hardware (Sadowski et al., 2016). Here we describe a generally applicable method for extracting the critical process parameters for media components needed for the manufacturing of a cellular product. Although our focus was on properly controlling pathways during a differentiation event, the HD-DoE process itself is amenable to identifying broader aspects, which may impact cellular identity such as basal media formulations, oxygen dependencies, seeding density, mechanical forces, and even timing of differentiation events as examples. We refer to this process as High-Dimensional Design of Experiments (HD-DoE).

The HD-DoE method circumvents many of the limitations imposed by the hypothesis-driven scientific process. By relying on computer-based experimental designs, such as D-optimal designs, it is possible to extract maximal information per cost unit from the system interrogated. From the experimenters' perspective, it is possible to do virtual experimentation (ISPA) in a highly predictive, fully data-driven manner. This approach allows for hypothesis testing to occur freely within the known space covered by the design geometry after the experiment has been performed. The HD-DoE method is compliant with the basic principles of Quality-by-Design (QbD). We argue that the HD-DoE approach is applicable for defining optimal differentiation conditions for almost any recalcitrant combinatorial problem related to mammalian cell

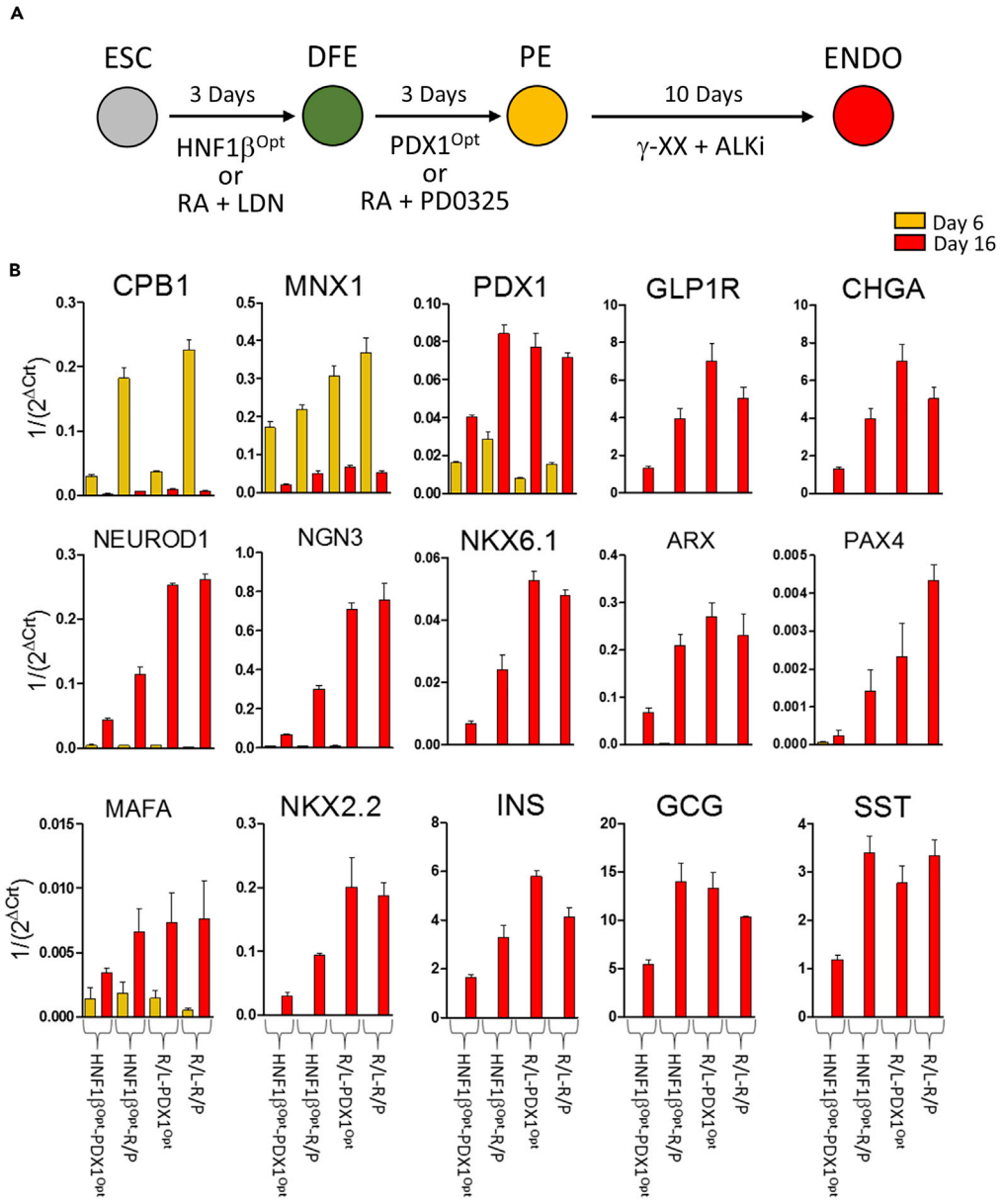


Figure 4. Biological Equivalence Testing of Small-Molecule-Based Pancreatic Endocrine Induction Protocols

(A) Schematic of the reactions performed in which the HNF1β^{Opt} and PDX1^{Opt} were replaced with only the CPP factors as identified as the effectors with the highest predicted factor contributions, retinoic acid and LDN3189 (R/L) or retinoic acid and PD0325901 (R/P), respectively.

(B) Transcript analysis of the four respective reaction conditions at either the PE stage (indicated in yellow) or the endocrine stage (indicated in red) and normalized to the averaged expression of *GAPDH*, *TBP*, and *YWHAZ*. Experiments consist of quadruplicate biological replicates performed in parallel experiments within a single TC plate.

culture. Industrial applications in need of process understanding include production of advanced biologics, specialized cells, and complex tissues for regenerative medicine. Drawbacks for process optimization using current methodologies, where individual components are tested serially, are that process understanding is lacking and that critical process parameters remain undefined, which could result in the process remaining unstable. Thus, for industrial manufacture of human cells, the current reductionist approach—iteratively moving from hypothesis to hypothesis—remains slow and fails to provide a statistical basis upon which manufacture can succeed. For example, Loh et al. used serial testing covering a very large number of experimental conditions (>3,200) to derive a logic for endoderm induction from pluripotency (Loh

et al., 2014). Comparatively, each modeling experiment performed using HD-DoE covers an experimental condition space of >4,000 possible conditions, and each condition is effectively arranged within the 12-dimensional design geometry. The labor intensity, inherent bias, and lack of design interactions coverage using serial experimentation makes machine-based experimental designs superior. Combining such with robotics and deep measurements provides the necessary systems-level interrogation and hereby enables the ISPA process.

Limitations of Study and Methodology

Although interrogations of human biology are amplified by implementation of high-dimensional experimentation, limitations exist. Generally, the ability to make good effector/response choices is highly dependent on developmental knowledge of the system being studied. The approach is not as useful for screening unknown effector inputs as compared with solving optimization problematics. Through our applications, we have understood that modeling experiments gives superior results when they are performed on homogeneous input cultures and that culture heterogeneity significantly blunts strength of response gene modeling. This is to be expected considering that multiple effector competencies would exist in the different cells, and therefore, the HD-DoE method is not ideal, and probably unable, to resolve problems at the subsequent stages for cultures that are highly heterogeneous. Also, interpretation of results relies deeply on the known behavior of chosen responses. Considering the fact that many genes are not limited to a single expression domain, but rather can be expressed in multiple tissues at different developmental stages, ambiguity follows during interpretation. In these cases, findings can only be resolved through understanding of the developmental system being studied. As an example, although PDX1 induction is usually chosen for determination of pancreatic fate, PDX1 expression is wider than in pancreas, occurring in the stomach, duodenum, and gall bladder. Only though examining multiple genes simultaneously and considering alternative PDX1-expressing fates was it possible to determine the best conditions for pancreatic fate. The most important tool for such analysis during ISPA is the dynamic profiling tool set that allows system inspection from the vantage point of, e.g., maximal expression of a given factor.

Although resting on large designs, the effector/response modeling always remains limited to factors being tested within the design space. Currently applied designs in this study were two-level D-optimal interaction designs, and therefore, non-linear responses within the concentration ranges tested would not be appropriately modeled. Inspecting the distribution of standardized residuals provides a means to detect non-linearity. Should a response gene display a response profile exceeding the capability of the second-degree polynomial fit method (Partial Least Squares), transformation of data is possible, and recommended, to achieve a more optimal fit. That said, the most critical limitation of the approach is that models cannot predict conditions that are not inherently embodied within the design space. Conclusions are also limited to the cellular stage tested (as cells typically change competency for input effectors upon differentiating). Consequently, creating robust forward differentiation processes for any specialized human cell using the HD-DoE method involves serial conduction of modeling experiments.

Mechanisms of DFE Specification

In consideration of these aforementioned methodology limitations we have demonstrated a novel protocol capable of rapidly converting pluripotent cells into a regionalized endodermal population (DFE). This endodermal population is competent to form dorsal pancreatic progenitors and undergoing endocrine conversion through a three-stage protocol relying on the use of only five small molecules. Whether this DFE population has an equivalent state during human development or whether differentiation of pluripotent cells to the DFE state occurs in absence of a gastrulation-like event is unknown. Considering that the developing embryo begins retinoic acid production during gastrulation (Ulven et al., 2000), and that retinoic acid patterns endoderm toward dorsal fates (Davenport et al., 2016), it is possible that locally produced retinoids pattern gastrulating cells during development. Because provision of the ALK5 inhibitor positively affects DFE generation, it is suggested that Nodal signaling does not occur in this protocol. Alternatively, since it is known that not all gut endoderm derives from gastrulating cells (Kwon et al., 2008, Mcdonald and Rossant, 2014) and that Nodal signaling is not needed for generating dorsal fates in more primitive developmental models (Rottinger et al., 2015), it is possible that dorsal endoderm arises from a non-migratory population adjacent to the node (Ulven et al., 2000). Given that plasticity is known to exist in undifferentiated endoderm (Kumar et al., 2003) the hypothesis that cell fate can be directed through positional cues is also plausible. Interestingly, many of the genes shown to be directly activated by RA in this study are known to be induced during Wnt/AA-mediated generation of definitive endoderm (most

notably *FOXA2* and *CXCR4*) and therefore potentially responding to RA stronger than they do to the WNT/AA inputs. NODAL induction of DE differs from AA-induced DE-induction (Chen et al., 2013), where NODAL is the physiologically relevant molecule expressed within the node. It is possible that the impact of AA in most currently used protocols creates ventral patterning due to unknown differences in the functional activity of these different morphogens. Others have argued that the descendant population of an AA-induction step is heterogeneous (Green et al., 2011), further substantiated by the use of AA at lower concentrations to attain mesodermal induction. Whether RA contributes early in AA induction of DE is not fully established. Future studies are needed to clarify to what extent APS-derived endoderm might be supported by trace levels of in-culture produced RA or if retinoids provided through serum during APS-type DE formation contributes to the endoderm induction (D'amour et al., 2005).

