SUPPLEMENTARY DATA

*Gene expression, pathway and network analyses*

Gene expression associated with CHI-D was determined using GSE32610 from the Gene Expression Omnibus (GEO: http://www.ncbi.nlm.nih.gov/geo/) (1). The transcriptomic datasets of CHI and control pancreas [Human 58K oligonucleotide array (GPL14670)] were downloaded into Qlucore Omics Explorer 2.3 (Qlucore, Lund, Sweden), normalized by multi-dimensional scaling (Isomapping) (2) and compared by ANOVA with a significance cut-off of $P<0.05$. Identification of biological pathways and functions associated with gene expression changes was undertaken using a right-sided Fisher's exact test within Ingenuity Pathway Analysis (IPA) software. Network analysis was also used to identify and prioritize key functional elements. In brief, to derive an interactome model differentially expressed genes were used as 'seeds' and all known protein:protein and protein:gene interactions between the seeds and their inferred immediate neighbours were calculated to generate a biological network using the Biogrid model of the human Interactome (31.2.101) (3). Network generation and processing was performed using Cytoscape 2.8.3. Clustering and 'community structure' within interactome models, known to be associated with function (4-6), was prioritized using the ModuLand plug-in for Cytoscape 2.8.3 to determine overlapping modules (7) and to identify hierarchical structure (8). Clusters were confirmed using MCODE algorithm (9) and module centrality was used to derive a priority index for genes of biological importance.

## Supplementary references


1. Michelsen NV, Brusgaard K, Tan Q, Thomassen M, Hussein K, Christesen HT: Investigation of Archived Formalin-Fixed Paraffin-Embedded Pancreatic Tissue with Whole-Genome Gene Expression Microarray. ISRN Pathology 2011;2011:Article ID 275102
2. Tenenbaum JB, de Silva V, Langford JC: A global geometric framework for nonlinear dimensionality reduction. Science 2000;290:2319-2323
3. Chatr-Aryamontri A, Breitkreutz BJ, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O'Donnell L, Reguly T, Breitkreutz A, Sellam A, Chen D, Chang C, Rust J, Livstone M, Oughtred R, Dolinski K, Tyers M: The BioGRID interaction database: 2013 update. Nucleic Acids Res 41:D816-823
4. Sun J, Zhao Z: A comparative study of cancer proteins in the human protein-protein interaction network. BMC Genomics 2010;11 Suppl 3:S5
5. Ravasz E, Barabasi AL: Hierarchical organization in complex networks. Phys Rev E Stat Nonlin Soft Matter Phys 2003;67:026112
6. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M: The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol 2007;3:e59
7. Szalay-Beko M, Palotai R, Szappanos B, Kovacs IA, Papp B, Csermely P: ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. Bioinformatics 28:2202-2204
8. Kovacs IA, Palotai R, Szalay MS, Csermely P: Community landscapes: an integrative approach to determine overlapping network module hierarchy, identify key nodes and predict network dynamics. PLoS One 2010;5
9. Bader GD, Hogue CW: An automated method for finding molecular complexes in large protein interaction networks. BMC Bioinformatics 2003;4:2
10. Piper K, Brickwood S, Turnpenny LW, Cameron IT, Ball SG, Wilson DI, Hanley NA: Beta cell differentiation during early human pancreas development. J Endocrinol 2004;181:11-23

SUPPLEMENTARY DATA

**Supplementary Table 1. CHI-D patient information.** Ten cases undergoing pancreatectomy for sustained hypoglycemia unresponsive to medical treatment. Mode of inheritance for gene defects; AR=autosomal recessive, AD=autosomal dominant. Details of mutations; p.? = intronic mutation resulting in unknown protein size.

| Patient | Age at Surgery | Sex | Birth Weight (kg) | Gene Defect | Inheritance AD/AR | Paternal Mutation | Maternal Mutation |
|---|---|---|---|---|---|---|---|
| CHI-D1 | 2m | M | 6.7 | *ABCC8* | AR | p.? (c.3992-9G>A) | p.? (c.3992-9G>A) |
| CHI-D2 | 2m | M | 4.3 | *KCNJ11* | AR | p.Q299R (c.896A>G) | p.Q299R (c.896A>G) |
| CHI-D3 | 2m | M | 4.4 | *ABCC8* | AR | p.A30V (c.89C>T) | p.A30V (c.89C>T) |
| CHI-D4 | 2m | F | 3.5 | *ABCC8* | AR | p.S581T (c.1741T>A) | p.? (c.3992-9G>A) |
| CHI-D5 | 3m | M | 4.5 | *ABCC8* | AR | p.? (c.1818-?_1923+?del) | p.T172fs c.512dup |
| CHI-D6 | 4m | M | 2.9 | *ABCC8* | AR | p.H36R (c.107A>G) | p. ? (c.1630+1G>T) |
| CHI-D7 | 5m | M | 3.5 | *ABCC8* | AR | p.L610P (C.1829T>C) | p.L610P (C.1829T>C) |
| CHI-D8 | 5m | M | 1.9 | *ABCC8* | AD | p.I1512T (c.4535T>C) | - |
| CHI-D9 | 6m | M | 2.9 | *ABCC8* | AR | p.? (c.148+1G>A) | p.? (c.148+1G>A) |
| CHI-D10 | 13m | F | 4.6 | *ABCC8* | AR | p.? (c.4612-1G>T) | p.A4V (c.11C>T) |

SUPPLEMENTARY DATA

**Supplementary Table 2. Primary antibodies**

The antibodies developed by O.D. Madsen (anti-NKX6.1) and T.M. Jessell (anti-NKX2.2 and anti-ISL1) were obtained from the Developmental Studies Hybridoma Bank developed under the auspices of the NICHD and maintained by the University of Iowa, Department of Biology, Iowa City, IA 52242.

