Insulin is expressed by enteroendocrine cells during human fetal development

Adi Egozi1, Dhana Llivichuzhca-Loja2, Blake T. McCourt2, Keren Bahar Halpern1, Lydia Farack1, Xiaojing An3, Fujing Wang3, Kong Chen3, Liza Konnikova2,3 and Shalev Itzkovitz1,3

Generation of beta cells via transdifferentiation of other cell types is a promising avenue for the treatment of diabetes. Here we reconstruct a single-cell atlas of the human fetal and neonatal small intestine. We identify a subset of fetal enteroendocrine K/L cells that express high levels of insulin and other beta cell genes. Our findings highlight a potential extra-pancreatic source of beta cells and expose its molecular blueprint.

Identifying insulin-expressing cells outside of the pancreatic islets of Langerhans has important implications for potential cell therapy in diabetes. The pancreas shares its developmental origins with the small intestine, budding from the foregut–midgut boundary at ~5 weeks gestational age (GA) in humans. Intestinal enteroendocrine cells share many of their transcriptional programs with pancreatic islet cells, including expression of hormones such as somatostatin and ghrelin, as well as common transcription factors (TFs). Enteroendocrine cells also share common stimulus-secretion mechanisms with pancreatic beta cells, including glucose-stimulated electrical activities and voltage-gated Ca2+ entry. Several studies have shown that perturbations of key TFs can result in insulin-expressing enteroendocrine cells. These included the in vivo inhibition of Foxo1 in Neurog3+ mouse enteroendocrine progenitors or in human intestinal organoids, as well as the simultaneous induction of Pdx1, Mafa and Neurog3 in the mouse small intestine. Fetal development is often accompanied by plasticity in cell identities. We, therefore, sought to explore whether insulin might be endogenously expressed in enteroendocrine cells of the human fetus.

We generated a single-cell atlas of human small intestinal (SI) samples of four fetal samples (two subjects at 21 weeks GA and two subjects at 23 weeks GA) and two neonatal samples (full term, 2 d old and 2 weeks old) (Fig. 1a and Extended Data Fig. 1a,b). The atlas included 36,359 cells, which clustered into 11 cell types (Supplementary Table 1): intestinal epithelial cells, enteroendocrine cells, fibroblasts, endothelial cells, lymphatic endothelial cells, neurons, B cells, T/natural killer (NK) cells, dividing T/NK cells, macrophages and dendritic cells. We performed differential gene expression analysis, highlighting differences between fetal and neonatal cells for each cell type (Extended Data Fig. 1c and Supplementary Table 2). Notably, we identified high expression levels of INS, encoding the insulin protein, in fetal enteroendocrine cells (Fig. 1a and Extended Data Fig. 1c).

To explore the molecular identity of the fetal INS+ cells, we re-clustered the enteroendocrine cells, revealing five distinct groups (Fig. 1b,c). These included enterochromaffin cells, expressing the serotonin catalyzing enzyme tryptophan hydroxylase 1 (TPH1), delta cells expressing somatostatin (SST), N/I cells expressing neurotensin (NTS) and cholecystokinin (CCK), X cells expressing ghrelin (GHRL) and K/L cells expressing gastric inhibitory polypeptide (GIP) and the gene encoding glucagon/glucagon-like peptides (GCG) (Fig. 1b,c). We identified distinct markers for each of these cell types (Fig. 1c and Supplementary Table 3). We further performed differential gene expression analysis that highlighted differences between fetal and neonatal cells for each of the enteroendocrine cell subtypes (Extended Data Fig. 2a and Supplementary Table 4).

INS was predominantly expressed in a subset of the fetal K/L cells (nine INS+ cells out of 37 fetal K/L cells, hypergeometric test, \( P < 4 \times 10^{-15} \); Fig. 1b). We denoted these cells as fetal INS + K/L (FIKL) cells. To determine if other beta cell genes were preferentially expressed in FIKL cells, we analyzed the expression of key beta cell genes in FIKL cells and other enteroendocrine lineages, as well as in fetal and adult human beta cells (Fig. 1d and Supplementary Table 5). FIKL cells, as well as other enteroendocrine lineages, expressed PCSK1, encoding the proprotein convertase 1, needed for conversion of pro-insulin to insulin. We observed similar broad expression among different enteroendocrine lineages for the endocrine cell granulogenic factor CHGA; for PDX1, encoding a TF associated with beta cell maturation and insulin transcription; and for UCN3, encoding a marker of beta cell maturation (Fig. 1d). Notably, expression of UCN3 in FIKL cells was higher than in fetal human pancreatic beta cells (Fig. 1d and Supplementary Table 5). FIKL cells also expressed high levels of GCK, encoding the glucokinase enzyme, which acts as a ‘glucose sensor’ in beta cells and other endocrine cells. In contrast, FIKL cells did not express IAPP, encoding the islet amyloid polypeptide, a hormone promoting satiety that is co-secreted with insulin in adult beta cells (Fig. 1d and Supplementary Table 5). Similarly, we observed only low levels of IAPP in fetal pancreatic beta cells (Fig. 1d and Supplementary Table 5).