Dorsal versus Ventral Pancreas Induction

Here we have demonstrated that it is possible to obtain patterned endoderm directly from pluripotency and have shown that it is competent for pancreatic induction. In comparison with previously published pancreatic protocols, we have demonstrated that three previously described differentiation stages can be reduced into a single stage. As a basis for our interpretation, RNA-seq data demonstrated that DFE is of a dorsal character using a set of 13 previously described genes defining a dorsal identity (Jennings et al., 2017). When compared with a pancreatic directed differentiation protocol (Xie et al., 2013) we concluded that DE-derived PE adopted a pancreatic phenotype during the published "PGT" stage as evident through the expression of *HHEX*, *HNF1β*, *RFX6*, *HNF4A*, *PDX1*, *PROM1*, and *PROX1* but had no discernible dorsal identity until the cultures were exposed to retinoic acid at the "PFG" stage. This implies that the dorsal phenotype previously attributed to this protocol (Jennings et al., 2017) was a result of redirecting the population toward a dorsal fate at stage 3 of the protocol. For the DFE, the dorsal identity carries forward to pancreas and this state is permissive for induction of all pancreatic lineages. Previous studies have demonstrated that DE-derived pancreatic progenitors suffer from stray hepatic fates controlled through BMP signaling (Mfopou et al., 2010), a phenomenon that could be attributed to the ventral pancreas having bipotential competency for liver induction (Angelo et al., 2012; Deutsch et al., 2001; Tremblay and Zaret, 2005; Bort et al., 2004) and that early lateral plate-derived BMPs instruct this precursor toward hepatic fates (Chung et al., 2008). DFE-derived PE showed a bipotential competence for antral stomach induction, and we propose it is controlled through SHH signaling. This observation is supported by developmental studies in chick (Hebrok et al., 1998). Regardless of the dorsal/ventral origin of the endodermal population, both DFE and APS-DE readily give rise to pancreatic endoderm capable of generating endocrine cells. As observed for APS-DE derived cells, the DFE-derived endocrine cells are functionally more similar to fetal β cells than to fully mature glucose-responsive β cells. Whether DFE-derived insulin-producing cells can undergo *in vivo* maturation, as shown for APS-DE-derived cells (Kroon et al., 2008), needs to be determined. From a diabetes cell therapy development perspective, functional maturation toward glucose-dependent insulin release *in vitro* is a desirable goal that can be addressed using the HD-DoE method.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Jan Jensen (jensenj2@ccf.org or jjensen@trailbio.com, Phone: + 001 216 445 0990/+001 216 479 9754).

Materials Availability

Not applicable.

Data and Code Availability

RNA-seq data generated in this study are available at the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra).

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101346.

ACKNOWLEDGMENTS

Funding for this study was provided by the Ohio Third Frontier grant number IPP 12-258 through assignment of demonstrator projects awarded by Trailhead Biosystems (www.trailbio.com) and through American Diabetes Association (ADA) grant number 1-16-ICTS-053. The authors thank Oliver Wessely and John McAfee for critical reading of the manuscript. The Chicago Diabetes Project is thanked for supporting the functional assays of stem cell derived insulin cells and its dedication towards continously supporting the quest towards developing a functional therapy for Type I diabetes. This paper is dedicated to the memory of Phyllis Probeck Brandon.

AUTHOR CONTRIBUTIONS

M.A.B. was responsible for experimental design, data analysis, and writing of the manuscript. Experiments were performed by M.A.B. (60%), A.M. (15%), K.F. (15%), K.S. (5%), and D.T. (5%). Y.W. and J.O. performed microfluidic assays and critical reading of the manuscript. J.J. provided project oversight and writing of the manuscript.

DECLARATION OF INTERESTS

J.J. is founder of and shareholder of Trailhead Biosystems, Inc., Cleveland, OH, USA. J.O. is a founder of SIGILON, Cambridge, MA, USA, advisory board member and shareholder, and is founder and shareholder of CellTrans Inc., Chicago, IL, USA. This work has been filed as international application number PCT/US20/34,201.

Received: May 28, 2019
Revised: April 15, 2020
Accepted: July 2, 2020
Published: August 21, 2020

REFERENCES

Adewola, A.F., Lee, D., Harvat, T., Mohammed, J., Eddington, D.T., Oberholzer, J., and Wang, Y. (2010). Microfluidic perfusion and imaging device for multi-parametric islet function assessment. Biomed. Microdevices *12*, 409–417.

Afelik, S., Qu, X., Hasrouni, E., Bukys, M.A., Deering, T., Nieuwoudt, S., Rogers, W., Macdonald, R.J., and Jensen, J. (2012). Notch-mediated patterning and cell fate allocation of pancreatic progenitor cells. Development *139*, 1744–1753.

Angelo, J.R., Guerrero-Zayas, M.I., and Tremblay, K.D. (2012). A fate map of the murine pancreas buds reveals a multipotent ventral foregut organ progenitor. PLoS One *7*, e40707.

Apelqvist, A., Li, H., Sommer, L., Beatus, P., Anderson, D.J., Honjo, T., Hrabe de Angelis, M., Lendahl, U., and Edlund, H. (1999). Notch signalling controls pancreatic cell differentiation. Nature *400*, 877–881.

Body-Bechou, D., Loget, P., D'herve, D., le Fiblec, B., Grebille, A.G., le Guern, H., Labarthe, C., Redpath, M., Cabaret-Dufour, A.S., Sylvie, O., et al. (2014). TCF2/HNF-1beta mutations: 3 cases of fetal severe pancreatic agenesis or hypoplasia and multicystic renal dysplasia. Prenat Diagn. *34*, 90–93.

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

Carinhas, N., Oliveira, R., Alves, P.M., Carrondo, M.J., and Teixeira, A.P. (2012). Systems biotechnology of animal cells: the road to prediction. Trends Biotechnol. *30*, 377–385.

Chakrabarty, A., Buzzard, G.T., and Rundell, A.E. (2013). Model-based design of experiments for cellular processes. Wiley Interdiscip. Rev. Syst. Biol. Med. *5*, 181–203.

Chen, A.E., Borowiak, M., Sherwood, R.I., Kweudjeu, A., and Melton, D.A. (2013). Functional evaluation of ES cell-derived endodermal populations reveals differences between Nodal and Activin A-guided differentiation. Development *140*, 675–686.

Chung, W.S., Shin, C.H., and Stainier, D.Y. (2008). Bmp2 signaling regulates the hepatic versus pancreatic fate decision. Dev. Cell *15*, 738–748.

D'amour, K.A., Agulnick, A.D., Eliazer, S., Kelly, O.G., Kroon, E., and Baetge, E.E. (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat. Biotechnol. *23*, 1534–1541.

Davenport, C., Diekmann, U., Budde, I., Detering, N., and Naujok, O. (2016). Anterior-posterior patterning of definitive endoderm generated from human embryonic stem cells depends on the differential signaling of retinoic acid, Wnt-, and BMP-signaling. Stem Cells *34*, 2635–2647.

de Vas, M.G., Kopp, J.L., Heliot, C., Sander, M., Cereghini, S., and Haumaitre, C. (2015). Hnf1b controls pancreas morphogenesis and the generation of Ngn3+ endocrine progenitors. Development *142*, 871–882.

Deutsch, G., Jung, J., Zheng, M., Lora, J., and Zaret, K.S. (2001). A bipotential precursor population for pancreas and liver within the embryonic endoderm. Development *128*, 871–881.

Gadue, P., Huber, T.L., Paddison, P.J., and Keller, G.M. (2006). 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 *103*, 16806–16811.

Gerin, S., Mathy, G., and Franck, F. (2014). Modeling the dependence of respiration and photosynthesis upon light, acetate, carbon dioxide, nitrate and ammonium in *Chlamydomonas reinhardtii* using design of experiments and multiple regression. BMC Syst. Biol. *8*, 96.

Green, M.D., Chen, A., Nostro, M.C., D'souza, S.L., Schaniel, C., Lemischka, I.R., Gouon-Evans, V., Keller, G., and Snoeck, H.W. (2011). Generation of anterior foregut endoderm from human embryonic and induced pluripotent stem cells. Nat. Biotechnol. *29*, 267–272.

Haumaitre, C., Barbacci, E., Jenny, M., Ott, M.O., Gradwohl, G., and Cereghini, S. (2005). Lack of TCF2/vHNF1 in mice leads to pancreas agenesis. Proc. Natl. Acad. Sci. U S A *102*, 1490–1495.

Hebrok, M., Kim, S.K., and Melton, D.A. (1998). Notochord repression of endodermal Sonic hedgehog Permits pancreas development. Genes Dev. *12*, 1705–1713.

Hrvatin, S., O'donnell, C.W., Deng, F., Millman, J.R., Pagliuca, F.W., Diiorio, P., Rezania, A., Gifford, D.K., and Melton, D.A. (2014). Differentiated human stem cells resemble fetal, not adult, beta cells. Proc. Natl. Acad. Sci. U S A *111*, 3038–3043.

Jennings, R.E., Berry, A.A., Gerrard, D.T., Wearne, S.J., Strutt, J., Withey, S., Chhatriwala, M., Piper Hanley, K., Vallier, L., Bobola, N., and Hanley, N.A. (2017). Laser capture and deep sequencing reveals the transcriptomic programmes regulating the onset of pancreas and liver differentiation in human embryos. Stem Cell Rep. *9*, 1387–1394.

Jensen, J., Heller, R.S., Funder-Nielsen, T., Pedersen, E.E., Lindsell, C., Weinmaster, G., Madsen, O.D., and Serup, P. (2000). Independent development of pancreatic alpha- and beta-cells from neurogenin3-expressing precursors: a role for the notch pathway in repression of premature differentiation. Diabetes *49*, 163–176.

Juran, J. (1993). Made in the U. S. A.: a renaissance in quality. Harv. Business Rev. *71* (4), 42–47.

Kitano, H. (2002). Computational systems biology. Nature *420*, 206–210.

Kroon, E., Martinson, L.A., Kadoya, K., Bang, A.G., Kelly, O.G., Eliazer, S., Young, H., Richardson, M., Smart, N.G., Cunningham, J., et al. (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nat. Biotechnol. *26*, 443–452.

Kumar, M., Jordan, N., Melton, D., and Grapin-Botton, A. (2003). Signals from lateral plate mesoderm instruct endoderm toward a pancreatic fate. Dev. Biol. *259*, 109–122.

Kumar, V., Bhalla, A., and Rathore, A.S. (2014). Design of experiments applications in bioprocessing: concepts and approach. Biotechnol. Prog. *30*, 86–99.

Kwon, G.S., Viotti, M., and Hadjantonakis, A.K. (2008). The endoderm of the mouse embryo arises by dynamic widespread intercalation of embryonic and extraembryonic lineages. Dev. Cell *15*, 509–520.

Li, H., Arber, S., Jessell, T.M., and Edlund, H. (1999). Selective agenesis of the dorsal pancreas in mice lacking homeobox gene Hlxb9. Nat. Genet. *23*, 67–70.

Lipsitz, Y.Y., Timmins, N.E., and Zandstra, P.W. (2016). Quality cell therapy manufacturing by design. Nat. Biotechnol. *34*, 393–400.

Loh, K.M., Ang, L.T., Zhang, J., Kumar, V., Ang, J., Auyeong, J.Q., Lee, K.L., Choo, S.H., Lim, C.Y., Nichane, M., et al. (2014). Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. Cell Stem Cell *14*, 237–252.

