Global expression analysis of gene regulatory pathways during endocrine pancreatic development

Guoqiang Gu[1,*,‡], James M. Wells[1,†,‡], David Dombkowski[2], Fred Preffer[2], Bruce Aronow[†] and Douglas A. Melton[1,§]

[1]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA
[2]Department of Pathology, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA
*Present address: Program of Developmental Biology, Department of Cell and Developmental Biology, Vanderbilt Medical Center, Nashville, TN 37232, USA
†Present address: Division of Developmental Biology, Cincinnati Children's Research Foundation, Cincinnati, OH 45229-3039, USA
‡These two authors contributed equally to this work and are listed alphabetically
§Author for correspondence (e-mail: dmelton@biohp.harvard.edu)

Accepted 30 September 2003

Development 131, 165-179
Published by The Company of Biologists 2004
doi:10.1242/dev.00921

Summary

To define genetic pathways that regulate development of the endocrine pancreas, we generated transcriptional profiles of enriched cells isolated from four biologically significant stages of endocrine pancreas development: endoderm before pancreas specification, early pancreatic progenitor cells, endocrine progenitor cells and adult islets of Langerhans. These analyses implicate new signaling pathways in endocrine pancreas development, and identified sets of known and novel genes that are temporally regulated, as well as genes that spatially define developing endocrine cells from their neighbors. The differential expression of several genes from each time point was verified by RT-PCR and in situ hybridization. Moreover, we present preliminary functional evidence suggesting that one transcription factor encoding gene (*Myt1*), which was identified in our screen, is expressed in endocrine progenitors and may regulate α, β and δ cell development. In addition to identifying new genes that regulate endocrine cell fate, this global gene expression analysis has uncovered informative biological trends that occur during endocrine differentiation.

Supplementary data available online

Key words: Myt1, endoderm, Pancreas, Endocrine, Islets, Microarray

Introduction

The vertebrate pancreas has two functions: producing digestive enzymes (exocrine), and regulating glucose homeostasis (endocrine). These separate functions are reflected in the complex architecture of the pancreas. The acini and ducts form the exocrine pancreas that produces and transports digestive enzymes into the duodenum. The endocrine islets contain four types of cells that secrete hormones to regulate glucose metabolism and other physiological processes (Slack, 1995). Thus, the developing pancreas presents a challenge for developmental biologists because of the complex morphogenetic processes underlying development of this organ. In addition, Type I or insulin dependent diabetes mellitus results from the autoimmune-mediated destruction of insulin-secreting β cells in islets, emphasizing the importance of understanding pancreas and β cell development (Mathis et al., 2001; Tisch and McDevitt, 1996).

The pancreas derives from the endoderm germ layer (Pictet et al., 1972; Slack, 1995), which in mouse is a cup of cells enveloping the mesoderm and ectoderm at embryonic day 7.5 (E7.5). At this time, the endoderm receives signals from adjacent mesoderm and ectoderm and becomes competent to respond to subsequent permissive signals that establish organ domains along the anterior-posterior axis (Wells and Melton, 1999). By E8.5, the endoderm begins to form a primitive gut tube, and the region destined to become the pancreas receives signals from the notochord and dorsal aorta, leading to the expression of essential pancreatic transcription factor genes such as pancreatic-duodenal homeobox 1 [*Pdx1*, also known as insulin-promoter factor 1 (Ipf1)] (Hebrok et al., 1998; Lammert et al., 2001). At E9.0, *Pdx1* expression marks both the dorsal and ventral domains of the developing pancreas, and defines where pancreatic buds will appear around E10 (Guz et al., 1995). As pancreatic buds expand and branch, signals from adjacent mesenchyme direct cells toward an endocrine or exocrine fate (Guz et al., 1995; Miralles et al., 1998a; Miralles et al., 1998b). Cells that have adopted an endocrine cell fate express the bHLH transcription factor neurogenin 3 (NGN3) (Gu et al., 2002).

Functional studies have identified several signaling pathways and transcription factors important for pancreatic development. Initial pancreatic specification of endoderm is mediated by the FGF, hedgehog, Notch and TGFβ/activin signaling pathways (Kim and Hebrok, 2001). These signals result in the expression of genes for several transcription factors in the developing pancreas including, *HNF1α* (*Tcf1α*), *HNF1β* (*Tcf1*β), *HNF4α* (*Tcf4α*), *Pdx1, NeuroD1, Ngn3, Pax4, Pax6* and others (Edlund, 1998). Mutations in some of these genes are associated with maturity onset

diabetes of the young (MODY 1, 3, 4, 5 and 6), and genetic analyses in mice have begun to elucidate how these transcription factors function during discrete stages of pancreas development (Stride and Hattersley, 2002). For example, loss of PDX1 results in defects of both early pancreatic specification and budding (Jonsson et al., 1994; Offield et al., 1996), whereas loss of NGN3 results in specific absence of endocrine cell development (Gradwohl et al., 2000). Moreover, cell lineage analysis supports the idea that PDX1 functions to establish the three basic lineages of the pancreas (ducts, acini, islets), whereas NGN3 functions specifically to establish the endocrine lineages (Gannon et al., 2000; Gu et al., 2002; Herrera, 2000; Herrera et al., 1998; Schwitzgebel et al., 2000).

Analyses of individual genes have begun to define some critical stages in the development of the endocrine pancreas, yet the complex interactions of extracellular signals and the responding genetic networks that control endocrine cell growth and differentiation are largely unstudied. For example, it is not known how *Pdx1* is induced and restricted to a defined region of the developing gut, nor is it known how *Ngn3* expression is temporally controlled resulting in the genesis of endocrine progenitor cells. Recently, 3,400 genes expressed in the pancreas were used to generate an endocrine pancreas microarray (PancChip), which is available through the β Cell Biology Consortium (Scearce et al., 2002). The PancChip will probably be a valuable diagnostic tool for the genetic analysis of pancreatic cell samples. However, the focus of the Endocrine Pancreas Consortium was not to provide a complete and quantitative analysis of the genes that are expressed during the formation of the endocrine pancreas. A transcriptional profile of pancreatic and endocrine progenitors would provide fundamental information about the processes regulating normal development of the endocrine pancreas. Moreover, regulatory factors identified in this screen might be used to promote regeneration of endocrine cells in vivo, or used to direct the differentiation of embryonic stem cells or adult stem/progenitor cells toward the β cell lineage in vitro.

We describe the fundamental gene expression profiles of several tissue or cell samples that define distinct stages during pancreatic and endocrine islet development. We used high-density microarrays from Affymetrix to systematically analyze the genes that are expressed at four key stages of pancreatic and endocrine development: E7.5 unspecified endoderm, E10.5 pancreatic cells that express *Pdx1*, E13.5 endocrine progenitor cells that express *Ngn3*, and mature islets of Langerhans. This genetic analysis is uniquely designed in several ways. First, we used a combination of dissection and cell-sorting using an eGFP reporter that was under the control of promoters of specific pancreatic genes to isolate highly purified cells from these well-defined stages of pancreatic development. Second, we compared both the temporal and spatial expression profile at each stage to more fully define these cell types. Third, we validated our profiles by demonstrating the cell-specific expression of several genes from each time point by RT-PCR and in situ hybridization (ISH). Finally, we demonstrated that one gene we identified, myelin transcription factor 1 (*Myt1*), might be a novel regulator of α, β and δ cell development in the pancreas.

Materials and methods

Transgene construction and transgenic mice generation

To generate the *Pdx1-eGFP* construct, a 1.5 kb DNA fragment containing the coding region of enhanced green fluorescence protein (*eGFP*) and a SV40 polyadenylation signal was amplified by PCR, digested by *Xba*I (via a site introduced in the 3′ end primer) and ligated to *Nco*I (blunt ended)-*Xba*I-digested plasmid *Pdx1-hsp68*-lacZ construct (a kind gift from C. E. V. Wright). The plasmid used as PCR template was pGreenlatern-1 (Clontech, Palo Alto, CA). The primers used are: forward: 5′-agcaagggcgaggaactgttc-3′ and reverse: 5′-catgatctagacatgataagatacattgatg-3′. The insert in the final construct (p#48) was released by *Sal*I digestion and used for transgenic animal production. The hybrid B6CBAF1 mouse strain was used to generate transgenic animals. Five transgenic lines were generated and the eGFP expression pattern was compared with that of PDX1 protein to ensure that eGFP expression mimics that of PDX1 protein. One line, P#48.9, whose expression recapitulates that of *Pdx1*, was used to obtain Pdx1-eGFP$^+$ cells.

An *Xba*I-*Sph*I (partial digestion) fragment that contains sufficient *Ngn3* enhancers (Gu et al., 2002) replaced the *Pdx1* enhancer region in *p#48* to generate the *Ngn3-eGFP* construct (p#63, Fig. 1). Insert was released by *Sal*I digestion to generate three transgenic lines. After verifying that eGFP expression mimics that of NGN3 by double ISH (Gu et al., 2002), one line P#63.1 was used to obtain embryos for cell sorting.

To generate function analyses constructs of *Myt1*, full length *Myt1a* or *Nzf2b* cDNA (a kind gift from L. D. Hudson) was inserted into the *Xho*I-*Sma*I site of the pCIG expression vector [under the control of CMV-beta actin promoter (Grapin-Botton et al., 2001)] to give plasmid p#116 and p#132. The dominant negative construct (*dnMyt1*) is generated using a similar approach, except the transcriptional activation sequence was deleted using a PCR approach. Specifically, two primers (p273: 5′-gacaattgaaggagcttctcacctgtcc-3′ and p252: 5′-ccatgtgtgcacctcagcatc-3′) were used to amplify a DNA fragment that had both the 5′ and 3′ ends of *Myt1a* cDNA and a vector sequence in the middle. This fragment was digested by *Eco*RI and *Mun*I and self-ligated. The resulting plasmid was a partial *Myt1* cDNA that had its transcription activation domain removed (*dnMyt1*). For transgenic animal production, *dnMyt1* was put under the control of the *Ngn3* promoter.

Tissue isolation and cell sorting

To obtain purified endoderm, mesoderm and ectoderm tissue, E7.5 embryos (90) were isolated from timed pregnant female ICR mice (Taconic, Germantown, New York) and the endoderm was manually dissected from the mesoderm and ectoderm with a polished tungsten needle (Wells and Melton, 2000). Isolated germ layers were combined into two pools. Each pool of isolated endoderm and mesoderm and ectoderm contained approximately 0.2-0.4 μg total RNA, which was used for cRNA probe generation (Baugh et al., 2001).

To isolate Pdx1-eGFP$^+$ cells, ICR or CD-1 mice were crossed with P#48 males, and the eGFP-expressing E10.5 embryos were identified under a fluorescence microscope. The pancreatic rudiments and the stomach and duodenum (Std) anlagen were separated by dissection. These tissues were trypsinized to single cells and sorted into eGFP$^+$ and eGFP$^-$ populations by FACS. From 350 eGFP$^+$ embryos, 1.3 and 1.8 million eGFP$^+$ cells were collected from the pancreatic or Std region, respectively. Meanwhile, five million eGFP$^-$ cells were also collected from both dissected samples. From these cells, 6, 8 and 14 μg total RNA was isolated and used for cRNA probe generation (each of these RNA were maintained in several small pools respectively). Ngn3-eGFP-expressing cells were isolated by a similar approach except that only the pancreatic rudiment was isolated, and the stage used was E13.5 (from 1300 eGFP$^+$ pancreata, 1.3 million Ngn3-eGFP$^+$ cells were collected and 5 μg total RNA was made and maintained as two pools).

Mouse islets were isolated by perfusing the pancreas with a collagenase solution (2 mg/ml), filtering the digested pancreas though a 300 μM wire mesh, and centrifugation on a histopaque 1077 cushion (Warnock et al., 1990). Islets were hand-picked to minimize contamination with exocrine tissue. For our analysis, pancreata from five adult animals were used to obtain 30 μg total RNA.

cRNA probe generation and hybridization to Affymetrix microarray chips

Total RNA samples were used to generate cRNA probes by two rounds of transcription (Baugh et al., 2001). Basically, a poly(dT) primer (with its 5′ end carrying T7 promoter sequence) was used to synthesize cDNA from total RNA. The cDNA were used to amplify cRNA using T7 polymerase. The cRNA product from this first round amplification was used to generate more cDNA by random priming, with the 3′ end carrying a T7 promoter sequence. This cDNA was used to transcribe biotinylated cRNA, which was used to hybridize to the Mu11K, Mu74Av1 or MU74Av2 microarrays produced by Affymetrix, following the manufacturer's protocol.

Data normalization and analysis

Two programs were used to analyze the data generated from the microarray hybridization.

First, using MicroArray Suite 5.0 (Affymetrix) image files were examined for uniform image quality without significant scratches or smudged fluorescence patterns. The images were processed into intensity data that was scaled per chip to a target intensity of 1500. Chip reports were examined for evidence of high quality and uniform RNA, RNA labeling, hybridization and scanning using approaches similar to those described at (http://cardiogenomics.med.harvard.edu/groups/proj1/pages/Method_QC.html). In brief, control oligonucleotide signal corresponding to spiked and constitutive RNAs were strong, uniform, sensitive and properly interpreted by the Affymetrix software. Background values were uniformly less than 100 and the scaling factor SF that is used to normalize the signal across the entire chip to 1500 signal units was within a twofold range for all chips. GeneSpring 5.0.1 (Silicon Genetics, Inc., Redwood City, CA) was used to analyze the resulting data values obtained from MicroArray Suite 5.0. The values used for filtering and clustering were 'Signal', 'Signal Confidence', 'Absolute Call' (Absent/Present). Data were normalized as follows: the 50th percentile of all measurements was used as a positive control for each array. Each measurement for each gene was divided by this synthetic positive control, assuming that this was at least 10. The bottom tenth percentile signal level was used as a test for correct background subtraction. The measurement for each gene in each sample was divided by the corresponding value in untreated samples, assuming that the value was at least 0.01. Throughout our analysis, only the genes that display more than threefold change between samples were listed and studied (P=0.01 in at least one statistical test).

Chick embryo electroporation

Chick embryo electroporations followed the reported protocol (Grapin-Botton et al., 2001). Briefly, electroporation was performed on embryos between the 18- and 25-somite stage (i.e., stage 13-15 HH). Eggs were windowed and DNA (2 μg/μl DNA in 1×PBS, 1 mM $MgCl_2$, 3 mg/ml carboxymethylcellulose, 50 μg/ml Nile Blue Sulfate) was injected in the blastocoel. A negative electrode was inserted below the embryo, and a positive electrode was held by a micromanipulator above the embryo and three square pulses of 17 volts for 50 mseconds each were applied (BTX T-820). After electroporation, eggs were incubated at 38°C for 48-60 hours, then collected and fixed in 4% paraformaldehyde/PBS, and sectioned for immunohistochemistry or in situ RNA analysis.

Immunohistochemistry

Electroporated embryos were sectioned and analyzed for hormone expression. Transgenic mouse embryos with the *Ngn3* promoter driving *dnMyt1* expression were analyzed by insulin and glucagon expression. The pancreata from five independently derived F_1 transgenic E14.5 embryos were fixed, completely sectioned (6 μm sections), immunostained with anti-insulin or glucagon antibodies, and the insulin$^+$ and/or glucagon$^+$ cells were counted on alternate paraffin sections. As a control, four littermate pancreata were counted in a similar fashion. Primary antibodies used were guinea pig anti-insulin (Dako, Carpinteria, CA), guinea pig anti-glucagon (Linco, St. Charles, MI) and rabbit anti-glucagon (Chemicon, Temecula, CA). Secondary antibodies used were peroxidase-conjugated donkey anti-guinea pig, FITC-conjugated donkey anti-guinea pig, and Cy3-conjugated donkey anti-rabbit (Jackson Immunoresearch, West Grove, PA). In order to obtain a significant number of insulin$^+$ or glucagon$^+$ cells, at least half of sections from each pancreas was counted.

RT-PCR and ISH

RT-PCR followed standard protocols. The primers used in our analyses were: ApoAIV forward: 5′-aaggtgaagggcaacacggaag-3′, reverse: 5′-cctcaagctggtacaagaagtgc-3′.

HPRT forward: 5′-gctggtgaaaggacctctc-3′, reverse: 5′-cacaggactagaacacctgc-3′ (Johansson and Wiles, 1995). Dkk1 forward: 5′-ggagatattccagcgctgtta-3′, reverse: 5′-ggtaagtgccacactgaggat-3′.

Prss12 forward: 5′-agagagaggccacagaaaacag-3′, reverse: 5′-ttgactccacatccatacccccc-3′.

Eya2 forward: 5′-ttactcccattacccacgggtc3′, reverse: 5′-gaagcctaaacaacgggcaaag-3′. Osteopontin forward: 5-gaagctttacagcctgcacccaga-3′; reverse: 5′-gcttttggttacaacggtgtttgc-3′; T7/osteopontin reverse: 5′-gtaatacgactcactatagggc aacagactaagctaagagccc-3′. Nkx2.2 forward: 5′-ccatgtcgctgaccaacacaaaga-3′; reverse; 5′-cgctcaccaagtccactgctgg-3′; T7/Nkx2.2 reverse: 5′-gtaatacgactcactatagggcggtgtgctgtcgggtactg-3′. Tm4sf3 (AF010499) forward: 5′-cagttccgctgtagcaatggctg-3′; reverse: 5′-cacacacactctaccactgagc-3′. T7/Tm4sf3 reverse: 5′-gtaatacgactcactatagggcagcacaaactacaaagaccca-3′. Spintz1 (AA57115): forward; 5′-gctgcaggcacacggatctctgc-3′; reverse: 5′-cagtgaatacctgtgaagatatc-3′. T7/Spintz1 reverse: 5′-gtaatacgactcactatagggcctcagtgagatacttcaataac-3′. Myt1 forward: 5′-gtctccggtggaagctcatggaca-3′; reverse: 5′-cttatggtgccctagtgtgtcatc-3′; T7/Myt1 reverse: 5′-gtaatacgactcactatagggccattaacataagagggtaa-3′. Rbp forward: 5′-ggctacatcataggtccctttcg-3′; reverse: 5′-tactgcctctctaggcacagctc-3′; T7/Rbp reverse: 5′-gtaatacgactcactatagggctgtctctgggctcaggc-3′. Galphao forward: 5′-gcatgcacgagtctctcatgctc-3′; reverse: 5′-ctagacagactagcctgacatg-3′; T7/Galphao reverse: 5′-gtaatacgactcactatagggcgaggcgccaggcccag-3′. Foxa3 forward: 5′-ataaccatggctattcagcaggct-3′; reverse: 5′-cacaggtcaatcaagattgccaac-3′; T7/Foxa3 reverse: 5′-gtaatacgactcactatagggccatccaacatcacgaccatc-3′; actin control forward: 5′-atgccaacac agtgctgtctggtgg-3′; reverse: 5′-gcgaccatcctcctcttaggagtg-3′

Sectioned in situ analysis was performed as described previously (Grapin-Botton et al., 2001). Paraffin sections (6 μm) were collected on glass slides (Superfrost Plus), dewaxed, treated with 1 μg/ml proteinase K for 7 minutes, and postfixed in 4% paraformaldehyde. Hybridization mix contained 1 μm/ml of probe, and hybridization was done overnight at 70°C. Sections were washed in maleic acid buffer and blocked with 20% lamb serum/2% Blocking Reagent (Boehringer Mannheim, Indianapolis, IN) and incubated overnight with anti-digoxigenin-alkaline phosphatase antibody (Boehringer Mannheim), 1:1000. Slides were washed again and developed with NBT and BCIP.

Whole-mount ISH was performed as described previously (Wilkinson and Nieto, 1993). Briefly, E7.5 embryos were fixed, dehydrated in methanol, rehydrated, treated with 6% hydrogen peroxide, proteinase K treated for 1.5 minutes, and postfixed in 4% paraformaldehyde. Embryos were hybridized in buffer containing 1 μg/ml probe overnight at 70°C. Embryos were washed and incubated overnight with an anti-digoxigenin antibody (1:1000). Embryos were developed with BM purple (Boehringer Mannheim). Probe templates for ApoAV, Dkk1, Prss12 and Eya2 were generated by PCR

amplification from an E7.5 endoderm library (Harrison et al., 1995) using a gene-specific forward primer (mentioned above), and a vector specific (pSPORT) reverse primer. The resulting amplified product contained the 3′ end of the gene and an SP6 polymerase site from the pSPORT vector. The amplified products were verified by sequencing and used in an in vitro transcription reaction to generate antisense probes. To generate cRNA probes for *Foxa3*, galphao, osteopontin, *Myt1*, *Nkx2.2*, *Rbp*, *Spintz1*, and *Tm4sf3*, T7-reverse primers (has T7 promoter sequence at the 5′ end, see above) were used to amplify cDNA fragments with corresponding forward primers.

Results

Isolation of cells and generation of cRNA

Our approach focused on two questions. (1) Which transcripts are up- or down-regulated as undifferentiated endoderm adopts a pancreatic and then endocrine cell fate (temporal gene expression)? (2) Which transcripts distinguish developing endocrine cells from adjacent cells at each stage of development (spatial gene expression)? We isolated tissue samples from four stages of the developing endocrine pancreas (Fig. 1), and separated the developing endoderm, pancreatic, or endocrine cells from their neighboring cells using manual dissection and/or cell sorting. These highly enriched cell samples were used to make targets for hybridization to Affymetrix microarrays that contain over 12,000 genes and ESTs (Fig. 1 and Materials and methods).

The stages shown in Fig. 1 were chosen for the following reasons. (1) E7.5 endoderm. At E7.5, the endoderm is a sheet of cells on the outside of the embryo. At this stage, endoderm cells are plastic and are not yet determined to form the pancreas (Wells and Melton, 2000). Analysis of undifferentiated endoderm should provide a genetic baseline and highlight genes involved in endoderm plasticity and pancreas differentiation. (2) *Pdx1*-expressing cells of the E10.5 pancreatic rudiment. PDX1$^+$ cells will yield both the exocrine and endocrine components of the adult pancreas and are therefore considered pancreatic progenitor cells (Gannon et al., 2000; Gu et al., 2002). At this stage, PDX1$^+$ cells are also found in the stomach and duodenum (Offield et al., 1996). A transcriptional analysis of PDX1$^+$ cells from the pancreas versus PDX1$^+$ cells from the stomach and duodenum, or from PDX1$^-$ cells, should highlight genes that specify pre-pancreatic cells from their gastrointestinal neighbors. (3) Endocrine progenitor cells (NGN3$^+$) of the E13.5 pancreas. The cells that express *Ngn3* at this stage will form only endocrine tissue (Gu et al., 2002). A comparison of the transcriptional profile of NGN3$^+$ cells with NGN3$^-$ cells was aimed at distinguishing the endocrine and exocrine compartments of the embryonic pancreas. (4) Adult islets. Adult islets represent mature, differentiated endocrine cells and will highlight the genes that need to be up-regulated, as well as down-regulated, in order to form the endocrine pancreas. This experimental approach was designed to quantitatively identify genes that are temporally and spatially regulated during endocrine development.