The correlation distance between the transcriptome of adult beta cells and FIKL cells was significantly smaller than the distances from other fetal enteroendocrine cell subtypes, even when excluding the INS gene (Extended Data Fig. 2b; Wilcoxon rank-sum \( P = 1 \times 10^{-31} \) between FIKL and K/L INS- cells, the second closest cell type). The most upregulated gene in the FIKL cells was RGS16, encoding a regulator of G-protein signaling that has been shown to promote insulin secretion (Extended Data Fig. 2c; average expression of 4.8×10^5 in FIKL cells versus 7.13×10^-6 in INS- fetal K/L cells, \( P < 0.0015 \)).

To further delineate the molecular signature of FIKL cells, we performed gene regulatory network analysis using the SCENIC tool.

1Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel. 2Department of Pediatrics and Department of Obstetrics, Gynecology and Reproductive Sciences, Human and Translational Immunology, Yale Medical School, New Haven, CT, USA. 3Department of Medicine, University of Pittsburgh Medical Center, Montefiore Hospital, Pittsburgh, PA, USA. ✉ e-mail: liza.konnikova@yale.edu; shalev.itzkovitz@weizmann.ac.il
We found that FIKL cells exhibited elevated activity levels of the TFs NR1H3 and HOXB5 (Extended Data Fig. 3a–c). FIKL cells exhibited reduced activities of multiple TFs, including PAX6 and MAFB, which peaked in activity in neonatal K/L cells (Extended Data Fig. 3d,e); ARX, which peaked in activity in X cells and neonatal K/L cells (Extended Data Fig. 3f); and the enterochromaffin cell TF LMX1A (Extended Data Fig. 3g).

To explore whether FIKL cells appear at earlier developmental time points, we analyzed a recent single-cell atlas of human SI cells during the first trimester14 (8–12 weeks GA; Extended Data Fig. 4a). We found that several enteroendocrine cells also express INS during this earlier developmental stage (Extended Data Fig. 4b). Notably, INS+ cells were enriched in the K/L cluster at these earlier time points as well (hypergeometric test, P < 1.3 × 10⁻⁵; Extended Data Fig. 4c).

To validate the existence of INS+ enteroendocrine cells in the fetal intestine, we next used single-molecule fluorescence in situ hybridization (smFISH) of insulin mRNA combined with immunofluorescence antibody staining of the insulin protein15 (Fig. 2 and Extended Data Fig. 5a–f). We hybridized and imaged sections from 14 fetal human subjects of GAs between 12 weeks and 23 weeks, as well as sections from three neonatal subjects. We identified multiple cells with high expression levels of insulin mRNAs and proteins in four fetal samples (12 weeks, 14 weeks, 18 weeks and 21 weeks; Extended Data Fig. 5c–f and Supplementary Table 7) and in none of the neonatal samples. To estimate the frequencies of INS+ enteroendocrine cells, we co-hybridized tissues with