Martin, M., Gallego-Llamas, J., Ribes, V., Kedinger, M., Niederreither, K., Chambon, P., Dolle, P., and Gradwohl, G. (2005). Dorsal pancreas agenesis in retinoic acid-deficient Raldh2 mutant mice. Dev. Biol. *284*, 399–411.

Matsuno, K., Mae, S.I., Okada, C., Nakamura, M., Watanabe, A., Toyoda, T., Uchida, E., and Osafune, K. (2016). Redefining definitive endoderm subtypes by robust induction of human induced pluripotent stem cells. Differentiation *92*, 281–290.

Matsuura, K., Katsumoto, K., Fukuda, K., Kume, K., and Kume, S. (2009). Conserved origin of the ventral pancreas in chicken. Mech. Dev. *126*, 817–827.

Mccracken, K.W., Cata, E.M., Crawford, C.M., Sinagoga, K.L., Schumacher, M., Rockich, B.E., Tsai, Y.H., Mayhew, C.N., Spence, J.R., Zavros, Y., and Wells, J.M. (2014). Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. Nature *516*, 400–404.

Mcdonald, A.C., and Rossant, J. (2014). Gut endoderm takes flight from the wings of mesoderm. Nat. Cell Biol. *16*, 1128–1129.

McConnell, J., Nunnally, B.K., and McGarvey, B. (2010). The forgotten origins of quality by design. J. Validation Technol. 30–34.

Mendes, L.F., Tam, W.L., Chai, Y.C., Geris, L., Luyten, F.P., and Roberts, S.J. (2016). Combinatorial Analysis of Growth Factors Reveals the Contribution of Bone Morphogenetic Proteins to Chondrogenic Differentiation of Human Periosteal Cells. Tissue Eng. Part C Methods *22*, 473–486.

Mercier, S.M., Diepenbroek, B., Dalm, M.C., Wijffels, R.H., and Streefland, M. (2013). Multivariate data analysis as a PAT tool for early bioprocess development data. J. Biotechnol. *167*, 262–270.

Mfopou, J.K., Chen, B., Mateizel, I., Sermon, K., and Bouwens, L. (2010). Noggin, retinoids, and fibroblast growth factor regulate hepatic or pancreatic fate of human embryonic stem cells. Gastroenterology *138*, 2233–2245, 2245 e1-14.

Molotkov, A., Molotkova, N., and Duester, G. (2005). Retinoic acid generated by Raldh2 in mesoderm is required for mouse dorsal endodermal pancreas development. Dev. Dyn. *232*, 950–957.

Pagliuca, F.W., Millman, J.R., Gurtler, M., Segel, M., van Dervort, A., Ryu, J.H., Peterson, Q.P., Greiner, D., and Melton, D.A. (2014). Generation of functional human pancreatic beta cells in vitro. Cell *159*, 428–439.

Poll, A.V., Pierreux, C.E., Lokmane, L., Haumaitre, C., Achouri, Y., Jacquemin, P., Rousseau, G.G., Cereghini, S., and Lemaigre, F.P. (2006). A vHNF1/TCF2-HNF6 cascade regulates the transcription factor network that controls generation of pancreatic precursor cells. Diabetes *55*, 61–69.

Rathore, A.S., Mittal, S., Pathak, M., and Arora, A. (2014). Guidance for performing multivariate data analysis of bioprocessing data: pitfalls and recommendations. Biotechnol. Prog. *30*, 967–973.

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

Roost, M.S., van Iperen, L., Ariyurek, Y., Buermans, H.P., Arindrarto, W., Devalla, H.D., Passier, R., Mummery, C.L., Carlotti, F., de Koning, E.J., et al. (2015). KeyGenes, a Tool to Probe Tissue Differentiation Using a Human Fetal Transcriptional Atlas. Stem Cell Rep. *4*, 1112–1124.

Rottinger, E., Dubuc, T.Q., Amiel, A.R., and Martindale, M.Q. (2015). Nodal signaling is required for mesodermal and ventral but not for dorsal fates in the indirect developing hemichordate, *Ptychodera flava*. Biol. Open *4*, 830–842.

Sadowski, M.I., Grant, C., and Fell, T.S. (2016). Harnessing QbD, Programming Languages, and Automation for Reproducible Biology. Trends Biotechnol. *34*, 214–227.

Sampaziotis, F., de Brito, M.C., Madrigal, P., Bertero, A., Saeb-Parsy, K., Soares, F.A.C., Schrumpf, E., Melum, E., Karlsen, T.H., Bradley, J.A., et al. (2015). Cholangiocytes derived from human induced pluripotent stem cells for disease modeling and drug validation. Nat. Biotechnol. *33*, 845–852.

Spence, J.R., Lange, A.W., Lin, S.C., Kaestner, K.H., Lowy, A.M., Kim, I., Whitsett, J.A., and Wells, J.M. (2009). Sox17 regulates organ lineage segregation of ventral foregut progenitor cells. Dev. Cell *17*, 62–74.

Spence, J.R., Mayhew, C.N., Rankin, S.A., Kuhar, M.F., Vallance, J.E., Tolle, K., Hoskins, E.E., Kalinichenko, V.V., Wells, S.I., Zorn, A.M., et al. (2011). Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. Nature *470*, 105–109.

Swain, S., Padhy, R., Jena, B.R., and Babu, S.M. (2018). Quality by design: concept to applications. Curr. Drug Discov. Technol. *16*, 240–250.

Tremblay, K.D., and Zaret, K.S. (2005). Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. Dev. Biol. *280*, 87–99.

Ulven, S.M., Gundersen, T.E., Weedon, M.S., Landaas, V.O., Sakhi, A.K., Fromm, S.H., Geronimo, B.A., Moskaug, J.O., and Blomhoff, R. (2000). Identification of endogenous retinoids, enzymes, binding proteins, and receptors during early postimplantation development in mouse: important role of retinal dehydrogenase type 2 in synthesis of all-trans-retinoic acid. Dev. Biol. *220*, 379–391.

Wang, Y., Lee, D., Zhang, L., Jeon, H., Mendoza-Elias, J.E., Harvat, T.A., Hassan, S.Z., Zhou, A., Eddington, D.T., and Oberholzer, J. (2012). Systematic prevention of bubble formation and accumulation for long-term culture of pancreatic islet cells in microfluidic device. Biomed. Microdevices *14*, 419–426.

Xie, R., Everett, L.J., Lim, H.W., Patel, N.A., Schug, J., Kroon, E., Kelly, O.G., Wang, A., D'amour, K.A., Robins, A.J., et al. (2013). Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. Cell Stem Cell *12*, 224–237.

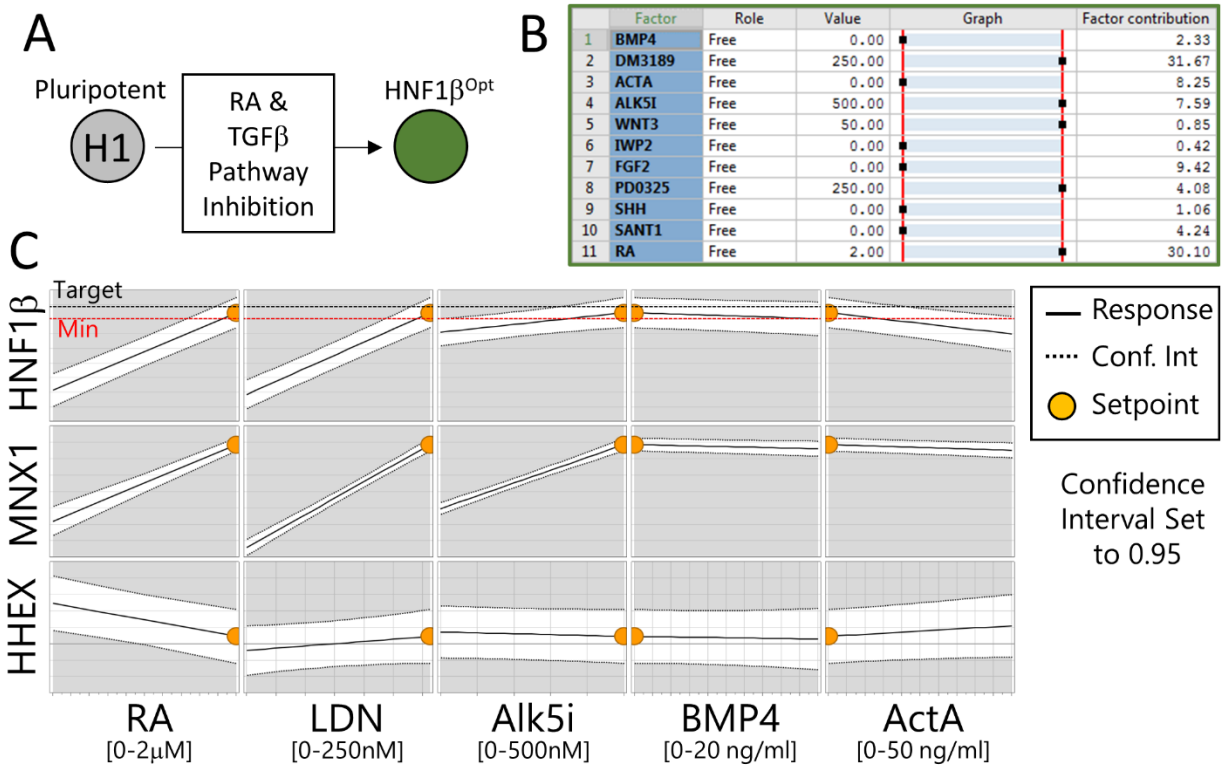
iScience, Volume *23*

Supplemental Information

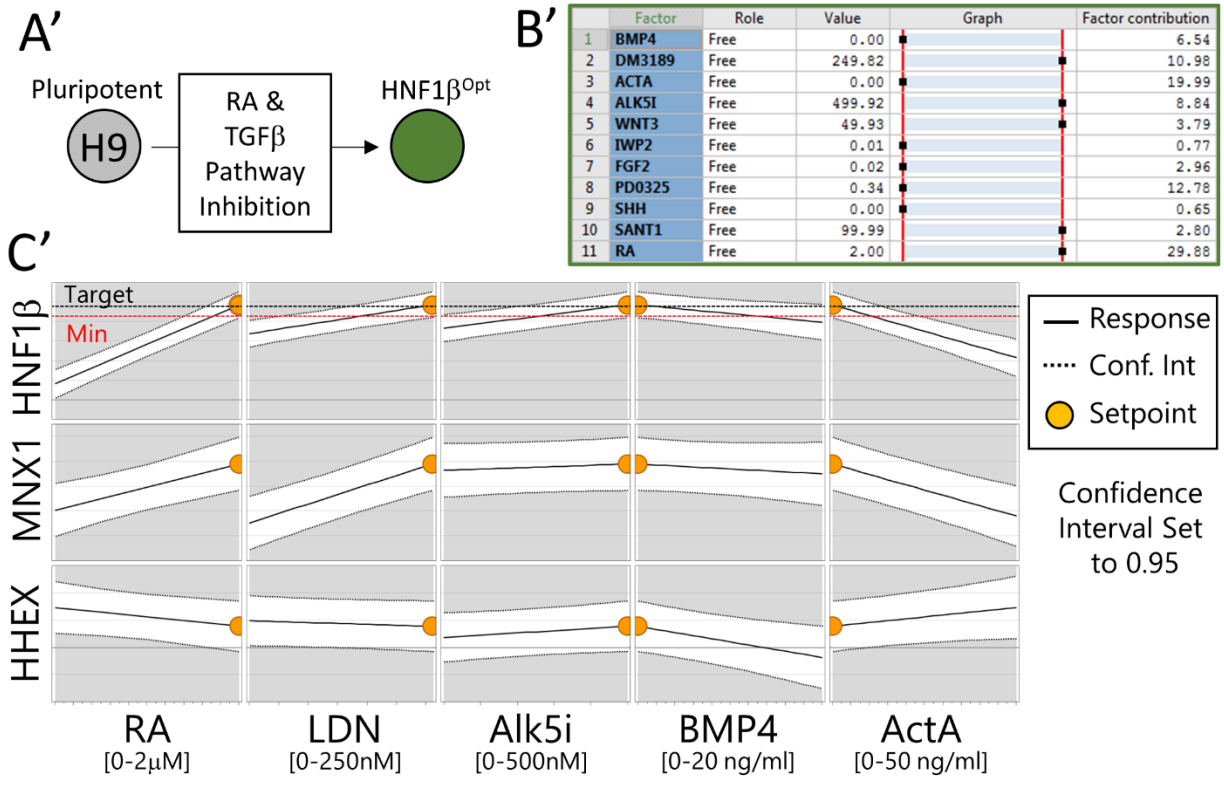
High-Dimensional Design-Of-Experiments Extracts Small-Molecule-Only Induction Conditions for Dorsal Pancreatic Endoderm from Pluripotency