| **Primary Antibody** | **Raised In** | **Dilution** | **Supplier** |
|---|---|---|---|
| Polyclonal anti-insulin | Rabbit | 1:1000 | Abcam |
| Polyclonal anti-insulin | Guinea Pig | 1:100 | Zymed |
| Polyclonal anti-glucagon | Rabbit | 1:50 | Zymed |
| Polyclonal anti-somatostatin | Rabbit | 1:50 | Zymed |
| Polyclonal anti-PP | Rabbit | 1:50 | Zymed |
| Polyclonal anti-ghrelin | Goat | 1:500 | Abcam |
| Polyclonal anti-gastrin | Rabbit | 1:200 | Cell Marque |
| Monoclonal anti-NKX6.1 | Mouse | 1:1000 | DSHB |
| Monoclonal anti-NKX2.2 | Mouse | 1:75 | DSHB |
| Monoclonal anti-Ki67 | Mouse | 1:100 | Novocastra |
| Monoclonal anti-PHH3 | Mouse | 1:200 | Cell Signaling |
| Polyclonal anti-phospho-γ-H2AX | Rabbit | 1:200 | Cell Signaling |
| Polyclonal anti-SOX9 | Rabbit | 1:5000 | Millipore |
| Polyclonal anti-GATA4 | Goat | 1:450 | Abcam |
| Polyclonal anti-CDK6 | Rabbit | 1:500 | Abcam |
| Polyclonal anti-CK19 | Mouse | 1:100 | Novocastra |
| Polyclonal anti- P27Kip1 | Rabbit | 1:200 | Santa Cruz |
| Polyclonal anti-pRb | Rabbit | 1:500 | Abcam |
| Polyclonal anti-FOXA2 | Goat | 1:800 | R & D |
| Polyclonal anti-PDX1 | Guinea pig | 1:500 | Abcam |
| Monoclonal anti-ISL1 | Mouse | 1:200 | DSHB |

SUPPLEMENTARY DATA

**Supplementary Table 3. qRT-PCR primers**
Forward and reverse primer sequences are given for each gene along with the accession number and product size. All primer pairs / products were intron-spanning wherever possible.

| Gene | Accession No. | Forward Primer | Reverse Primer | Size (bp) |
|---|---|---|---|---|
| *β-ACTIN* | NM_001101.3 | CCAACCGCGAGAAGATGA | CCAGAGGCGTACAGGGATAG | 138 |
| *HPRT* | NM_000194.2 | TGACCTTGATTTATTTTGCATACC | CGAGCAAGACGTTCAGTCCT | 102 |
| *PDX1* | NM_000209.3 | AAAGGCCAGTGGGCAGGCGG | GCGCGGCCGTGAGATGTACT | 135 |
| *FOXA2* | NM_021784.4 | GAAGATGGAAGGGCACGAGC | GTACGTGTTCATGCCGTTCA | 115 |
| *SOX9* | NM_000346.3 | GTACCCGCACTTGCACAAC | TCGCTCTCGTTCAGAAGTCTC | 72 |
| *NKX6.1* | NM_006168.2 | GGCCTGTACCCCTCATCAAG | TCCGGAAAAAGTGGGTCTCG | 79 |
| *NEUROD1* | NM_002500.4 | GAGGCCCCAGGGTTATGAGA | TGGTCATGTTTCGATTTCCTTTGT | 70 |
| *MAFA* | NM_201589.3 | AGAGCGAGAAGTGCCAACTC | GTACAGGTCCCGCTCTTTGG | 83 |
| *NKX2.2* | NM_002509.3 | ATGTCGCTGACCAACACAAAG | GATGTCCTTGACCGAAAACCC | 45 |
| *INSULIN* | NM_001185097.1 | TGGCTTCTTCTACACACCCA | TCTAGTTGCAGTAGTTCTCCA | 984 |
| *PAX4* | NM_006193.1 | AAGAAAGCAGCTTGCGTTGAC | GGGCGTGAGACAGAGGTATC | 165 |
| *PAX6* | NM_000280.4 | CCCGGCAGAAGATTGTAGAG | GCTAGCCAGGTTGCGAAGAA | 323 |
| *ARX* | NM_139058.2 | ATTCGGCAGGCTCTTTTCCA | ATGTTGGAGTTGGAGCGAGG | 502 |
| *MAFB* | NM_005461.3 | TGAACTTTGCGCGTTAAGCC | CACGCAGCCGCCGGAGTTTC | 436 |
| *CDKN1B* | NM_004064.4 | TCCGGCTAACTCTGAGGACA | GAAGAATCGTCGGTTGCAGG | 120 |

SUPPLEMENTARY DATA

**Supplementary Table 4. Top 20 functional modules following network analysis ranked by the index of centrality of their most central gene**