Fig. 1 | Fetal human K/L enteroendocrine cells contain a subset of INS+ cells. a. UMAP of human SI cells colored by INS expression. Cluster names are labeled. Blue—high expression. b. UMAP of the enteroendocrine cells in the atlas of a. Gray cells are fetal; black cells are neonatal; and yellow-red hues correspond to INS expression (log₁₀ of the sum-normalized UMI counts). c. Expression of the top markers for each enteroendocrine cell type. d. Expression of beta cell genes. K/L cells were split into neonatal, fetal INS− and fetal INS+ (FIKL) cells. Fetal and adult beta cell expression signatures were extracted from refs. 8,9. Dot color corresponds to average expression (log₁₀ of the sum-normalized UMI counts); size corresponds to the fractions of cells with non-zero reads. EC, enterochromaffin cell; UMAP, uniform manifold approximation and projection.
smFISH probes for CHGA, encoding a gene broadly expressed by all enteroendocrine cells (Fig. 1d). INS⁺ cells were much scarcer than cells positive for CHGA (fraction of 0.02 ± 0.01 of CHA⁺ cells, averaged over 47 imaging fields from four fetal samples). This fraction, based on in situ validation, was similar to the fraction of INS⁺ cells in the single-cell atlas (11 INS⁺ cells out of 453 enteroendocrine cells, a fraction of 0.024). INS⁺ cells exhibited a distinct intra-cellular polarization with insulin mRNA localized toward the apical sides of the cells and insulin protein polarized to the basal sides (Fig. 2a,b and Extended Data Fig. 5). A similar polarization was previously shown in adult mouse beta cells and was also apparent when imaging insulin mRNA and protein in adult human pancreatic islet sections (Fig. 2c). We further used smFISH to validate co-expression of INS and GCG (Fig. 2d,e). Most INS⁺ cells appeared as isolated cells (Extended Data Fig. 6a–f), with an exception of two of 47 imaging fields, in which they appeared to be clustered (Fig. 2a,b). Using smFISH, we could not detect INS⁺ cells in the mouse embryonic and neonatal small intestine (six mice at embryonic day (E) 16.5 and five mice at P0; Extended Data Fig. 6a,b).

Fig. 2 | In situ validation for the expression of INS⁺ cells in the fetal human small intestine. a, Intestinal section of 12-week GA fetal small intestine simultaneously stained with smFISH probe library for INS-mRNA (magenta) and an antibody for INS-protein (green). Scale bar, 50 μm. b, Blowup of INS⁺ cells showing the apical polarization of INS-mRNA (magenta) and basal polarization of INS-protein (green). Scale bar, 5 μm. Autofluorescent blobs are marked with gray arrows. c, Pancreatic human islet (nPOD6447) stained for INS-mRNA (magenta) and INS-protein (green) showing a similar apical–basal mRNA–protein polarization to that observed in the intestine. Scale bars, 20 μm (5 μm for blowups). d, e, An INS⁺ enteroendocrine cell stained with INS-mRNA smFISH probe library (magenta) and antibody against INS-protein (green, d) and INS-mRNA (magenta) and GCG-mRNA (green, e). Scale bars, 5 μm. In all images, nuclei are stained by DAPI (blue). Subject 1170 (12 weeks GA) (a, b); subject 1147 (14 weeks GA) (d, e).
The physiological role of the intestinal INS+ enteroendocrine cells is unclear and could have either a local paracrine effect or a systemic endocrine effect. A systemic effect on the mother's glucose homeostasis is unlikely because insulin cannot transfer through the placenta (Methods, 'Estimation of the number of INS+ cells'). Moreover, the numbers of intestinal fetal INS+ cells are around four orders of magnitude smaller than the number of beta cells in the mother's pancreas. Notably, the numbers of FIKL cells are less than two orders of magnitude smaller than the estimated number of fetal pancreatic beta cells, and so one cannot exclude a systemic effect on fetal glucose homeostasis (Methods, 'Estimation of the number of INS+ cells'). To assess the potential for local paracrine effects, we analyzed the expression of the insulin receptor (INSR) in all cells included in our single-cell atlas, as well as in a recent human fetal cell atlas of the first trimester small intestine41. We found particularly elevated expression of INSR in intestinal epithelial cells, in enteroendocrine cells and, most prominently, in endothelial cells (Extended Data Fig. 7a,b). Insulin has been shown to affect vasculature by stimulating the production of the vasodilator molecule nitric oxide and by promoting vessel sprouting18. Insulin can also act as a general growth factor19, potentially stimulating intestinal tissue expansion at the fetal stages. Notably, the fact that FIKL cells were not observed in all samples indicates that they might not carry essential physiological functions. Bi-hormonal pancreatic cells, expressing both insulin and glucagon, are commonly seen in the human and mouse fetal pancreas and are thought to be eliminated rather than maturing into either alpha cells or beta cells17–19. The FIKL cells might be the intestinal equivalent of these pancreatic bi-hormonal cells.