Michael A. Bukys, Alexander Mihas, Krystal Finney, Katie Sears, Divya Trivedi, Yong Wang, Jose Oberholzer, and Jan Jensen

Supplemental Figure 1

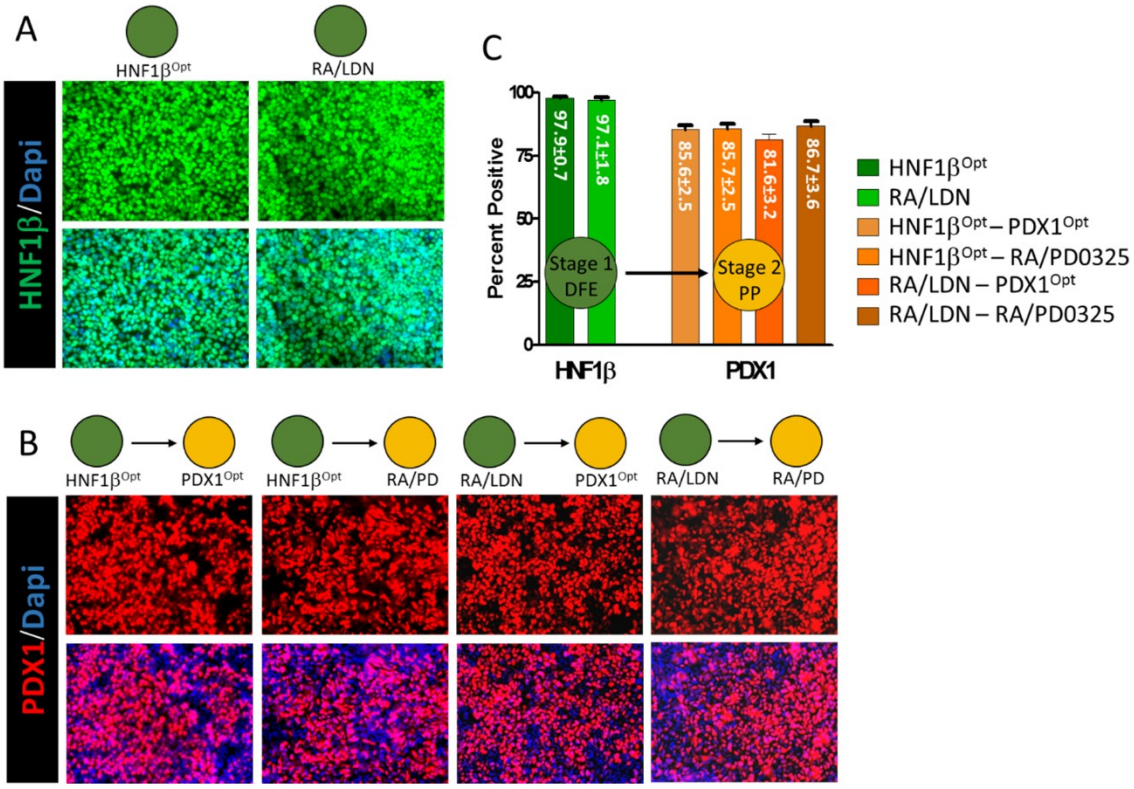


	Factor	Role	Value	Graph	Factor contribution
1	BMP4	Free	0.00		2.33
2	DM3189	Free	250.00		31.67
3	ACTA	Free	0.00		8.25
4	ALK5I	Free	500.00		7.59
5	WNT3	Free	50.00		0.85
6	IWP2	Free	0.00		0.42
7	FGF2	Free	0.00		9.42
8	PD0325	Free	250.00		4.08
9	SHH	Free	0.00		1.06
10	SANT1	Free	0.00		4.24
11	RA	Free	2.00		30.10

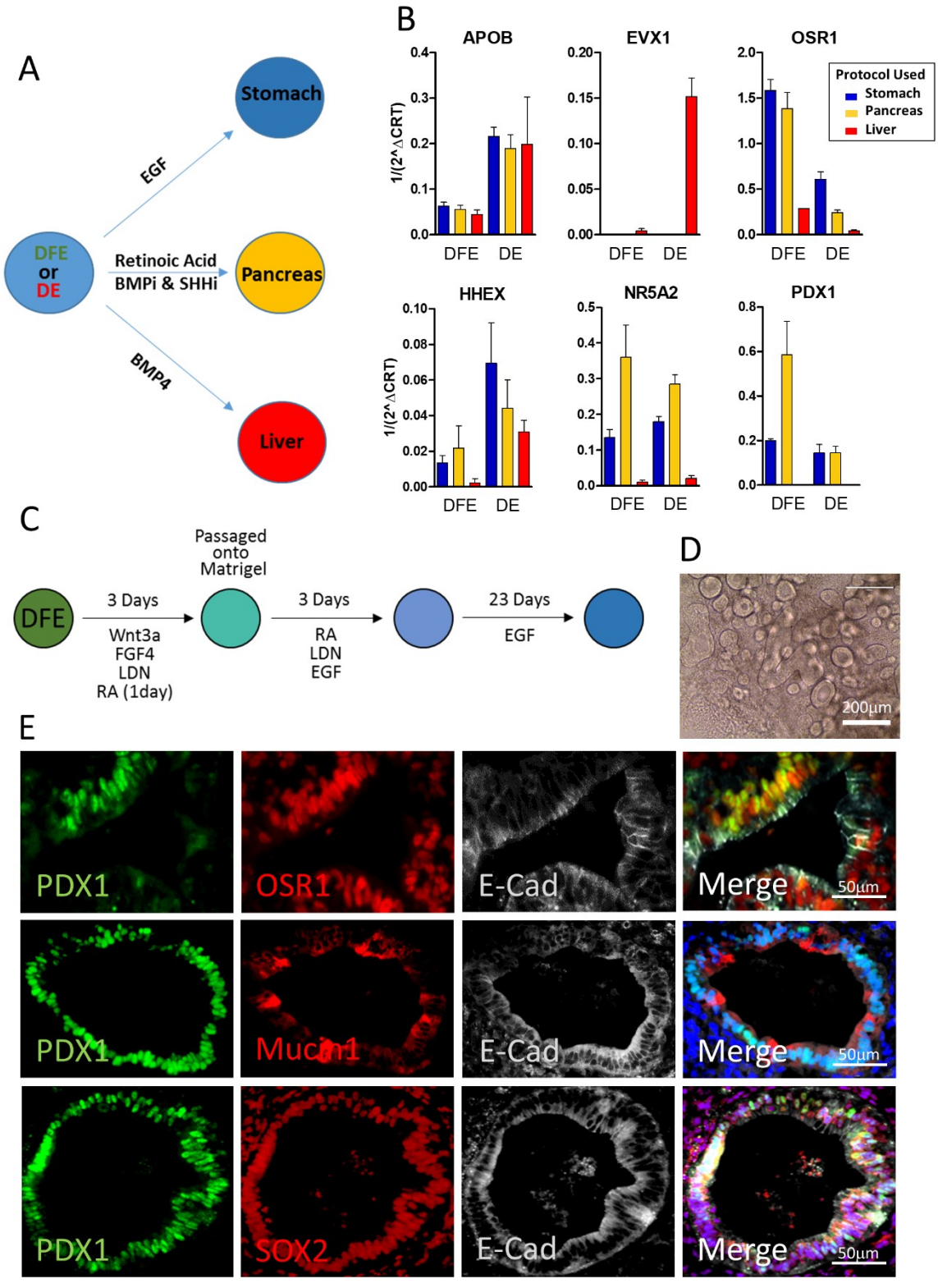


	Factor	Role	Value	Graph	Factor contribution
1	BMP4	Free	0.00		6.54
2	DM3189	Free	249.82		10.98
3	ACTA	Free	0.00		19.99
4	ALK5I	Free	499.92		8.84
5	WNT3	Free	49.93		3.79
6	IWP2	Free	0.01		0.77
7	FGF2	Free	0.02		2.96
8	PD0325	Free	0.34		12.78
9	SHH	Free	0.00		0.65
10	SANT1	Free	99.99		2.80
11	RA	Free	2.00		29.88