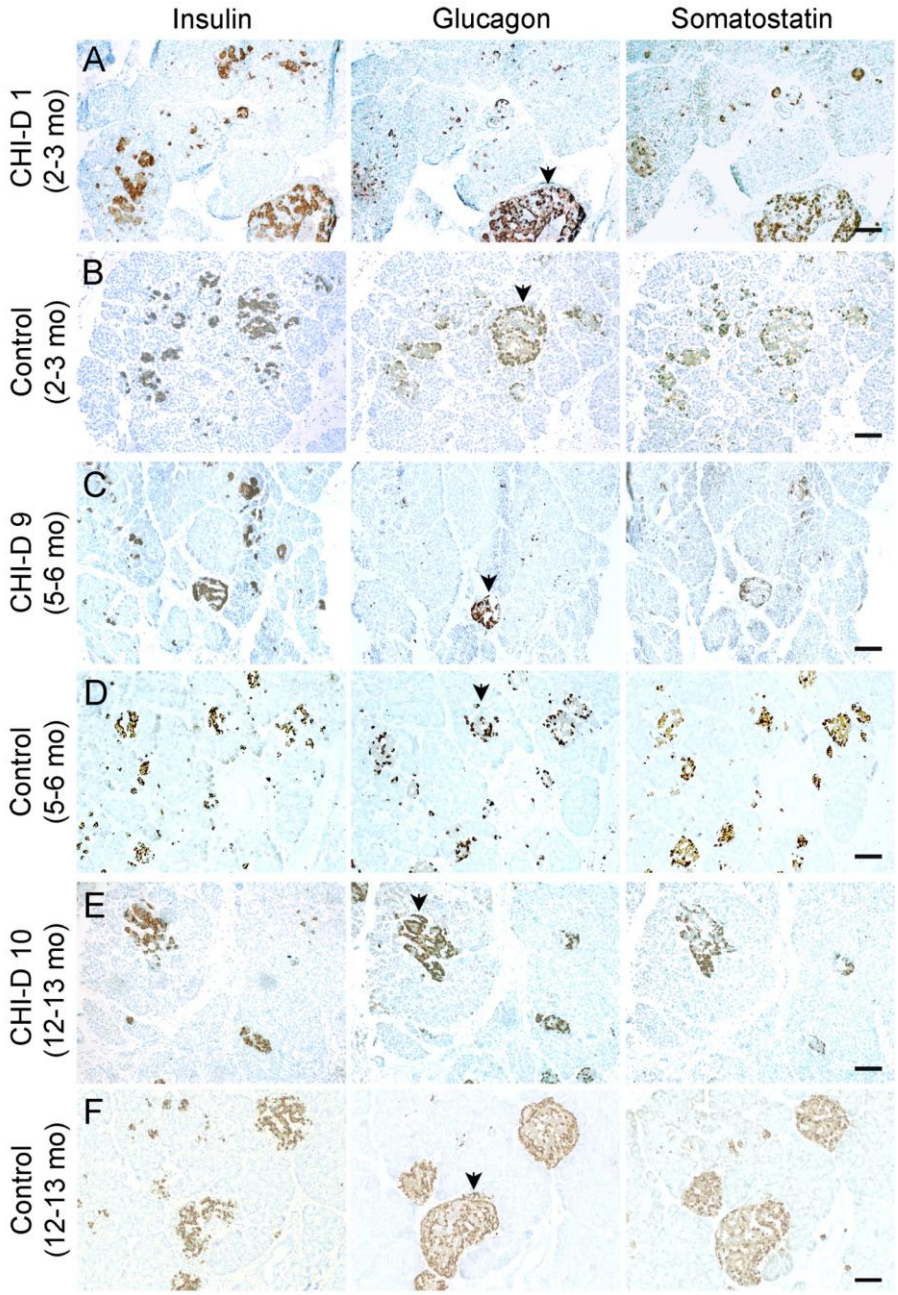
The most central gene in the top 20 modules is shown in bold. In parentheses are the other genes in the top ten for each module.

| **Module rank** | **Most central gene in the module** (others genes in the top 10) |
|---|---|
| **1** | ***EP300*** (*UBD, SIRT7, KAT2B, TP53, FN1, PRMT1, ARRB1, VCAM1, TRIM28*) |
| **2** | ***PML*** (*UBD, DAXX, SUMO2, TRIM28, FN1, SUMO1, PIAS1, ARRB1, VCAM1*) |
| **3** | ***CDKN1B*** (*CDK2, CDK7, CDK4, SKP2, CCT8, UBD, CCNE1, COPS5, VCAM1*) |
| **4** | ***NCOA2*** (*ESR1, EP300, AR, GNAI2, PRMT1, FN1, SIRT7, SAFB, ACTR2*) |
| **5** | ***CAND1*** (*CUL1, COPS3, FN1, CUL3, PHKG2, CUL4A, RNF7, SIRT7, UBD*) |
| **6** | ***PIK3R1*** (*CRK, FYN, CBL, VAV1, ITSN1, ERBB3, GRB2, EGFR, BCAR1*) |
| **7** | ***TBP*** (*TAF9, TAF6, TAF10, TAF5, TCEA1, TAF1, TAF4, ELAVL1, GTF2B*) |
| **8** | ***HDAC5*** (*AR, UBD, TBL1XR1, NCOR1, PHKG2, ZBTB16, ARRB1, HNF4A, HSP90AA1*) |
| **9** | ***CDK7*** (*TCEA1, CCNH, GTF2H2, POLR2A, SIRT7, ERCC3, MNAT1, GTF2H1, APP*) |
| **10** | ***COPS3*** (*COPS5, COPS6, ERCC8, TK1, SIRT7, CUL3, DDB1, CUL4B, TOR1AIP2*) |
| **11** | ***MRE11A*** (*SIRT7, RAD50, NBN, DDX1, TERF1, BRCA1, NRF1, HNRNPD, VCAM1*) |
| **12** | ***ZBTB16*** (*HDAC1, HDAC5, TRIM28, FN1, PHB, DNMT3B, TK1, EHMT2, VCAM1*) |
| **13** | ***SMURF1*** (*SMAD1, SMAD7, SMAD5, STRAP, NEDD4, APP, DCTN2, PHKG2, UBE2D3*) |
| **14** | ***FN1*** (*SIRT7, HNRNPD, VCAM1, GAPDH, VHL, TK1, PHKG2, CCT8, ELAVL1*) |
| **15** | ***BAG1*** (*HSPA8, HSPA4, TTC1, ARRB1, NRF1, PHKG2, STUB1, TERF1, ACTR2*) |
| **16** | ***NEDD4*** (*UBE4B, UBE2D2, UBE2D3, UBE2L3, LAPTM5, MKRN3, MGRN1, ARRB1, UBE2D1*) |
| **17** | ***UBE2M*** (*APP, NEDD8, PDIA3, NRF1, RBX1, CLU, NDUFS6, DCUN1D1, UBA3*) |
| **18** | ***PSMA6*** (*PSMA5, UBD, VCAM1, PHKG2, FKBP8, FN1, PSMA2, PSMA3, STK4*) |
| **19** | ***HEXIM1*** (*BRD4, CDK9, CCNT1, EAF1, TERF1, RN7SK, NRF1MED12, STK4*) |
| **20** | ***RING1*** (*BMI1, PHC1, RNF2, PCGF2, PHC2, INTS6, CBX4, TERF1, APP*) |