The following methods were used to obtain tissue samples for transcriptional analysis. (1) The endoderm from 90 E7.5 embryos was manually separated from mesoderm/ectoderm. (2) The mouse *Pdx1* promoter, which recapitulates the endogenous *Pdx1* expression (Gu et al., 2002; Wu et al., 1997), was used to drive expression of eGFP, and eGFP expression

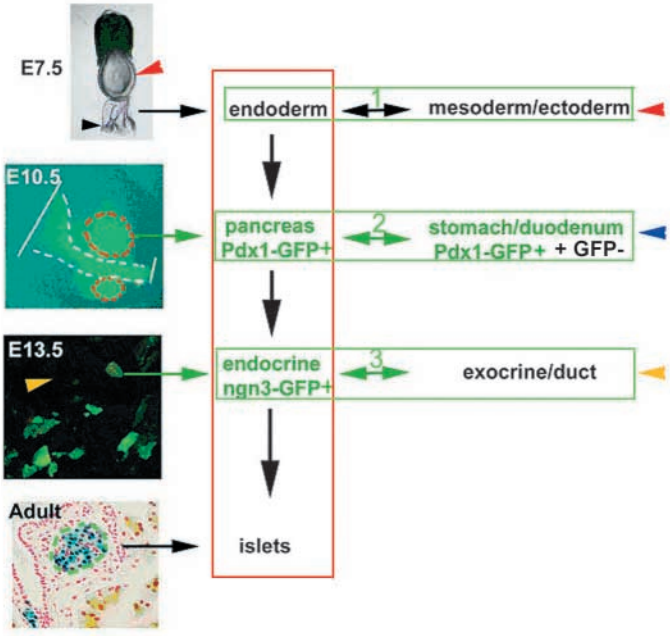

Fig. 1. Scheme of the temporal and spatial gene expression analysis for four stages of islet development. The red box highlights the temporal gene expression comparison and the green boxes highlight the spatial expression comparisons. (1) At E7.5, endoderm (black arrowhead) and mesectoderm (red arrowhead) were manually separated and analyzed. (2) At E10.5, the eGFP$^+$ cells from pancreatic buds (red dashed lines), stomach/duodenum (white dashed lines; two solid white lines indicate the approximate borders of dissection), and eGFP$^-$ cells from *Pdx1-eGFP* transgenic animals were separated by FACS and analyzed. (3) At E13.5, endocrine progenitor cells expressing Ngn3-eGFP were separated by FACS from eGFP$^-$ cells (yellow arrowhead, primarily exocrine and ductal cells) and both eGFP$^+$ and eGFP$^-$ cells were analyzed. Adult islets were hand picked and used for direct analysis (green dashed lines. The blue staining within the islet is from *Pdx1-lacZ*). cRNA probes for each sample were hybridized to Affymetrix microarrays Mu11K, Mu74Av1 or Mu74Av2 (Materials and methods). Temporal analysis (the red box) compared the gene expression patterns of endoderm, pancreatic progenitors (Pdx1-eGFP$^+$), endocrine precursors (Ngn3-eGFP$^+$), and adult islets. Spatial analysis (green boxes 1, 2 and 3) compared different samples from the same stages, e.g. genes expressed in E7.5 endoderm versus mesoderm + ectoderm.

was used to FACS sort PDX1$^+$ from PDX1$^-$ from dissected pancreas, stomach and duodenum. A total of 1.3×10^6 or 1.8×10^6 Pdx1-eGFP$^+$ cells (from pancreatic or stomach/duodenumal regions, respectively) were isolated from 350 E10.5 embryos. The trypsinization of tissue before cell sorting did not alter the ability of these cells to differentiate into insulin-producing cells in vitro (G.G. and D.A.M., unpublished data), nor did it dramatically alter the presence or absence of predicted gene expression in this analysis. However, we cannot rule out the possibility that the expression levels of some genes were altered by this isolation method. (3) The *Ngn3* promoter, which recapitulates endogenous *Ngn3* expression (Gu et al., 2002), was used to drive eGFP expression in endocrine progenitor cells. 1.3×10^6 Ngn3-eGFP$^+$ cells were collected from 1,300 E13.5 embryos. (4) Islets were isolated from 10 adult mice. All tissue or cell samples were separated into duplicates and used to generate labeled cRNA samples using

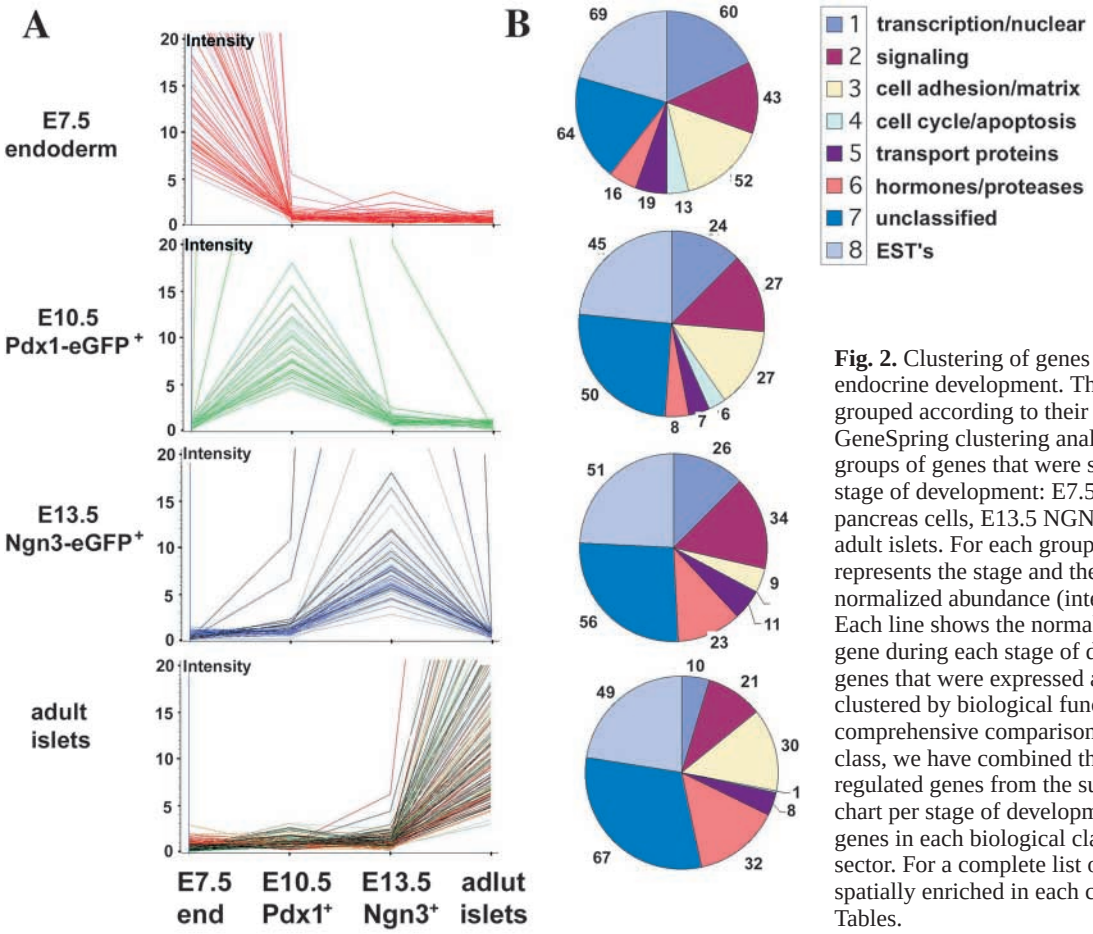

Fig. 2. Clustering of genes expressed at each stage of endocrine development. These clustered genes are also grouped according to their biological function. (A) A GeneSpring clustering analysis that identified four groups of genes that were specifically expressed in each stage of development: E7.5 endoderm, E10.5 PDX1$^+$ pancreas cells, E13.5 NGN3$^+$ endocrine progenitors, and adult islets. For each group of genes, the X-axis represents the stage and the Y-axis represents the normalized abundance (intensity) of each transcript. Each line shows the normalized expression level of one gene during each stage of development. (B) Pie charts of genes that were expressed at each stage of development, clustered by biological function. For a more comprehensive comparison of genes in each biological class, we have combined the temporally and spatially regulated genes from the supplemental tables into one pie chart per stage of development. The absolute number of genes in each biological class is shown next to each pie sector. For a complete list of genes either temporally or spatially enriched in each cell sample, see Supplementary Tables.

an in vitro transcription-based linear amplification protocol (Baugh et al., 2001). Amplified RNA samples were hybridized to the Affymetrix microarrays (Materials and methods), and the data were analyzed using GeneSpring and Resolver clustering analysis software. Genes expressed at each stage of development were grouped according to biological function (Fig. 2B and tables), and separated into classes that are temporally or spatially regulated during endocrine development. Genes that were expressed in the pancreas, but were not temporally or spatially regulated were generally not listed in the tables (see supplemental data for a complete listing of genes expressed in these samples: http://dev.biologists.org/supplemental).

Analysis of temporal gene expression during endocrine islet development

We used Affymetrix software (M.A.S.5) to identify genes expressed at significant levels within each sample. We found that 47, 38, 35 and 46% of the genes present on the microarrays are expressed in the E7.5 endoderm, E10.5 pancreatic progenitor cells, E13.5 endocrine precursors, and islets of Langerhans, respectively (data not shown). We used GeneSpring software (Silicon Genetics) to group genes whose expression is temporally restricted a specific stage of development. Three different statistical group comparisons were used (Student's *t*-test, Welch *t*-test and a nonparametric test). In order to have high confidence that selected genes are differentially expressed, we focused on genes that exhibit at least a threefold expression difference between samples. Raw data are available at www.genet.chmcc.org (contact G.G. for details). We identified 193, 60, 71 and 217 genes whose expression is enriched in E7.5 endoderm, E10.5 PDX1$^+$ pancreatic cells, E13.5 endocrine progenitors, and endocrine islets respectively (Fig. 2).

Endoderm cells express many transcripts involved in pattern-formation

E7.5 endoderm expresses 193 genes or ESTs (out of the ~12,000 on the microarray) at greater than threefold higher levels than cells at later stages of pancreas development. These include 25 growth factors or other signaling-related molecules and 44 transcription factors or other nuclear proteins (Fig. 2). Many of these factors were previously implicated in embryonic pattern formation. For example, endoderm expresses molecules involved in TGFβ signaling, including Nodal, cerberus 1, and follistatin and the *Wnt* antagonist dickkopf (Bouwmeester et al., 1996; Conlon et al., 1994; Mukhopadhyay et al., 2001). Endoderm-expressed transcription factors including *Cdx1*, *Hesx1*, *Irx3*, *Gata3*, *MespI* and *Sox17* (see Table S1, http://dev.biologists.org/supplemental). In addition, we have implicated several new signaling pathways in endoderm and pancreatic development by virtue of their abundant expression. Some examples include the *cKit* ligand, *Edg2* (G-protein coupled receptor) and *Epha2* (Eph receptor A2). The *cKit*

Table 1. A partial list of abundant genes both spatially and temporally restricted in each sample (for a complete list of these genes, see Table S1*)

Endoderm (E7.5)		PDX1+ cells (E10.5)		Endocrine progenitors (E13.5)		Mature islet	
Signaling molecules							
Cer1	Cerberus 1 homolog	*Dlk1*	Delta-like homolog 1	*Alk6*	Alk-6	*Notch4*	Notch homolog 4
Dab2	disabled homolog 1	*IGFbp5*	IGF binding protein 5	*Dlk2C*	Threonine/serine protein kinase Dlk	*Inha*	Inhibin alpha
Dkk1	Dickkopf homolog 1	*SfrP1*	Secreted frizzled-related protein 1	*Mfng*	Manic fringe	*Thra*	thyroid hormone receptor alpha
Edg2	Endothelial differentiation receptor 2			*Pim2*	Pim 2 kinase	*Prlr*	Prolactin receptor
EphA2	Eph receptor A2						
Kitl	c-Kit ligand						
Transcription factors							
Cdx1	Caudal type homeobox 1†	*Barx1*	BarH-like homeobox 1	*Brn4*	Brain pou-domain 4	*Atf5*	Activating transcription factor 5
Hesx1	Homeobox gene expressed in ES cells	*Nkx6.2*	NK related transcription factor 2	L-*Myc*	Murine L-myc	*Cpeb*	Cytoplasmic polyadenylation element binding protein
Irx3	Iroquouis related protein 3	*Ocut1*	One cut domain, member 1	*MafB*	v-Maf oncogene	*Myt1*	Myelin transcription factor 1
Gata3	GATA binding protein 3	*Sox11*	SRY-box containing gene 11	*Myt1L*	Myelin transcription factor 1-like	*Stat5B*	Stat5B
Msep1	Mesoderm posterior 1	*Zac1*	zinc finger protein regulator of apoptosis and cell cycle arrest	*NeuroD1*	Neural differentiation 1		
Sox17	SRY-box containing gene 17			*Pax4*	paired box-homeo-protein 4		
Cell adhesion/matrix protein							
Cubn	Cubilin	*Col1a1*	Collagen 1, alpha 1 subunit	*Anxa4*	Annexsin A4	*CD84*	CD84 antigen
EndoA	Cytokeratin endoA	*Col1a2*	Collagen 1, alpha 2 subunit	*Mtap1b*	Microtubule associated protein 1b	*Crpd*	Crp-ductin
		Col5a2	Collagen 5, alpha 2 subunit				
		Tnc	Tenascin C				
		Vnn1	Vanin 1				

*http://dev.biologists.org/supplemental/

pathway is known to function during hematopoiesis and germ cell migration and development (Ueda et al., 2002) and both of these processes involve interactions with endoderm. Thus, the role of endodermally expressed *cKit* may be restricted to hematopoietic and germ cell development.

Gene expression complexity decreases as cells become restricted to the pancreatic lineage

The PDX1$^+$ cells of the E10.5 pancreas (precursors to all components of the developing pancreas) expressed 60 genes at enriched levels (Fig. 2), a smaller number than the endoderm-specific genes. This is consistent with the PDX1$^+$ cells being a fate-restricted population while the endoderm cells contain progenitors for all endoderm-derived organs (Wells and Melton, 1999).

Examination of these genes suggested that Notch activity and Wnt signaling might play roles in promoting endoderm to adopt a pancreatic fate, since the genes for the Notch ligand *Dlk1* and Wnt signaling antagonist *Sfrp1* were highly expressed in these PDX1$^+$ cells (Table 1). In addition, genes for several transcription factors, including *Barx1, Nkx6.2, Onecut1*, *Sox11* and a few other zinc finger proteins were highly expressed in the PDX1$^+$ cells. Several ECM proteins, including collagens Iα1, Iα2, Vα2, tenascin and vinin 1 were also highly enriched in the PDX1$^+$ cells, suggesting that these molecules could be involved in the budding process of the early pancreatic epithelium (reviewed by Kim and Hebrok, 2001).

We identified 71 transcripts that are enriched in NGN3$^+$ endocrine progenitors (Fig. 2). *Manic fringe, IGFbp*, an activin-receptor-like kinase (*Alk6*), and two serine/threonine protein kinase transcripts are abundantly expressed (Tabe 1; Table S1, http://dev.biologists.org/supplemental). *Manic fringe* encodes a glycosyl transferase and is an important modifier of Notch signaling (Johnston et al., 1997; Shimizu et al., 2001). Its expression only in the endocrine progenitors suggested its involvement in endocrine development. Several transcription factors, including mouse brain-2 Pou domain protein and myelin transcription factor 1 are also expressed in the NGN3$^+$ progenitors, suggesting their involvement in development of the endocrine pancreas. Relative to the other stages of pancreatic development, the number of extracellular matrix/cell adhesion molecules is low in endocrine precursors. This finding is consistent with the idea that endocrine progenitor cells are not part of the epithelium, but rather have delaminated and remain apart from the branching exocrine pancreas (Kim and Hebrok, 2001).

Genes expressed in adult islets

The islet preparation contained the four major endocrine cell types, endothelial cells, some exocrine cells, and other cells that contaminated the islet preparations. We found that the expression of 217 genes (Fig. 2; Table S1, http://dev.biologists.org/supplemental) were enriched at this stage, and most of these are associated with the function of the adult organ. Among these, the transcripts for four endocrine hormones, hormone processing enzymes, secretory apparatus, prolactin receptor, REG1 and REG3, were found at very high levels. In addition, we identified the novel expression of numerous regulatory molecules in adult islets (Table S1, http://dev.biologists.org/supplemental). Genes for the transcription factors that were expressed include activating transcription factor 5 (*Atf5*), myelin transcription factor 1-like (*Myt1l*), putative homeodomain transcription factor (*Phtf*), and short stature homeobox 2 (*Shox2* also *Prx3*). Although the role of these transcription factors in islet function or maintenance is not known, *Atf5*, *Mytl1* and *Shox2* are all expressed in the

CNS, implicating them in neuroendocrine as well as pancreatic endocrine development and function (Angelastro et al., 2003; Kim et al., 1997a; van Schaick et al., 1997). There were also components of several signaling pathways expressed, including Notch 4, inhibin α, Wnt4, leukemia inhibitory factor receptor, and epidermal growth factor, to name a few. These molecules and pathways are possibly involved in regulation of islet size, function and perhaps maintenance.

The adult islets also expressed many of the same transcription factors that function in embryonic pancreatic development. One example is *Pdx1*, which is expressed in entire embryonic pancreas, but is restricted to β cells in the islets. PDX1 was shown to regulate expression of several genes in islets including insulin, glucagon, somatostatin, islet amyloid polypeptide (*Iapp*), glucokinase and *Glut2* (Brissova et al., 2002; Perfetti et al., 2001). PDX1 is also implicated in β cell maintenance in the adult (Sharma et al., 1999; Wells and Melton, 1999), suggesting that one additional role of some embryonic transcription factors might be maintain progenitor cells in the adult.

Gene expression levels as an indicator of differentiation, plasticity and transformation

As endocrine progenitor cells differentiate and form islets, the number of transcriptional and growth factor molecules expressed in endocrine cells decreased. These data suggest that maintenance of progenitor cell plasticity may depend on low-level expression of multiple regulatory genes. Alternatively, the fact that progenitor cells expressed numerous regulatory genes at low levels could reflect the heterogeneity of the progenitor pools. Analyses of genes expressed in single cells of the E10.5 pancreas suggested that *Pdx1*-expressing cells are a relatively heterogeneous population (Chiang and Melton, 2003). Another interpretation of that data is that only a subset of PDX1$^+$ cells are specified toward pancreatic lineages and the remainder are still plastic. This idea is supported by cell lineage studies which demonstrated that many of the cells of the embryonic pancreas, once thought to be pancreatic progenitor cells, never actually contribute to the adult organ (Herrera, 2000).

To identify additional genes that might regulate early cell plasticity, we performed a clustering analysis to identify genes that were down-regulated as a function of differentiation. This cluster of genes contains many known regulators of differentiation, proliferation and plasticity during development (Fig. S1; Table S2, http://dev.biologists.org/supplemental). Included in this cluster of 'down-regulated genes' were numerous tumor-associated genes such as *Tera* (teratocarcinoma expressed, serine rich), *Tacc3* (transforming acidic coiled coil containing protein 3), *Ptov1* (prostate tumor over expressed 1), *Tacstd2* (tumor-associated calcium signal transducer 2), *Trap1a* (tumor rejection antigen 1), *Frat1* (frequently arranged in advanced T-cell lymphomas), and *Lag* (leukemia associated gene). Although the function of these factors in pancreas development is unknown, they were all identified by their abundant expression in different types of tumors and are thus implicated in cellular transformation.

Analysis of genes that are spatially restricted during islet cell development

Our temporal analysis of gene expression identified genes that were known to regulate temporal cell differentiation during endocrine cell development. However, it is equally important to identify the genes that define developing pancreatic cells from their neighbors. For example, how are PDX1$^+$ cells of the pancreas different from the PDX1$^+$ cells of stomach or duodenum, and how do the NGN3$^+$ cells differ from NGN3$^-$ cells? To catalog the genes that control these cell fate decisions, we have generated a transcriptional profile from developing endocrine progenitor cells and from adjacent cells at each stage of development (Fig. 1, green boxes).

Genes differentially expressed in the early endoderm as compared to mesoderm and ectoderm

In order to identify the genes whose expression is spatially restricted to endoderm at E7.5, we compared gene expression profiles between E7.5 endoderm and mesoderm plus ectoderm (Fig. 1, green box 1). We identified 203 transcripts that are greater than threefold enriched in endoderm, while 262 were enriched in the mesoderm plus ectoderm (Fig. 3, Table 2; Table S3, http://dev.biologists.org/supplemental). We have verified endodermal expression of 25 genes by RT-PCR, and 17 of these were further analyzed by ISH analysis (Fig. 3; Table S6, http://dev.biologists.org/supplemental). The gene expression patterns shown in Fig. 3 (*ApoAIV, Dkk1, Prss12*, and *Eya2*) are representative examples of genes that were expressed in endoderm. The expression of these genes in endoderm validates that our approach was successful in identifying endodermally expressed genes.

Several transcription factor mRNAs such as *Foxa2* (*HNFf3β*) and *Sox17* were known to be expressed in endoderm and were detected using microarrays. We also found that several genes with homologs in *Drosophila* such as *Hes1* (hairy enhancer of split) and *Klf5* (kruppel-like factor 5) were enriched in endoderm (Table 2; Table S3a, Table S3b, http://dev.biologists.org/supplemental). Genes for two transcription factors, *EYA2* and *Six1*, that were characterized by their function during eye development, are expressed by E7.5 endoderm cells (Fig. 3), but their function here is unknown. There were several signaling molecule genes that were more abundantly expressed in endoderm, as compared to mesoderm + ectoderm, including *Wnt11, IgfII, chordin,* cerberus 1 (*Cer1*) and genes encoding proteins that enhance growth factor activity, such as, *Fgfbp1* and *Igfbp5*. The co-expression of factors with opposite activities in endoderm highlights the complex nature of signals involved in patterning the endoderm and the adjacent germ layers at this stage of development (Beddington and Robertson, 1999).

Genes differentially expressed in PDX1$^+$ cells of the pancreas, stomach and duodenum

Pdx1 expression marks all pancreatic progenitors of the E8.5-10.5 pancreas (Gannon et al., 2000; Gu et al., 2002). Yet, *Pdx1* is also expressed in cells in rostral stomach, and the mucosal cells of the duodenum (Offield et al., 1996), demonstrating that additional factors are necessary to specify pancreatic fate. We isolated PDX1$^+$ cells of the pancreas, stomach and duodenum to identify the genes that are specifically expressed in pancreatic progenitor cells (Fig. 1, green box 2). Cell lineage analyses have demonstrated that PDX1$^+$ cells in the pancreatic buds at E10.5 give rise to all pancreatic tissues whereas the PDX1$^+$ cells in the stomach/duodenum rudiment do not give rise to pancreatic tissues (Gu et al., 2002). We also analyzed PDX1$^-$ cells from the mesoderm surrounding the pancreas,

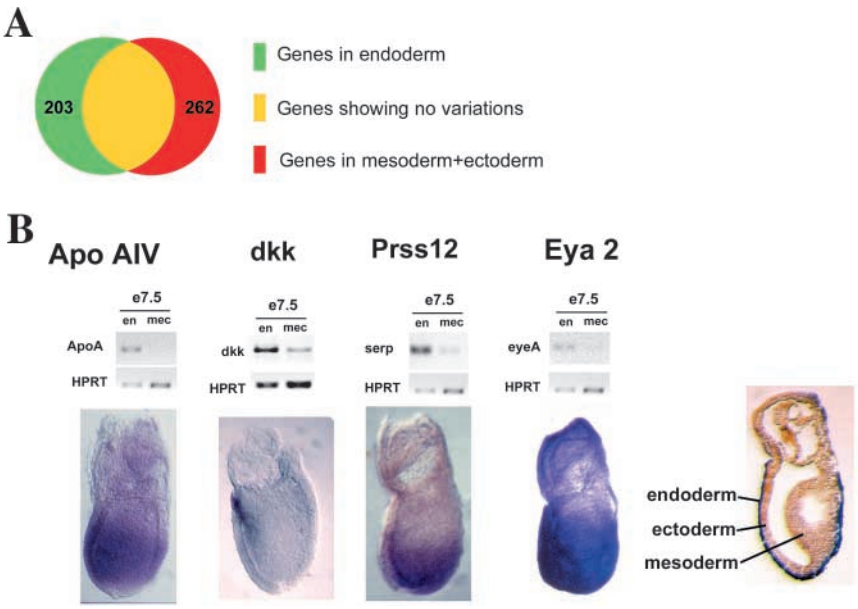

Fig. 3. Genes expressed in endoderm. (A) Venn diagram illustrating genes that were at least threefold enriched in unspecified endoderm as compared to adjacent mesoderm and ectoderm. 203 genes were enriched in the endoderm, whereas 262 genes were enriched in the mesectoderm. (B) The endodermal expression of four genes, *Apo AIV, Dkk1, Prss12* and *Eya2*, identified from our microarray analysis, was verified by RT-PCR (end, endoderm; mec, mesoderm + ectoderm; HPRT is used as a control for RT-PCR) and whole mount ISH. These genes were not previously shown to be expressed in endoderm. The endodermal expression of *Eya2* was verified by sectioning the stained embryo. ect, ectoderm; end, endoderm; mes, mesoderm.

stomach and duodenum. These include the mesenchymal cells surrounding the endoderm and PDX1$^-$ epithelial cells.

We identified the transcripts that are enriched in the pancreatic PDX1$^+$ cells by comparing the expression profile of these cells with that of the combined expression profile of the PDX1$^+$ cells of the stomach and duodenum, as well as that of the PDX1$^-$ cells. This clustering analysis identified 158 genes that are enriched in the PDX$^+$ pancreatic buds. 208 transcripts were enriched in the stomach, duodenum, and the PDX1$^-$ cells (Fig. 4A and Table 3; Table S4, http://dev.biologists.org/supplemental). We verified the expression pattern of 25 candidate genes whose transcripts were enriched in the PDX1$^+$ pancreatic cells by RT-PCR (25-30 cycles) and 12 by ISH. We determined that the transcripts of 21 of the 25 genes were highly enriched in the pancreatic epithelium, compared to that of the duodenum or stomach and surrounding mesenchymes. Expression of the remaining four candidates was not detectable in any tissue (Fig. 4; Table S6, http://dev.biologists.org/supplemental). Similarly, 15 of the 18 candidate genes whose transcripts were enriched in the non-pancreatic cells were found by RT-PCR and/or ISH to be enriched only in the mesenchyme, stomach or duodenum (Table S6, http://dev.biologists.org/supplemental). The remaining three transcripts were not detected in any tissues (Fig. 4B-E and data not shown). We increased the number of PCR cycles in our analysis to 45 and found that we could detect the seven low-abundance transcripts. Data from our Affymetrix analysis predicted these seven genes to be expressed at low levels.

Genes or signaling pathways known to function for pancreas development are expressed in the pancreatic PDX1$^+$ cells

Several genes that are known to be involved in pancreatic function were detected only in the PDX1$^+$ cells in the pancreatic buds. Examples include glucagon, *App* (amyloid precursor proteins), *Glut2* transporter, the vesicle forming proteins *Cop4*, and clathrin coating protein *AP47*. In addition, transcripts of different components for signaling pathways known to function for pancreas development were also detected at enriched level in the pancreatic PDX1$^+$ cells. These include *Notch1* and its ligand *Delta-like 1*, and several FGF receptors (Table 3). We also confirmed that genes that are known to play a role in stomach or duodenum development, including *Rab8, IGFBP2, Shh, Ihh* (Ramalho-Santos et al., 2000), and those of several transcription factors, including *Elf3, Eklf, KlF4, Pax1, Sox2* and *Sox11*, are substantially enriched in PDX1$^+$ cells in the stomach or duodenum and/or mesenchymal cells.

Identification of new pathways or factors that are expressed in pancreatic PDX$^+$ cells

Several genes that were not known to be involved in pancreatic development were found to be expressed by the pancreatic PDX1$^+$ cells. Examples include a G protein (*RhoB*), a related signaling member [diaphonos homolog 1 *(Dab1)*] and calmodulin (*Cldn*, Table 3). Because Rho plays an essential role in focal adhesion formation, another molecule, FAK (focal adhesion kinase), also detected in these cells, (data not shown) may interact with the four above-mentioned molecules to control the morphogenesis of the pancreatic rudiment.