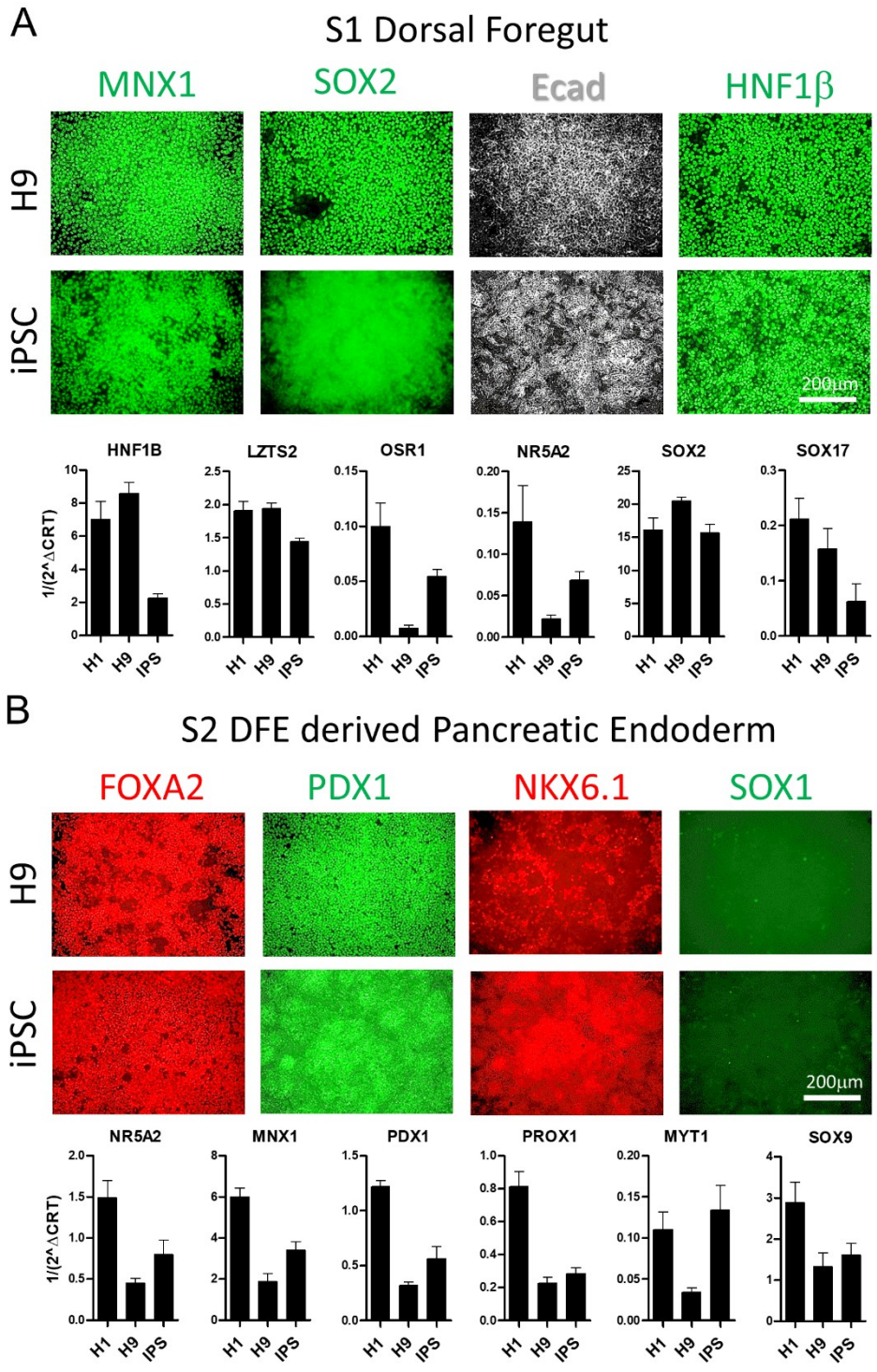
Supplemental Figure 2



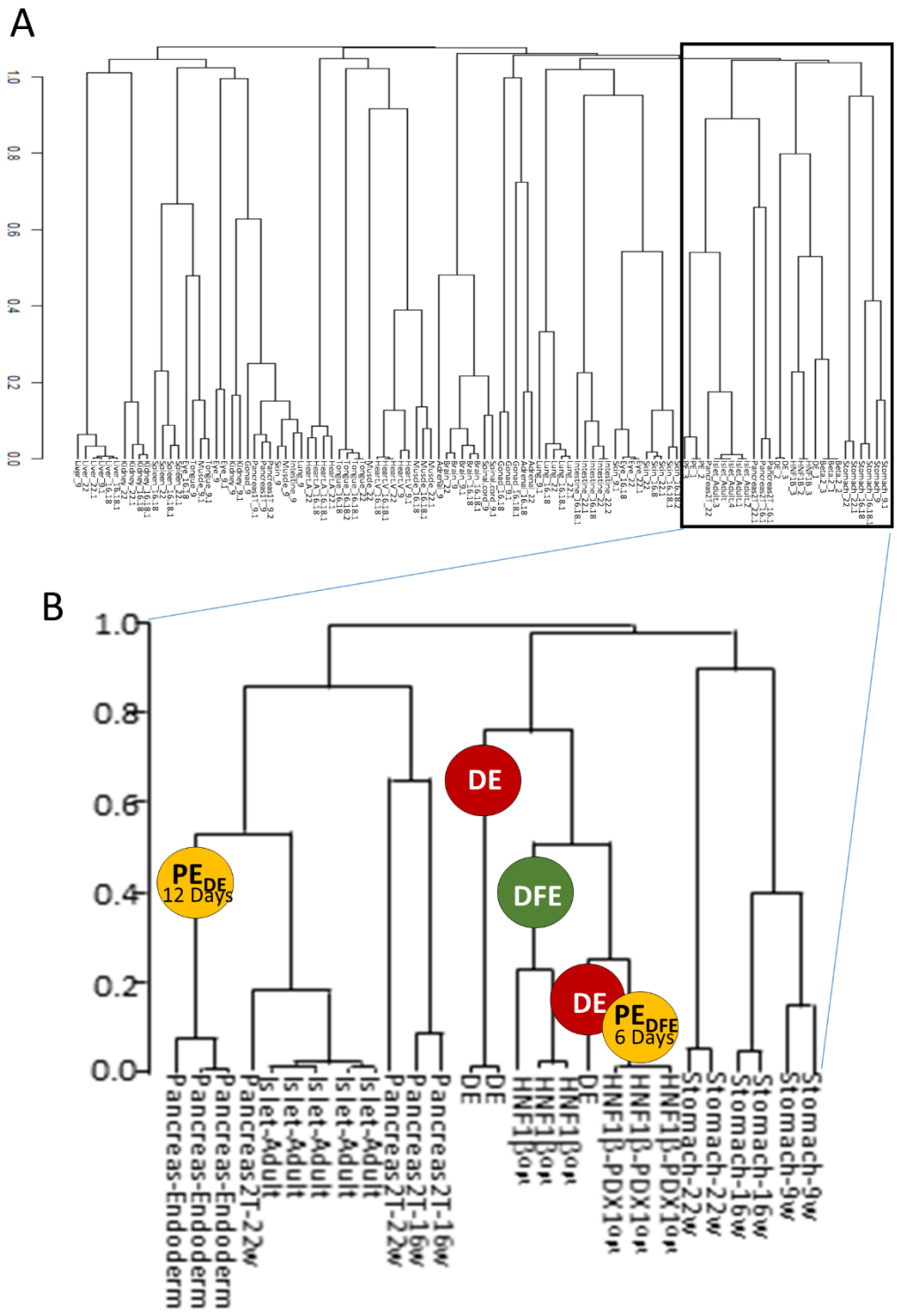
Supplemental Figure 3



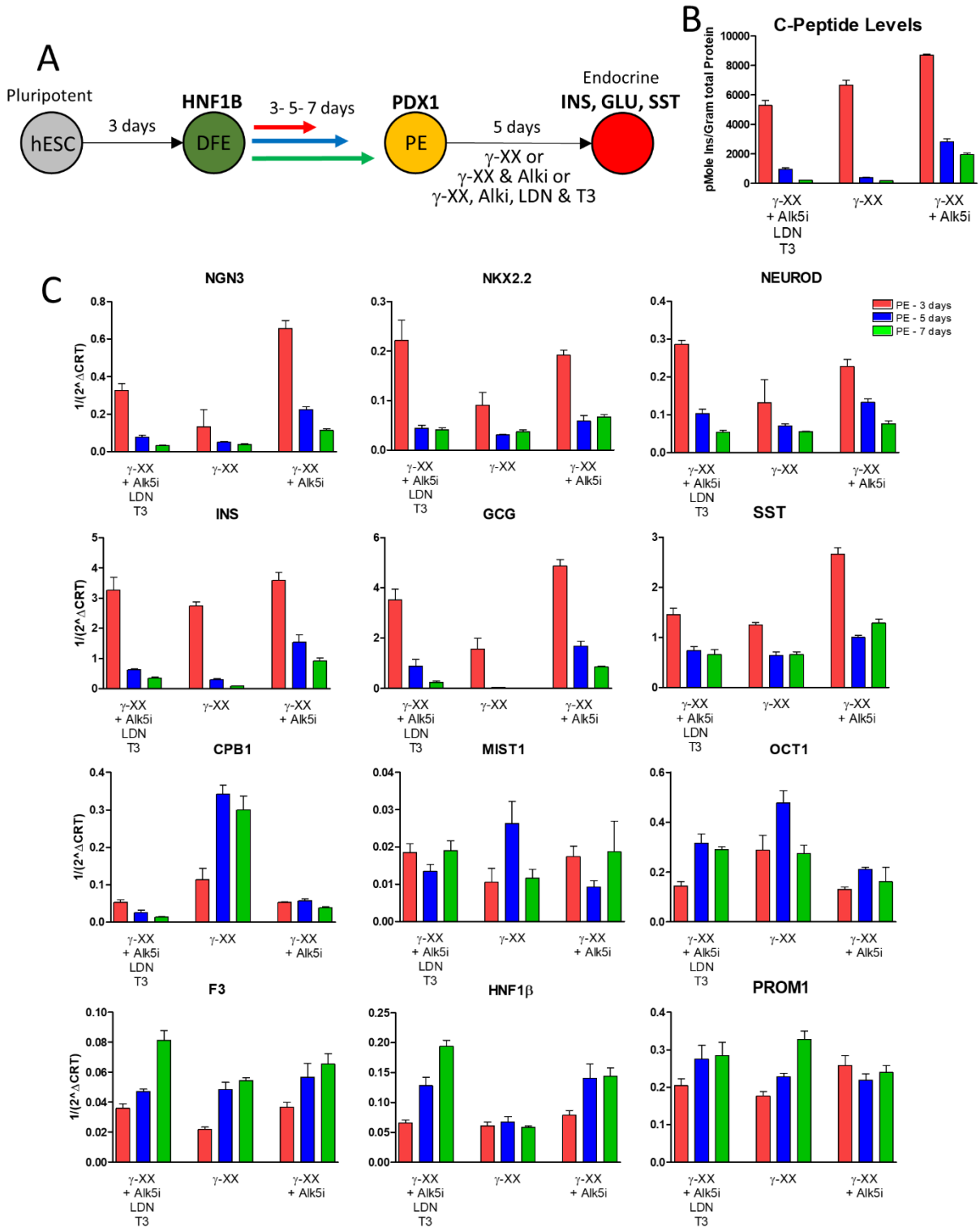
Supplemental Figure 4



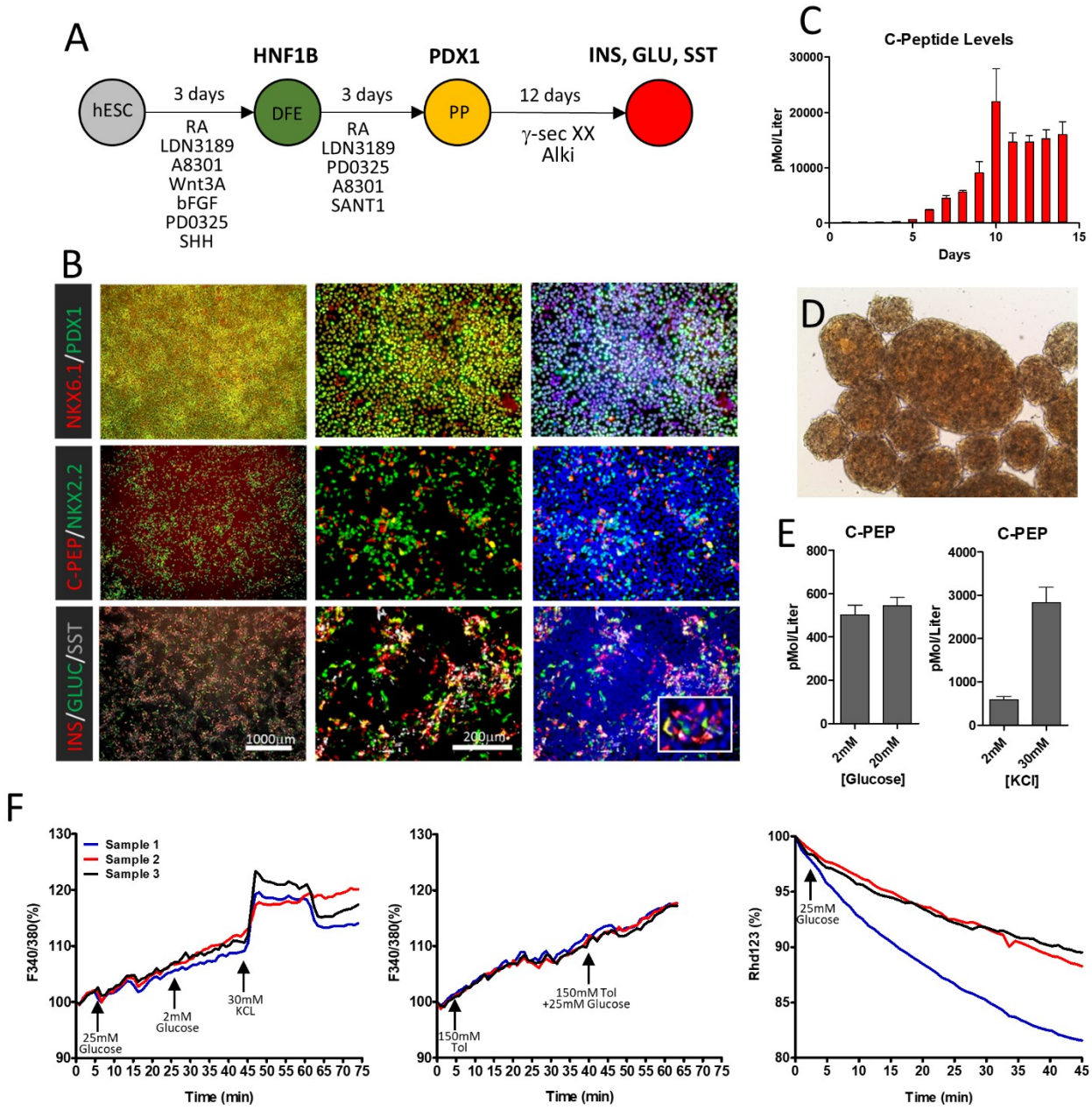
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



Supplemental Figure 1 – Reproducibility of Modeling Experiments between different biological cell lines. Related to Fig. 1 A: Schematic showing the critical inputs needed for DFE induction. **B:** Actual HNF1β^{Opt} predicted from HD-DoE experiment performed as outlined in Supp. Table 1. **C:** Dynamic profiles demonstrating the split between dorsal and ventral endodermal fields. The Y-axis provides the response dynamics to the listed effector at the set point, ranging from minimal to maximal gene expression within the experimental space; white coloring represents the 95% confidence interval band of the effect. The X-axis is the concentration of the effector used. The setpoint depicted within the dynamic profile is the concentration predicted in the respective HNF1β^{Opts}. All prime letters are the same as detailed above but from a replicate experiment using the H9 pluripotent cell line.

Supplemental Figure 2 – Quantification of optimizers. Related to Fig 1 and Fig. 3. A: Panel showing a representative image of the HNF1β^{Opt} expression as compared to the defined CPP components RA and LDN. **B:** Panel showing a representative image of the PDX1^{Opt} expression as compared to the defined CPP components RA and PD0325901 on DFE generated in both previous defined conditions. All representative images shown in panels A and B are taken from a stitched region within a 100 images collage. **C:** Graph totaling the percentage of the culture positive for either HNF1β (for DFE generating conditions) or PDX1 (for PE generating conditions). Error bars were generated from reactions being performed in triplicate parallel biological replicates.

Supplemental Figure 3 – Patterning competence of DFE differs from DE-derived endoderm. Related to Fig. 2. A: Shows a schematic of how either an HNF1β Optimized Culture or a culture differentiated to definitive endoderm respond to being challenged to differentiate towards stomach, pancreas or liver fates. **B:** Relative expression of select markers for genes indicated between the different protocols. Blue bars represent cultures that were provided conditions designed to differentiate towards stomach. Yellow bars are the cultures that were provided conditions that were designed to generate pancreas. Red bars represent the cultures that were provided conditions designed to generate liver. Experiments consist of quadruplicate biological replicates performed in parallel experiments within a single TC plate. Expression levels were normalized to the averaged expression of *GAPDH*, *TBP* and *YWHAZ*. **C:** Schematic of a follow up experiment in which organoids were grown from HNF1β^{opt} cultures challenged with stomach inducing conditions. **D:** DFE-Derived organoids that have been differentiated and grown for 23 days. **E:** Representative organoid staining for gene markers of PSC-derived presumptive antral PDX1+/SOX2+/OSR1+ stomach.

Supplemental Figure 4 - Robust and homogenous induction of pancreatic fate from DFE using iPSC and female hESC. Related to Fig. 1 and Fig. 3. A: IHC and corresponding transcript analysis validating the HNF1β^{opt} protocol between three different pluripotent cell lines. **B:** IHC and corresponding transcript analysis validating the PDX1^{Opt}

performed on DFE cultures comparing three different pluripotent cell lines. Error bars represent quadruplicate biological replicates performed in parallel experiments within a single TC plate. Expression levels were normalized to the averaged expression of *GAPDH*, *TBP* and *YWHAZ*.

Supplemental Figure 5 – Hierarchical clustering comparing pluripotent derived cells to several primary tissues. Related to Fig 3. A: Hierarchical clustering was performed on the KeyGenes data set in R. **B:** Since all pluripotent derived samples fell within the same cluster containing the human fetal stomach, fetal pancreas and adult islets samples this region of the cluster is shown.

Supplemental Figure 6 - Effective and rapid induction of pancreatic endocrine fate from pluripotency via DFE endoderm using small molecules. Related to Fig. 4. A: Shows a schematic of the reactions performed varying the pancreatic induction timing. **B:** C-peptide levels detected throughout the different experimental conditions as determined through triplicate biological samples quantified through ELISA. **C:** Transcript expression analysis for several key pancreatic and endocrine specific genes normalized to the averaged expression of *GAPDH*, *TBP* and *YWHAZ*. Error bars represent quadruplicate biological replicates performed in parallel experiments within a single TC plate.

Supplemental Figure 7 – DFE derived endocrine cells have an immature phenotype. Related to Fig. 4. A: Schematic showing the directed differentiation protocol used to generate endocrine fields. **B:** IHC Evaluation of the end point culture for various endocrine specific factors. **C:** A 14-day endocrine induction was assayed daily for c-peptide generation using triplicate biological samples quantified through ELISA. **D:** Retention of dithizone by pluripotent derivatives grown as aggregates. **E:** Glucose stimulated insulin release assay (GSIS) evaluation of endocrine function of endocrine cultures. **F:** Microfluidic analysis of internal calcium release in response to glucose and potassium challenges.