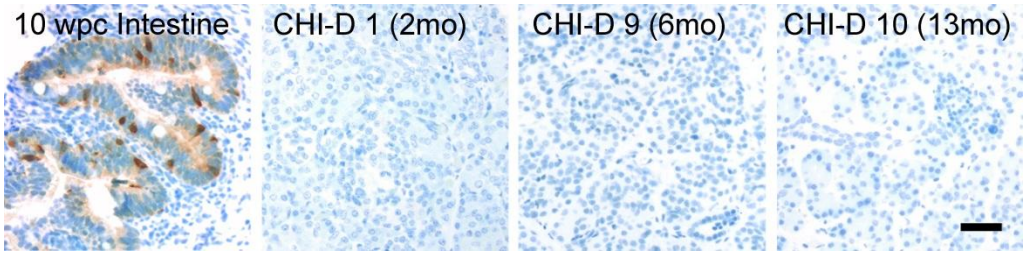
SUPPLEMENTARY DATA

**Supplementary Figure 1. β-cells, α-cells and δ-cells in CHI-D and control pancreas during the early postnatal period**

Consecutive 5 μm sections from three examples of CHI-D and age-matched control pancreas are stained for insulin (β-cells), glucagon (α-cells) and somatostatin (δ-cells) (all brown) counter-stained with toluidine blue. **(A-B)** 2-3 months (mo); **(C-D)** 5-6 months; and **(E-F)** 12-13 months. Arrowheads in each glucagon panel point to an islet at each age that is also apparent in sections either side stained for insulin and somatostatin. At 2-3 months, note the extra-islet, scattered hormone-positive cells in both CHI-D and controls. At all ages, note that glucagon-positive cells are more evenly distributed throughout the islets in CHI-D compared to the peripheral α-cells in control islets. The consistently central insulin staining and, in older CHI-D samples, more peripheral somatostatin in adjacent sections exemplifies that this feature was not due to the plane of sectioning (also see Fig. 1 in main text). Gastrin was not observed in CHI-D or age-matched control pancreas compared to a positive control of human fetal intestine at 10 weeks post-conception (**image below**). Scale bar represents 200 μm.