Genes differentially expressed in early endocrine (NGN3$^+$) progenitors

Pancreata from E13.5 embryos were dissected from animals expressing eGFP from the *Ngn3* promoter, and cells were dissociated and separated into NGN3$^+$ and NGN3$^-$ cells based on their eGFP expression [The *Ngn3* promoter used in these experiments recapitulates endogenous *Ngn3* expression (Gu et al., 2002)]. We determined that 204 genes were enriched in endocrine progenitors, as compared to 256 genes that were enriched in non-*Ngn3*-expressing cells (Fig. 5A, Table 4; Table S5, http://dev.biologists.org/supplemental). All genes known to be important for islet development were detected at high levels only in the NGN3$^+$ cell samples (Table S5a). In addition, transcripts of many genes not previously identified as playing roles in endocrine development were also enriched in the Ngn3-eGFP$^+$ cells. In the Ngn3-eGFP$^-$ cells, *Ngn3* transcripts were not detected, demonstrating that our sorted Ngn3-eGFP$^-$ pool was devoid of *Ngn3*-expressing cells. We used ISH to analyze

Table 2. A partial list of genes enriched in early endoderm compared with mesoderm/ectoderm (for a complete list, see Table S3a,b*)

E7.5 endoderm		E7.5 mesoderm and ectoderm	
Signaling molecules			
Cer1	cerberus 1 homolog	*Crabp2*	Cellular RA bp2
Chrd	Chordin	*Fst*	Follistatin
Dkk1	Dickkopf homolog 1	*Tssc3*	Tumor surpressing
FGFbp1	FGF binding protein 1	*Gng3*	G-protein gamma 3
IGF11	Insulin growth factor like 1		
Igfbp5	IGF binding protein 5		
Wnt11	Wnt11 signaling molecule		
Transcription factors			
Eya2	eyes absent 2 homolog	*Gbx2*	gastrulation brain homeobox2
Fos	Fos oncogene		
Foxa2	Forkhead box a2	*Hoxa1*	homeobox a1
Hes1	Hairy enhancer of split 1	*Pou5f1*	POU class 5 TF 1
Klf5	Krupple factor 5	*Sox 2, 3*	SRY-box 2, 3
Six1	sine oculis homeobox 1	*TF1*	Transcrption factor 1
Sox17	SRY-box containing gene 17	*Zic2*	Zinc finger of the cerebellum 1
Cell adhesion/matrix/protease protein			
Prss12	protease, serine, 1	*Actc1*	Cardiac alpha actin
		LamR1	Laminin receptor
Unclassified			
Apoa4,e	Apolipo protein a4, e	*Bcat1*	Branched chain amino transfer
		UTF1	Undifferentiated in ES cells

*http://dev.biologists.org/supplemental/

the expression pattern of 18 candidate genes whose transcripts were only present in endocrine progenitor (NGN3$^+$) cells. 12/18 candidates analyzed were expressed in a scattered cell population in the E10.5, E12.5 and E15.5 pancreat ic rudiments (Fig. 5B-E; Table S6, http:// dev.biologists.org/supplemental), an expression pattern that is highly similar to that of *Ngn3* (Gradwohl et al., 2000). Six of the 18 candidates cannot be detected by ISH, possibly because they are expressed at low levels, which would be consistent with their low hybridization intensity on the microarray (data not shown).

Endocrine progenitors only transiently express *Ngn3* prior to differentiating into mature endocrine cells (Gu et al., 2002). Since eGFP protein is very stable, we anticipated that eGFP$^+$ cells isolated from the *Ngn3-eGFP* transgenic animals would contain some cells that had differentiated toward mature endocrine cells, yet still had eGFP. It is therefore not surprising that substantial levels of somatostatin, glucagon and insulin transcripts were detected in the Ngn3-eGFP$^+$ cell pool. The expression levels of these hormones in the Ngn3-eGFP$^+$ cells were less than 5% of the expression levels in adult islets (data not shown). This finding is consistent with the idea that Ngn3-eGFP$^+$ cells are the endocrine precursors that eventually give rise to mature endocrine islets.

Several G-protein signaling components were enriched in endocrine progenitors

Transcripts encoding several G protein-coupled receptors *(GPR27* and *GPR56*) and multiple guanine nucleotide-binding proteins, including *Gα0, Rab3D, Rab7* and a *GDP* dissociation inhibitor (Table 4), were highly enriched in the NGN3$^+$ cells.

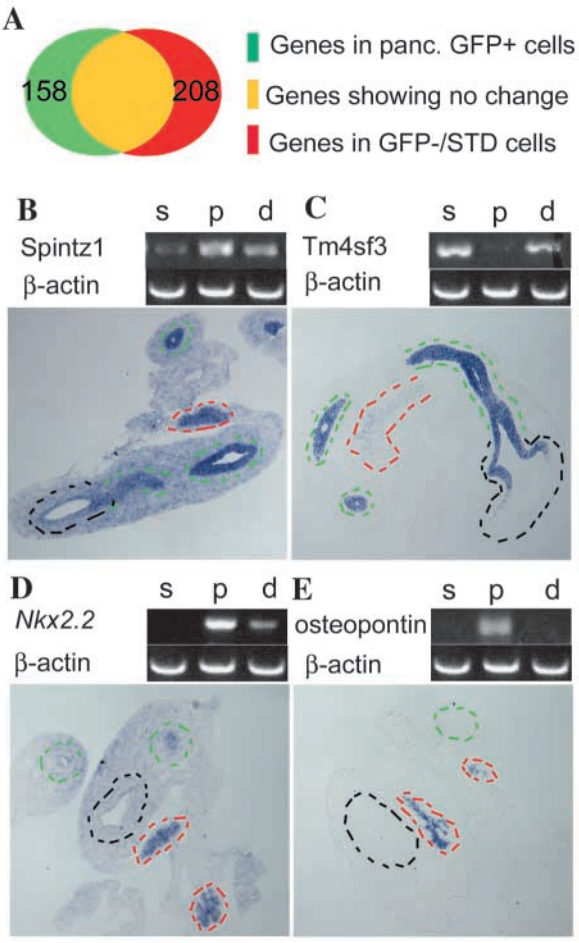

Fig. 4. A summary of genes expressed in the pancreatic Pdx1-eGFP$^+$ cells and non-pancreatic cells at E10.5 (including Pdx1-eGFP+ cells of the stomach + duodenum as well as PDX1-GFP$^-$ cells). (A) Venn diagram showing that 158 genes (green) were enriched in the eGFP$^+$ cells within the pancreatic region whereas 208 (red) genes were enriched in the PDX1$^+$ cells in the duodenum and stomach or eGFP- cells. (B-E) Expression analyses of four genes by RT-PCR and in situ hybridization. β actin expression was used as a control. (B) Spintz1(Af010499) was highly expressed in the pancreas and duodenum but not the stomach. (C) Tm4sf3 (AA571115) was expressed at a high level in the stomach and duodenum but not the pancreas. (D) *Nkx2.2* was expressed at a higher level in the pancreas. (E) Osteopointin was only expressed in the pancreas. Black dashed lines, stomach; green dashed lines, duodenum; red dashed lines, pancreas.

Transcripts for several calcium signaling-related molecules, a calcium-binding protein (*ALG2*), a calcium-dependent activator (*Cadps*), calcium-dependent kinase II (*Camk2b*), and a calcium-independent phospholipase A II (*Pla2g6*) were also highly enriched in endocrine progenitors. The presence of these molecules suggests that G-protein-mediated signaling, through receptor GPR27 or GPR56, and calcium mediated signaling might participate in endocrine development or function.

Components of the notch-signaling pathway are expressed by endocrine progenitor cells

Our screen not only revealed the presence of the transcripts for

Table 3. List of genes differentially expressed in PDX1$^+$ pancreatic (but not stomach) cells compared with those in PDX1$^+$ stomach and duodenum (but not pancreatic) and PDX1$^-$ cells (for detailed description of these genes and all candidates, see Table S4*)

PDX1$^+$ pancreatic cell		PDX1$^+$ stomach and duodenum cells	
Signaling molecules			
Cldn	Calmodulin	*Rab8*	Rab protein 8
Dl1	Delta like 1	*IGFBP2*	IGF binding protein 2
Dab1	Diaphonos homolog 1	*Ihh* and *Shh*	Indian and Sonic hedgehog
Notch1	Notch homolog 1		
RohB	Guanine nucleotide exchange factor		
FGFR2, 4	FGF receptor 2 or 4		
Transcription factors			
Hoxc5	Homeobox protein c5	*Cut1-L*	Cut1-like
C/EBP	CCAAT/Enhancer-binding protein	*EKLF*	Erythrocyte kluppel-like factor
		Elf3	E47-like factor 3
		KLF4	Kluppel factor 4
		Pax1	Paired box gene 1
		Sox2, 11	SRY-box containing gene 2 and 11
Cell adhesion/matrix protein			
AP47	Adaptor protein 47		
Cop4s	COP 9 homolog, 4		

*http://dev.biologists.org/supplemental/

Notch signaling members, but we also discovered that of a Notch modifier, manic fringe *(Mfng)* and a transcription factors, *Myt1* (Bellefroid et al., 1996) that participate in Notch signaling (Table 4). This finding suggests that *Mfng* and *Myt1* could be involved in endocrine cell development.

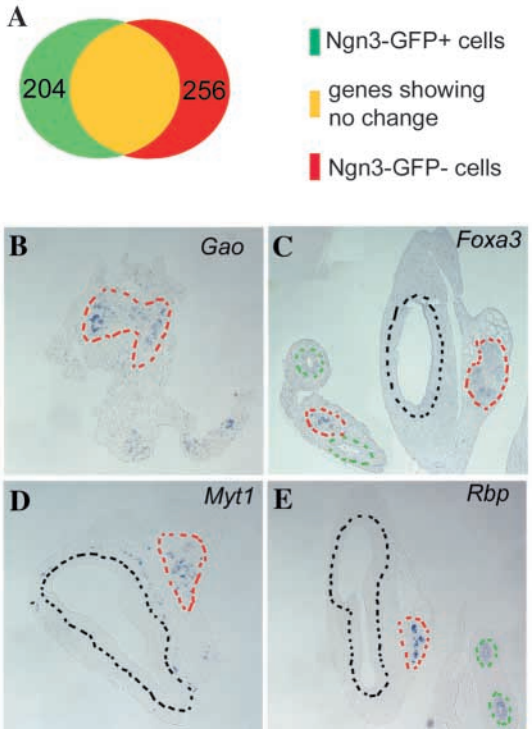

Table 4. A partial list of genes differentially expressed in endocrine compared with nonendocrine progenitor cells (for detailed description of these genes and all candidates, see Table S5*)

Endocrine ngn3+ progenitor cells (E13.5)		Nonendocrine ngn3– cells (E13.5)	
Signaling molecules			
Alg2	Calcium binding protein	*EdoR*	Endothelin R
Cadps	Ca dependent activator for secretion	*GPR26*	G-protein coupled receptor 26
Cmk2b	Ca/Calmodulin dependent kinase 2b	*IGF1*	Insulin like growth factor 1
G0α1	Guanine nucleotide binding, alpha O	*IGF2*	Insulin like growth factor 2
Gdpd1	GDP dissociation inhibitor1	*IGFBP5*	IGF binding protein 5
GPR27	G-protein coupled receptor 27	*Notch1*	Notch 1 homolog
GPR56	G-protein coupled receptor 56	*PDGFR*	PDGF receptor
Mfng	*Manic fringe*	*Sfrp1*	Secreted frizzled protein 1
Pla2g6	Phospholipase 2, group 6		
Rab3d	Member of ras oncoprotein 3d	*Sfrp2*	Secreted frizzled protein 2
Rab7	Member of ras oncoprotein 7	*Thrombin R*	Thrombin receptor
Transcription factors			
Myt1	Myelin transcription factor 1		

*http://dev.biologists.org/supplemental/

Signaling molecules expressed by NGN3$^-$ cells

The NGN3$^-$ cells included several tissue types, such as progenitor cells that had not been specified toward the endocrine cell fate (by virtue of its *Ngn3* expression), precursors that give rise to the exocrine pancreas, and mesodermally derived tissues within the pancreas. Consequently, diverse signaling pathways were found to be expressed by the NGN3$^-$ cells. Transcripts enriched in Ngn3-eGFP$^-$ cells included the endothelin receptor, PDGFR, thrombin receptor, which are known for hematopoietic development. However it is still possible that these genes are important for endocrine development.

Analysis of *Myt1* function during endocrine cell development

One goal of our gene expression analysis was to identify new genes that are functionally involved in endocrine islet development. Of the genes whose transcripts are enriched in the endocrine progenitors, one gene, *Myt1*, is a promising candidate regulator of endocrine development. In *Xenopus laevis*, xMyt1 has been shown to cooperate with xNgn1 to induce neurogenesis (Bellefroid et al., 1996). Because islet development has many similarities with that of neuronal development (Gu et al., 2003), we wanted to determine whether *Myt1* is involved in endocrine differentiation.

Fig. 5. Genes expressed in the Ngn3-eGFP$^+$ and Ngn3-eGFP$^-$ cells at E13.5. (A) Venn diagram showing that 204 genes (green) were enriched at least threefold in the Ngn3-eGFP$^+$ cells as compared to 256 (red) genes that were enriched in the Ngn3-eGFP$^-$ cells. (B-E) Expression analysis of four genes by ISH. *Gαo* (B), *Foxa3* (C), *Myt1* (D), and *Rbp* (E) were only detected at high levels in a set of scattered pancreatic cells, similar to *Ngn3*. Black dashed lines, stomach; green dashed lines, duodenum; red dashed lines, pancreas.

Table 5. The number of insulin$^+$ and glucagon$^+$ cells in wild-type and transgenic animals carrying dominant-negative *Myt1 (dnMyt1)* transgene

	Insulin$^+$ cells		Glucagon$^+$ cells	
	Number	% reduction	Number	% reduction
Wild type	2968, 2846, 2359, 2157	NA	1342, 981, 1192, 1280	NA
dnMyt1	2209, 1498, 1353, 1306, 1193	39%	647, 1286, 872, 587, 661	32%

Wild type, n=4; transgenic animals, n=5.

The *Myt1* locus produces two isoforms by utilizing alternative transcriptional starts, *Myt1* (noted as *Myt1a*) and *Nzf2b*, both containing C_2HC zinc fingers and a transcriptional activation domain. These two isoforms differ in their N-terminal 100 amino acid residues (Matsushita et al., 2002), such that NZF2b has an extra zinc finger (MYT1a has 6 zinc fingers and NZF2b has 7 zinc fingers). For simplicity, we refer to both RNA isoforms from the *Myt1* locus as *Myt1*, and we refer to the 6-zinc-finger *Myt1* cDNA as *Myt1a* (Kim et al., 1997a; Matsushita et al., 2002). Our semi-quatitative RT-PCR results showed that *Myt1a* and *Nzf2b* are both expressed in the developing pancreas, with *Nzf2b* being expressed at much higher levels (data not shown). In situ analysis using probes common to both isoforms demonstrated that *Myt1* is expressed in a few cells of the developing gut (E8.5) where the pancreatic buds will form [between the seventh and ninth somites, adjacent to the dorsal aorta (Fig. 6B and data not shown)], as well as in the nervous system (Fig. 6B). As embryogenesis proceeds, *Myt1* is expressed in the pancreas in a similar fashion to that of *Ngn3*, i.e. in a scattered subset of epithelial cells that are adjacent to or within characteristic duct-like structures (Fig. 6C). After E15.5, *Myt1* transcripts were considerably reduced (Fig. 6D), yet not abolished (longer exposure of these tissue sections yields positive *Myt1* mRNA hybridization signals, data not shown). The expression pattern of *Myt1* suggests that it functions, like *Ngn3*, during the early stages of endocrine cell specification. We utilized gain-of-function and loss-of-function approaches to determine if *Myt1* was involved in development of the endocrine pancreas, using both mouse and chicken embryos as model systems.

In mouse, we broadly misexpressed *Myt1a* in pancreatic buds during embryogenesis using a *Pdx1* promoter. We found that *Myt1a* ectopic expression did not affect exocrine or endocrine cell development at E14.5 or E16.5, in terms of pancreatic morphology or molecular marker expressions (data not shown). We next tested whether MYT1 is necessary for mouse endocrine differentiation. We constructed a dominant negative (dn) *Myt1* to inhibit MYT1 function during endocrine development by deleting the transcriptional activation domain from *Myt1a*. This strategy was previously used to inhibit *Myt1* function during neural development in *Xenopus* (Bellefroid et al., 1996). We used the *Ngn3* promoter to specifically over express *dnMyt1* in endocrine progenitor cells of first generation, transient transgenic E14.5 embryos, and characterized the pancreatic phenotype. We found that the total number of insulin- and glucagon-expressing cells were reduced on average by 39% and 32%, respectively, in five transient transgenic E14.5 embryos (Fig. 6E,F and Table 5). Although

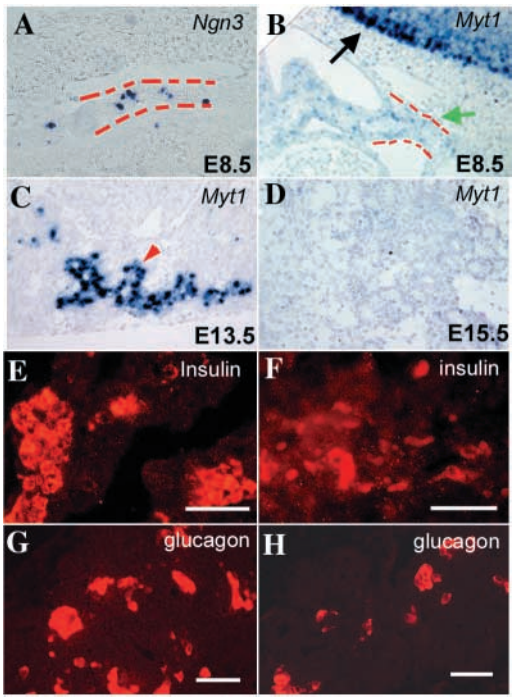

Fig. 6. Functional characterization of *Myt1* during mouse endocrine development. (A-D) In situ hybridization showing *Ngn3* and *Myt1* expression at several stages of embryogenesis. A is a control showing *Ngn3* expression in the developing pancreatic region (red dashed lines) at E8.5. (B) At E8.5 *Myt1* is expressed at a low level in the prospective pancreatic region in a manner similar to *Ngn3* (red dashed line). The prospective pancreatic region is recognized by its position below somite 8-10 (total of 15 somites), and the position of the dorsal aorta (green arrow). Strong *Myt1* expression was also detected in the neural tube (black arrow). (C) At E10.5, *Myt1* was detected in a scattered fashion in the pancreatic bud, within or close to duct-like structures (red arrowheads). (D) At E15.5, *Myt1* expression was much reduced (see text). (E-H) Transgenic expression of *dnMyt1* in endocrine progenitor cells inhibits endocrine differentiation. (E) Insulin$^+$ cells in a representative wild-type pancreas. (F) Insulin$^+$ cells in a representative pancreas section of a *Ngn3-dnMyt1* littermate. Note fewer insulin-expressing cells in F. Similarly, the number of glucagon$^+$ cells is reduced in the *Ngn3-dnMyt1* expressing embryo (H) compared with that of wild type embryo (G). In E-H, Cy3-conjugated antibodies were used. Scale bar: 25 μm.

the *Ngn3* promoter is cell specific to endocrine progenitor cells (Gu et al., 2002), the level of transgene expression may be too low for a dominant negative approach to totally abolish endocrine cell differentiation. Because somatostatin and pancreatic polypeptide are not yet expressed by E14.5, the effect of DnMYT1 on δ and PP cell development could not be determined.

As an alternative approach to test the function of *Myt1* in developing endocrine cells, we overexpressed *Myt1* (*Myt1a* and *Nzf2b*) and *dnMyt1* in chicken embryonic endoderm at approximately the 25-somite stage using electroporation (Materials and methods). This approach was previously used to demonstrate that misexpression of NGN3 in hindgut endoderm results in the differentiation of a large number of glucagon-expressing cells and a small number of somatostatin-expressing cells, but not insulin and pancreatic polypeptide-expressing

cells (Grapin-Botton et al., 2001). We found that misexpression of *Myt1a* in the chicken gut endoderm did not result in ectopic expression of any pancreatic markers (data not shown). However, misexpression of *Nzf2b* induced ectopic expression of glucagon and somatostatin, but not significant amount of insulin and pancreatic polypeptide, in the stomach and duodenum cells (Fig. 7). These results suggest that NZF2b is sufficient to partially initiate endocrine development in endoderm. Currently, it is not known why MYT1a and NZF2b have different activity in inducing endocrine marker expression. It is also not clear why *Ngn3* or *Myt1* fail to induce the formation of insulin and pancreatic polypeptide expressing cells.

Finally, we determined whether inhibition of *Myt1* activity suppresses the ability of NGN3 to induce ectopic endocrine differentiation in chick endoderm (Grapin-Botton et al., 2001). For this purpose, we co-expressed *dnMyt1* together with *Ngn3* in chicken embryonic hindgut endoderm and examined the expression of endocrine markers. Our results demonstrated that full-length *Myt1a* (Fig. 7E) or *Nzf2b* (data not shown) did not affect the ability of NGN3 to induce glucagon expression. Yet DnMYT1 significantly reduced the ability of NGN3 to induce ectopic glucagon expression (Fig. 7F). Since NGN3 has a very limited ability to induce formation of somatostatin-expressing cells, we were not able to use this dominant negative approach to determine if DnMYT1 could inhibit somatostatin expression. These results, combined with the transgenic mouse data, suggest that the ability of NGN3 to promote α and β cell differentiation depends, directly or indirectly, on *Myt1* function.

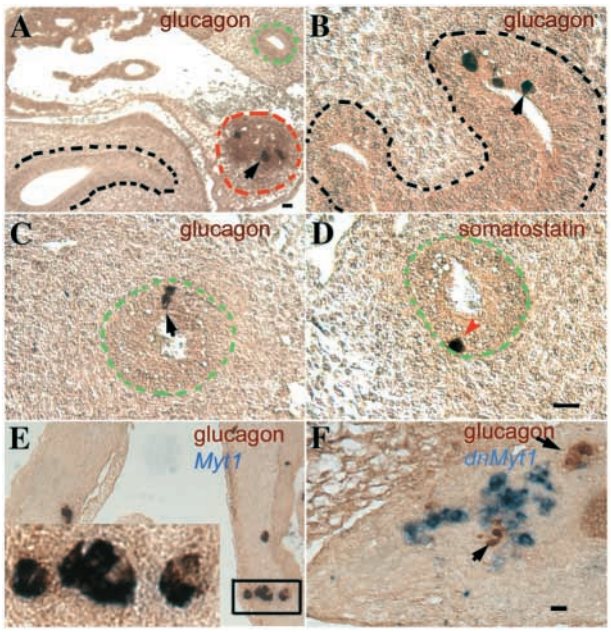

Fig. 7. *Myt1* is involved in generation of glucagon and somatostatin-expressing cells in chicken embryonic endoderm and may be in the same pathway as *Ngn3*. (A-C) In normal chicken embryos (A) glucagon (brown staining, black arrow) and somatostatin-expressing cells (data not shown) are absent in chicken stomach and duodenum. When *Nzf2b* was ectopically expressed in the chicken embryonic gut endoderm around stage 18, glucagon expression was induced in the stomach (B) and duodenum (C). A small, yet significant, number of somatostatin-expressing cells were also induced (D, red arrowhead). (E,F) dnMYT1 inhibits NGN3-mediated induction of glucagon$^+$ cell formation in chicken gut endoderm. *Ngn3* + *Myt1a* (E) or *Ngn3* + *dnMyt1* (F) were co-electroporated into chicken gut endoderm respectively. Glucagon expression (brown antibody staining) and *Myt1* or *dnMyt1* expression (blue, in situ hybridization) was measured. In E, co-expression of *Ngn3* and *Myt1a* (data not shown) (Grapin-Botton et al., 2001) resulted in induction of glucagon$^+$ cells (brown and blue color overlap resulting in dark brown/purple). In F, cells that co-express *Ngn3* and *dnMyt1* almost never express glucagon (brown and blue cells do not overlap) suggesting that the presence of dnMYT1 inhibited the ability of NGN3 to promote the generation of glucagon-expressing cells. Black dashed line, stomach; green dashed line, duodenum; red dashed line, pancreas.

Discussion

During organogenesis, specialized cell types are generated from progenitor cell populations and are precisely organized into the elaborate structure of the adult organ. This process involves numerous cell-cell communications and initiation of complex inter-regulating genetic networks to ensure fidelity of organogenesis. We have used a transcriptional profiling approach to begin to characterize the expression of regulatory or functional components during the development of early endoderm to pancreatic precursors, then to endocrine progenitors, and eventually to functional islet cells. The strength of this approach is that we started with enriched cell populations at each chosen stage of development, which we predict to greatly increase the sensitivity of the microarray analysis. The success of our analysis was immediately apparent since we detected the transcripts of most of the genes previously implicated in endocrine development in the developing endocrine islets. We subsequently have catalogued a large number of new candidate genes that may participate in islet cell development at different stage. These genes include cell-cell signaling molecules (receptors, growth factors), signal modifiers, transcription factors, transporters, ECM proteins, and many others. These analyses provide us with a global expression profile of genes that may interact to dictate the sequence of cellular development, from an unspecified progenitor to precursors whose fates are restricted to specific organs and finally to mature, functional cells.

New signaling molecules in endocrine development

In addition to those genetic pathways known to play a role in pancreatic development, our results have newly implicated several additional pathways (Table 6). In endoderm, we detected the expression of a genetic network that has been well studied in eye development. This network includes the genes *Eyes absent 2 (Eya2)* and *Sine oculis-related homeobox 1 (Six1)* that genetically interact during eye development in flies and mice (Heanue et al., 1999; Pignoni et al., 1997; Ridgeway and Skerjanc, 2001). In addition, we have identified multiple components of the *Wnt* pathway, including *Wnt* ligands, *Wnt* receptors (*Fzds*), *Wnt* receptor antagonists [secreted frizzled-related 1 and 3 (*SfrPs*)], and its downstream targets (*Dvls*) in early endoderm, general pancreatic progenitors and endocrine progenitors. In PDX1$^+$ cells in the pancreatic region (E10.5), *RhoB, Dab1, Cldn* and *FAK* are all expressed at enriched level. These genes have been shown to function in modifying cell cytoskeleton and they might be involved in pancreatic epithelia morphogenesis. In the endocrine progenitors, we found the

specific expression of G-protein cascade, *GPR14, GPR27, GPR56, Ga0, Rab3d, Rab7* and a *GDP dissociation factor* genes. These factors might interact with each other and participate in endocrine lineage differentiation. In addition, members of the calcium-activated signaling cascade may also participate in islet formation or function.

In addition to identifying new signaling pathways that possibly regulate islet formation, we identified new members of pathways that are known to function in islet differentiation. For example, we found that *Myt1* and *Mfng* are specifically expressed in the endocrine progenitors. These molecules have previously been linked to Notch signaling, either as a mediator or a modifier. Their presence in the endocrine progenitors suggests that these gene products may participate in islet formation. Our preliminary demonstration that DnMYT1 inhibits the generation of insulin and glucagon-expressing cells in mouse and/or chick supports this hypothesis.