Transparent Methods

Experimental Model and Subject Details

Cell lines used throughout this study included the male H1 and female H9 cell lines, both of which were obtained from WiCell and the male IPS-282CW cell line generated and obtained from Paul Tesar at Case Western Reserve. Pluripotent cultures were grown at 37^{o}C, cultured in Essential 8 media and grown on vitronectin coated plates. Differentiation experiments were seeded at 75,000 cells per cm^{2} and maintained in E8 medium for 48 hours until cultures were approximately 90% confluent before initiating differentiation experiments. All growth factors and small molecules used throughout this study were reconstituted according to supplier's recommendations and aliquots were stored at -80^{o}C for up to a year. Growth medium was changed daily. The basal medium used in all differentiation experiments was CDM2 (Loh et al., 2014) with the exception of the DE protocol which used RPMI.

Experiments performed using the H9 Cell line were performed as detailed above with the exception that the pluripotent culture was maintained in mTeSR growth medium. Experiments performed using the IPS-282CW cell line were performed as described above with the exception that the pluripotent cell line was maintained on feeders and the initial seeding for modeling experiments was maintained in feeder-conditioned medium for 48 hours before initiation of validation experiments. Growth media was formulated using DMEM/F12 as the basal media supplemented with 20% Knock Serum Replacer, 1mM L-glutamine, 0.5% 2-mercaptoethanol, 1% non-essential amino acids and 4ng/ml bFGF. Conditioning of growth media was accomplished by incubating ES media (not supplemented with bFGF) on top of mouse embryonic fibroblast cultures for 24 hours at which time media was collected and FGF2 was added at a concentration of 4ng/ml. Mouse embryonic fibroblast were generated at the LRI, assayed for mycoplasma contamination and irradiated before use for either co-culture with IPS cells or conditioning of growth media.

Method Details

Generating DoE designs and Perturbation Matrixes

All DoE experimental designs were computer generated using D-optimal interaction designs in MODDE software (Sartorius Stedim Data Analytical Solutions, SSDAS). All factors tested and genes measured were manually inputted into Design Wizard within MODDE software and screening was selected within the Objective window. Factors known to initiate differentiation in pluripotent cells were chosen for our initial design, whereas factors known to influence the differentiation of endodermal progenitors were chosen for our second DoE design. In most cases agonist molecules were paired with a corresponding antagonist within the same design, however constraints were manually inserted disallowing the occurrence of reaction conditions in which both agonists and antagonists of the same pathway were present. Reagent concentrations were either matched to previous publications using the

same reagent or ran 3-5 times greater than known IC50 values. Designs runs were chosen to have up to 93 reactions conditions with the addition of 3 center point conditions. Once DoE designs were generated the one with the highest G-efficiency was chosen. The selected DoE design was then used as a template for the generation of perturbation media matrix. Perturbation matrixes were generated (96 independent experimental runs) on a Freedom Evo150 liquid handling robot (TECAN, CH). MODDE-based DoE designs were exported into Excel software which were then used as an instructional template for the Evo150 liquid handling system. The freedom Evo150 was enclosed within a X-Vivo unit and the system was cleaned using a 10% bleach solution before and after each perturbation matrix generation to ensure sterility. All factor dilutions and basal growth medium for perturbation matrixes used CDM2 (Loh et al., 2014) . The DoE designs used throughout this study are shown in Supp. Table 1 & 3.

Growth Conditions and cDNA Generation

Pluripotent cells were seeded into 96 well plates at a concentration of 20,000 cells per well and incubated for 48 hours after which growth media was removed and replaced with perturbation matrixes for the initial DoE design. Whereas the second DoE design were initially exposed to the HNF1β^{Opt} for 3 days after initial seeding before being exposed to perturbation matrix. Cells were incubated within perturbation matrix for 3-days with daily media exchanges. All cell culture, including manual preparatory expansion and seeding as well as any robotic growth media exchanges were contained in a modular X-Vivo system (Biospherix, NY, USA) providing Process Analytical Technology (PAT) of the cell culture conditions. Cultures were maintained in 90% N_2, 5% O_2 and 5% CO_2, at 37^{o}C with continuous monitoring of these experimental parameters. RNA extractions for DoE based modeling experiments were performed using a MagMax-96 Total RNA Isolation Kit and performed according to manufacturer's protocol. Quantification of RNA was performed on an epoch. Reverse transcription of RNA samples was performed using the reaction conditions provided with the High Capacity cDNA RT Kit.