SUPPLEMENTARY DATA

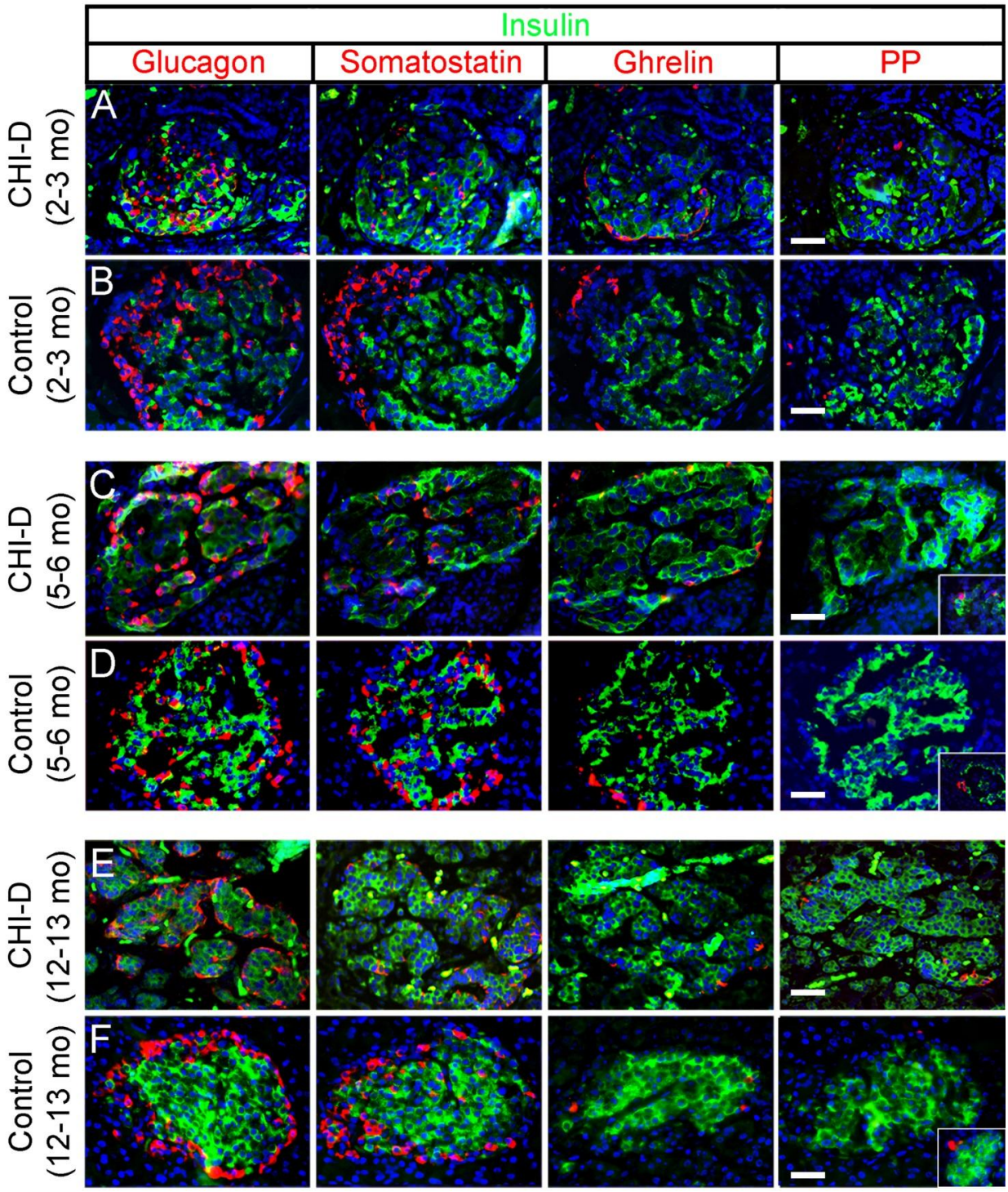



Size bar = 100 μm

SUPPLEMENTARY DATA

**Supplementary Figure 2. Endocrine lineages and islet composition in CHI-D and age-matched controls**

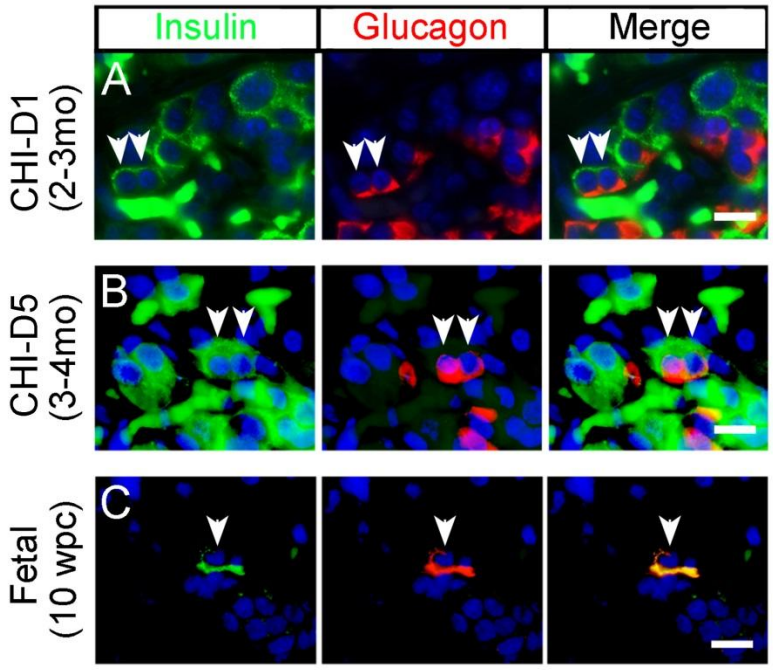
Dual confocal immunofluorescence of consecutive sections from CHI-D and age-matched control pancreas counterstained with DAPI. (**A-B)**, 2-3 months (mo); (**C-D)**, 5-6 months; and **(E-F)**, 12-13 months. Insets demonstrate PP staining elsewhere within the same section. At early stages α-cells and δ-cells are more diffusely scattered throughout islets rather than arranged as a mantle. Ghrelin staining was located peripherally along with occasional PP-cells in both CHI-D and control Islet structure remains less organized and compact in CHI-D. Scale bars represent 100 μm.



SUPPLEMENTARY DATA

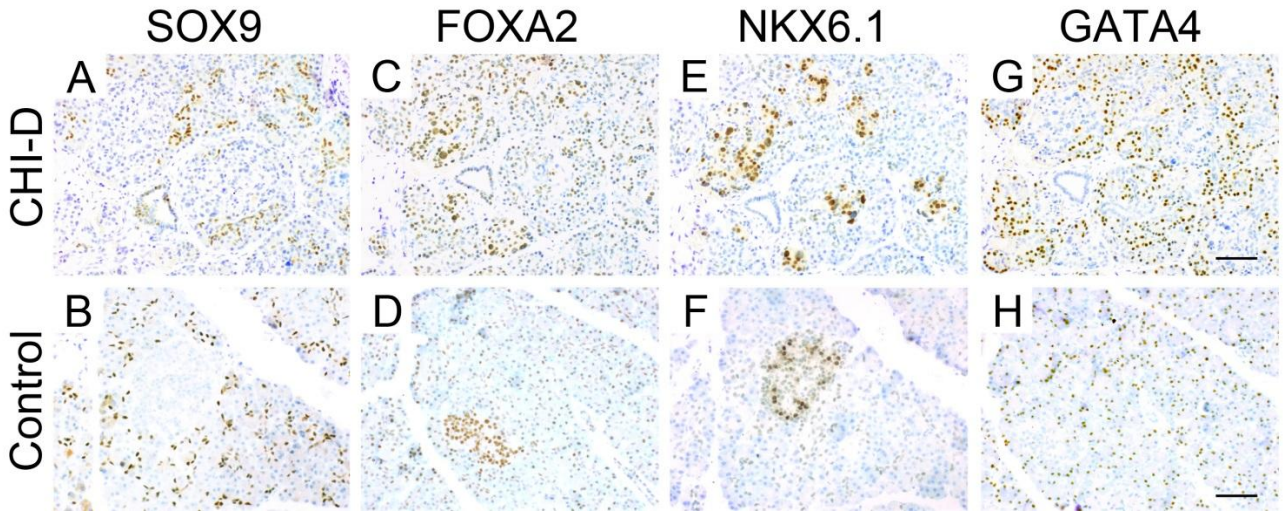
**Supplementary Figure 3. Localization of insulin and glucagon within the same endocrine cells in CHI-D**