The same signaling molecules regulate different cell fate decisions

We detected components of all common signaling pathways in each cell population representing different stages of islet generation (data not shown). Several specific growth factor receptors are expressed at each stage of development, yet probably direct the expression of different target genes, depending on the cell in which it is expressed. For example, cells at all stages analyzed expressed the activin receptor 2b. Activin/TGFβ signaling can be regulated by extracellular modifiers like Cer1, and receptors can transduce a signal via several different downstream *Smads*. It is therefore easy to speculate that the response of any given cell to activin signaling depends on many other cell-intrinsic and extrinsic factors according to the levels of signal strength and/or the competence factors present in the cells. This highlights the belief that a relatively small number of regulatory molecules can be used to determine the eventual cell type.

Global trends in gene expression to study complex biological processes

Our transcriptional profile of the developing endocrine pancreas has generated a quantitative gene expression database that can be used to analyze complex gene expression networks that would be impossible to study by other strategies. For example, our analysis suggests that the most plastic cell type, E7.5 endoderm, expressed many genes involved in cell fate specification, and the number of these genes becomes progressively fewer as endocrine cells begin to differentiate. Adult endocrine cells expressed the fewest number of cell fate regulatory genes but abundantly expressed genes associated with the adult function of the islets. The progressive decrease in the number of cell-fate regulators during endocrine development is consistent with the hypothesis that differentiation is a function of cells becoming progressively restricted toward one lineage. We identified another group of genes that were down regulated as a function of differentiation (Fig. S1 and Table S2, http://dev.biologists.org/supplemental). There is a significant number of tumor associated genes associated with this gene cluster, suggesting that the genetic machinery underlying cell plasticity in the embryo might overlap with the genes involved in the 'de-differentiation' that occurs during oncogenesis.

Table 6. A summary of new pathways detected by microarray analysis

Stage of development	Pathway members
E7.5 endoderm	*Eya2, Six1, Kit* receptor and ligand
E10.5 PDX1+ cells	*RhoB, Dab1, Caln, FAK*
E13.5 NGN3+ cells	*GpR27, GpR56, Ga0, Rabb3D, Rab7, GDPd1*
	Alg-2, Ca activator, CadKII, plpII
	Arx, Mfng, Myt1, Notch2

Expression data from these experiments will be available at www.genet.cchcc.org and can be directly compared to the expression profiles generated from other studies to look for informative biological trends between cell types and across organ systems. For example, a comparison of expression profiles between the developing and regenerating pancreas, or between two branching organs such as the pancreas and kidney could potentially uncover molecular trends associated with these processes.

Gene discovery: markers and regulators of developing pancreatic progenitor cells

Given the current research emphasis on deriving functional islets from stem or other cell types, the identification of new endocrine regulatory genes and markers is timely. There is ample evidence suggesting that many of the genes involved in endocrine pancreatic development also function in the homeostasis of the adult islet (Wells, 2003). It was our intention that a transcriptional profile of the developing endocrine pancreas would be an important resource for the diabetes research community. The genes identified in this study should facilitate analysis of the putative stem cells identified in the pancreatic ducts (Abraham et al., 2002; Cornelius et al., 1997; Ramiya et al., 2000; Zulewski et al., 2001). In addition, this catalog of signaling molecules and transcription factors expressed during endocrine development will expedite attempts to promote stem cell, embryonic or adult, differentiation into the islet cell lineage (Hori et al., 2002; Lumelsky et al., 2001).

Our temporal and spatial analysis of genes expressed in embryonic endocrine cells has generated a database of potential progenitor cell markers. For example, we have cross referenced our spatial and temporal analysis of genes expressed in E7.5 endoderm and identified 60 genes that were both spatially and temporally restricted to E7.5 endoderm (Table S1, http://dev.biologists.org/supplemental). These include genes of known endodermally expressed factors (*Sox17, Foxa2, Dkk1, Cer1*), and novel markers of endoderm (*Eya2*, cKit ligand, *Prss12*). Similar analyses revealed temporally and spatially restricted expression of genes in pancreatic and endocrine progenitor cells. We identified 16 genes that are highly enriched or only expressed in E10.5 $PDX1^+$ cells of the pancreatic rudiment (Table 4; Table S1, http://dev.biologists.org/supplemental), and 36 genes whose expression was temporally and spatially enriched in $NGN3^+$ endocrine precursors (Table 4; Table S1, http://dev.biologists.org/supplemental).

Thus far, we have not identified any genes that are exclusively restricted to developing endocrine cells. For example, *Eya2* and *Kit* ligand are expressed in E7.5 endoderm and in other tissues at later stages of development (Godin et

al., 1991; Motro et al., 1991; Xu et al., 1997). *Ngn3* and *Myt1* are both expressed in endocrine progenitor cells, as well as a set of neural progenitors in the developing nervous system (Apelqvist et al., 1999; Gradwohl et al., 2000; Kim et al., 1997b). It is possible that some of the ESTs in our database are truly expressed in a cell-specific manner. Alternatively, our results suggest that embryonic precursor cells seem to express many genes as a function of maintaining plasticity, where as adult islets expressed cell-type-specific genes. Nonetheless, the expression of a combination of several genes within each group may provide us with a diagnostic standard to determine whether cells are of endoderm, general pancreatic progenitor, endocrine precursors or mature islets.

Myt1 function might be necessary for endocrine islet development

Other than revealing general gene expression trends during islet development, our analysis also uncovered many candidate genes whose function could be required for islet development. One such example is *Myt1*. Our results suggested that the *NZF2b* isoform of *Myt1* promotes the formation of glucagon and somatostatin-expressing cells when ectopically expressed in chicken embryonic gut endoderm when it is expressed in developing endoderm (before endogenous pancreatic cells appear). MYT1a and NZF2b seemingly have different activity with regards to regulating the formation of hormone-expressing cells (insulin, glucagon and somatostatin). Differences in their activity could be due to differential stability of the proteins, or differences in post-translational modification or nuclear localization, or their different DNA binding activity.

We have also demonstrated that a dominant negative MYT1 partially inhibits endocrine cell differentiation in transgenic mouse embryos and efficiently inhibits NGN3 activity in chicken gut endoderm. Combine with the ectopic gene expression analysis, these results suggest that *Myt1* is involved in endocrine islet differentiation, and may function along the same pathway as NGN3. Although the β- and α-differentiation-inhibitory effect of dnMYT1 could result from the functional inhibition of other MYT1-like molecules, the other two *Myt1* homologues, *Myt1l* (Kim et al., 1997a) and *Myt3* (GeneBank acc. no.: BC032273) are not expressed in the developing pancreas (data not shown).

We are grateful to Kristin Stanley and Sara Rankin for assistance with Genespring and the Affymetrix software, and to Jennifer Kordich, Steve Potter, Maureen Gannon and Chris Wright for comments on the manuscript. This research was supported by grants from the Howard Hughes Medical Institute and the National Institutes of Health. J.M.W. was supported by a postdoctoral fellowship from Juvenile Diabetes Foundation.

References

Abraham, E. J., Leech, C. A., Lin, J. C., Zulewski, H. and Habener, J. F. (2002). Insulinotropic hormone glucagon-like peptide-1 differentiation of human pancreatic islet-derived progenitor cells into insulin-producing cells. *Endocrinology* **143**, 3152-3161.

Angelastro, J. M., Ignatova, T. N., Kukekov, V. G., Steindler, D. A., Stengren, G. B., Mendelsohn, C. and Greene, L. A. (2003). Regulated expression of ATF5 is required for the progression of neural progenitor cells to neurons. *J. Neurosci.* **23**, 4590-4600.

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.

Baugh, L. R., Hill, A. A., Brown, E. L. and Hunter, C. P. (2001). Quantitative analysis of mRNA amplification by in vitro transcription. *Nucleic Acids Res.* **29**, E29.

Beddington, R. S. and Robertson, E. J. (1999). Axis development and early asymmetry in mammals. *Cell* **96**, 195-209.

Bellefroid, E. J., Bourguignon, C., Hollemann, T., Ma, Q., Anderson, D. J., Kintner, C. and Pieler, T. (1996). X-MyT1, a Xenopus C2HC-type zinc finger protein with a regulatory function in neuronal differentiation. *Cell* **87**, 1191-1202.

Bouwmeester, T., Kim, S., Sasai, Y., Lu, B. and De Robertis, E. M. (1996). Cerberus is a head-inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature* **382**, 595-601.

Brissova, M., Shiota, M., Nicholson, W. E., Gannon, M., Knobel, S. M., Piston, D. W., Wright, C. V. and Powers, A. C. (2002). Reduction in pancreatic transcription factor PDX-1 impairs glucose-stimulated insulin secretion. *J. Biol. Chem.* **277**, 11225-11232.

Chiang, M. K. and Melton, D. A. (2003). Single-cell transcript analysis of pancreas development. *Dev. Cell* **4**, 383-393.

Conlon, F. L., Lyons, K. M., Takaesu, N., Barth, K. S., Kispert, A., Herrmann, B. and Robertson, E. J. (1994). A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**, 1919-1928.

Cornelius, J. G., Tchernev, V., Kao, K. J. and Peck, A. B. (1997). In vitro-generation of islets in long-term cultures of pluripotent stem cells from adult mouse pancreas. *Horm. Metab. Res.* **29**, 271-277.

Edlund, H. (1998). Transcribing pancreas. *Diabetes* **47**, 1817-1823.

Gannon, M., Herrera, P. L. and Wright, C. V. (2000). Mosaic Cre-mediated recombination in pancreas using the pdx-1 enhancer/promoter. *Genesis* **26**, 143-144.

Godin, I., Deed, R., Cooke, J., Zsebo, K., Dexter, M. and Wylie, C. C. (1991). Effects of the steel gene product on mouse primordial germ cells in culture. *Nature* **352**, 807-809.

Gradwohl, G., Dierich, A., LeMeur, M. and Guillemot, F. (2000). neurogenin3 is required for the development of the four endocrine cell lineages of the pancreas. *Proc. Natl. Acad. Sci. USA* **97**, 1607-1611.

Grapin-Botton, A., Majithia, A. R. and Melton, D. A. (2001). Key events of pancreas formation are triggered in gut endoderm by ectopic expression of pancreatic regulatory genes. *Genes Dev.* **15**, 444-454.

Gu, G., Dubauskaite, J. and Melton, D. A. (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. *Development* **129**, 2447-2457.

Gu, G., Brown, J. R. and Melton, D. A. (2003). Direct lineage tracing reveals the ontogeny of pancreatic cell fates during mouse embryogenesis. *Mech. Dev.* **120**, 35-43.

Guz, Y., Montminy, M. R., Stein, R., Leonard, J., Gamer, L. W., Wright, C. V. and Teitelman, G. (1995). Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. *Development* **121**, 11-18.

Harrison, S. M., Dunwoodie, S. L., Arkell, R. M., Lehrach, H. and Beddington, R. S. (1995). Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo. *Development* **121**, 2479-2489.

Heanue, T. A., Reshef, R., Davis, R. J., Mardon, G., Oliver, G., Tomarev, S., Lassar, A. B. and Tabin, C. J. (1999). Synergistic regulation of vertebrate muscle development by Dach2, Eya2, and Six1, homologs of genes required for Drosophila eye formation. *Genes Dev.* **13**, 3231-3243.

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

Herrera, P. L. (2000). Adult insulin- and glucagon-producing cells differentiate from two independent cell lineages. *Development* **127**, 2317-2322.

Herrera, P. L., Orci, L. and Vassalli, J. D. (1998). Two transgenic approaches to define the cell lineages in endocrine pancreas development. *Mol. Cell Endocrinol*. **140**, 45-50.

Hori, Y., Rulifson, I. C., Tsai, B. C., Heit, J. J., Cahoy, J. D. and Kim, S. K. (2002). Growth inhibitors promote differentiation of insulin-producing tissue from embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **99**, 16105-16110.

Johansson, B. M. and Wiles, M. V. (1995). Evidence for involvement of activin A and bone morphogenetic protein 4 in mammalian mesoderm and hematopoietic development. *Mol. Cell Biol.* **15**, 141-151.

Johnston, S. H., Rauskolb, C., Wilson, R., Prabhakaran, B., Irvine, K. D.

and Vogt, T. F. (1997). A family of mammalian Fringe genes implicated in boundary determination and the Notch pathway. *Development* **124**, 2245-2254.

Jonsson, J., Carlsson, L., Edlund, T. and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. *Nature* **371**, 606-609.

Kim, J. G., Armstrong, R. C., v Agoston, D., Robinsky, A., Wiese, C., Nagle, J. and Hudson, L. D. (1997a). Myelin transcription factor 1 (Myt1) of the oligodendrocyte lineage, along with a closely related CCHC zinc finger, is expressed in developing neurons in the mammalian central nervous system. *J. Neurosci. Res.* **50**, 272-290.

Kim, S. K. and Hebrok, M. (2001). Intercellular signals regulating pancreas development and function. *Genes Dev.* **15**, 111-127.

Kim, S. K., Hebrok, M. and Melton, D. A. (1997b). Pancreas development in the chick embryo. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 377-383.

Lammert, E., Cleaver, O. and Melton, D. (2001). Induction of pancreatic differentiation by signals from blood vessels. *Science* **294**, 564-567.

Lumelsky, N., Blondel, O., Laeng, P., Velasco, I., Ravin, R. and McKay, R. (2001). Differentiation of embryonic stem cells to insulin-secreting structures similar to pancreatic islets. *Science* **292**, 1389-1394.

Mathis, D., Vence, L. and Benoist, C. (2001). beta-Cell death during progression to diabetes. *Nature* **414**, 792-798.

Matsushita, F., Kameyama, T. and Marunouchi, T. (2002). NZF-2b is a novel predominant form of mouse NZF-2/MyT1, expressed in differentiated neurons especially at higher levels in newly generated ones. *Mech. Dev.* **118**, 209-213.

Miralles, F., Battelino, T., Czernichow, P. and Scharfmann, R. (1998a). TGF-beta plays a key role in morphogenesis of the pancreatic islets of Langerhans by controlling the activity of the matrix metalloproteinase MMP-2. *J. Cell Biol.* **143**, 827-836.

Miralles, F., Czernichow, P. and Scharfmann, R. (1998b). Follistatin regulates the relative proportions of endocrine versus exocrine tissue during pancreatic development. *Development* **125**, 1017-1024.

Motro, B., van der Kooy, D., Rossant, J., Reith, A. and Bernstein, A. (1991). Contiguous patterns of c-kit and steel expression: analysis of mutations at the W and Sl loci. *Development* **113**, 1207-1221.

Mukhopadhyay, M., Shtrom, S., Rodriguez-Esteban, C., Chen, L., Tsukui, T., Gomer, L., Dorward, D. W., Glinka, A., Grinberg, A., Huang, S. P. et al. (2001). Dickkopf1 is required for embryonic head induction and limb morphogenesis in the mouse. *Dev. Cell* **1**, 423-434.

Offield, M. F., Jetton, T. L., Labosky, P. A., Ray, M., Stein, R. W., Magnuson, M. A., Hogan, B. L. and Wright, C. V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* **122**, 983-995.

Perfetti, V., Vignarelli, M. C., Casarini, S., Ascari, E. and Merlini, G. (2001). Biological features of the clone involved in primary amyloidosis (AL). *Leukemia* **15**, 195-202.

Pictet, R. L., Clark, W. R., Williams, R. H. and Rutter, W. J. (1972). An ultrastructural analysis of the developing embryonic pancreas. *Dev. Biol.* **29**, 436-467.

Pignoni, F., Hu, B., Zavitz, K. H., Xiao, J., Garrity, P. A. and Zipursky, S. L. (1997). The eye-specification proteins So and Eya form a complex and regulate multiple steps in Drosophila eye development. *Cell* **91**, 881-891.

Ramalho-Santos, M., Melton, D. A. and McMahon, A. P. (2000). Hedgehog signals regulate multiple aspects of gastrointestinal development. *Development* **127**, 2763-2772.

Ramiya, V. K., Maraist, M., Arfors, K. E., Schatz, D. A., Peck, A. B. and Cornelius, J. G. (2000). Reversal of insulin-dependent diabetes using islets generated in vitro from pancreatic stem cells. *Nat. Med.* **6**, 278-282.

Ridgeway, A. G. and Skerjanc, I. S. (2001). Pax3 is essential for skeletal myogenesis and the expression of Six1 and Eya2. *J. Biol. Chem.* **276**, 19033-19039.

Scearce, L. M., Brestelli, J. E., McWeeney, S. K., Lee, C. S., Mazzarelli, J., Pinney, D. F., Pizarro, A., Stoeckert, C. J., Jr, Clifton, S. W., Permutt, M. A. et al. (2002). Functional genomics of the endocrine pancreas: the pancreas clone set and PancChip, new resources for diabetes research. *Diabetes* **51**, 1997-2004.

Schwitzgebel, V. M., Scheel, D. W., Conners, J. R., Kalamaras, J., Lee, J. E., Anderson, D. J., Sussel, L., Johnson, J. D. and German, M. S. (2000). Expression of neurogenin3 reveals an islet cell precursor population in the pancreas. *Development* **127**, 3533-3542.

Sharma, A., Zangen, D. H., Reitz, P., Taneja, M., Lissauer, M. E., Miller, C. P., Weir, G. C., Habener, J. F. and Bonner-Weir, S. (1999). The homeodomain protein IDX-1 increases after an early burst of proliferation during pancreatic regeneration. *Diabetes* **48**, 507-513.

Shimizu, K., Chiba, S., Saito, T., Kumano, K., Takahashi, T. and Hirai, H. (2001). Manic fringe and lunatic fringe modify different sites of the Notch2 extracellular region, resulting in different signaling modulation. *J. Biol. Chem.* **276**, 25753-25758.

Slack, J. M. (1995). Developmental biology of the pancreas. *Development* **121**, 1569-1580.

Stride, A. and Hattersley, A. T. (2002). Different genes, different diabetes: lessons from maturity-onset diabetes of the young. *Ann. Med.* **34**, 207-216.

Tisch, R. and McDevitt, H. (1996). Insulin-dependent diabetes mellitus. *Cell* **85**, 291-297.

Ueda, S., Mizuki, M., Ikeda, H., Tsujimura, T., Matsumura, I., Nakano, K., Daino, H., Honda Zi, Z., Sonoyama, J., Shibayama, H. et al. (2002). Critical roles of c-Kit tyrosine residues 567 and 719 in stem cell factor-induced chemotaxis: contribution of src family kinase and PI3-kinase on calcium mobilization and cell migration. *Blood* **99**, 3342-3349.

van Schaick, H. S., Smidt, M. P., Rovescalli, A. C., Luijten, M., van der Kleij, A. A., Asoh, S., Kozak, C. A., Nirenberg, M. and Burbach, J. P. (1997). Homeobox gene Prx3 expression in rodent brain and extraneural tissues. *Proc. Natl. Acad. Sci. USA* **94**, 12993-12998.

Warnock, G. L., Kneteman, N. M., Evans, M. G. and Rajotte, R. V. (1990). Isolation of purified large mammal and human islets of Langerhans. *Horm. Metab. Res.* **Supplement 25**, 37-44.

Wells, J. M. (2003). Genes expressed in the developing endocrine pancreas and their importance for stem cell and diabetes research. *Diabetes Metab. Res. Rev.* **19**, 191-201.

Wells, J. M. and Melton, D. A. (1999). Vertebrate endoderm development. *Annu. Rev. Cell. Dev. Biol.* **15**, 393-410.

Wells, J. M. and Melton, D. A. (2000). Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* **127**, 1563-1572.

Wilkinson, D. G. and Nieto, M. A. (1993). Detection of messenger RNA by ISH to tissue sections and whole mounts. *Methods Enzymol.* **225**, 361-373.

Wu, K. L., Gannon, M., Peshavaria, M., Offield, M. F., Henderson, E., Ray, M., Marks, A., Gamer, L. W., Wright, C. V. and Stein, R. (1997). Hepatocyte nuclear factor 3beta is involved in pancreatic beta-cell-specific transcription of the pdx-1 gene. *Mol. Cell Biol.* **17**, 6002-6013.

Xu, P. X., Woo, I., Her, H., Beier, D. R. and Maas, R. L. (1997). Mouse Eya homologues of the Drosophila eyes absent gene require Pax6 for expression in lens and nasal placode. *Development* **124**, 219-231.

Zulewski, H., Abraham, E. J., Gerlach, M. J., Daniel, P. B., Moritz, W., Muller, B., Vallejo, M., Thomas, M. K. and Habener, J. F. (2001). Multipotential nestin-positive stem cells isolated from adult pancreatic islets differentiate ex vivo into pancreatic endocrine, exocrine, and hepatic phenotypes. *Diabetes* **50**, 521-533.

Table S1. Lists of genes temporally enriched in each of the four tissue samples respectively (genes without an asterisk(*) are both temporally and spatially enriched. An * denotes that a gene is expressed in other tissues at this stage of development)

1a: list of genes enriched in E7 endoderm relative to cells from other stages of endocrine pancreas development

Growth factors, receptors and signaling molecules

Bcap37*	NM_007531	B-cell receptor-associated protein 37
Cer1	NM_009887	cerberus 1 homolog
Dab2	NM_023118	disabled homolog 2
Dkk1	XM_123389	dickkopf homolog 1
Ddt*	AF068199	macrophage migration inhibitory factor homolog
Edg2	NM_010336	Endothelial differentiation related receptor 2
Epha2	NM_010139	Eph receptor A2
Eya2	AI327163	eyes absent 2 homolog
Fgfbp1	U49641	fibroblast growth factor binding protein 1
Folr1	NM_008034	folate receptor 1
Fst*	NM_008046	follistatin
Igf2bp1*	NM_009951	insulin-like growth factor 2, binding protein 1
Igfbp2*	NM_008342	insulin-like growth factor binding protein 2
Int-2*	Y00848	int-2
Kitl	NM_013598	kit ligand
Mlf1*	NM_010801	myeloid leukemia factor 1
Nodal*	X70514	nodal
Rbp1*	NM_011254	retinol binding protein 1, cellular
S100a10*	NM_009112	S100 calcium binding protein A10 (calpactin)
ScyA5	U07602	small inducible cytokine A5 (ScyA5)
Tacstd2	NM_020047	tumor-associated calcium signal transducer 2
Ttk	NM_009445	Ttk protein kinase
Traf4*	NM_009423	Tnf receptor associated factor 4
Trh	NM_009426	Thyrotropin releasing hormone
Tyro3*	U18343	TYRO3 protein tyrosine kinase 3

Transcriptional regulators and other nuclear factors

Cdx-1	M80463	caudal type homeobox-1
Dri2*	NM_019689	dead ringer homolog 2
Eif4ebp1*	AK013033	eukaryotic translation initiation factor 4E binding protein 1
Eomes*	AB031037	eomesodermin homolog
Foxa1	NM_008259	forkhead box A1
Foxb1*	X92592	forkhead box B1
Foxc2*	NM_013519	forkhead box C2
Foxd4	X86368	forkhead box D4
Foxj1*	L13204	forkhead box J1
Gata3*	NM_008091	GATA binding protein 3
Hist1*	NM_013548	histone gene complex 1
H2A*	Z35401	histone H2A.X.
H3*	M32459	histone H3
H4*	Y12290	histone H4
Hesx1*	NM_010420	homeo box gene expressed in ES cells
hox-2.9*	X53063	Hox-2.9 homeodomain protein
Idb1*	AK008264	inhibitor of DNA binding 1
Idb3*	NM_008321	inhibitor of DNA binding 3
Irx3*	NM_008393	Iroquois related homeobox 3
junB	U20735	junB transcription factor
Khdrbs2*	NM_133235	KH domain containing, RNA binding, signal transduction associated 2
Lmo4*	NM_010723	LIM only 4
Mesp1*	D83674	mesoderm posterior 1
Mlf1	NM_010801	myeloid leukemia factor 1
Msx1	NM_010835	msh-like 1 homeo box
Mybbp1a*	NM_016776	MYB binding protein (P160) 1a
Mybl2*	NM_008652	myeloblastosis oncogene-like 2
Na*	BC002306	PROBABLE DNA-DIRECTED RNA POLYMERASES I, II, AND III
Pem	NM_008818	placentae and embryos oncofetal homeobox
ppan*	BC014688	peter pan homolog
PolI*	BC002306	RNA POLYMERASES I
Np95*	NM_010931	nuclear protein 95
Polg*	NM_017462	polymerase, gamma
Rbpms*	BC011288	RNA binding protein gene with multiple splicing
Rnf10*	NM_016698	ring finger protein 10
Six1	X80339	sine oculis-related homeobox 1 homolog (Drosophila)
Snrpd*2	AK007389	small nuclear ribonucleoprotein D2
Snrpd2*	AK007389	small nuclear ribonucleoprotein D2
Sox17	AK004781	SRY-box containing gene 17
T	NM_009309	T brachyury
TAFII30*	AJ249987	mTAFII30 protein

Tcf7*	NM_009331	transcription factor 7, T-cell specific
Tcfap2c*	NM_009335	transcription factor AP-2, gamma
Zic3*	NM_009575	zinc finger protein of the cerebellum 3

Cell surface/adhesion/matrix/cytoskeletal proteins

Actc1*	NM_009608	actin, alpha, cardiac
Actn4*	NM_021895	actinin alpha 4
Adam19	D50410	a disintegrin and metalloproteinase domain 19 (meltrin beta)
ColIXa2*	M63649	M.musculus alpha2 (IX) collagen
Cnn2	NM_007725	calponin 2
Cubn	AF197159	cubilin (intrinsic factor-cobalamin receptor)
Emb*	NM_010330	embigin
EndoA	X15662	cytokeratin endo A
EndoB*	M22832	cytokeratin (endoB)
Gja1*	BC006894	gap junction membrane channel protein alpha 1
Gjb3	NM_008126	gap junction membrane channel protein beta 3
Gp38*	NM_010329	glycoprotein 38
k19	M36120	keratin 19
Klk8*	NM_008940	kallikrein 8
Lamb3	NM_008484	laminin, beta 3
Mylc2a*	NM_022879	myosin light chain, regulatory A
Nid2	NM_008695	nidogen 2
Pls2	BC022943	plastin 2, L
Tm4sf7	NM_053082	transmembrane 4 superfamily member 7
Vil2*	NM_009510	villin 2

Cell death

Tdag	NM_009344	T-cell death associated gene
Tssc3	NM_009434	tumor surpressing subdhromosomal transferable fragment 3