The scarcity of the INS+ enteroendocrine cells currently prohibits secretion measurements in organoids. An open question is what leads to the cessation of insulin expression in the neonatal small intestine. Our analysis uncovered some molecular differences between fetal and neonatal K/L cells, including lower expression of the K/L TFs ARX and PAX6. The loss of insulin expression post-natally might be a result of processes occurring at the rare NEUROG3+ enteroendocrine progenitors, which were not captured in our single-cell RNA sequencing data. Endocrine cell identity is thought to be shaped by the repression of ‘disallowed genes’ by mechanisms such as microRNAs20. It will be interesting to consider potential microRNA regulators that might be differentially expressed in FIKL cells. The SI tissues analyzed in our study were not annotated for the precise segment (duodenum, jejunum or ileum). It will be interesting to explore whether FIKL cells exist throughout the small intestine or, rather, are limited to specific segments. The induction of insulin-expressing enteroendocrine cells is a promising avenue for cell therapy in diabetes1. Our finding of endogenous expression of insulin in fetal K/L cells provides an important demonstration of the ability of such cells to surpass lineage barriers toward beta cells and exposes the molecular blueprint that is compatible with a beta cell phenotype.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41591-021-01586-1.

Received: 27 May 2021; Accepted: 22 October 2021; Published online: 9 December 2021

References
1. Merrell, A. J. & Stanger, B. Z. Adult cell plasticity in vivo: de-differentiation and transdifferentiation are back in style. Nat. Rev. Mol. Cell Biol. 17, 413–425 (2016).
2. Jennings, R. E., Berry, A. A., Strutt, J. P., Gerrard, D. T. & Hanley, N. A. Human pancreas development. Development 142, 3126–3137 (2015).
3. Habib, A. M. et al. Overlap of endocrine hormone expression in the mouse intestine revealed by transcriptional profiling and flow cytometry. Endocrinology 153, 3054–3065 (2012).
4. Tulkch, C., Xuan, S., Kitamura, T., DePinho, R. A. & Accili, D. Generation of functional insulin-producing cells in the gut by Foxo1 ablation. Nat. Genet. 44, 406–412 (2012).
5. Bouchi, R. et al. FOXO1 inhibition yields functional insulin-producing cells in human gut organoid cultures. Nat. Commun. 5, 4242 (2014).
6. Chen, Y.-J. et al. De novo formation of insulin-producing ‘neo-β’ cell islets' from intestinal crypts. Cell Rep. 6, 1046–1058 (2014).
7. Sissa, Y. et al. Gastrin: a distinct fate of neurogenin3 positive progenitor cells in the embryonic pancreas. PLoS ONE 8, e70397 (2013).
8. Cao, J. et al. A human cell atlas of fetal gene expression. Science 370, eaau7721 (2020).
9. Baron, M. et al. A single-cell transcriptomic map of the human and mouse pancreas reveals inter- and intra-cell population structure. Cell Syst. 3, 346–360 (2016).
10. Blum, B. et al. Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. Nat. Biotechnol. 30, 261–264 (2012).
11. Matschinsky, F. M. & Wilson, D. F. The central role of glucokinase in glucose homeostasis: a perspective 50 years after demonstrating the presence of the enzyme in islets of Langerhans. Front. Pharmacol. 10, 148 (2019).
12. Vivot, K. et al. The regulator of G-protein signaling RGS16 promotes insulin secretion and β-cell proliferation in rodent and human islets. Mol. Metab. 5, 988–996 (2016).
13. Gross, S. et al. Lmx1a functions in intestinal serotonin-producing enterochromaffin cells downstream of Nkx2.2. Development 143, 2616–2628 (2016).
14. Elmentaitre, R. et al. Single-cell sequencing of developing human gut reveals transcriptional links to childhood Crohn's disease. Dev. Cell 55, 771–783 (2020).
15. Farack, L. et al. Transcriptional heterogeneity of beta cells in the intact pancreas. Dev. Cell 48, 115–125 (2019).
16. Nowak-Sliwinska, P. et al. Oncofetal insulin receptor isoform A marks the tumour endothelium; an underestimated pathway during tumour angiogenesis and angiostatic treatment. Br. J. Cancer 120, 218–228 (2019).
17. Menon, R. K. & Sperling, M. A. Insulin as a growth factor. Endocrinol. Clin. North Am. 25, 633–647 (1996).
18. Herrera, P. L. Adult insulin- and glucagon-producing cells differentiate from intestinal crypts. Development 127, 2317–2322 (2000).
19. Teitelman, G., Alpert, S., Polak, J. M., Martinez, A. & Hanahan, D. Precursor cells of mouse endocrine pancreas coexpress insulin, glucagon and the neuronal proteins tyrosine hydroxylase and neuropeptide Y, but not pancreatic polypeptide. Development 118, 1031–1039 (1993).
20. Kaspi, H., Pasvolsky, R. & Hornstein, E. Could microRNAs contribute to the maintenance of β cell identity? Trends Endocrinol. Metab. 25, 285–292 (2014).
Brief Communication