QuantStudio Data Analysis

Data collection was performed using an Open Array (QuantStudio, Life Technologies) with custom design gene cards (53 genes/sample). The initial designs included genes responding to all major sub-lineage fates downstream of pluripotency, while the later design focused more on endodermal lineages with an emphasis on pancreatic fates. Samples were loaded onto a custom design Quant Studio Card using an OpenArray AccuFill System and ran on a QuantStudio 12k Flex Real-Time PCR System. QuantStudio runs were performed according to manufactories protocol. The resulting QuantStudio gene expression data was analyzed using Expression SuiteTM software (Life Tech). Three criteria were used for exclusions of data points from the raw data set. First, exported images of the QS cards were visualized using Image J software and probesets that were not amplified had their corresponding Crt values removed from further analysis. Second, Expression Suite software was used to identify

samples that had an undetermined Crt value, these samples were excluded from further analysis. Third, all samples with an Amp Score less than 0.8 or greater than 1.24 were manually inspected using Expression Suite software and if they had abnormal amplification curves their corresponding Crt values were excluded from further analysis. The data set was then exported to Excel and normalized against three internal standard housekeeping genes present on the QS card. Internal standards were changed between designs and standards used for the individual experiments are stated within figure legends. Final expression levels were expressed as $1/(2^{\Delta Crt})$ x 1000 and imported to MODDE. Each modeling experiment generated ~10,000 individual gene expression response data points on which MODDE optimization was based on.

Generating Computer Gene Models

Modeling of differentiation space was performed using MODDE software. Once data is imported into MODDE primary 'Summary of Fit Plots' are automatically generated. The Summary of Fit Plots contain R2 and Q2 measurements for the models of each gene. R2 is a measurement of the data fit to the gene model, while Q2 is an estimate of the precision of the prediction of the gene model. Both range from 0 to 1 with 0 being a non-significant gene model and 1.0 being a perfect gene model in that all of the data fits perfectly within the model. To improve predictive models all primary gene models were maximized for Q2 within the Analysis Wizard feature within MODDE software. Q2 maximization was accomplished by removing non-significant interaction terms from the individual gene models Coefficient Plots. Coefficient Plots are a primary read-out from MODDE that are graphical representations of the model terms significance. The Normalized mode of Coefficient Plots were always chosen so that direct comparison between the influences of the different effectors on the various genes was possible. Coefficients Plots error bars represent confidence intervals and significant terms are identified as terms that have confidence intervals that do not overlap the Y-axis. Initially generated models include interaction terms for all occurring combination of effectors. However since not all of the effectors tested had significant or synergistic interactions inclusion of these interactions add noise to optimization models. Double and triple interaction terms were deemed insignificant when their confidence intervals crossed the Y-axis and when deleted the gene models Q2 and R2 values increased. When both of these circumstances occurred coefficient terms were removed from the primary Coefficient Plot. No primary coefficients were ever removed from any gene model regardless of whether they were significant or not. Only gene optimizers which had $Q2 > 0.5$ and $R2 > 0.5$ were considered to be valid models and investigated further.

Computer Guided Gene Optimization

Once gene models underwent Q2 maximization the software predictive tool Optimizer was chosen to predict optimal conditions for maximization of any chosen gene. This consisted of choosing the criteria of Maximize for the desired gene followed by Run Optimizer selection. The default Maximized Optimizer is the setpoint with the

lowest predicted probability of failure and is the one used throughout this manuscript though additional considerations were evaluated for optimizers. For each gene model optimizers predict relative expression levels for all genes measured throughout the experiment as a whole. This enables the evaluation of gene optimizers in association within a network of genes that are optimized within any given gene optimizer. As an example, in the case of the HNF1β^{Opt} we were initially able to evaluate it as a potential endodermal population due to the fact that several other endodermal genes were maximized within the experimental space at the same time as HNF1β. Additionally optimization models were evaluated using Dynamic Profile analysis. The Dynamic Profile feature enables the prediction of how the individual effectors influence the expression levels of all of the genes measured. As an example this feature was crucial in understanding how the pancreatic progenitor was different from a stomach progenitor.

Optimizers and Differentiation Protocols Used

APS^{Opt} was composed of 50 ng/ml AA and 15 ng/ml Wnt3a. Daily growth medium exchanges were performed for three consecutive days.

HNF1β^{Opt} was supplemented with 250nM LDN3189, 500nM A8301, 50ng/ml Wnt3a, 250nM PD0325901 and 2uM retinoic acid. Growth medium was changed daily for three days.

DE was generated using conditions previously described (D'Amour et al., 2005). Briefly, on the first day of differentiation the culture was exposed to RPMI supplemented with 100ng/ml AA and 20ng/ml Wnt3a. The following two days the culture was exposed to RPMI supplemented with 100ng/ml AA and 2% FBS.

PDX1^{Opt} was composed of 500nM A8301, 250nM LDN3189, 100nM Sant1, 250nM PD0325901, 2μM retinoic acid and 1% B27 supplement and with the exception of where it is stated differently in the text was used for 3 sequential days on differentiating cells.

Endocrine push medium was supplemented with 100nM γ-secretase inhibitor XX and 500nM A8301. This usually was performed with adherent cultures. However for cultures that were subjected to dithizone staining or microfluidics analysis were grown in suspension. The ratio of one 6-well plate per T75 Flask was used. Briefly on the fifth day of the stage 3 endocrine push cultures were washed with PBS and incubated in the presence of Accutase for 5 minutes at 37^{o}C. Cells were then washed off of plates, collected through centrifugation and re-suspended within T75 flasks in growth medium containing 100nM γ-secretase inhibitor XX, 500nM A8301 and 10μM Y-27632. Subsequent growth media exchanges were performed using demi depletion and the ROCK inhibitor was only used during the initial passage event.

Stomach organoids were generated using a modified protocol (McCracken et al., 2014). Briefly cultures were incubated for three days within the HNF1β^{Opt}. This was followed by a three day incubation within a stage 2

medium supplemented with 500 ng/ml Wnt3a, 500ng/ml FGF4, 250 nM LDN3189 and 2μM retinoic acid (retinoic acid was only used for the first day of the stage 2 media). Cultures were passaged 1:6 with seeding occurring in a 1:4 Matrigel/CDM2 mixture and incubated for 30 minutes at 37^{o}C before a three day application of stage 3 medium composed of 2μM retinoic acid, 250 nM LDN3189 and 100 ng/ml EGF. Cultures were then supplemented with a stage 4 medium supplemented with 100 ng/ml EGF for 23 days before organoids were recovered and fixed for sectioning.

RNA and cDNA Preparation

RNA extractions for RNA Seq and qPCR validation were performed using a Trizol based method according to manufacturer's protocol. Reverse transcription of RNA samples was performed using the reaction conditions provided with the High Capacity cDNA RT Kit. For OpenArray validation (not modeling) cDNA was subjected to QuantStudio analysis on the same custom design used for modelling experiments. All validation experiments were performed in quadruplicates of biological replicates which were performed in parallel experiments. The resulting data was then normalized to housekeeping genes and graphed using GraphPad Prism 5.02 software. Samples subjected to RNA Sequencing were shipped overnight on dry ice to the Genomics Facility at The University of Chicago for sequencing.

Immunohistochemical Analysis of Cultures

Histological characterization occurred by treating cell cultures with 4% paraformaldehyde solution at ambient temperature for 15 minutes. Samples were then blocked and permeabilized using 0.1M Tris/HCl pH 7.5 supplemented with 0.5% blocking reagent and 0.1% TrintonX-100 for 1 hour. All antibody solutions were diluted in 0.1M Tris/HCl pH 7.5 and primary antibodies were incubated over-night while secondary antibodies were incubated for an hour. All incubations occurred at ambient temperature. Cultures were treated with a 1:2 dilution of Vectashield mounting medium. Cultures were generally stained within the TC culture vesicle they were differentiated in (usually 96 well plates) while aggregates or organoids were fixed and embedded in OCT followed with cryostat sectioning before staining. The staining procedure was the same whether cells were stained within TC plates or sectioned and on slides.

Quantification of HNF1β^{Opt} and PDX^{Opt}

A Keyence BZ-X710 fluorescence microscope was used to take a stitched image measuring 5.25mm^2 from the center of triplicate biological replicated wells within a 96-well TC plate. All images had the background set to black and images were collected in stitch mode. Expression patterns were then quantified using BZ-X Analyzer software. This was accomplished by opening up the first image of the respective groups within the software and then performing a Hybrid Cell Count using the settings for Cell Separation with uniform brightness. These conditions

were saved and used as a template for a sequential Macro Cell Count in Image Stitching Mode of 100 consecutive images. Resulting analysis were saved and total dapi counts were compared to total HNF1β and/or PDX1 counts of the same region to determine percentage of the culture expressing the protein of the respective target genes. Three biological replicates from parallel experiments performed within the same 96 well plates were counted in this manner resulting in more than 30,000 cells per well and greater than 100,000 cells total being counted per condition.

Endocrine Cell Characterization

GSIS assays were performed by incubating cells in Krebs-Ringer buffer supplemented with 2mM glucose for 30 minutes for determination of basal C-peptide levels. This was followed by changing reaction buffer to a Krebs Ringer buffer either supplemented with 20mM glucose or 30mM KCl for an additional 30 minute incubation. Buffer samples were collected and C-Peptide levels were quantified using ELISA according to manufacturer's protocol. Microfludic assays were performed as previously described (Adewola et al., 2010, Wang et al., 2012).

RNA Seq & Key Gene Analysis

RNA samples were sequenced (Illumina HiSeq4000) by the University of Chicago Genomics Core. FASTQ files were aligned against the reference hg19 using HISAT2 version 2.0.5 software. Gene counts for the aligned reads were produced by HTSeq version 0.8.0 using RefSeq annotation file. Differential expression analysis was performed using edgeR version 3.18.1 with a cutoff of 4 counts per million.

KeyGenes analysis was performed using a provided training set according to published protocol (Roost et al., 2015). Hierarchical clustering was conducted using RStudio software with R packages gplots and ggplot2 on 87 KeyGenes classifier genes with only the portion containing all the hESC derivatives shown. Differential expression was determined with the EdgeR package and heatmaps were generated using Microsoft Excel software.

The heatmap comparing HNF1β^{Opt} to definitive endoderm used newly generated RNA Sequencing data for both data sets and was processed as described above. The heatmap comparing both the HNF1β^{Opt} and the PDX1^{Opt} to the published viacyte protocol used the same HNF1β^{Opt} data set and a newly generated PDX1^{Opt} which was processed as described above. The viacyte protocol used in this analysis was a deposited data set (Xie et al).