Transport proteins

Apoa4	NM_007468	apolipoprotein A-IV
Apom	NM_018816	apolipoprotein M
Rbp1*	NM_011254	retinol binding protein 1, cellular
Slc19a1*	NM_031196	solute carrier family 19 (sodium/hydrogen exchanger), member 1
Slc2a1*	M23384	solute carrier family 2 (facilitated glucose transporter), member 1
Slc7a7	NM_011405	solute carrier family 7 (cationic amino acid transporter, y+ system)
Slc9a3r1*	NM_012030	solute carrier family 9 (sodium/hydrogen exchanger), isoform 3 regulator 1
Tomm40	NM_016871	translocase of outer mitochondrial membrane 40 homolog

Hormones and proteases and secretory apparatus

Adam19	D50410	a disintegrin and metalloproteinase domain 19 (meltrin beta)
CA-IV	U37091	carbonic anhydrase IV
Ctsc	NM_009982	cathepsin C
Ctsh	NM_007801	cathepsin H
F10	NM_007972	coagulation factor X
Klk8	NM_008940	kallikrein 8
MT-II	K02236	metallothionien II
Prss12	NM_008939	protease, serine, 12 neurotrypsin, (motopsin)
Trh	NM_009426	thyrotropin releasing hormone
Tmprss2	NM_015775	transmembrane protease, serine 2

Others

Abcd4*	NM_008992	ATP-binding cassette, sub-family D (ALD), member 4
Acate2*	NM_019736	acyl-Coenzyme A thioesterase 2, mitochondrial
Ard1*	NM_019870	N-acetyltransferase ARD1 homolog
Arsa	X73230	arylsulfatase A
Bcat1*	AK013888	branched chain aminotransferase 1, cytosolic
Bcrp1*	NM_011793	breakpoint cluster region protein 1
Birc5*	BC004702	baculoviral IAP repeat-containing 5
Calca*	BC028771	calcitonin/calcitonin-related polypeptide, alpha
Cyp26a1	NM_007811	cytochrome P450, 26, retinoic acid A1
Dgcr6	AU044152	DiGeorge syndrome critical region 6
Ehd1*	AF099186	EH-domain containing 1
Eno1*	NM_023119	enolase 1, alpha non-neuron
Esg1*	NM_025274	embryonal stem cell specific gene 1
Gfpt2*	NM_013529	glutamine fructose-6-phosphate transaminase 2
Glk*	NM_016905	galactokinase
Gpi1*	L09104	glucose phosphate isomerase 1
Hdc	NM_008230	histidine decarboxylase
HkII*	Y11666	hexokinase II
Ier2*	NM_010499	immediate early response 2
Lerepol	AV025472	immediate early response, erythropoietin 1
Ldh1*	NM_010699	lactate dehydrogenase 1, A chain

Lgmn	NM_011175	legumain
Ltb	NM_008518	lymphotoxin B
Mcmd*	X62154	mini chromosome maintenance deficient
MT-2*	K02236	metallothionien II
Mvk*	NM_023556	mevalonate kinase
Noc4*	NM_010926	neighbor of Cox4
Nudt1*	NM_008637	nudix (nucleoside diphosphate linked moiety X)-type motif 1
Oat*	NM_016978	ornithine aminotransferase
Pdxk	BC027745	pyridoxal kinase
Pdzk1	NM_021517	PDZ domain containing 1
Pk3*	NM_011099	pyruvate kinase 3
Pls2*	BC022943	plastin 2, L
Ptdss2*	NM_013782	phosphatidylserine synthase 2
Ptov1*	NM_133949	prostate tumor over expressed gene 1
Ris2*	NM_026014	retroviral integration site 1
Rpl36*	NM_018730	ribosomal protein L36
Soat2	NM_011433	sterol O-acyltransferase 2
Srm*	Z67748	M. musculus spermidine synthase gene
Tex19	NM_028602	testis expressed gene 19
Tex20*	NM_029395	testis expressed gene 20
Tex271*	BC010716	testis expressed gene 271
Thop1*	NM_022653	thimet oligopeptidase 1
Tkt*	NM_009388	transketolase
Trap1a	NM_011635	tumor rejection antigen P1A
Trh	NM_009426	thyrotropin releasing hormone
Upp*	NM_009477	uridine phosphorylase
Ywhaz*	NM_011740	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein

ESTs

1010001J12Rik
1110032C13Rik
1110033J19Rik
1110061O04Rik
1200008D14Rik
1600025H15Rik
1600029D21Rik
1810035L17Rik
2310008B08Rik
2310076D10Rik
2410004C24Rik
2410026K10Rik
2410071B14Rik
2410130M07Rik
2610012O22Rik
2700066J21Rik
3110023F10Rik
3300001G02Rik
4833422P03Rik
4933419D20Rik
5730408K05Rik
5730591C18Rik
6720465F12Rik
AA408729
AA755260
AI315052
AI327276
AI467657
BF607517
BQ180406
C81439
D15Wsu97e
D19Wsu57e
D2Wsu23e
D3Wsu161e
D7Wsu180e

1b: genes enriched in mPDX1+ cells

Growth factors, receptors and signaling molecules

Basp2*	NM_008083	brain abundant, membrane attached signal protein 2
Calmbp1	NM_009791	calmodulin binding protein 1
Dlk1	NM_010052	delta-like 1 homolog
Gas1	NM_008086	growth arrest specific 1
Igfbp5*	L12447	insulin-like growth factor binding protein 5
Sfrp1*	NM_013834	secreted frizzled-related sequence protein 1

Transcriptional regulators and other nuclear factors

Barx1	NM_007526	BarH-like homeobox 1
Mfap2	NM_008546	microfibrillar-associated protein 2
Meis1	NM_010789	myeloid ecotropic viral integration site 1
Myb	NM_033597	myeloblastosis oncogene
Nkx6-2	AK008173	NK transcription factor related, locus 2
Onecut1	NM_008262	one cut domain, family member 1
Snai2*	NM_011415	snail homolog 2
Sox11*	NM_009234	SRY-box containing gene 11
Sp4	NM_009239	trans-acting transcription factor 4
Tcf21*	NM_011545	transcription factor 21
Zfp125	AJ005350	zinc finger protein 125
Zac1*	AF147785	zinc finger protein regulator of apoptosis and cell cycle arrest

Cell surface/adhesion/matrix/cytoskeletal proteins

Dnahc11*	NM_010060	dynein, axonemal, heavy chain 11
Gja7	NM_008122	gap junction membrane channel protein alpha 7
Itm2a*	NM_008409	integral membrane protein 2A
Mfap2*	NM_008546	microfibrillar-associated protein 2
Col1a1*	U08020	procollagen, type I, alpha 1
Col1a2*	NM_007743	procollagen, type I, alpha 2
Col5a2*	NM_007737	procollagen, type V, alpha 2
Tac2	NM_009312	tachykinin 2
Thy1.2	X99915	Thy-1.2 glycoprotein gene
Tnc*	NM_011607	tenascin C
Vnn1	NM_011704	vanin 1

Cell cycle related

CDC7	AB019388	gene for muCdc7
Cdkn1c*	NM_009876	cyclin-dependent kinase inhibitor 1C (P57)
Mki67*	X82786	antigen identified by monoclonal antibody Ki 67

Transport proteins

Kcnj8*	NM_008428	potassium inwardly-rectifying channel, subfamily J, member 8

Hormones proteases and secretory apparatus

Capn6	NM_007603	calpain 6

Others

Agc1	NM_007424	aggrecan 1
Hba-x*	NM_010405	hemoglobin X, alpha-like embryonic chain in Hba complex
Hbb-y*	NM_008221	hemoglobin Y, beta-like embryonic chain
Gld1	AF039021	glucosidase I gene
Peg1/MEST	AF017994	Peg1/MEST protein
Mgst1	NM_019946	microsomal glutathione S-transferase 1
Nr2f1	NM_010151	nuclear receptor subfamily 2, group F, member 1
Peg3	AB003040	paternally expressed 3
Ptn	NM_008973	pleiotrophin
Sez6	NM_021286	seizure related gene 6
Unc5h3	NM_009472	unc5 homolog (C. elegans) 3

ESTs

1200003E16Rik
1700066C05Rik
2410012A13Rik
2410012A13Rik
2610042L04Rik
2610042L04Rik
2700038M07Rik
6330403K07Rik
AA987150
AF039021
AU016842
AU020206
AW121776
C76256
V00714

1c: genes enriched Ngn3+ cells versus pancreatic tissue of other stages

Growth factors. Receptors and signaling molecules

Alk-6	L35029	ALK-6 mRNA
DLK2C	U14636	serine/threonine protein kinase DLK
Fosb	NM_008036	FBJ osteosarcoma oncogene B

Gfra3	NM_010280	glial cell line derived neurotrophic factor family receptor alpha 3
Gyk*	NM_008194	glycerol kinase
Mfng	NM_008595	manic fringe homolog
pim-2	L41495	protein-serine/threonine kinase (pim-2)
PIRA1	Y13361	immunoglobulin-like receptor
Plc2C	AB007798	protein phosphatase 2C beta

Transcriptional regulators and other nuclear factors

Brn2	M88301	brain-2 class III POU-domain protein
Isl1	NM_021459	ISL1 transcription factor, LIM/homeodomain
L-myc*	X13945	Murine L-myc gene
Mafb	L36434	v-maf musculoaponeurotic fibrosarcoma oncogene family, protein B
Myt1	NM_008665	myelin transcription factor 1
Neurod1	AK005073	neurogenic differentiation 1
Pax4	AF031150	paired box gene 4
Rpo1-1	NM_009085	RNA polymerase 1-1

Cell surface/adhesion/matrix/cytoskeletal proteins

Ap1g1	NM_009677	adaptor protein complex AP-1, gamma 1 subunit
Col9a3	AF349718	procollagen, type IX, alpha 3
Mtap1b	NM_008634	microtubule-associated protein 1B

Cell cycle related

Cdkn1a	AK007630	cyclin-dependent kinase inhibitor 1A (P21)
GAD-45*	U00937	GADD45 protein (gadd45)

Hormones, proteases and secretory apparatus

Anxa4*	NM_013471	annexin A4
Cfi	NM_007686	complement component factor I
Gast	NM_010257	gastrin

Others

Adfp	NM_007408	adipose differentiation related protein
Apg12l	NM_026217	autophagy 12-like (S. cerevisiae)
Cbfa2t1h*	NM_009822	CBFA2T1 identified gene homolog (human)
Epb4.2	NM_013513	erythrocyte protein band 4.2
Evi-2	Z22923	Mouse ecotropic viral integration site 2 (Evi-2) ORF
Gip*	NM_008119	gastric inhibitory polypeptide
Glud	NM_008133	glutamate dehydrogenase
Nedd4b*	NM_031881	neural precursor cell expressed, developmentally down-regulated gene 4b
Ogn	NM_008760	osteoglycin
Pik3r1	U50413	phosphatidylinositol 3-kinase, regulatory subunit, polypeptide 1
Prrg2	AK005453	proline-rich Gla (G-carboxyglutamic acid) polypeptide 2
Glk*	L41631	putative; Mus musculus glucokinase gene, complete cds.
Rhomr1	NM_139228	rhomboid related gene 1
Selenbp2	NM_019414	selenium binding protein 2
Siat10	NM_018784	sialyltransferase 10 (alpha-2,3-sialyltransferase VI)
Sudd	AK004748	sudD, suppressor of bimD6 homolog (Aspergillus nidulans)
Tekt2	NM_011902	tektin 2
Tieg	NM_013692	TGFB inducible early growth response
Txnip*	NM_023719	thioredoxin interacting protein
TIS21	M64292	TIS21; Mouse TIS21 gene, complete cds
Usp18	NM_011909	ubiquitin specific protease 18

ESTs

1700031C13Rik
2210012L08Rik
2310061B02Rik
2610301I15Rik
2810011G06Rik
2810011G06Rik
3830417M17Rik
4933407C03Rik
5830427H10Rik
AA189214
AA407619
AF035263
AI425994
AI448995
AW553439
BE283402
BF168859
BF581594
BI155328
C81234

C81612
D6Ertd365e
L35029
NM_007408
U01170
X04120
Z22923

1d: genes enriched in mature islets versus pancreatic tissue of early stages

Growth factors, receptors and signaling molecules

egf	V00741	polyprotein; epidermal growth factor
notch4*	AF030001	Notch4
Gnat2	NM_008141	guanine nucleotide binding protein, alpha transducing 2
Ikbke	NM_019777	inhibitor of kappaB kinase epsilon
Il1r1	NM_008362	interleukin 1 receptor, type I
Il6ra	X53802	interleukin 6 receptor, alpha
Inha	NM_010564	inhibin alpha
Lifr	NM_013584	leukemia inhibitory factor receptor
Mapk10	NM_009158	mitogen activated protein kinase 10
P2ry1	NM_008772	purinergic receptor P2Y, G-protein coupled 1
Pip5k1a	NM_008846	phosphatidylinositol-4-phosphate 5-kinase, type 1 alpha
Prkca	NM_011101	protein kinase C, alpha
Prkcz	NM_008860	protein kinase C, zeta
Prlr	NM_011169	prolactin receptor
Procr	NM_011171	protein C receptor, endothelial
Ptprn	NM_008985	protein tyrosine phosphatase, receptor-type, N
Rgs11	AF061934	regulator of G-protein signaling 11
Rgs2*	NM_009061	regulator of G-protein signaling 2
Scya19	NM_011888	small inducible cytokine A19
Thra	NM_011584	thyroid hormone receptor alpha
Tie1*	NM_011587	tyrosine kinase receptor 1

Transcriptional regulators and other nuclear factors

Ang	U22516	angiogenin precursor (Ang) ribonuclease
Atf5	NM_030693	activating transcription factor 5
Creg	AK010947	cellular repressor of E1A-stimulated genes
Cpeb	NM_007755	cytoplasmic polyadenylation element binding protein
Ell	NM_138953	ELL-related RNA polymerase II, elongation factor
Myt1l	NM_008666	myelin transcription factor 1-like
Nr1h3	NM_013839	nuclear receptor subfamily 1, group H, member 3
Nucb2	NM_016773	nucleobindin 2
Nupr1	NM_019738	nuclear protein 1
STAT5B	AJ237939	Mus musculus partial STAT5B gene, exons 6-9

Cell surface/adhesion/matrix/cytoskeletal proteins

Anxa1	NM_010730	annexin A1
Cd36l2	NM_007644	CD36 antigen-like 2
Cd84	NM_013489	CD84 antigen
Clu	NM_013492	clusterin
Crpd	NM_007769	crp-ductin
Ctnnd2	NM_008729	catenin delta 2
Emcn	NM_016885	endomucin
Epb7.2t	NM_013515	erythrocyte protein band 7.2
Grn	NM_008175	granulin
Hb5	AF029982	keratin Hb5; Mus musculus hair keratin basic 5 (Hb5)
Itga7	NM_008398	integrin alpha 7
Itmap1t	NM_008411	integral membrane-associated protein 1
Lamp2	NM_010685	lysosomal membrane glycoprotein 2
Ldc	AF013262	KSPG; corneal clarity protein
Lgals6	AF026799	galectin-6 (Lgals6) gene, exons 7-8
Lipo 1	M69260	lipocortin I gene, exon 13
Lman1	NM_027400	lectin, mannose-binding, 1
Ly64	NM_010739	lymphocyte antigen 64
Ly6a	NM_010738	lymphocyte antigen 6 complex, locus A
Mglap	BC009120	matrix gamma-carboxyglutamate (gla) protein
Myh11	NM_013607	myosin heavy chain 11, smooth muscle
rab7*	D14011	Rab7 pseudogene.
Sema6a	NM_018744	sema domain, transmembrane domain (TM), and cytoplasmic domain
syngr4	D78382	Mus musculus mRNA for synaptogyrin 4 protein.
Syt4	U10355	synaptotagmin 4
Sytl4	NM_013757	synaptotagmin-like 4
Tgoln2	NM_009444	trans-golgi network protein 2
THBS1	M62470	thrombospondin (THBS1)
Thbd	NM_009378	thrombomodulin
Vsnl1	NM_012038	visinin-like 1

Transport proteins

Apobec1	NM_031159	apolipoprotein B editing complex 1
Fabp1	NM_017399	fatty acid binding protein 1, liver
Fabp2	NM_007980	fatty acid binding protein 2, intestinal
Glrb	NM_010298	glycine receptor, beta subunit
Slc25a5	BF384626	solute carrier family 25, member 5
Slc2a2*	NM_031197	solute carrier family 2 member 2
Slc7a8*	NM_016972	solute carrier family 7, member 8
Snap25	NM_011428	synaptosomal-associated protein, 25 kDa

Hormones proteases and secretory apparatus

Amy1	NM_007446	amylase 1, salivary
Amy2*	NM_009669	amylase 2, pancreatic
Chga*	NM_007693	chromogranin A
Ctrl	NM_023182	chymotrypsin-like
Dnajb9	NM_013760	DnaJ (Hsp40) homolog, subfamily B, member 9
Ela1*	AK007392	elastase 1, pancreatic
Ela2	NM_007919	elastase 2
Gcg*	AF276754	glucagon
GK-6	M13500	mGK-6 kallikrein
Hgfac	NM_019447	hepatocyte growth factor activator
Iapp	NM_010491	islet amyloid polypeptide
Itih1	NM_008406	inter-alpha trypsin inhibitor, heavy chain 1
Itih4	NM_018746	inter alpha-trypsin inhibitor, heavy chain 4
Klk16	AB039276	kallikrein 21
Klk26	NM_010644	kallikrein 26
Klk5	NM_008456	kallikrein 5
Klk9	NM_010116	kallikrein 9
klk22	V00829	kallikrein epidermal growth factor binding protein type 1
Pcsk1*	NM_013628	proprotein convertase subtilisin/kexin type 1
Pcsk1n*	NM_013892	proprotein convertase subtilisin/kexin type 1 inhibitor
Pcsk2	NM_008792	proprotein convertase subtilisin/kexin type 2
Pgcp-pending	NM_018755	plasma glutamate carboxypeptidase
Ppicap	NM_011150	peptidylprolyl isomerase C-associated protein
Ppy	NM_008918	pancreatic polypeptide
Psmb8	NM_010724	proteosome (prosome, macropain) subunit, beta type 8
Psmb9	NM_013585	proteosome (prosome, macropain) subunit, beta type 9
Ribl	X60103	Ribonuclease, pancreatic
Scg3	NM_009130	secretogranin III
SgII	X68837	M. musculus SgII gene for secretogranin II, exon 2
Sgne1	NM_009162	secretory granule neuroendocrine protein 1, 7B2 protein
Try2	NM_009430	trypsin 2

Cell death

Cpp32	X16202	CPP32 apoptotic protease

Others

Abcb3	NM_011530	ATP-binding cassette, sub-family B (MDR/TAP), member 3
aj4	AJ235940	IgVk aj4 gene
AL033311	BC003898	NADH dehydrogenase (ubiquinone) Fe-S protein 2
Alox8	C85523	M arachidonate 8(S)-lipoxygenase
AU01957	BC005577	MSimilar to hypothetical protein FLJ11110
AU022351	D26352	cytochrome c gene (MC1)
bdfn1	AF003525	beta-defensin 1
C1qb*	NM_009777	complement component 1, q subcomponent, beta polypeptide
C3	K02782	complement component 3
Cel*	NM_009885	carboxyl ester lipase
Chic1	Y11896	cysteine-rich hydrophobic domain 1
Cox6a2	NM_009943	cytochrome c oxidase, subunit VI a, polypeptide 2
Cyp3a11	AK004861	cytochrome P450, steroid inducible 3a11
D14Ertd813e	BC016131	CG8726 gene product
Ddc	AF071068	dopa decarboxylase
Dio1	NM_007860	deiodinase, iodothyronine, type I
Enpep	NM_007934	glutamyl aminopeptidase
Entpd1	NM_009848	ectonucleoside triphosphate diphosphohydrolase 1
Fkbp1b	NM_016863	FK506 binding protein 1b (12.6 kDa)
G6pc-rs	NM_021331	glucose-6-phosphatase, catalytic, related sequence
Gad1	NM_008077	glutamic acid decarboxylase 1
Gatm*	NM_025961	glycine amidinotransferase (L-arginine:glycine amidinotransferase)
Gdap1	NM_010267	ganglioside-induced differentiation-associated-protein 1
Ggh	NM_010281	gamma-glutamyl hydrolase
H2-Aa	NM_010378	histocompatibility 2, class II antigen A, alpha
H2-Ab1	NM_010379	histocompatibility 2, class II antigen A, beta 1
H2-D1	NM_010380	histocompatibility 2, D region locus 1

H-2KB	V00746	H-2K gene for MHC class I antigen H-2KB
H2-L	M34961	histocompatibility 2, L region
Hadhsc	NM_008212	L-3-hydroxyacyl-Coenzyme A dehydrogenase, short chain
Herpud1	NM_022331	homocysteine-inducible, ubiquitin-like domain member 1
Ifi203	NM_008328	interferon activated gene 203
Ifi204	NM_008329	interferon activated gene 204
Iga	J00475	Ig germline D-J-C region alpha gene and secreted tail
Igfals	U66900	cell membrane ecto-ATPase
Igk-V28	U68543	immunoglobulin kappa chain variable 28 (V28)
IgM	M80423	IgK chain gene, C-region, 3′ end
Ii	BC003476	Ia-associated invariant chain
Kl	NM_013823	klotho
LOC56628	NM_019909	MHC (A.CA/J(H-2K-f) class I antigen
Lpl*	NM_008509	lipoprotein lipase
Ly9	NM_008534	lymphocyte antigen 9
Lyzs	M21050	lysozyme M gene, exon 4
Msln	NM_018857	mesothelin
mSP	U78770	trefoil protein
Ndr4	BC006595	N-myc downstream 4
Npn3	BF120858	neoplastic progression 3
Pah*	NM_008777	phenylalanine hydroxylase
Pap	NM_011036	pancreatitis-associated protein
Papss2	NM_011864	3′-phosphoadenosine 5′-phosphosulfate synthase 2
Pcnt	NM_008787	pericentrin
Pcp4*	NM_008791	Purkinje cell protein 4
Pdyn	AF026537	prodynorphin
Pftk1	NM_011074	PFTAIRE protein kinase 1
Pnliprp1*	NM_018874	pancreatic lipase related protein 1
Pnliprp2	NM_011128	pancreatic lipase-related protein 2
Prp	M23236	precursor; Mouse proline-rich protein (M14) gene
Reg1	NM_009042	regenerating islet-derived 1
Reg3a	NM_011259	regenerating islet-derived 3 alpha
Reg3g	D63362	Regenerating islet derived 3 gama; mRNA for p80 Ku autoantigen
SPRR2A	AJ005559	Mus Musculus SPRR2A gene
Sqrdl	NM_021507	sulfide quinone reductase-like
Tapbp	BC015074	TAP binding protein
TCRBV8S3	AE000664	Cluster Incl AE000664:Mus musculus TCR beta locus
Thea	AF416923	thioesterase, adipose associated

ESTs

AB001489
AC002397
AI327354
AI447560
AI461839
AI604013
AI604013
AI848610
AI848715
AI893437
AL033311
AU019574
AU022351
AW061306
AW109744
AW122677
BB646062
BC031768
BF534966
BI697602
BI729512
C76990
C78582
C78788
C81363
M22679
V00829
X51468
0910001K16Rik
1110003E01Rik
1110013J02Rik
1200002G13Rik
1200015P04Rik
1300007C21Rik
1300011P19Rik

1700015E05Rik
1810004P07Rik
1810010M01Rik
1810015C04Rik
1810049E02Rik
2010305L05Rik
2200003J09Rik
2210416J16Rik
2300002F06Rik
2310016A09Rik
2610102M01Rik
2900019I22Rik
5730403E06Rik
5830406C15Rik
9030401P18Rik
9030401P18Rik

Table S2. Genes down-regulated as endocrine cells develop

Growth factors, receptors and signaling molecules		
Borg4	NM_020006	binder of Rho GTPase 4
Dll1	NM_007865	delta-like 1 (Drosophila)
Fgfr3	M81342	fibroblast growth factor receptor 3
Fzd2	NM_020510	frizzled homolog 2 (Drosophila)
Gas11	NM_018855	growth arrest specific 11
Gnb2-rs1	NM_008143	guanine nucleotide binding protein, beta 2, related sequence 1
Igf2bp1	NM_008341	insulin-like growth factor binding protein 1
Il1a	X01450	Interleukin 1 alpha
Mlp	NM_010807	MARCKS-like protein
notch4	AF030001	Notch4
Plk	NM_011121	polo-like kinase homolog, (Drosophila)
Ppp2r4	NM_138748	protein phosphatase 2A, regulatory subunit B (PR 53)
Ran	NM_009391	RAN, member RAS oncogene family
Shd	NM_009168	src homology 2 domain-containing transforming protein D
Smoh	AF089721	smoothened homolog (Drosophila)
Tpbg	AJ012160	-13 receptor alpha 2 mRNA, complete cds.
Tyro3	U18343	TYRO3 protein tyrosine kinase 3
Transcriptional regulators and other nuclear factors		
Chaf1a	NM_013733	chromatin assembly factor 1, subunit A (p150)
PBX1B	M33988	PBX1B mRNA, complete cds.
Pole2	AF036898	DNA polymerase epsilon, subunit 2
Rpl18	NM_009077	ribosomal protein L18
Rpl8	NM_012053	ribosomal protein L8
Rrm1	NM_009103	ribonucleotide reductase M1
Rxra	NM_011305	retinoid X receptor alpha
Snai1	NM_011427	snail homolog 1 (Drosophila)
Cell surface/adhesion/matrix proteins		
Becn1	NM_019584	beclin 1 (coiled-coil, myosin-like BCL2-interacting protein)
Coil	NM_016706	coilin
Col18a1	NM_009929	procollagen, type XVIII, alpha 1
Cryz	S70056	crystallin, zeta
Eda	NM_010099	ectodysplasin-A
Fn1	BC025521	fibronectin 1
Gpc1	NM_016696	glypican 1
Hyal2	NM_010489	hyaluronidase 2
Krt1-18	NM_010664	keratin complex 1, acidic, gene 18
Lama1	NM_008480	laminin, alpha 1
Lmnb2	NM_010722	lamin B2
Myh7	AY056464	myosin, heavy polypeptide 7, cardiac muscle, beta
Myh7	NM_080728	myosin, heavy polypeptide 7, cardiac muscle, beta
Myh9	AJ312390	Myosin heavy chanin IX; expressed sequence C80049
Myhca	NM_010856	myosin heavy chain, cardiac muscle, adult
Nnat	NM_010923	neuronatin
Tm4sf7	NM_053082	transmembrane 4 superfamily member 7
Tncs	NM_009394	troponin C, fast skeletal
Hormones and proteases		
Psmc3ip	AB000121	proteasome 26S subunit, ATPase 3, interacting protein
Prep	AB022053	Mus musculus gene for prolyl oligopeptidase, exon 11, 12, 13, 14, 15
Ppil2	AK012472	peptidylprolyl isomerase (cyclophilin)-like 2
Ube2v1	BC003449	ubiquitin-conjugating enzyme E2 variant 1
Psmd7	NM_010817	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7
Prep	NM_011156	prolyl endopeptidase
Cell cycle associated genes		
Gas11	NM_018855	growth arrest specific 11
Cell death		
Bid	BC002031	BH3 interacting domain death agonist
Tumor associated proteins		
Frat1	NM_008043	frequently rearranged in advanced T-cell lymphomas
Lag	NM_019641	leukemia-associated gene
Ptov1	NM_133949	prostate tumor over expressed gene 1
Shd	NM_009168	src homology 2 domain-containing transforming protein D
Tacc3	NM_011524	transforming, acidic coiled-coil containing protein 3
Tacstd2	NM_020047	tumor-associated calcium signal transducer 2
Tera	NM_019643	teratocarcinoma expressed, serine rich
Trap1a	NM_011635	tumor rejection antigen P1A
Xrcc1	NM_009532	X-ray repair complementing defective repair in Chinese hamster cells 1