Dual confocal immunofluorescence for insulin and glucagon counterstained with DAPI in tissue sections from two cases of CHI-D **(A-B)** and human fetal pancreas **(C)** at 10 weeks post-conception (wpc). In CHI-D the arrowheads point to cells which contain both insulin and glucagon but where each hormone localizes to a discrete area of the cell. In contrast, detection of the two hormones completely overlaps in human fetal pancreas (10). The associated supplementary video confirms the cellular co-detection by moving through a cell on confocal Z-stack. Scale bar represents 50 µm.



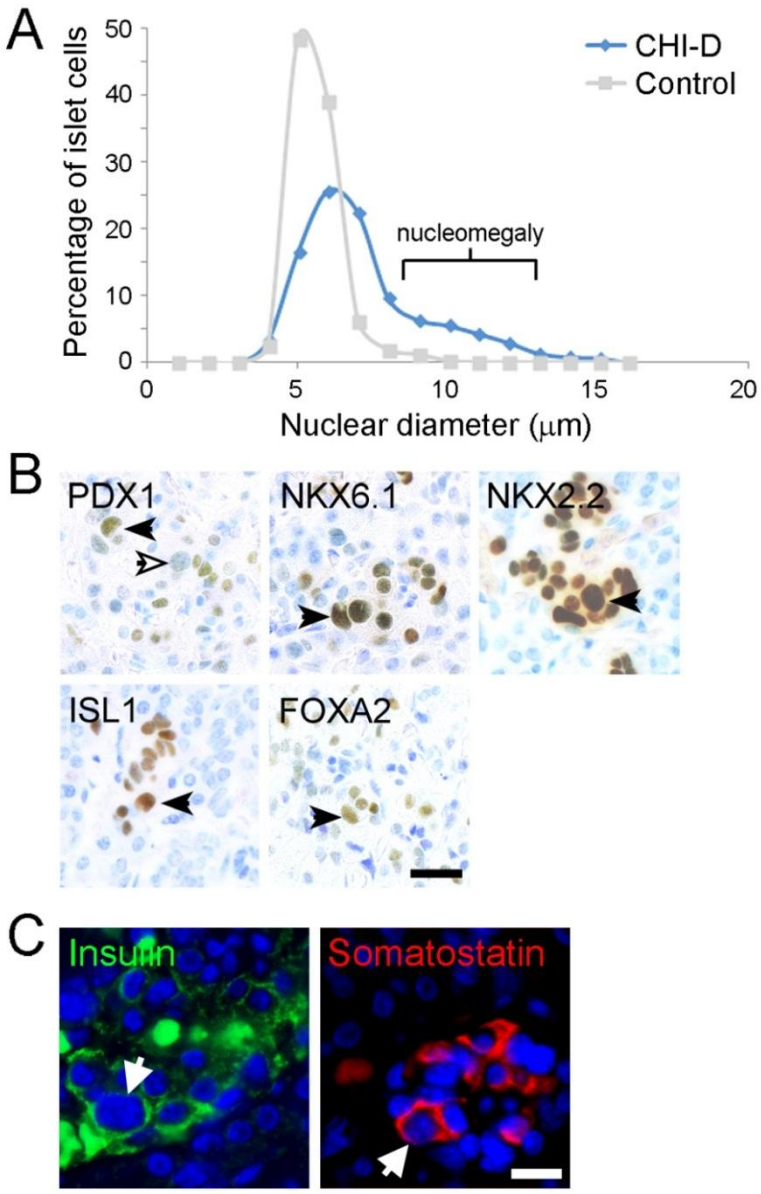
SUPPLEMENTARY DATA

**Supplementary Figure 4. SOX9, FOXA2, NKX6.1 and GATA4 in CHI-D and control pancreas**
Brightfield immunohistochemistry counterstained with toluidine blue shows SOX9 (duct cells, **A-B**), FOXA2 (duct cells and β-cells**, C-D**), NKX6.1 (β-cells, **E-F**) and GATA4 (acinar cells, **G-H**) in an example of CHI-D pancreas and its aged-matched control. The staining profile was identical in all CHI-D samples (2-13 months). Note examples of nucleomegaly amongst the FOXA2 and NKX6.1 staining in CHI-D but not CHI-D duct or acinar cells or any of the control cells. NKX6.1 also illustrates the diffuse nature of β-cells in CHI-D compared to control pancreas. Note also the increased GATA4 staining in CHI-D compared to control consistent with the major acinar cell proliferation observed in CHI-D (Fig. 6C, main text). Along with satisfactory tissue morphology, the sensitive detection of nuclear transcription factors argues against significant pancreatic autolysis. Scale bars represent 200 μm.