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat-anti-C-Peptide	DSHB	Cat#GN-ID4: RRID:AB_2255626
Rat-anti-ECAD	Invitrogen	Cat#13-1900: RRID:AB_2533005
Goat-anti-FOXA2	Santa Cruz Biotechnologies	Cat#sc-6554: RRID:AB_2262810
Mouse-anti-GATA4	Santa Cruz Biotechnologies	Cat#sc-25310: RRID:AB_627667
Rabbit-anti-GATA6	Santa Cruz Biotechnologies	Cat#sc-9055: RRID:AB_2108768
Mouse-anti-GCG	Sigma	Cat#G2654: RRID:AB_259852
Rabbit-anti-HNF1β	Santa Cruz Biotechnologies	Cat#sc-22840: RRID:AB_2279595
Guinea Pig-anti-INS	Dako	Cat#A0564: RRID:AB_10013624
Rabbit-anti-MNX1	Invitrogen	Cat#PA5-23407: RRID:AB_2540929
Rabbit-anti-MUC1	Cell Signaling Technologies	Cat#14161: RRID:AB_2798408
Rabbit-anti-NKX2.2	Abcam	Cat#AB191077: RRID:AB_2811076
Mouse-anti-NKX6.1	R&D	Cat#MAB5857: RRID:AB_10642178
Rabbit-anti-NR5A2	Abcam	Cat#AB189876: RRID:AB_2732890
Rabbit-anti-OSR1	Cell Signaling Technologies	Cat#3729S: RRID:AB_2157610
Goat-anti-PDX1	R&D	Cat#AF2419: AF2419, RRID:AB_355257
Goat-anti-SOX1	Santa Cruz Biotechnologies	Cat#sc-17318: RRID:AB_2195365
Rabbit-anti-SOX2	Millipore	Cat#AB5603: RRID:AB_2286686
Rabbit-anti-SOX9	Millipore	Cat#AB5535: RRID:AB_2239761
Mouse-anti-SOX17	R&D	Cat#MAB1924: RRID:AB_2195646
Rabbit-anti-SST	Dako	Cat#A0566: RRID:AB_2688022
Alexa Fluor 488 Donkey-anti-Goat	Jackson Immuno Research	Cat#705-546-147: RRID:AB_2340430
Alexa Fluor 488 Donkey-anti-Mouse	Jackson Immuno Research	Cat#715-546-151: RRID:AB_2340850
Alexa Fluor 488 Donkey-anti-Rabbit	Jackson Immuno Research	Cat#711-546-152: RRID:AB_2340619
Alexa Fluor 488 Donkey-anti-Rat	Jackson Immuno Research	Cat#712-545-150: RRID:AB_2340683
Alexa Fluor 594 Donkey-anti-Goat	Jackson Immuno Research	Cat#705-005-003: RRID:AB_2340384
Alexa Fluor 594 Donkey-anti-Guinea Pig	Jackson Immuno Research	Cat#706-586-148: RRID:AB_2340475
Alexa Fluor 594 Donkey-anti-Mouse	Jackson Immuno Research	Cat#715-585-150: RRID:AB_2340854
Alexa Fluor 594 Donkey-anti-Rabbit	Jackson Immuno Research	Cat#711-586-152: RRID:AB_2340622
Alexa Fluor 594 Donkey-anti-Rat	Jackson Immuno Research	Cat#712-586-153: RRID:AB_2340691
Alexa Fluor 647 Donkey-anti-Rabbit	Jackson Immuno Research	Cat#711-605-152: RRID:AB_2492288
Cy5 Donkey-anti Rat	Jackson Immuno Research	Cat#712-175-153: RRID:AB_2340672
Chemicals, Peptides, and Recombinant Proteins		
Essential 8 media	Gibco	A15169-01
Vitronectin	Gibco	A14700
mTeSR1	Stemcell Technologies	85850
IMDM	Life Technologies	21056-023
F12	Life Technologies	12660-012
RPMI	Lerner Research Institute Media Core	12-500p
Polyvinyl Alcohol	Sigma Aldrich	P8136
Chemically Defined Lipid Concentrate	Gibco	11905-31
Monothioglycerol	Sigma Aldrich	M6145
Insulin	Roche	1376497
Transferrin	Roche	652202
DMEM/F-12	Invitrogen	11330-032
Knock Out Serum Replacer	Invitrogen	10828-028
L-Glutamine Solution	Invitrogen	25030-081

2-Mercaptoethanol	Sigma	M-7522
Non-Essential Amino Acids	Lerner Research Institute Media Core	NEAA-(100X)
LDN3189	Sellekchem	S2618
Activin A	Peprotech	120-14
FGF2	Gibco	13256029
Sant1	Sellekchem	S7092
A8301	Biogems	4463325
Wnt3a	R&D	5036WN/CF
PD0325	Selleckchem	S1036
SHH	Peprotech	100-45
Retinoic Acid	Sigma Aldrich	R2625
B27 supplement	Gibco	12587
γ-secretase inhibitor XX	EMD Millipore	565789
FGF4	R&D	7460-F4
Matrigel	Fisher Scientific	354230
EGF	R&D	236-EG
Trizol	Life Technologies	15596018
4% Paraformaldehyde Solution	EMD	30525-89-4
Tris/HCl	Promega	H5123
Blocking Reagent	Perkin Elmer	FP1012
TrintonX-100	Fisher Scientific	BP151
Vectashield Mounting Medium	Vector Laboratories	H-1200
OCT	Sakura	4583
Dithizone	Sigma-Aldrich	D5130
Accutase	Innovative Cell Technologies	AT-104
Y-27632	Calbiochem	688000
Critical Commercial Assays		
C-Peptide ELISA	Mercodia	10-1141-01
MagMax-96 Total RNA Isolation Kit	Life Technologies	AM1830
High Capacity cDNA RT Kit	Life Technologies	4368814
Deposited Data		
RNA Seq	NBCI SRA	SRA: TBD
Experimental Models: Cell Lines		
H1 (WA-01) hESC line	WiCell Research Institute	NIHhESC-10-0043: RRID:CVCL_9771
H9 (WA-09) hESC line	WiCell Research Institute	NIHhESC-10-0062: RRID:CVCL_9773
IPS-282CW Primary Fibroblast	Paul Tesar, Case Western	N/A
Oligonucleotides		
APOB (Hs00181142_m1)	Life Technologies	Cat# 4331182
ARX (Hs00292465_m1)	Life Technologies	Cat# 4331182
CDX2 (Hs01078080_m1)	Life Technologies	Cat# 4331182
CHGA (Hs00900375_m1)	Life Technologies	Cat# 4331182
CPB1 (Hs00157026_m1)	Life Technologies	Cat# 4331182
CXCR4 (Hs00607978_m1)	Life Technologies	Cat# 4331182
EOMES (Hs00172872_m1)	Life Technologies	Cat# 4331182

EPCAM (Hs00901885_m1)	Life Technologies	Cat# 4331182
EVX1 (Hs00231104_m1)	Life Technologies	Cat# 4331182
F3 (Hs01076029_m1)	Life Technologies	Cat# 4331182
Software and Algorithms		
MODDE	Sartorius Stedim Data Analytical Solutions	https://umetrics.com/product/modde-pro
Expression SuiteTM	Life Technologies	https://www.thermofisher.com/us/en/home/technical-resources/software-downloads/expressionsuite-software.html
Image J	NIH	https://imagej.nih.gov/ij/download.html
BZ-X Analyzer software	Keyence	https://www.keyence.com/products/microscope/fluorescence-microscope/bz-x700/index_pr.jsp
GraphPad Prism 5.02	GraphPad	https://www.graphpad.com/scientific-software/prism/
HISAT2 version 2.0.5	John Hopkins University	https://ccb.jhu.edu/**software/hisat2**
HTSeq version 0.8.0	doi:10.1093/bioinformatics/btu638	https://htseq.readthedocs.io/en/release_0.11.1/search.html?q=download&check_keywords=yes&area=default
RStudio	GNU Project	https://rstudio-data-recovery.en.lo4d.com/windows
Other		
OpenArray AccuFill System	Life Technologies	4471021
QuantStudio 12k Flex Real-Time PCR System	Life Technologies	4471090
Xvivo System	BioSpherix	N/A
Freedom Evo150	Tecan	N/A
Epoch	BioTek Instruments Inc	N/A
T100 Thermal Cycler	BioRad	1861096
Keyence BZ-X710	Keyence	N/A

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
FOXA2 (Hs00232764_m1)	Life Technologies	Cat# 4331182
GCG (Hs01031536_m1)	Life Technologies	Cat# 4331182
GLIS3 (Hs00541450_m1)	Life Technologies	Cat# 4331182
GLP1R (Hs00157705_m1)	Life Technologies	Cat# 4331182
GLUT2 (Hs01096908_m1)	Life Technologies	Cat# 4331182
GSC (Hs00906630_g1)	Life Technologies	Cat# 4331182
HEY1 (Hs050447713_s1)	Life Technologies	Cat# 4331182
HHEX (Hs00242160_m1)	Life Technologies	Cat# 4331182
HNF1β (Hs01001602_m1)	Life Technologies	Cat# 4331182
INS (Hs00355773_m1)	Life Technologies	Cat# 4331182
LEFTY1 (Hs00764128_s1)	Life Technologies	Cat# 4331182
LZTS2 (Hs00383274_m1)	Life Technologies	Cat# 4331182
MAFA (Hs01651425_s1)	Life Technologies	Cat# 4331182
MESP1 (Hs05050983_s1)	Life Technologies	Cat# 4331182
MIST1 (Hs00703572_s1)	Life Technologies	Cat# 4331182
MIXL1 (Hs05060541_s1)	Life Technologies	Cat# 4331182
MNX1 (Hs00907365_m1)	Life Technologies	Cat# 4331182
MYT1 (Hs01027966_m1)	Life Technologies	Cat# 4331182
NEUROD (Hs01922995_s1)	Life Technologies	Cat# 4331182
NGN3 (Hs01875204_s1)	Life Technologies	Cat# 4331182
NKX2.2 (Hs00159616_m1)	Life Technologies	Cat# 4331182
NKX6.1 (Hs00232355_m1)	Life Technologies	Cat# 4331182
NR5A2 (Hs00187067_m1)	Life Technologies	Cat# 4331182
ONECUT1 (Hs00413554_m1)	Life Technologies	Cat# 4331182
ONECUT2 (Hs04986540_s1)	Life Technologies	Cat# 4331182
OSR1 (Hs01586544_m1)	Life Technologies	Cat# 4331182
PAX4 (Hs00173014_m1)	Life Technologies	Cat# 4331182
PDX1 (Hs00236830_m1)	Life Technologies	Cat# 4331182
PROM1 (Hs01009250_m1)	Life Technologies	Cat# 4331182
PROX1 (Hs00896294_m1)	Life Technologies	Cat# 4331182
SFRP5 (Hs00169366_m1)	Life Technologies	Cat# 4331182
SOX17 (Hs00751752_s1)	Life Technologies	Cat# 4331182
SOX2 (Hs01053049_s1)	Life Technologies	Cat# 4331182
SOX9 (Hs01001343_g1)	Life Technologies	Cat# 4331182
SST (Hs00356144_m1)	Life Technologies	Cat# 4331182
T (Hs00610080_m1)	Life Technologies	Cat# 4331182
YWHAZ (Hs03044281_g1)	Life Technologies	Cat# 4331182
GAPDH (Hs02758991_g1)	Life Technologies	Cat# 4331182
EEF1A1 (Hs00265885_g1)	Life Technologies	Cat# 4331182
TBP (Hs00427620_m1)	Life Technologies	Cat# 4331182