Others

Abcf2	BC003300	ATP-binding cassette, sub-family F (GCN20), member 2
Aco1	X61147	aconitase 1
Acp1	AK014603	acid phosphatase 1, soluble
Adprt1	NM_007415	ADP-ribosyltransferase (NAD+; poly (ADP-ribose) polymerase) 1
Apeg1	AF215896	aortic preferentially expressed gene 1
Bat1a	NM_019693	HLA-B-associated transcript 1A
Bysl	AK019239	bystin-like
Cct3	NM_009836	chaperonin subunit 3 (gamma)
Cops6	AF071315	COP9 (constitutive photomorphogenic) homolog, subunit 6
Cox4a	NM_009941	cytochrome c oxidase, subunit IVa
Cox7c	NM_007749	cytochrome c oxidase, subunit VIIc
Csrp2	NM_007792	cysteine-rich protein 2
Dbn1	NM_019813	drebrin 1
Eif2s2	NM_026030	eukaryotic translation initiation factor 2, subunit 2 (beta, 38kDa)
Gng3	U25051	Gng3lg gene, and G protein gamma 3 subunit (Gng3) gene
H19	X58196	H19 mRNA
Hdac1	NM_008228	histone deacetylase 1
Hn1	NM_008258	hematological and neurological expressed sequence 1
Hsp60	BC016400	heat shock protein, 60 kDa
Incenp	NM_016692	inner centromere protein
Mcmd	X62154	mini chromosome maintenance deficient (S. cerevisiae)
Mcmd7	NM_008568	mini chromosome maintenance deficient 7 (S. cerevisiae)
Mdk	M34094	retinoic acid-responsive protein (MK) precursor; Mouse retinoic acid-responsive protein (MK) gene
MGC38424	NM_144818	hypothetical protein MGC38424
Mre11a	NM_018736	meiotic recombination 11 homolog A (S. cerevisiae)
Mrpl45	NM_025927	mitochondrial ribosomal protein L45
Mrpl45	NM_025927	mitochondrial ribosomal protein L45
Mrps25	NM_025578	mitochondrial ribosomal protein S25
Nudt1	NM_008637	nudix (nucleoside diphosphate linked moiety X)-type motif 1
Pafah1b3	NM_008776	platelet-activating factor acetylhydrolase, isoform 1b, alpha1
Pin1	NM_023371	protein (peptidyl-prolyl cis/trans isomerase) NIMA-interacting 1
Rabggta	NM_019519	Rab geranylgeranyl transferase, a subunit
Rev1l	NM_019570	REV1-like (S. cerevisiae)
Sap18	NM_009119	Sin3-associated polypeptide 18
Scr59	NM_139140	serine-rich spermatocytes and round spermatid protein, 59kDa
Set	AK019960	SET translocation
Slbp	M_009193	stem-loop binding protein
Smap	NM_019772	small acidic protein
Tead2	NM_011565	TEA domain family member 2
Tex20	C78563	testis expressed gene 20
Tex271	BC010716	testis expressed gene 271
Tmpo	NM_011605	thymopoietin
Traip	BC006929	TRAF-interacting protein
Trip13	AK010336	thyroid hormone receptor interactor 13
Txnl2	NM_023140	thioredoxin-like 2
Uba52	NM_019883	ubiquitin A-52 residue ribosomal protein fusion product 1
Ube2v1	BC003449	ubiquitin-conjugating enzyme E2 variant 1

ESTs

RIKEN cDNA 0610009C03 gene
RIKEN cDNA 1110004G16 gene
RIKEN cDNA 1200009I24 gene
RIKEN cDNA 1210002E11 gene
RIKEN cDNA 1300017C10 gene
RIKEN cDNA 1600023A02 gene
RIKEN cDNA 1810030M08 gene
RIKEN cDNA 2010009J12 gene
RIKEN cDNA 2310042P20 gene
RIKEN cDNA 2310076D10 gene
RIKEN cDNA 2410018C03 gene
RIKEN cDNA 2410041F14 gene
RIKEN cDNA 2410080P20 gene
RIKEN cDNA 2410088E07 gene
RIKEN cDNA 2510027N19 gene
RIKEN cDNA 2610200G18 gene
RIKEN cDNA 2610200G18 gene
RIKEN cDNA 2610511O17 gene
RIKEN cDNA 2700038C09 gene
RIKEN cDNA 2700069E09 gene
RIKEN cDNA 2810406K24 gene
RIKEN cDNA 2810410M20 gene
RIKEN cDNA 2810453H10 gene
RIKEN cDNA 4930455J02 gene

RIKEN cDNA 4933426K21 gene
RIKEN cDNA 5730507C05 gene
RIKEN cDNA 5830450F21 gene
RIKEN cDNA 9030402K04 gene
AA408554
AA409373
AI427066
AI507533
BC003335
BC010711
BC021499
BF46265
BG066293
BG066610
BI410067
BI690444
C80208
C81206
C87313
D9Wsu18e
Z38128

Table S3. Genes differentially expressed in the 3 germ layers of the e7.5 embryo

3a: list of genes enriched in E7.5 endoderm versus mesoderm and ectoderm

Growth factors, receptors and signaling molecules

Cer	NM_009887	cerberus 1
Chrd	NM_009893	chordin
Cmkor1	NM_007722	chemokine orphan receptor 1
Dkk1	XM_123389	dickkopf
Edg2	NM_010336	endothelial differentiation G-protein-coupled receptor, 2
Epha2	NM_010139	Eph receptor A2
Esrra	NM_007953	estrogen related receptor, alpha
Eya2	BC003755	eyes absent 2
Fgfbp1	U49641	fibroblast growth factor binding protein 1
Folr1	NM_008034	folate receptor 1
Gab1	BC007483	growth factor receptor bound protein 2-associated protein 1
Gpr49	NM_010195	G protein-coupled receptor 49
IGF-II	X71922	IGF-II
Igfbp5	L12447	insulin-like growth factor binding protein 5
IlgfRII	U04710	insulin-like growth factor 2 receptor
JAK1	BM460951	Janus kinase 1
Kitl	NM_013598	kit ligand
Lnx1	NM_010727	ligand of numb-protein X 1
Ltb	NM_008518	lymphotoxin B; TNF family of ligands
Lyn	M57696	v-yes-1 oncogene homolog
Mapk12	NM_013871	MAP kinase-activated protein kinase 12
Nrp	NM_008737	neuropilin; VEGF receptors
Nedd9	NM_017464	neural precursor cell, developmentally down-regulated gene 9
Pcbd	NM_025273	6-pyruvoyl-tetrahydropterin synthase, TCF1 cofactor
Ptpn16	NM_013642	tyrosine phosphatase, non-receptor type
Pthr	NM_011199	parathyroid hormone related peptide receptor
Rhoip3	U73200	Rho interacting protein 3
ScyA5	U07602	Small chemoline, Ephrin B1
Tgtp	NM_011579	T-cell specific GTPase
Trh	NM_009426	thyrotropin releasing hormone
Wnt11	NM_009519	wingless-related MMTV integration site 11

Transcriptional regulators and other nuclear factors

Ah	M94623	Aryl-hydrocarbon receptor, DNA binding transcription factor
Cited1	NM_007709	Cbp/p300-interacting transactivator, Glu/Asp-rich C-term. Dom. 1
Crtr1	NM_023755	Tcfcp2-related transcriptional repressor 1
c-fos	V00727	cellular oncogene, fos
Dab2	NM_023118	disabled homolog 2
Elf3	NM_007921	E74-like factor 3
Foxa1	NM_008259	forkhead box A1
Foxa2	NM_010446	forkhead box A2
Foxa3	NM_008260	forkhead box A3
Foxd4	X86368	forkhead box D4
Fhl1	NM_010211	Four and a half LIM domains 1
Hey1	NM_010423	hairy/enhancer-of-split related with YRPW motif 1
junB	U20735	junB (junB)
Lhx1	NM_008498	lim homeobox protein 1
Klf5	BC006646	Kruppel-like factor 5
Msx1	NM_010835	Homeobox msh-like 1, Hox7.1
N10	X16995	N10 nuclear hormonal binding receptor
Nfkb1	NM_008689	nuclear factor of kappa light chain gene enhancer
Nr1h3	NM_013839	nuclear receptor subfamily 1, group H, member 3
Nr2f2	NM_009697	nuclear receptor subfamily 2, group F, member 2
Pem	NM_008818	placentae and embryos oncofetal homeobox gene
Polg	NM_017462	polymerase, gamma
Sox17	AK004781	SRY-box containing gene 17
Six1	X80339	sine oculis-related homeobox 1 homolog
Xbp1	NM_013842	X-box binding protein 1

Cell surface/adhesion/matrix/cytoskeletal proteins

Ambp	NM_007443	alpha 1 microglobulin/bikunin
Apbb2	AK004792	amyloid beta (A4) precursor protein-binding, family B, member 2
APP	U82624	amyloid precursor protein
COL18A1	U03715	alpha-1(XVIII) collagen
Cldn4	NM_009903	claudin 4; tight junction membrane associated proteins
Cldn6	NM_018777	claudin 6
Cldn7	NM_016887	claudin 7
Cnn2	NM_007725	calponin 2; actin binding cytoskeleton organization
Col4a1	J04694	procollagen, type IV, alpha 1
Col4a2	J04695	procollagen, type IV, alpha 2

Clcn3	NM_007711	chloride channel 3
Csrp	NM_007791	cysteine rich protein; LIM double zinc finger protein
Cubn	AF197159	cubilin
Dsc2	NM_013505	desmocollin 2
dystrophin	U56724	Duchenne muscular dystrophy locus
endoA	X15662	cytokeratin endo A
EndoB	M22832	cytokeratin (endoB)
Emb	NM_010330	embigin; Ig family integral membrane protein
F10	NM_007972	coagulation factor X
Fgb	AK011118	fibrinogen, B beta polypeptide
Gjb3	NM_008126	gap junction membrane channel protein beta 3
Gpc4	NM_008150	glypican 4
Gsn	NM_010354	gelsolin/actin depolymerizing factor; tumor supressor
Img	NM_008377	integral membrane glycoprotein
Itm2b	NM_008410	integral membrane protein 2B
Krt1-19	NM_008471	keratin complex 1, acidic, gene 19
Krt2-8	XM_122854	keratin complex 2, basic gene 8
Lamb3	NM_008484	laminin, beta 3
Ltb	NM_008518	lymphotoxin b
Myo5b	NM_008661	myosin Vb
Nid2	NM_008695	nidogen 2
Pls2	BC022943	plastin 2
Pvs	NM_008990	poliovirus sensitive gene; integral membrane protein
Rdx	NM_009041	radixin; actin binding, caps actin filaments
Tacstd2	NM_020047	tumor-associated calcium signal transducer 2
Tekt1	NM_011902	tektin 1; filament forming proteins
Tncc	NM_009393	troponin C, cardiac/slow skeletal
Vil	NM_009509	villin
Vtn	NM_011707	vitronectin
Xpr1	NM_011273	xenotropic and polytropic retrovirus receptor 1

Hormones and proteases and secretory apparatus

Adam19	D50410	a disintegrin and metalloproteinase domain 19 (meltrin beta)
Cpa3	NM_007753	carboxypeptidase A3, mast cell
Cpd	NM_007754	carboxypeptidase D
Cpn1	NM_030703	carboxypeptidase N, polypeptide 1, 50kD
Ctsc	NM_009982	cathepsin C
Ctsh	NM_007801	cathepsin H
Ctsz	NM_022325	cathepsin Z
Lgmn	NM_011175	legumain; lysosomal cystein proteinase
Prg	NM_011157	proteoglycan, secretory granule
Prss12	NM_008939	protease, serine, 12 neurotrypsin, (motopsin)
Serpinh1	NM_009825	serine (or cysteine) proteinase inhibitor
Tfpi	AF016313	tissue factor pathway inhibitor
Tmprss2	NM_015775	transmembrane protease, serine 2
tPA	NM_008872	plasminogen activator, tissue

Cell cycle associated genes

Cd59a	NM_007652	CD59a antigen
Cdkn1c	NM_009876	cyclin-dependent kinase inhibitor 1C (P57)
Cdkn2b	NM_007670	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)
Gadd45b	NM_008655	growth arrest and DNA-damage-inducible 45 beta
Gas6	NM_019521	growth arrest specific 6
Sipa1	D11374	signal-induced proliferation associated gene 1

Cell death

Capn6	NM_007603	calpain 6; cystein type endopeptidase
Casp11	NM_007609	caspase 11
Pea15	NM_008556	phosphoprotein enriched in astrocytes 15
Perp	NM_022032	p53 apoptosis effector related to Pmp22
Rad51	NM_011234	RAD51 homolog (S. cerevisiae);nucleus DNA repair
Tdag	NM_009344	T-cell death associated gene

Transport proteins

Afp	NM_007423	alpha fetoprotein
Apoa4	NM_007468	apolipoprotein A-IV
ApoAI	U79573	apolipoprotein A-I
APOC2	Z22216	apolipoprotein c2
ApoE	D00466	apolipoprotein e
Apom	NM_018816	apolipoprotein M
Lrp10	NM_022993	low-density lipoprotein receptor-related protein 10
MUSPA3	D00073	Prealbumin
Rbp4	AK004839	retinol binding protein 4, plasma
Slc16a1	NM_009196	solute carrier family 16 member 1
Slc2a2	NM_031197	solute carrier family 2 member 2

Slc2a3	NM_011401	solute carrier family 2 member 3
Slc34a2	NM_011402	solute carrier family 34 (sodium phosphate), member 2
Slc7a7	NM_011405	solute carrier family 7 member 7
Slc7a9	NM_021291	solute carrier family 7, member 9

Others

Abcd4	NM-008992	ATP-binding cassete, sub-family D (ALD), member 4
Arsa	X73230	arylsulfatase A
Bhmt2	NM_022884	betaine-homocysteine methyltransferase 2
Car4	U37091	carbonic anhydrase IV gene, complete cds.
Cyp26	NM_007811	cytochrome P450, 26, retinoic acid
EDEM	AB042828	EDEM, similar to alpha-mannosidase
Ggta1	NM_010283	glycoprotein galactosyltransferase alpha 1, 3
Gpx2	X91864	glutathione peroxidase
H19	X58196	H19 mRNA
Hdc	NM_008230	histidine decarboxylase cluster
Lerepo1	AV025472	immediate early response, erythropoietin 1
Pctpl	NM_019990	phosphatidylcholine transfer protein-like
Pdzk1	NM_021517	PDZ domain containing 1
Pon2	NM_008896	paraoxonase 2; arylesterase
Reck	NM_016678	reversion-inducing-cysteine-rich protein with kazal motifs
Rrbp1	NM_024281	ribosome binding protein 1
Sat	NM_009121	spermidine/spermine N1-acetyl transferase
Scel	NM_022886	sciellin; keratinocyte expressed; in epithelium, LIM domain
Soat2	NM_011433	sterol O-acyltransferase 2
Tex19	NM_028602	testis expressed gene 19
Tex264	AK009326	testis expressed gene 264
Trap1a	NM_011635	tumor rejection antigen P1A
Ugt1a6	U16818	UDP glycosyltransferase 1 family, polypeptide A6
Xlr3b	NM_011727	X-linked lymphocyte-regulated 3b
Xlr4	NM_021365	X-linked lymphocyte-regulated 4

ESTs

0610011M24Rik
1300002E07Rik
1300006M19Rik
1600029D21Rik
2210012L08Rik
2610019F03Rik
2700053F16Rik
2900060B22Rik
3010002H13Rik
3110001A13Rik
4931426K16Rik
4933419D20Rik
AA755260
AA407887
AA408729
AA409659
AA536743
AA589382
AI159700
AI173274
AI173996
AI315052
AI326010
AI415009
AI462105
AI447817
AI467657
AI848508
AL024016
AU041093
AV025472
AV270913
AW228162
BF583462
BF607517
BG067993
BG175611
BG293144
BM461191
C75969
C81439
R74815

3b: list of genes enriched in mesoderm + ectoderm versus endoderm at E7

Growth factors, receptors and signaling molecules

adrb2	X15643	beta-2-adrenergic receptor
Arhc	NM_007484	ras homolog 9 (RhoC)
Arpp19-pend	NM_021548	cyclic AMP phosphoprotein, 19 kDa
Btrc	NM_009771	beta-transducin repeat containing protein
Cmkbr10	AK019478	chemokine (C-C) receptor 10
Epha8	NM_007939	Eph receptor A8
Gh	NM_008117	growth hormone
Ghrhr	L07379	growth hormone releasing hormone receptor
Grb2	NM_008163	growth factor receptor bound protein 2
Gtpi-pend	NM_019440	interferon-g induced GTPase
Grn	NM_008175	granulin
Il1r2	NM_010555	interleukin 1 receptor, type II
Il4	U01310	interleukin 4
Kdr/VEGFR2	NM_010612	kinase insert domain protein receptor
Nttp1	NM_008748	neuronal tyrosine/threonine phosphatase 1
or37c	AJ133427	or37olfactory receptor
Opn1mw	NM_008106	opsin 1
Pkp1	X97675	plakophilin 1
Plxnb2	BC007481	plexin B2; semaphorin receptor
Procr	NM_011171	protein C receptor, endothelial
Rgs10	NM_026418	regulator of G-protein signalling 10
Rock1	NM_009071	Rho-associated protein serine threonine kinase 1
Scya27	NM_011336	small inducible cytokine A27
Sdf1	NM_013655	stromal cell derived factor 1
Sfrp2	NM_009144	secreted frizzled-related 2
Stmn4	NM_019675	stathmin-like 4
Hrh1	D50095	histamine H1 receptor

Transcriptional regulators and other nuclear factors

Brunol4	AF314173	bruno-like 4
Cbx3	AK002910	chromobox homolog 3
Dbp	NM_016974	D site albumin promoter binding protein
Dtx1	NM_008052	deltex 1
Eif4e	NM_007917	eukaryotic translation initiation factor 4E
Ercc1	NM_007948	excision repair cross-complementing 1
Foxc2	NM_013519	forkhead box C2
Gbx2	NM_010262	gastrulation brain homeobox 2
Hox1.11	M93148	homeobox protein (Hox-1.11)
Hoxc8	X07439	homeo box C8
Hist1	NM_013548	histone gene complex 1
H3	M32459	Mouse histone H3
Klf1	NM_010635	Kruppel-like factor 1
Ldb2	NM_010698	LIM domain binding 2
MTH2B	X90778	M. musculus H2B gene
Nmyc	M12731	N-myc protein
Orc2	NM_008765	origin recognition complex, subunit 2 homolog
Pttg1	NM_013917	pituitary tumor-transforming 1; nuclear and cytoplasmic
Rb1	NM_009029	retinoblastoma 1
Rbpsuhl	NM_009036	recombining binding protein suppressor of hairless
Sall3	X97581	sal-like 3
Sox3	NM_009237	SRY-box containing gene 3
Sox7	NM_011446	SRY-box containing gene 7
Tbx6	NM_011538	T-box 6

Cell surface/adhesion/matrix/cytoskeletal proteins

Actc1	NM_009608	actin, alpha, cardiac
Cdh3	X06340	cadherin 3
Gabrg3	NM_008074	gamma-aminobutyric acid (GABA-A) receptor, subunit gamma 3
Ifitm3l	NM_030694	interferon induced transmembrane protein 3-like
Ina	L27220	neuronal intermediate filament protein (alpha-internexin)
Itm2a	NM_008409	integral membrane protein 2A
Kcnj12	NM_010603	potassium inwardly-rectifying channel, subfamily J, member 12
Kcns1	NM_008435	K+ voltage-gated channel, subfamily S, 1
Kip2	NM_019686	kinase interacting protein 2
Krt1-1	NM_010659	keratin complex 1, acidic, gene 1
Ktn1	NM_008477	kinectin 1
Lamr1	NM_011029	laminin receptor 1 (67kD, ribosomal protein SA)
mlc1f	X12973	myosin alkali light chain
Pcdha4	NM_007766	protocadherin alpha 4
Pvrl3	NM_021495	poliovirus receptor-related 3; nectin, cell adhesion
Tuba1	NM_011653	tubulin alpha 1
Tubb3	NM_023279	tubulin, beta 3

Pvrl3	NM_021495	poliovirus receptor-related 3; nectin, cell adhesion
PERF 15	X87128	lipid binding protein PERF 15
Cell cycle associated genes		
Cetn3	NM_007684	centrin 3
Cdk4	NM_009870	cyclin-dependent kinase 4
Map3k5	NM_008580	mitogen activated protein kinase kinase kinase 5
Ndr2	NM_013864	N-myc downstream regulated 2; cell growth
Pttg1	NM_013917	pituitary tumor-transforming 1
Rbbp7	NM_009031	retinoblastoma binding protein 7
Roc2	NM_019692	GTP-binding protein Roc2; cell cycle
Cell death		
Gadd45	U00937	Mus musculus GADD45 protein (gadd45) gene
Pdcd6	NM_011051	programmed cell death 6
Psmb3	NM_011971	proteasome (prosome, macropain) subunit, beta type 3
Transport proteins		
Atp5j2	NM_020582	ATP synthase, H+ transporting, mitochondrial F2
Atp5l	NM_013795	ATP synthase, H+ transporting, mitochondrial Fg
Fabpe	AJ223066	other name of mRNA: mal1; Mus musculus Fabpe gene
Fxyd6	AK003888	FXYD domain-containing ion transport regulator 6
lox-1	AF303744	oxidized LDL receptor (Lox-1)
Rangnrf	NM_021329	RAN guanine nucleotide release factor
Rae1	AK014325	RAE1 RNA export 1 homolog
Slc22a1l	NM_008767	solute carrier family 22 member 1-like
Slc27a2	NM_011978	solute carrier family 27 member 2
Slc34a1	NM_011392	solute carrier family 34 member 1
Slc7a3	NM_007515	solute carrier family 7 member 3
Slc8a1	NM_011406	solute carrier family 8 member 1
Hormones, proteases and secretory apparatus		
Nspl1	AF093624	Neuroendocrine specific protein gene 1, reticulon 2
Klk8	NM_008940	kallikrein 8; serine proteinase, neuropsin
Phex	NM_011077	phosphate regulating neutral endopeptidases on the X chromosome
Prep	NM_011156	prolyl endopeptidase
Timp	NM_011593	tissue inhibitor of metalloproteinase
Others		
Abcb4	NM_008830	ATP-binding cassette, sub-family B member 4
Aldo3	AK005077	aldolase 3, C isoform
Atox1	NM_009720	ATX1 antioxidant protein 1
Bcat1	AK013888	branched chain aminotransferase 1, cytosolic
Bche	NM_009738	butyrylcholinesterase
Btbd1	AF363030	BTB (POZ) domain containing 1
Cd160	NM_018767	CD160 antigen; NK cell receptor
Ckb	NM_021273	creatine kinase, brain
Cox7c	NM_007749	cytochrome c oxidase, subunit VIIc
Crmp1	NM_007765	collapsin response mediator protein 1
Cyp2a12	BC018356	cytochrome P450, 2a12
Cyp7b1	NM_007825	cytochrome P450, 7b1
Dio2	NM_010050	deiodinase, iodothyronine, type II
Dnmt3a	NM_007872	DNA methyltransferase 3A
Fah	NM_010176	fumarylacetoacetate hydrolase
Fkbp7	NM_010222	FK506 binding protein 7 (23 kDa)
Gatm	NM_025961	glycine amidinotransferase
Glra1	X75832	alternative spliced transcript; Glra1 gene
Gstm2	NM_008183	glutathione S-transferase, mu 2
Gstm5	NM_010360	glutathione S-transferase, mu 5
Gstt2	NM_010361	glutathione S-transferase, theta 2
Hbb-b1	NM_008220	hemoglobin, beta adult major chain
Hip2	NM_016786	huntingtin interacting protein 2; ubiquitin ligase
Hist1	NM_013548	histone gene complex 1
Histone H3	M32459	Mouse histone H3 (H3.2-221) gene, complete cds.
Idh3g	NM_008323	isocitrate dehydrogenase 3 (NAD+), gamma
Igj	BC006026	immunoglobulin joining chain
Lpl	NM_008509	lipoprotein lipase
Matr3	NM_010771	matrin 3; nuclear inner membrane structural protein
Mcmd5	NM_008566	mini chromosome maintenance deficient 5
Mgst1	NM_019946	microsomal glutathione S-transferase 1
Mrp64	NM_025595	mitochondrial ribosomal protein 64
Mrpl55	AK012885	mitochondrial ribosomal protein L55
Mrps10	AK004151	mitochondrial ribosomal protein S10
MTH2B	X90778	M.musculus H2B gene.
Mtm1	NM_019926	X-linked myotubular myopathy gene 1