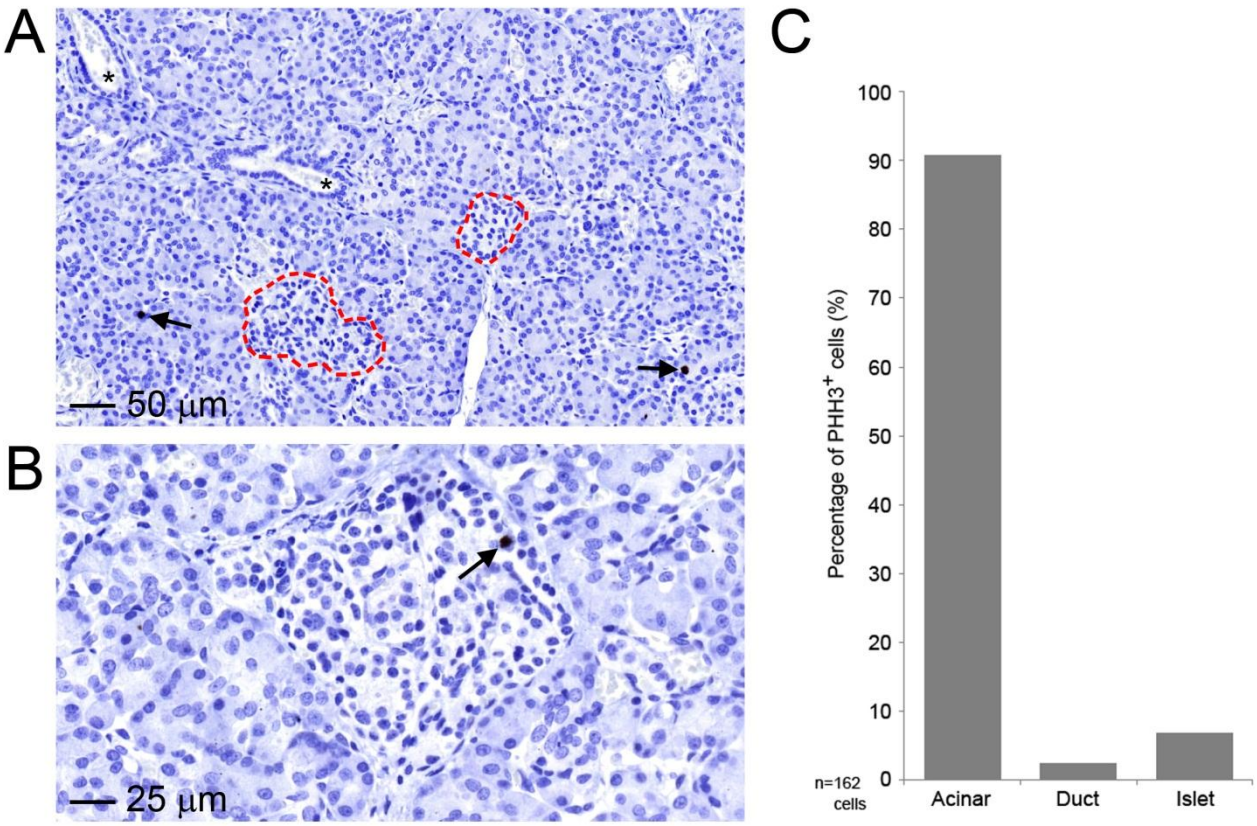
SUPPLEMENTARY DATA

**Supplementary Figure 5. Nucleomegaly in CHI-D affects δ-cells as well as β-cells**
**(A)** Nuclear diameter of islet cells in CHI-D samples compared to age-matched controls. CHI-D cells had larger nuclei on average with a subset having especially large nuclei beyond a normal distribution (nucleomegaly; marked on graph). Diameter was measured using digitization data obtained from histological slides using the HistoQuant and Pannoramic Viewer software. (**B)** Brightfield immunohistochemistry for transcription factors (stained brown) in CHI-D pancreas demonstrating examples of nucleomegaly (arrowheads). Not all nucleomegalic cells stained for PDX1 (e.g. white arrowhead). Size bar represents 20 μm. (**C)** Immunofluorescence counterstained with DAPI. Arrows demonstrate examples of nucleomegalic β-cells (stained for insulin) and δ-cells (stained for somatostatin). Scale bar represents 10 μm.