Nature Medicine

Methods

Intestinal tissue acquisition and storage. Human fetal SI tissue was obtained from the University of Pittsburgh Biospecimen Core after institutional review board (IRB) approval (IRB no. PRO180101491; Supplementary Table 7). The reasons for elective fetal terminations were not disclosed. Human fetal SI tissue was collected in an unidentified fashion through a biobank intermediary who obtained signed informed consent from the mother. Explicit signed permission was obtained for the specimen to be used for general research purposes. All specimen dissections were performed by the biobank. No compensation was provided to the participant for the donated tissue. Fresh SI tissue from human neonatal samples was collected from surgical resections in infants with IRB approval (IRB no. PRO17072226; Supplementary Table 7). No consent was obtained for neonatal samples as they were collected without any identifying information under a discarded specimen protocol that were deemed non-human research by the University of Pittsburgh IRB. For single-cell sequencing and cryoblocks, tissue was cryopreserved23. In brief, intestinal tissue samples were cut into sub-centimeter pieces and cryopreserved in freezing media (10% DMSO in FBS) in a slow-cooling container (Mr. Frosty) at −80°C for 24 h and then transferred into liquid nitrogen for long-term storage. For paraffin blocks, tissue was fixed in 4% formalin for 48 h and transferred to ethanol until embedded in paraffin. Blocks were stored in paraffin until sectioned for analysis.

Intestinal tissue digestion. Cryopreserved samples were processed as previously described23. In brief, intestinal tissue samples were quickly thawed and washed in RPMI medium plus 10% FBS (Corning), 1 mM sodium pyruvate, 100 IU ml−1 of streptomycin. Next, intestinal tissue was incubated overnight in the same media the specimen to be used for general research purposes. All specimen dissections were performed using a 100 bootstrapping iterations, in which cells were sampled with replacement from each cell type, and distances were re-calculated. Enrichment of INS+ cells in individual cell clusters was calculated using hypergeometric tests. For differential gene expression, analyses excluded all cytoplasmic and mitochondrial ribosomal genes and re-normalized all gene expression to the sum of the remaining genes. P values were calculated using two-sided Wilcoxon rank-sum tests. Q values were computed using the Benjamini–Hochberg false discovery rate correction.

Gene regulatory network analysis was performed using the SCENIC tool (version 0.11.2) with default parameters25. Area under the curve (AUC) scores per each cell were computed (Supplementary Table 6) and used for downstream analyses. Mean AUC scores were computed for each cell type. A subset of differentially active TFs was extracted by identifying genes with a AUC above 0.01 and higher than two-fold the mean in any other cell type. If more than six TFs conformed to this criterion, the six TFs with the highest ratio were retained. Clustering was performed on the z-score-transformed matrix of the mean AUC values of these TFs (Extended Data Fig. 3a). For the first trimester human intestinal cells in Extended Data Figs. 4 and 7 was obtained from Elmentaita et al.23 using cells from the small intestine. In Extended Data Fig. 4, the two datasets were integrated using the Seurat IntegrateData command, and cells with mitochondrial fraction above 10% were filtered out. INS+ cells were defined as cells with INS sum-normalized expression above 5×10−6. Cluster annotations in Extended Data Fig. 7a were validated using the published metadata and data, and clusters containing fewer than 10 cells were filtered out. Expression was normalized using Seurat log normalization; feature counts for each cell were divided by the total counts, multiplied by 104, incremented with a count of 1 and natural-log transformed.

Mouse experiments. All mice experiments were approved by the Institutional Animal Care and Use Committee of the Weizmann Institute of Science (protocol no. 06530821-2) and performed in accordance with institutional guidelines. Pregnant C57BL/6 female mice (age 8–10 weeks) were purchased from Envigo. Female pregnant mouse at E16.5 was killed by cervical dislocation, the same day of arrival from Envigo. Before arrival, mice were housed at five per cage, maintained at a constant temperature of 20–24°C and humidity of 50–80%, exposed at all times to a 12-h light/dark cycle and had access to food and water ad libitum. Uterus was extracted and embryos were kept in the embryonic sac in PBS until processed. Each embryo was decapitated, and whole intestine was extracted to 4% formaldehyde (FA). For P0 pups, each pup was decapitated, and whole intestine tissue was extracted to 4% FA. Tissues were fixed for 4 h at 4°C, followed by overnight incubation in 30% sucrose in 4% FA. Fixed tissues were embedded in OCT (Scigen, 4586). Then, 7-μm-thick sections of fixed intestinal tissues were sectioned onto poly-l-lysine-covered coverslips and fixed again in 4% FA in PBS for 15 min, followed by 70% ethanol dehydration for 2h in 4°C. Eleven mice were imaged: six E16.5 embryos and five P0.