Ndufc1	NM_025523	NADH dehydrogenase (ubiquinone) 1, subcomplex unknown, 1
Nme3	NM_019730	expressed in non-metastatic cells 3
Nspl1	AF093624	Neuroendocrine specific protein gene 1, reticulon 2
Nufip1	NM_013745	nuclear fragile X mental retardation protein interacting protein
Orc2	NM_008765	origin recognition complex, subunit 2
Pcp4	NM_008791	Purkinje cell protein 4
Phgdh	L21027	3-phosphoglycerate dehydrogenase
Phka1	NM_008832	phosphorylase kinase alpha 1, muscle metabolism
Pla2g10	AF166097	secreted phospholipase A2, group X
Pmm1	AK004631	phosphomannomutase 1
Ppp1r3c	U89924	protein phosphatase 1, regulatory (inhibitor) subunit 3C
Pso	NM_008952	peroxisomal sarcosine oxidase
Ptgis	NM_008968	prostaglandin I2 (prostacyclin) synthase
Rae1	AK014325	RAE1 RNA export 1 homolog
Ril	NM_019417	reversion induced LIM gene
Rpl37a	NM_009084	ribosomal protein L37a
Sclip	NM_009133	Scgn10 like-protein
Slfn4	AF099977	schlafen4
Sult1a1	BC005413	sulfotransferase family 1A, phenol-preferring, member 1
Sult-x1	NM_020564	sulfotransferase-related protein SULT-X1
Tbxas1	NM_011539	thromboxane A synthase 1, platelet
Tcte3	NM_011560	t-complex-associated testis expressed 3
Timm10	NM_013896	translocase of inner mitochondrial membrane 10 homolog
Tmk	NM_023136	thymidylate kinase
TSA-1	U47737	thymic shared antigen-1
U1-C	X79214	U1-C gene
Ubce7ip3	NM_019705	ubiquitin conjugating enzyme 7 interacting protein 3
Usmg5	NM_023211	upregulated during skeletal muscle growth 5
Xpa	NM_011728	xeroderma pigmentosum, complementation group A
Zfp260	D45210	zinc finger protein 260
Zfp37	X89264	zinc finger protein 37
Zfp61	NM_009561	zinc finger protein 61

ESTs

AA589625
AB023564
AF069953
AF099808
AF110520
AI225904
AI317193
AI447783
AI450287
AI646709
AI848081
AI848392
AI853864
AI854235
AL024256
BB749215
BC005775
BC019504
BC023090
BE852123
BF181466
BG066634
C76683
C77168
C78811
C79777
C81070
D14Ertd231e
D17Wsu51e
D2Ertd93e
D4Ertd335e
D10627
M12379
U35108
X15789
X82688
X87128
R74862
0610006I08Rik
0610009I16Rik
0610010I12Rik

0610016J10Rik
0610033H09Rik
0610043B10Rik
0710001O03Rik
1010001M12Rik
1110001H19Rik
1110001J03Rik
1110017C15Rik
1110018L13Rik
1110018O12Rik
1110036B12Rik
1110046L09Rik
1110058B13Rik
1200006I17Rik
1200009C21Rik
1300011P19Rik
1500019O16Rik
1500031N16Rik
1500040F11Rik
1700013L23Rik
1810008O21Rik
1810009A15Rik
1810010N17Rik
1810030M08Rik
1810045K17Rik
2010012C24Rik
2010107E04Rik
2010319C14Rik
2210039B01Rik
2210408F11Rik
2310021G01Rik
2310022L06Rik
2310065H03Rik
2410015J15Rik
2410015N17Rik
2410043G19Rik
2510002A14Rik
2510006M18Rik
2510039O18Rik
2510049I19Rik
2610001J05Rik
2610039C10Rik
2610101M19Rik
2610313E07Rik
2610511E03Rik
2610511F02Rik
2700099C19Rik
2810025M15Rik
2810432L12Rik
2810443J12Rik
2900010M23Rik
2900092E17Rik
3110052F15Rik
4930563P03Rik
5730408K05Rik
7420700M05Rik
9130022B02Rik

Table S4. Genes differentially expressed in PDX1+ and PDX1– cells

4a: list of genes enriched only in PDX1+ pancreatic cells (E10.5) versus PDX1+ cells in stomach and duodenum of the same stages

Growth factors, receptors and signaling molecules

Admr	D17292	adrenomedullin receptor
Amfr	AA108866	autocrine motility factor receptor
Arhb	AA052286	ras homolog
Bmp7*	X56906	bone morphogenetic protein 7
Caldn*	M27844	Calmodulin
Calm2	W55117	calmodulin 2
Cdrap	U88328	cytokine inducible SH2-containing protein 3
Dab1	U96963	Diaphnous homolog 1
Dlk1	W89849	delta-like 1 homolog (Drosophila)
Eif2ak3	AA104316	eukaryotic translation initiation factor 2 alpha kinase 3
Egr2	x06746	early growth response 2
Fgfr4*	X59927	fibroblast growth factor receptor 4
Fgfrp	U04204	fibroblast growth factor regulated protein
Gdnf	u37459	glial cell line derived neurotrophic factor
Gnb2-rs1	AA097202	guanine nucleotide binding protein, beta 2, related sequence 1
Jak1	aa561503	Janus kinase 1
Melk	x95351	maternal embryonic leucine zipper kinase
Notch1*	aa271199	Notch gene homolog 1 (Drosophila)
Ppp2r2a	AA168363	protein phosphatase 2
Prkab1	AA098332	protein kinase, AMP-activated, beta 1 non-catalytic
Ptk2b	AA143956	protein tyrosine kinase 2 beta
RhoB*	AJ010045	Rho, Guanine nucleotide exchange factor
Rgs2	U67187	regulator of G-protein signaling 2
Smoc1	aa015322	secreted modular calcium binding protein 1
Sos2	Z11664	Son of sevenless homolog 2
Ssr4	C78062	signal sequence receptor, delta

Transcriptional regulators and other nuclear factors

ABPabpc4l	AA068153	poly(A)-binding protein, cytoplasmic 4-like
Atf2	w15790	activating transcription factor 2
Bin1	u86405	myc box dependent interacting protein 1
Cebpd	X61800	CCAAT/enhancer binding protein (C/EBP)
deltaId	b3M60523	inhibitor of DNA binding 3
Eif2ak3	AA104316	eukaryotic translation initiation factor 2 alpha kinase 3
Idb1	M31885	inhibitor of DNA binding 1
Idb3	W88041	inhibitor of DNA binding 3
MGC6685	AA068153	similar to poly(A)-binding protein, cytoplasmic 4
Pbx3	af020200	pre B-cell leukemia transcription factor 3
Recc1	x72711	replication factor C, 140 kDa
Rpo2-3	d83999	RNA polymerase II 3
Sf3a2	X83733	splicing factor 3a, subunit 2, 66kD

Cell surface/adhesion/matrix/cytoskeletal proteins

Anxa4*	AA108947	annexin A4
Ap47	M62419	Clathrin associated protein
App*	U82624	Amyloid protein precursor
Col15a1*	U03715	procollagen, type XV
Gja1*	x61576	gap junction membrane channel protein alpha 1
Glc*	Z46845	Glucagon
Krt1-3	X75650	keratin complex 1, acidic, gene 3
Krt2-10	M92088	keratin complex 2, basic, gene 10
Prkcb	x53532	protein kinase C, beta
Lamc1	NM010683	laminin, gamma 1
Sept9	W13793	septin 9
Spnb3	AF026489	beta-spectrin 3
Sypl	AA124405	synaptophysin-like protein
Tln	aa009154	talin
Vim	x51438	vimentin
Vtn*	m77123	vitronectin
Zyx	y07711	zyxin

Cell cycle associated genes

Ccnd1	w08016	cyclin D1
Cdkn1c*	U22399	cyclin-dependent kinase inhibitor 1C (P57)
Dp1	U28168	deleted in polyposis 1

Hormones, proteases and secretory apparatus

Chgb	x53028	chromogranin B
Cops4	z31238	COP9 (constitutive photomorphogenic) homolog, subunit 4

Ctsd	x52886	cathepsin D Diap1
Spink3	X06342	serine protease inhibitor, Kazal type 3
Dab1	96963	diaphanous homolog 1 (Drosophila)
Cell death		
Ddx16	AA153354	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 16
Transport proteins		
Abca1	aa690738	ATP-binding cassette, sub-family A (ABC1), member 1
Apoa2	W35864	apolipoprotein A-II
Apoa4	M64250	apolipoprotein A-IV
Apoe	AA036067	apolipoprotein E
Ap3m1	AA073984	adaptor-related protein complex 3, mu 1 subunit
Atp1a	W546381	ATPase, Na+/K+ transporting, alpha 1 polypeptide
Cbg*	X70533	corticosteroid binding globulin
Pltp	u37226	phospholipid transfer protein
Slc2a2*	x15684	solute carrier family 2, member 2
Glut2*	M23383	transporter 2, facilitate glucose transportation
Others		
Adfp	M93275	adipose differentiation related protein
Adm	D67076	Adamts1
Adam19	aa726223	a disintegrin and metalloproteinase domain 19
Afp	M16395	alpha fetoprotein
Ap1m2	W57084	adaptor protein complex AP-1, mu 2 subunit
Ap3m1	AA073984	adaptor-related protein complex AP-3, mu 1
Aplp1	L04538	amyloid beta (A4) precursor-like protein 1
BC003940	w82831	phosphotyrosyl phosphatase activator
Cdk5rap3	aa386606	CDK5 regulatory subunit associated protein 3
Clps	aa611440	colipase, pancreatic
Ela1*	M27347	elastase 1, pancreatic
D7Rp2e	x04097	DNA segment, Chr 7, Roswell Park 2 complex,
D11Wsu68e	aa110660	DNA segment, Chr 11, Wayne State University 68
Deb1	U59417	differentially expressed in B16F10 1
Ddx19	L25125	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 19
Dnajb3	U95607	DnaJ (Hsp40) homolog, subfamily B, member 3
Emu1	w80149	Emu1 gene
Enpp2	aa178190	ectonucleotide pyrophosphatase/phosphodiesterase2
Fxr1h	X90875	fragile X mental retardation gene 1
Fkbp8	AF030635	FK506 binding protein 8 (38 kDa)
Fkbp9	AA163272	FK506 binding protein 9
Gc*	M55413	group specific component
Gcg	46845	glucagon
Gsta3	M73483	glutathione S-transferase, alpha 3
Hpgd	U44389	hydroxyprostaglandin dehydrogenase 15 (NAD)
Hrc	W29533	histidine rich calcium binding protein
Hspca	W51433	heat shock protein 1, alpha
Ifitm3l	w41552	interferon induced transmembrane protein 3-like
Ly6c	AA027619	lymphocyte antigen 6 complex, locus C
Mest	D16262	mesoderm specific transcript
Mgll	AJ001118	monoglyceride lipase
Mbp	X67319	myelin basic protein
Mpdu1	w53527	mannose-P-dolichol utilization defect 1
Myd116	X51829	myeloid differentiation primary response gene 116
Ndufab1	AA023445	NADH dehydrogenase (ubiquinone) 1, alpha/beta subcomplex, 1
P2rx1	X84896	purinergic receptor P2X, ligand-gated ion channel, 1
Pah*	X51942	phenylalanine hydroxylase
Per1	af022992	period homolog 1 (Drosophila)
Phgdh	afl21027	
Pphn	AA124405	pantophysin
Rai12	AF004107	retinoic acid induced 12
Rdx	X60672	radixin
Schip1	AA250009	schwannomin interacting protein 1
Sdpr	aa673143	serum deprivation response
Serpinf1	W08269	serine (or cysteine) proteinase inhibitor, clade F), member 1
Sparc	x04017	ecreted acidic cysteine rich glycoprotein
Spp1	X16151	secreted phosphoprotein 1
Srp9	X78304	signal recognition particle 9 kDa
Tex271	X81059	testis expressed gene 271
Tm4sf3	aa571115	transmembrnae
Ube2l3	w78245	ubiquitin-conjugating enzyme E2L 3
Uck2	u53547	uridine-cytidine kinase 2
Uqcrc1	aa239369	ubiquinol-cytochrome c reductase core protein 1
Vars2	W55743	valyl-tRNA synthetase 2
Ywhah	aa415028	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta

ESTs
1110021H02Rik
2900008M13Rik
4922501H04Rik
5730409F23Rik
9530006B08Rik
C130041E03Rik
D330037A14Rik
AA009154
AA063844
AA066782
AA110660
AA120507
AA145181
AA168866
AA177753
AA386606
AA546047
AA590086
AA638262
AL024210
C81234
W04097
W10325
W13793
W29889
W29920
W48989
W53547
W57249
W70398
W80149
W82831

4b: list of genes enriched in PDX1– cells (E10.5)

Growth factors, receptors and signaling molecules

Arhgdib	L07918	Rho, GDP dissociation inhibitor (GDI) beta
Bmp1	L24755	bone morphogenetic protein 1
Calm3	AA103356	calmodulin 3
Calml4	aa709719	calmodulin-like 4
Ccrl2	AA034646	chemokine (C-C motif) receptor-like 2
Cmkar4	X99581	chemokine (C-X-C) receptor 4
Cmkbr1l2	AA034646	chemokine (C-C) receptor 1-like 2
Cmkor1	AF000236	chemokine orphan receptor 1
cPpp2cb	AA065915	protein phosphatase 2a, catalytic subunit, beta
Crabp1	X51715	cellular retinoic acid binding protein I
Csf1r	X06368	colony stimulating factor 1 receptor
Csk	U05247	c-src tyrosine kinase
Dlk1	AA002627	delta-like 1 homolog (Drosophila)
Dab2	u18869	disabled homolog 2 (Drosophila)
EphRA4	X65138	Eph receptor A4
F2r	aa596794	coagulation factor II (thrombin) receptor
Fstl	M91380	follistatin-like
Gnb4	M63658	guanine nucleotide binding protein, beta 4
Gng11	aa529056	guanine nucleotide binding protein (G protein), gamma 11
Ihh	u85610	Indian hedgehog homolog, (Drosophila)
Il11ra1	x74953	interleukin 11 receptor, alpha chain 1
Nr2f1	x74134	nuclear receptor subfamily 2, group F, member 1
Nrk	w82359	Nik related kinase
Ogfr	aa266897	opioid growth factor receptor
Pdgfra	m84607	platelet derived growth factor receptor, alpha
Ptgis	AB001607	prostaglandin I2 (prostacyclin) synthase
Rab5c	aa260094	member RAS oncogene family
Rras	M21019	Harvey rat sarcoma oncogene, subgroup R
Snai2	u79550	snail homolog 2 (Drosophila)
Sfrp1	U88566	secreted frizzled-related sequence protein 1
Shc1	aa119070	src homology 2 domain-containing transforming protein C1
Sfrp2	af017989	secreted frizzled-related sequence protein 2
Sdf1	l12030	stromal cell derived factor 1
S100a6	m37761	S100 calcium binding protein A6 (calcyclin)
S100a8	m83218	S100 calcium binding protein A8 (calgranulin A)

Shc1	aa119070	src homology 2 domain-containing transforming protein C1
Shh	X76290	sonic hedgehog homolog, (Drosophila)
Tie1	x73960	tyrosine kinase receptor 1
Tgfbi	L19932	transforming growth factor, beta induced, 68 kDa
Tgfb1i1	L22482	transforming growth factor beta 1 induced transcript 1
Trrp2	AA109109	transient receptor protein 2
ThrombinR	Z48043	Coagulation factor 2 (thrombin receptor)

Transcriptional regulators and other nuclear factors

Aebp1	X80478	AE binding protein 1
CutL-1	X75014	Cut-1 like
Elf3	AF016294	E74-like factor 3
Fhl1	aa522276	four and a half LIM domains 1
Eif5a	W17411	eukaryotic translation initiation factor 5A
EklF	M97200	Erythrocyte Kruppel-like factor
Foxa1	u44752	forkhead box A1
Gata4	u85046	GATA binding protein 4
Klf4	u20344	Kruppel-like factor 4 (gut)
Klf5	AA014295	Kruppel-like factor 5
Lmo4	AA103457	LIM domain only 4
Mab21l2	AA245183	mab-21-like 2 (C. elegans)
Myog	d90156	myogenin
Osf2	D13664	osteoblast specific factor 2 (fasciclin I-like)
Pax1	U03873	paired mesoderm homeobox 1
Pea3	X63190	polyomavirus enhancer activator 3
Pbx1	aa275198	pre B-cell leukemia transcription factor 1
Pitx2	u80036	paired-like homeodomain transcription factor 2
Rbms2	W82026	RNA binding motif, single stranded interacting protein 2
Ril	Y08361	reversion induced LIM gene
Rnf19	X71642	ring finger protein (C3HC4 type) 19
Snrp116	AA089187	U5 small nuclear ribonucleoprotein 116 kDa
Sox2	U31967	SRY-box containing gene 2
Sox11	AF009414	SRY-box containing gene 11
Tcfe2a	W41716	transcription factor E2a
Tcf21	AF029753	transcription factor 21
Zfhx1a	d76432	zinc finger homeobox 1a

Cell surface/adhesion/matrix/cytoskeletal proteins

Acta2	X13297	actin, alpha 2, smooth muscle, aorta
Actn4	AA039109	alpha actinin 4
Bgn	l20276	biglycan
Cd81	W98255	CD 81 antigen
Col1a2	X58251	procollagen, type I, alpha 2
Col5a1	aa607513	procollagen, type V, alpha 1
Col5a2	L02918	procollagen, type V, alpha 2
Cdh11	x77557	cadherin 11
Dctn6	aa217493	References dynactin
Fbn2	L39790	fibrillin 2
Flna	AA038347	filamin, alpha
Efnb	2l38847	ephrin B2
Eng	X77952	endoglin
Eln	aa152671	elastin
Ermelin	aa217411	endoplasmic reticulum membrane protein
Flna	W29429	filamin, alpha
Itgax	AA161818	integrin alpha X
Itm2a	L38971	integral membrane protein 2A
Klc1	w81858	kinesin light chain 1
Klk8	D30785	kallikrein 8
Krt1-19	m28698	keratin complex 1, acidic, gene 19
Lgals4	AA590686	lectin, galactose binding, soluble 4
Myl9	AA109109	myosin, light polypeptide 9, regulatory
Nid1	x14194	nidogen 1
Nid2	AA063985	nidogen 2
Prg	x16133	proteoglycan, secretory granule
Tnc	x56304	tenascin C
Tpm1	AA153343	tropomyosin 1, alpha
Tubb2	AA059763	tubulin, beta 2
Tubb3	w34733	tubulin, beta 3

Cell cycle associated genes

Ccnd2	m83749	cyclin D2
Clk3	AF033565	CDC-like kinase 3
Nup50	aa286303	nucleoprotein 50
Psmc3	AA409481	proteasome (prosome, macropain) 26S subunit, ATPase 3

Transport proteins

Abcg2	aa616278	ATP-binding cassette, sub-family G 2 (WHITE)
Kcnj8	D88159	potassium inwardly-rectifying channel, J8
Sec23	aa237364	SEC23A (S. cerevisiae)
Slc20a1	M73696	solute carrier family 20, member 1
Slc25a19	AA086942	solute carrier family 25 (member 19

Others

Adcy6	m93422	adenylate cyclase 6
Agr2	aa518397	anterior gradient 2 (Xenopus laevis)
Aip	U85489	aryl-hydrocarbon receptor-interacting protein
Akr1c13	aa592828	aldo-keto reductase family 1, member C13
Aldo3	W53351	aldolase 3, C isoform
Alas2	M63244	aminolevulinic acid synthase 2, erythroid
Ap1b1	AA003413	adaptor protein complex AP-1, beta 1 subunit
Basp1	Z31269	brain abundant, membrane attached signal protein 1
BBag3	aa407278	Bcl2-associated athanogene 3
C81439	c81439	expressed sequence C81439
Car2	k00811	carbonic anhydrase 2
Car14	aa032866	carbonic anhydrase 14
Cav	W13196	caveolin, caveolae protein, 22 kDa
Cdipt	aa027723	CDP-diacylglycerol--inositol 3-phosphatidyltransferase
ce1	C77702 B	lymphocyte gene 1
Centb1	W51687	centaurin, beta 1
C1qb	X16874	complement component 1, q subcomponent, beta
C1qa	X58861	complement component 1, q subcomponent, alpha
Cenpb	X55038	centromere autoantigen B
Crip2 :	aa028770	cystin rich protein 2
Dnajb3	W13189	DnaJ (Hsp40) homolog, subfamily B, member 3
Ehd1	W40735	EH-domain containing 1
Eplin	aa606536	epithelial protein lost in neoplasm
F2rl1	Z48043	coagulation factor II (thrombin) receptor-like
Efnb1	U07602	ephrin B11
Fcer1g	W41745	Fc receptor, IgE, high affinity I, gamma
Eplin	aa606536	epithelial protein lost in neoplasm
F2rl1	Z48043	coagulation factor II (thrombin) receptor-like 1
Galgt1	L25885	UDP-N-acetyl-alpha-D-galactosamine
Glipr2	aa289661	GLI pathogenesis-related 2
Gsta4	aa638086	glutathione S-transferase, alpha 4
Gsto1	U80819	glutathione S-transferase omega 1
Hbb-y	m26897	hemoglobin Y, beta-like embryonic chain
Hsp1	AA034638	heat shock protein 1
Hsp70-3	m12572	heat shock protein, 70 kDa 3
Klk26	x63327	kallikrein 26
Loxl	W98413	lysyl oxidase-like
Ldh2	X51905	lactate dehydrogenase 2, B chain
Ltb4dh	aa591235	leukotriene B4 12-hydroxydehydrogenase
Lypla3	W65635	lysophospholipase 3
Macs	aa546670	myristoylated alanine rich protein kinase C substrate
Mpeg1	L20315	macrophage expressed gene 1
Msh2	x81143	mutS homolog 2 (E. coli)
Mpo	W44075	myeloperoxidase
Mrc1	Z11974	mannose receptor, C type 1
Nsg1	W55727	neuron specific gene family member 1
Hspcb	AA118062	heat shock protein 1, beta
Nt5e	L12059	5′ nucleotidase, ecto
Oc90	W50767	otoconin 90
Onzin	aa638539	onzin
Pde6g	Y00746	phosphodiesterase 6G, cGMP-specific, rod, gamma
Penk1	M13227	preproenkephalin 1
Plat	aa426892	plasminogen activator, tissue
Ppic	M74227	peptidylprolyl isomerase C
Ppicap	x67809	peptidylprolyl isomerase C-associated protein
Pp6	aa261047	placental protein 6
Pfkm	W11082	phosphofructokinase, muscle
Pxf	W66916	peroxisomal farnesylated protein
Pfkfb3	AA163244	inducible 6-phosphofructo-2-kinase
Rrr-	aa543365	regulator for ribosome resistance homolog
Scin	u04354	scinderin
Sdccag28	W41070	serologically defined colon cancer antigen 28
Sema3c	X85994	sema domain, immunoglobulin domain
Sepp1	X99807	selenoprotein P, plasma, 1
Sqle	D42048	squalene epoxidase
Sh3bgr	aa060862	SH3-binding domain glutamic acid-rich protein
Siat8c	X80502	sialyltransferase 8 (alpha-2, 8-sialytransferase)

Smoc2	AA030421	SPARC related modular calcium binding 2
Spints1	W08269	serine (cysteine) proteinase inhibitor, clade F
STs,	aa387631	Weakly similar to T00374 hypothetical protein
Sultn	AA109782	N-sulfotransferase
Sultn	AF026073	N-sulfotransferase
Smy	X05260	spermiogenesis gene
Tenc1	w10598	tensin like C1 domain-containing phosphatase
Tfpi	AA111494	tissue factor pathway inhibitor
Ugcgl	aa285635	UDP-glucose ceramide glucosyltransferase-like

ESTs

0610007H10Rik
1110054N06Rik
1110055E19Rik
1300012P22Rik
1700093E07Rik
1810028B20Rik
1810058I24Rik
2300002G02Rik
2310058J06Rik
2310075E07Rik
2810462M08Rik
5730494N06Rik
6330403K07Rik
E230009N18Rik
AAI173274
AA117665
AA199380
AA266385
AA589382
AA409140
W98059
BC026432
C77685
W50866
W55004
W75072

Table S5. List of genes differentially expressed in endocrine or non-endocrine progenitors

5a: list of genes enriched in endocrine progenitors (Ngn3-GFP+) versus non-endocrine progenitor cells (Ngn3-GFP–, E13.5)

Growth factors, receptors and signaling molecules

Alg-2	U49112	Calcium binding protein
Cadps	D86214	Ca<2+>dependent activator protein for secretion
Camk2b	X63615	calcium/calmodulin-dependent protein kinase II, beta
Celsr2	AI586083	cadherin EGF LAG seven-pass G-type receptor 2
Ddr1	L57509	discoidin domain receptor family, member 1
Fgfr4	X59927	fibroblast growth factor receptor 4
Gip	U34295	gastric inhibitory polypeptide
Gpr14	AI326966	G protein-coupled receptor 14
Gpr27	AF027955	G protein coupled receptor 27
Gpr56	AI841654	G protein-coupled receptor 56
Gfra3	AB008833	glial cell line derived neurotrophic factor family
Gnai2	AI841629	guanine nucleotide binding protein, alpha inhibiting 2
Gnao	M36777	guanine nucleotide binding protein, alpha o
Gdm1	D50430	glycerol phosphate dehydrogenase 1, mitochondrial
Hpca	AB015200	hippocalcin
Irf6	U73029	interferon regulatory factor 6
Isg20	AW122677	interferon-stimulated protein
Dm15	Z38015	dystrophia myotonica kinase, B15
Mfng	AF015769	manic fringe homolog (Drosophila)
Mafb	L36435	v-maf musculoaponeurotic fibrosarcoma oncogene
Napa	AW125607	N-ethylmaleimide sensitive fusion protein attachment
Pla2g6	U88624	phospholipase A2, group VI
Pld3	AF026124	phospholipase D3
Ptov1	AI838337	prostate tumor over expressed gene 1
Ptpn21	D37801	protein tyrosine phosphatase, non-receptor type 21
Rab3d	M89777	RAB3D, member RAS oncogene family
Roh-GDI2	U73198	GDP dissociation factor
Rps6ka2	AJ131021	ribosomal protein S6 kinase, 90kD, polypeptide 2
Rab7	X89650	RAB7, member RAS oncogene family
Stxbp1	D45903	syntaxin binding protein 1
Vdr	AW061016	vitamin D receptor