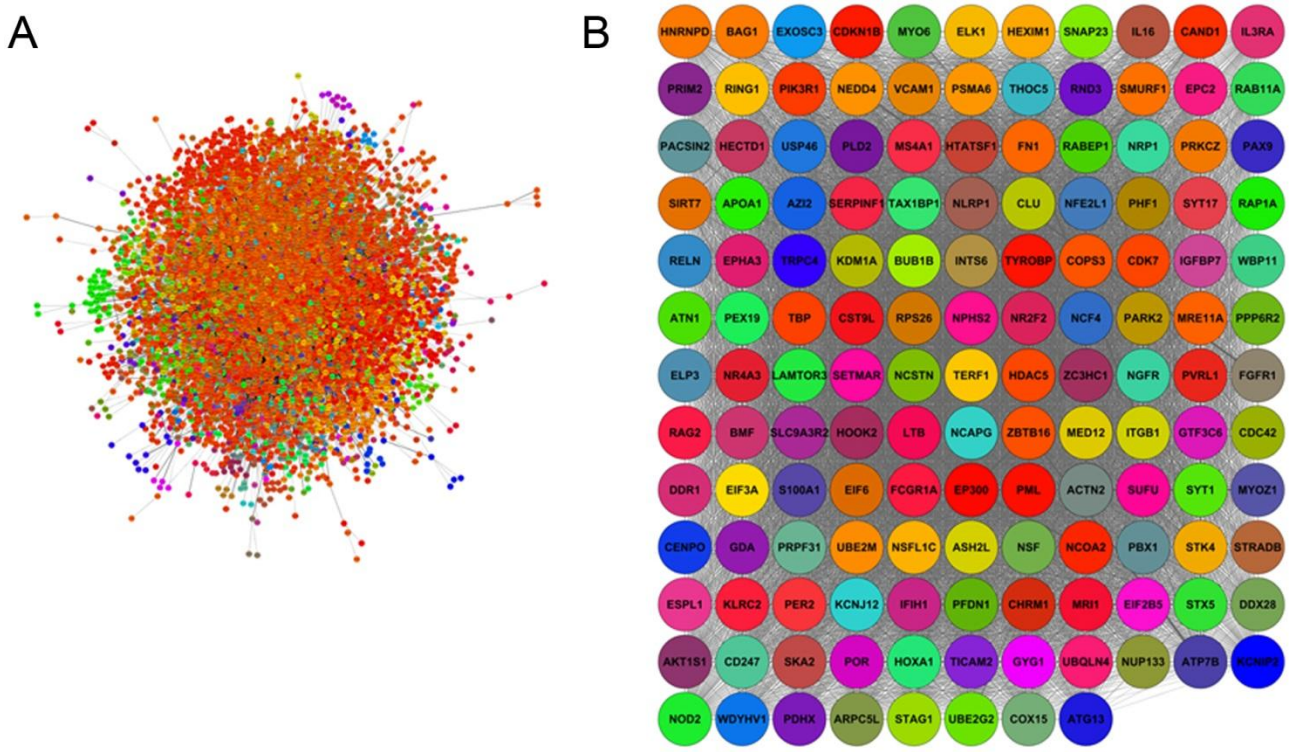
SUPPLEMENTARY DATA

**Supplementary Figure 6. Relative detection of PHH3 in CHI-D pancreas.**
**(A)** PHH3 was predominantly detected in acinar cells (arrows), rather than in ducts (asterisk) or islets (red hatched circles). **(B)** Higher magnification example to show PHH3 detection within an islet. **(C)** Summary data of PHH3-positive cells in two cases of CHI-D. Approximately 90% of PHH3-positive cells were acinar, 7% were islet and 3% were duct (n=162).



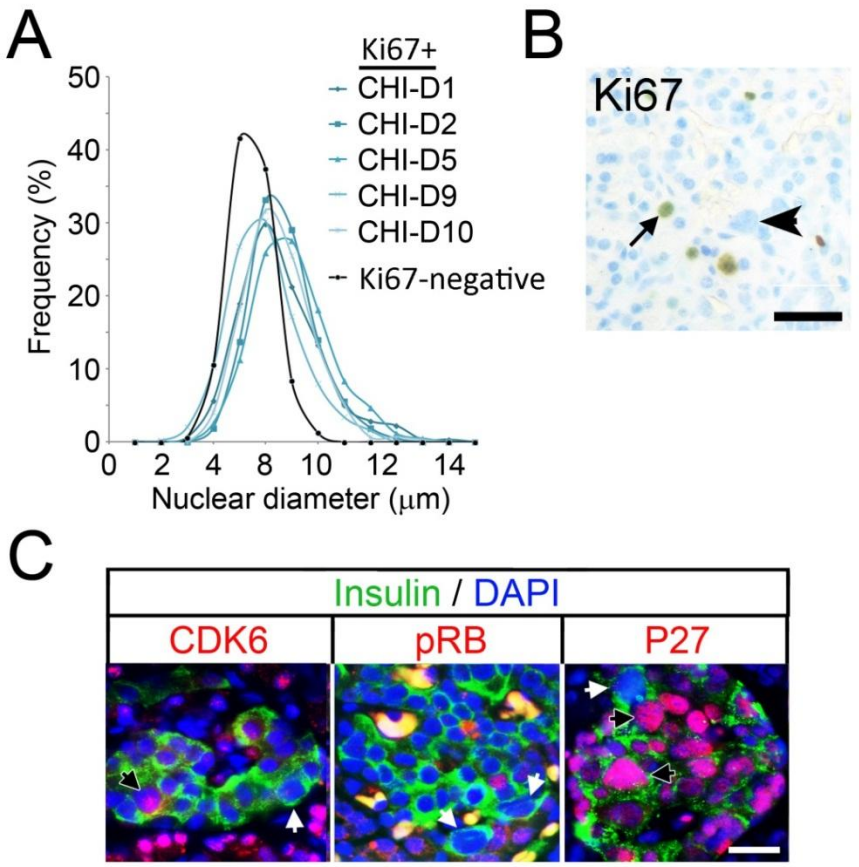
SUPPLEMENTARY DATA

**Supplementary Figure 7. Clusters of genes identified as functional modules within a network model of CHI** **(A)** Interactome model derived from the BioGRID database of significantly altered gene expression in CHI tissue (1). Cluster modules (coloured) were identified with the interactome model using the Moduland algorithm. **(B)** Network of the cluster modules showing the most central gene (top 20 ranking shown in Supplementary Table 4).



SUPPLEMENTARY DATA

**Supplementary Figure 8. Correlation between cell-cycle markers and β-cell nucleomegaly**
**(A)** Ki67-positive nuclei in CHI-D are slightly larger. Frequency histograms showing the range of nuclear diameters recorded in Ki67-positive cells from each of the five cases of CHI-D (blue curves) and the mean diameter from Ki67-negative cells from the same cases (black curve). Comparing nuclear diameter in Ki67-positive and Ki67-negative populations was significant at $P<0.01$. **(B)** Example of nuclear Ki67 staining (arrow) showing not all nucleomegalic cells are proliferative (arrowhead). **(C)** Dual immunofluorescence for CDK6, retinoblastoma protein (pRb) and P27 proteins with insulin counterstained with DAPI showing not all nucleomegalic cells are positively stained. Scale bar represents 20 μm.




**Supplementary Video. Insulin and glucagon colocalization in CHI-D**
Animation captured at x630 magnification moving through a series of cross-sectional z-stack images for a CHI-D cell demonstrating codetection of insulin (green) and glucagon (red).