Transcriptional regulators and other nuclear factors

Arx	AB006103	aristaless related homeobox gene (Drosophila)
Baz2a	AW122821	bromodomain adjacent to zinc finger domain, 2A
Cbfa2t1h	D32007	CBFA2T1 identified gene homolog (human)
Cutl1	U66249	cut-like 1 (Drosophila)
Edr2	AF060076	early development regulator 2 (homolog of polyhomeotic 2)
Ell2	AI197161	ELL-related RNA polymerase II, elongation factor
FoxaA2	L10409	forkhead box A2
Foxa3	X74938	forkhead box A3
Gtf2f2	AI841186	general transcription factor IIF, polypeptide 2
H2-Ke6	AF100956	H2-K region expressed gene 6
Hes6	AW048812	hairy and enhancer of split 6, (Drosophila)
Hist2	U62674	histone gene complex 2
Isl1	AJ132765	ISL1 transcription factor, LIM/homeodomain
Lmyc1	X13945	lung carcinoma myc related oncogene 1
Math4b	Y09167	Math4b gene
Myt1	AF004294	myelin transcription factor 1
Neurod1	U28068	Neurogenic differentiation 1
Neurog3	Y09167	neurogenin 3
Nkx2-2	U31566	K2 transcription factor related, locus 2
Pax4	AB010557	paired box gene 4
Pax6	X63963	paired box gene 6
Pdx1	X74342	pancreatic and duodenal homeobox gene 1
Pcbd	AW046590	6-pyruvoyl-tetrahydropterin synthase/dimerization
Pou3f4	M88301	POU domain, class 3, transcription factor 4
Rest	AI449034	RE1-silencing transcription factor
Sox18	L35032	SRY-box containing gene 18
Stat3	AI837104	signal transducer and activator of transcription 3
Zfp288	AI987985	zinc finger protein 288

Cell surface/adhesion/matrix/cytoskeletal proteins

Actb	M1248	1actin, beta, cytoplasmic
Adrm1	AW123694	adhesion regulating molecule 1
Anxa4	U7294	1annexin A4
Centg3	AW124150	centaurin, gamma 3
Cldn4	AB000713	claudin 4
Cplx2	D38613	complexin 2
Emb	AW061330	embigin

Krt2-19	AA763244	keratin complex 2, basic, gene 19
Krt2-7	A755126	keratin complex 2, basic, gene 7
Mapt	M18775	microtubule-associated protein tau
Pclo	Y19186	piccolo (presynaptic cytomatrix protein)
Tekt2	AB027138	tektin 2
Tm4sf2	D26483	transmembrane 4 superfamily member 2
Vamp2	AI838601	vesicle-associated membrane protein 2

Cell cycle associated genes

Clk	AF033565	DC-like kinase 3
Cdkn1a	AW048937	cyclin-dependent kinase inhibitor 1A (P21)
Cdk8	AW049619	cyclin-dependent kinase 8
Gadd45g	AF055638	growth arrest and DNA-damage-inducible 45 gamma
Npdc1	X67209	neural proliferation, differentiation and control gene 1
Scyd1	U92565	small inducible cytokine subfamily D, 1
Snrpd3	AA796831	small nuclear ribonucleoprotein D3

Hormones, proteases and secretory apparatus

App	U82624	Amyloid protein precursor
Capn10	AW049679	calpain 10
Chga	M64278	chromogranin A
Chgb	X51429	chromogranin B
Cpe	X61232	carboxypeptidase E
Ctsf	AJ131851	cathepsin F
Ica1	U37186	islet cell autoantigen 1, 69 kDa
Ins1	X04725	insulin I
Ins2	AF044669	Insulin II
Pcsk2	M55669	proprotein convertase subtilisin/kexin type 2
Gast	U34293	gastrin
Gck	L41631	glucokinase
gne1	X15830	secretory granule neuroendocrine protein 1, 7B2
Gcg	Z46845	glucan
Glc	Z46845	Glucagon
Pcsk1	M69196	proprotein convertase subtilisin/kexin type 1 inhibitor
Resp18	L34214	regulated endocrine-specific protein 18
Sct	X73580	secretin
Smst	X51468	somatostatin
Spint1	AW230369	serine protease inhibitor, Kunitz type 1
Vamp2	AI838601	vesicle-associated membrane protein 2

Cell death

Dri2	AF116847	dead ringer homolog 2 (Drosophila)
Rap2ip	U73941	Rap2 interacting protein

Transport proteins

Abcc8	AA967624	ATP-binding cassette, sub-family C (CFTR/MRP), member 8
Atp2a3	AI504474	ATPase, Ca++ transporting, ubiquitous
Atp6v0b	AI842889	ATPase, H+ transporting, lysosomal 21kDa, V0 subunit B
Clcn3	AI849432	chloride channel 3
Fxyd3	X93038	FXYD domain-containing ion transport regulator 3
G6pt1	AF080469	glucose-6-phosphatase, transport protein 1
Kcnh2	AF012871	potassium voltage-gated channel, subfamily H member 2
Lrp10	AI839286	low-density lipoprotein receptor-related protein 10
Rbp4	U63146	retinol binding protein 4, plasma
Slc12a7	AI182203	solute carrier family 12, member 7
Vdr	AW061016	vitamin D receptor

Others

Akr1c13	AB027125	aldo-keto reductase family 1, member C13
Apg12l	AI851598	autophagy 12-like
Aig1	AW208098	acupuncture induced gene 1
Arg1	U51805	arginase 1, liver
Arfgap3	AI844507	ADP-ribosylation factor GTPase activating protein 3
Btg2	M64292	B-cell translocation gene 2, anti-proliferative
Btrc	AF099932	beta-transducin repeat containing protein
Cbfa2t3h	AF038029	core-binding factor, runt domain, alpha subunit2
Cck	X59520	cholecystokinin
Chkl	AA204010	choline kinase-like
Cdipt	AI838053	CDP-diacylglycerol--inositol 3-phosphatidyltransferase
Cyfip2	AI835274	cytoplasmic FMR1 interacting protein 2
Cyp4b1	D50834	cytochrome P450, subfamily IV B, polypeptide 1
Cyb561	AI846517	cytochrome b-561
Dia1	AW122731	diaphorase 1 (NADH)
Dp1l1	AA755260	deleted in polyposis 1-like 1
Dscr1l2	AI847661	Down syndrome critical region gene 1-like 2

DXCch3	L41495	DNA segment, Chr X, Celltech Chiroscience 3
Elovl1	AI842813	elongation of very long chain fatty acids Cpt2
Fbxo9	AI853035	f-box only protein 9
Galnt2	AI841884	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2
Gars	AI839918	glycyl-tRNA synthetase
Gch	L09737	GTP cyclohydrolase 1
Glipr2	AA983101	GLI pathogenesis-related 2
Gpx3	U13705	glutathione peroxidase 3
Gspt2	AB003503	G1 to phase transition 2
Gdi1	U07950	guanosine diphosphate (GDP) dissociation inhibitor 1
Gtdh	X57024	glutamate dehydrogenaseapss2
Hap1	AJ002272	huntingtin-associated protein 1
Hdac6	AF006603	histone deacetylase 6
Hmgcr	M62766	3-hydroxy-3-methylglutaryl-Coenzyme A reductase
Hip1r	AI846182	huntingtin interacting protein 1 related
loxe3	Y14695	arachidonate lipoxygenase 3
Lrp10	AI839286	low-density lipoprotein receptor-related protein 10
Mtlrp	AJ243503	motilin-related peptide
Mlgll2	AA733674	Mus musculus lethal giant larvae-like protein 2
Mrpl49	AI854652	mitochondrial ribosomal protein L49
Myoz1	AA733946	myozenin 1
Napa	AW125607	N-ethylmaleimide sensitive fusion protein attachment protein alpha
Nedd4b	AI840868	neural precursor cell expressed, developmentally down-regulated gene 4b
Ndrl	U52073	N-myc downstream regulated-like
Qdpr	AI845337	quininoid dihydropteridine reductase
Pam	U79523	peptidylglycine alpha-amidating monooxygenase
Pcbd	AW046590	6-pyruvoyl-tetrahydropterin synthase/dimerization cofactor of hepatocyte nuclear factor 1 alpha (TCF1)
Pdk3	AI842259	pyruvate dehydrogenase kinase, isoenzyme 3
Peg10	AI836610	paternally expressed 10
Ppicap	X67809	peptidylprolyl isomerase C-associated protein
Prodh2	AA675075	proline dehydrogenase (oxidase) 2
Prrg2	AW122679	proline-rich Gla (G-carboxyglutamic acid) polypeptide 2
Prnp	M18070	prion protein
Ptov1	AI838337	prostate tumor over expressed gene 1
Sdccag28	AW049732	serologically defined colon cancer antigen 28
Sdnsf	AI837599	neural stem cell derived neuronal survival protein
Siat8c	X80502	sialyltransferase 8 (alpha-2, 8-sialytransferase)
Siat10	AI153959	sialyltransferase 10 (alpha-2,3-sialyltransferase VI)
Stard10	AW049732	serologically defined colon cancer antigen 28
Ubxdc2	AW120609	UBX domain-containing 2
Tat	AI255353	tyrosine aminotransferase
Tcp11	X52128	t-complex protein 11
Thea	AI847418	hioesterase, adipose associated
Tor2a	AI841457	torsin family 2, member A
Tapbp	AI836367	TAP binding protein
Ubqln2	AI836406	ubiquilin 2
Ubxdc2	AW120609	UBX domain-containing 2
Usp18	AW047653	ubiquitin specific protease 18
wbp2	U40826	W domain binding protein 2
Wfs1	AA832886	Wolfram syndrome 1 homolog (human)
Zap70	AI386093	zeta-chain (TCR) associated protein kinase

ESTs

0610010O12Rik
0610011F06Rik
0610011I04Rik
0610043B10Rik
1110008P14Rik
1110014L17Rik
1110014H17Rik
1300013G12Rik
2610016K11Rik
3110038L01Rik
4930415K17Rik
5830471E12Rik
9130006H11Rik
9430077D24Rik
E230009N18Rik
AA798246
AA607767
AI849432
AI845823
AI842259
AI850887

AI838337

5b: genes present or enriched in NGN3– cells versus NGN3+ cells (E13.5)

Growth factors, receptors and signaling molecules

Basp1	AW124113	brain abundant membrane attached signal protein 1
Basp2	AI841303	brain abundant membrane attached signal protein 2
Calu	U81829	calumenin
Cckar	D85605	cholecystokinin A receptor
Cla3	AI849333	cerebellar ataxia 3
Dab2	U18869	disabled homolo g 2 (Drosophila)
EndoR	U32329	Endothelin receptor
Epha3	M68513	Eph receptor A3
Fin14	U42386	fibroblast growth factor inducible 14
Fkbp10	L07063	FK506 binding protein 10 (65 kDa)
Fstl	M91380	follistatin-like
Fzd2	AW123618	frizzled homolog 2 (Drosophila)
Grb10	AF022072	growth factor receptor bound protein 10
Igfbp5	L12447	insulin-like growth factor binding protein 5
Igfbp4	X76066	insulin-like growth factor binding protein 4
Il11ra2	U69491	interleukin 11 receptor, alpha chain 2
Igf2r	U04710	insulin-like growth factor 2 receptor
Igf1	X04480	insulin-like growth factor 1
Igf2	X71922	insulin-like growth factor 2
Igfbp3	AI842277	insulin-like growth factor binding protein 3
Igfbp4	AI838737	insulin-like growth factor binding protein 4
Ksr	AI664600	kinase suppressor of ras
Ltbp4	AA838868	latent transforming growth factor beta binding protein 4
Basp1	AW124113	brain abundant, membrane attached signal protein 1
Notch1	Z11886	Notch gene homolog 1, (Drosophila)
Nr2f2	X76653	nuclear receptor subfamily 2, group F, member 2
Nrk	AA866768	Nik related kinase
PdgfraM	57683	platelet derived growth factor receptor, alpha polypeptide
Ppfia4	AI838513	protein tyrosine phosphatase, receptor type 4a
Ptgis	AB001607	prostaglandin I2 (prostacyclin) synthase
Racgap1	AW122347	Rac GTPase-activating protein 1
Sfrp1	U88566	secreted frizzled-related sequence protein 1
S100a8	M83218	S100 calcium binding protein A8 (calgranulin A)
Snai2	U79550	snail homolog 2 (Drosophila)
Sdf1	L12030	stromal cell derived factor 1
Sgk	AW046181	serum/glucocorticoid regulated kinase
Sfrp2	U88567	secreted frizzled-related sequence protein 2
Stk5	D21099	serine/threonine kinase 5
Tgfbi	L19932	transforming growth factor, beta
ThrombinR	Z48043	Coagulation factor 2 (thrombin receptor)
Trrp2	AI842649	transient receptor protein 2

Transcriptional regulators and other nuclear factors

Eif4ebp1	U28656	eukaryotic translation initiation factor 4E binding protein 1
Fen1	L26320	flap structure specific endonuclease 1
Hmga2	X99915	high mobility group AT-hook 2
Hmgb2	X67668	high mobility group box 2
Mad3	U32394	Max dimerization protein 3
Lmo4	AF074600	LIM only 4
Nde1	AW120739	nuclear distribution gene E homolog 1
Nfib	Y07686	nuclear factor I/B
Nol5a	AW121447	nucleolar protein 5A
Osf2	D13664	osteoblast specific factor 2 (fasciclin I-like)
Pold1	AF024570	DNA polymerase delta 1, catalytic domain
Rnps1	X70067	ribonucleic acid binding protein S1
Rfc5	AA667128	replication factor C (activator 1) 5
Rnf2	AI043016	ring finger protein 2
Sfpq	AA690583	splicing factor proline/glutamine rich
Snrpd3	AA796831	small nuclear ribonucleoprotein D3
Tbx2	U15566	T-box 2
Tcf3	AI841235	transcription factor 3
Tcf21	AF035717	transcription factor 21
Top2a	U01915	topoisomerase (DNA) II alpha
Zac1	X95503	zinc finger protein regulator of apoptosis andcell cycle arrest
Zfhx1a	D76432	zinc finger homeobox 1a
Zfp275	AI153693	Zinc finger protein 275
Zfp36l1	M58566	zinc finger protein 36, C3H type-like 1
Zfp422	AW209414	zinc finger protein 422

Cell surface/adhesion/matrix/cytoskeletal proteins

Acta2	X13297	actin, alpha 2, smooth muscle, aorta
Col1a1	AA763466	procollagen, type I, alpha 1
Col1a2	X58251	procollagen, type I, alpha 2
Col18a1	U03715	procollagen, type XVIII, alpha 1
Col3a1	AA655199	procollagen, type III, alpha 1
Col5	a2L02918	procollagen, type V, alpha 2
Col5a1	AB009993	procollagen, type V, alpha 1
Col6a2	Z18272	procollagen, type VI, alpha 2
Col6a3	AF064749	procollagen, type VI, alpha 3
Ckap2	AI121796	cytoskeleton associated protein 2
Cks2	AA681998	CDC28 protein kinase regulatory subunit 2
Cspg2	D45889	chondroitin sulfate proteoglycan 2
Dcn	X53929	decorin
Dnahc11	AI314784	dynein, axon, heavy chain 11
Eln	AA919594	elastin
Emp1	X98471	epithelial membrane protein 1
Emp3	U87948	epithelial membrane protein 3
Eplin	AI836140	epithelial protein lost in neoplasm
Eng	X77952	endoglin
Fbln1	X70853	fibulin 1
Fbn1	L29454	fibrillin 1
Eln	AA919594	elastin
Gap43	AI841303	growth associated protein 43
Ggta1	M85153	glycoprotein galactosyltransferase alpha 1,3
Gsn	J04953	gelsolin
Incenp	AA823653	inner centromere protein
Islr	AB024538	immunoglobulin superfamily containing leucine-rich repeat
Itm2a	L38971	integral membrane protein 2A
Knsl1	AJ223293	kinesin-like 1
Kif23	AI591702	kinesin family member 23
Lama1	M36775	laminin, alpha 1
Lama4	69176	laminin, alpha 4
Lgals1	X15986	lectin, galactose binding, soluble 1
Mbnl	AI854802	muscleblind-like
Mfap2	L23769	microfibrillar-associated protein 2
Mpp1	U38196	membrane protein, palmitoylated (55 kDa)
Myl9	AI842649	myosin, light polypeptide 9
Myoz1	AA733946	myozenin 1
Nrp	50086	neuropilin
P4ha2	U16163	procollagen-proline, 2-oxoglutarate 4-dioxygenase alpha II polypeptide
Sdc4	D89571	syndecan 4
Spin	AA681862	spindlin
Tagln	Z68618	transgelin
Tm4sf7	AW124470	transmembrane 4 superfamily member 7
Tubb5	X04663	tubulin, beta 5
Vcl	AI462105	vinculin
Vtn	M77123	vitronectin

Cell cycle associated genes

Csnk1d	AI846289	casein kinase 1, delta
Ccna2	X75483	cyclin A2
Cdc2a	M38724	cell division cycle 2 homolog A (S. pombe)
Cdkn1c	U22399	cyclin-dependent kinase inhibitor 1C (P57)
Cks1	AB025409	CDC28 protein kinase 1
Gas1	X65128	growth arrest specific 1
Mad2l1	U83902	MAD2 (mitotic arrest deficient, homolog)-like 1
Mki67	X82786	Ki 67
Pcna	X57800	proliferating cell nuclear antigen
Trp53	B021961	transformation related protein 53

Hormones, proteases and secretory apparatus

Amy2	X02578	amylase 2, pancreatic
Capn6	Y12582	calpain 6
Pnliprp1	AA674409	pancreatic lipase related protein 1
Sec23a	AI843665	SEC23A

Transport proteins

Cbg	X70533	corticosteroid binding globulin
Gja1	M63801	gap junction membrane channel protein alpha1
Kcnj8	D88159	potassium inwardly-rectifying channel, J8
Slc38a4	AA795541	solute carrier family 38, member 4
Slc25a19	AI852682	solute carrier family 25

Others

AA407323	AA866768	expressed sequence AA407323
Abcf2	AI837302	Mrpl18
Ahcy	L32836	S-adenosylhomocysteine hydrolase
AI173274	AI642389	expressed sequence AI173274
AI788959	AI788959	expressed sequence AI788959
AI853703	AI853703	expressed sequence AI853703
Akap12	AB020886	A kinase (PRKA) anchor protein (gravin) 12
AI987814	AW121801	expressed sequence AI987814
Anp32e	AI849333	acidic (leucine-rich) nuclear phosphoprotein 32 family, member E
Anxa2	M14044	annexin A2
AW121776	AW121776	expressed sequence AW121776
AW122030	AW122030	Mus musculus, Similar to phosphoserine aminotransferaseclone
Birc5	AB013819	baculoviral IAP repeat-containing 5
Chaf1b	AI173038	chromatin assembly factor 1, subunit B (p60)
Csrp2	D88792	cysteine-rich protein 2
Cyr61	M32490	cysteine rich protein 61
Dnmt1	F036008	DNA methyltransferase (cytosine-5) 1
Dtymk	AA624336	deoxythymidylate kinase
Dusp9	AA285446	dual specificity phosphatase 9
Ech1	F030343	enoyl coenzyme A hydratase 1, peroxisomal
Edg2	U13370	endothelial differentiation, lysophosphatidic acid G-2Ela1
Els1	M27347	elastase 1, pancreatic
Enpp5	AW048581	ectonucleotide pyrophosphatase/phosphodiesterase 5
F2r	L03529	coagulation factor II (thrombin) receptor
Gamt	AF010499	guanidinoacetate methyltransferase
Glk	AB027012	galactokinase
Glipr2	AA983101GLI	pathogenesis-related 2
Gstm2	J04696	glutathione S-transferase, mu 2
Gstm5	AA241764	glutathione S-transferase, mu 5
Hba-a1	V00714	hemoglobin alpha, adult chain 1
Hba-x	M13125	hemoglobin X, alpha-like embryonic chain in Hba complex
Hbb-b1	J00413	hemoglobin, beta adult major chain
Hbb-b2	V00722	hemoglobin, beta adult minor chain
Hbb-bh1	X14061	hemoglobin Z, beta-like embryonic chain
Hbb-y	V00726	hemoglobin Y, beta-like embryonic chain
Hic1	AW048074	hypermethylated in cancer 1
Hsd17b12	AF064635	hydroxysteroid (17-beta) dehydrogenase 12
Igsf7	AW060457	immunoglobulin superfamily, member 7
Incenp	AA823653	inner centromere protein
Gc	M55413	group specific component
Enc1	AI848479	ectodermal-neural cortex 1
Ldh1	M17516	lactate dehydrogenase 1, A chain
Lsp1	D49691	lymphocyte specific 1
Lum	AF013262	lumican
Maoa	AI848045	monoamine oxidase A
Mcmd	X62154	mini chromosome maintenance deficient
Mcmd4	D26089	mini chromosome maintenance deficient 4
Mcmd5	D26090	mini chromosome maintenance deficient 5
Meis1	U33629	myeloid ecotropic viral integration site 1
Nap1l1	X61449	nucleosome assembly protein 1-like 1
Nasp	AF034610	nuclear autoantigenic sperm protein
Np95	D87908	nuclear protein 95
nrsf/rest	AI449034	Mus musculus NRSF/REST mRNA for neural-restrictive
Oat	X64837	ornithine aminotransferase
Oxct	AI843232	3-oxoacid CoA transferase
Penk1	M55181	preproenkephalin 1
Pmf1	AW060657	polyamine-modulated factor 1
Pmp22	Z38110	peripheral myelin protein, 22 kDa
Ppic	M74227	peptidylprolyl isomerase C
Pnliprp1	AA674409	pancreatic lipase related protein 1
Prc1	AA856349	protein regulator of cytokinesis 1
Prdx5	AF093853	peroxiredoxin 5
Pscd3	AI846077	pleckstrin homology, Sec7 and coiled-coil domains 3
Ptgds	AI840733	prostaglandin D2 synthase (brain)
Pygb	AI846739	brain glycogen phosphorylase
Qk	AI846695	quaking
Rrm1	K02927	ribonucleotide reductase M1
S100	a983219	S100 calcium binding protein A9 (calgranulin B)
S100a6	X66449	S100 calcium binding protein A6 (calcyclin)
Sec23a	AI843665	SEC23A (S. cerevisiae)
Shmt1	AA913994	serine hydroxymethyl transferase 1 (soluble)
Smc4l1	AA032310	SMC4 structural maintenance of chromosomes 4-like 1
Sparc	X04017	secreted acidic cysteine rich glycoprotein
Stmn2	AI839868	stathmin-like 2

Tapbp	AF110520	TAP binding protein
Tcp1-rs1	M35797	t-complex protein 1, related sequence 1
Tmpo	U39074	thymopoietin
Tgif	X89749	TG interacting factor
Tmk	AW121709	thymidylate kinase
Tyms	AU044050	thymidylate synthase
Wdr5	AI463460	WD repeat domain 5

ESTs

0610038P07Rik
1110003O22Rik
1110031K21Rik
1110007H17Rik
1500001L20Rik
1500019O16Rik
1600023A02Rik
1810014L12Rik
2310021G01Rik
2310058J06Rik
2310075E07Rik
2400003B06Rik
2410008J01Rik
2610034N03Rik
2610312E17Rik
2810417H13Rik
2810429C13Rik
2810442O16Rik
2900097C17Rik
3110001O07Rik
5033405D03Rik
5730403J10Rik
5730409F23Rik
6330403K07Rik
6330514M23Rik
9330200A01Rik
9430077D24Rik
C130041E03Rik
E230009N18Rik
AA285446
AA795541
AI314784
AI642389
AI647612
AI788959
AI852734
AI853668
AI853703
AI836140
AA856349
AI853899
AW121031
AW122331
AW122860
AW124049
AW125453
AW212532

Table S6. List of genes whose expression has been investigated by RT-PCR or in situ hybridization (+:PCR/in situ result is consistent with Microarray data; ?: Microarray data cannot be confirmed; ND: not determined)

Enriched in endoderm			enriched in pancreatic PDX1+ cells			enriched in PDX1– cells			Enriched in NGN3+ cells	
gene	R.T.	I.S.	gene	R.T.	I.S.	gene	R.T.	I.S.	gene	I.S.
ApoAIV	+	+	Bin1	+	ND	Aip	+	ND	APP	+
Cer1	+	+	Cbg	+	+	Ctkl-19	+	+	Cldn4	+
Ckit-l	+	+	Ccnd1	?	ND	Elf3	+	ND	Col1a1	+
Cubillin	+	ND	Cebpd	+	+	Foxa1	+	+	Foxa3	+
Cux	+	+	Dlk1	+	+	Gata4	+	+	Fzd2	?
Disabled	+	ND	Dp1	+	+	Klf4	+	ND	Gao	+
Dkk1	+	+	Fgfr4*	+	ND	Ril	+	ND	Hmgb2	?
Epha2	+	+	Fgfrp	?	ND	S100a6	?	ND	Lmyc-1	?
Eya2	+	+	Fxr1h	?	ND	Scin	?	ND	Mafb	+
FGFbp1	+	+	Idr	+	+	Sfrp1	+	ND	Mfng	+
Fhl1	ND	+	Myd116	+	ND	Sh3bgr	?	ND	Myt1	+
Gas6	+	?	Melk	+	ND	Sox2	+	+	Npdc1	+
Irx3	+	+	Nkx2.2	+	+	Spintz1	+	+	Peg10	?
Klf5	+	+	RhoB	+	+	Tcf21	+	ND	RBP1	+
Lhx1	+	+	Rp2	+	+	Tm4sf3	+	+	Stat3	+
Pdzk1	+	?	Smoc1	?	ND	Col1a2	+	ND	Wfs1	?
Prss12	+	+	Sos2	+	ND	Ihh	+	ND	Aa015322	?
Six1	+	+	Spp1	+	+	AI854235	+	ND	AI448995	+
Tfp1	+	ND	Vim	+	+	W13423	+	ND		
Hey1	+	?	Vtn	+	+					
Sox17	+	+	AA009154	+	ND					
Villin	+	+	W04097	+	+					
Zic3	+	+	W12941	+	ND					
AA408729	+	ND	w29920	+	ND					
AA623587	+	ND	W80149	+	ND					