Immunofluorescence combined with smFISH. smFISH combined with immunofluorescence was performed on both paraffin-embedded tissue sections and frozen sections, with a modified smFISH protocol that was optimized for human intestinal tissue based on a protocol by Farack et al.23 and Massalha et al.24. For frozen sections, intestines were fixed in 4% FA (T.Baker, JT2106) in PBS for 1-2 h and subsequently agitated in 30% sucrose and 4% FA in PBS overnight at 4°C. Fixed tissues were embedded in OCT (Scigen, 4586). Then, 5-8-μm-thick sections of fixed intestinal tissues were sectioned onto poly-l-lysine-coated coverslips and fixed again in 4% FA in PBS for 15 min, followed by 70% ethanol dehydration for 2h in 4°C.

For paraffin sections, paraffin blocks were cut into 3-μm sections. Sections were deparaffinized in xylene (Sigma, 183164) twice for 10 min and then moved to 1:1 limonene and ethanol solution for 10 min. Tissues were incubated in 100% ethanol for 5 min and then 10 min, followed by 95% ethanol for 10 min. Finally, the tissues were transferred to 70% ethanol for 2h in 4°C.

For both frozen and paraffin sections, human and mouse tissues were incubated for 10 min with proteinase K in 50°C or 37°C, respectively (10 μg·ml−1 of Ambion AM2546), and washed twice with 2x SSC (Ambion AM9765). Tissues were incubated in wash buffer (20% formamide, Ambion AM9342, and 2x SSC) for 15 min (for frozen sections) or 5 h (for paraffin sections) and mounted with the hybridization mix above. Hybridization mix contained hybridization buffer (10% dextran sulfate, Sigma D8906; 30% formamide; 1 ml·1 of Escherichia coli tRNA, Sigma R1753; 2x SSC; 0.02% BSA, Ambion AM2616; and 2 mM vanadylribonucleoside complex, New England Biolabs S1402S) mixed with 1:3000 dilution of probes and 1:900 dilution of guinea pig anti-insulin antibody (polyclonal, Dako, cat. A0564, lot 10199333). Hybridization mix was incubated with tissues for overnight in a
30°C incubator. smFISH probe libraries (Supplementary Table 9) were coupled to Cy5, TMR or Alexa Fluor 594. After the hybridization, tissues were washed with wash buffer and then incubated with secondary antibody donkey anti-guinea pig Cy5 (1:300, polyclonal, Jackson Laboratories, cat. 706-165-148, lot 1554462) or goat anti-guinea pig FITC (1:300, polyclonal, Abcam, cat. ab6904, lot GR3345928-1) for 20 min at room temperature.

Tissues were moved to GLOX buffer (0.4% glucose, 1% Tris and 10% SSC) containing 50 ng mL−1 of DAPI (Sigma, D9542) for 10 min, followed by GLOX buffer wash. Probe libraries were designed using Stellaris FISH Probe Designer software (Biosearch Technologies) (Supplementary Table 9).

Imaging. Imaging was performed on Nikon eclipse Ti2 inverted fluorescence microscopes equipped with x100 and x60 oil-immersion objectives and a Photometrics Prime 95B 25-mm EMCCD camera. Image stacks were collected with a z spacing of 0.3–0.5 μm. Identification of positive cells was performed using Fiji15. All images included an additional fluorescence channel to enable identification of autofluorescent elements, which appear in multiple channels. Only cells with signal for both INS protein and mRNA and no signal in the autofluorescence channel were counted as INS cells.

Estimation of the number of INS+ cells. A typical adult human has 10⁹ pancreatic beta cells16. An adult human intestine contains around 10¹⁰ epithelial cells17. Therefore, a fetus at GA 21 weeks, weighing 420 g,19 would have an order of magnitude of 10¹⁰ × 0.420/70=6 × 10⁶ intestinal epithelial cells (assuming the reference man weight of 70 kg)20. In our imaging, we estimated that the fraction of INS+ cells out of the total intestinal epithelial cells is around 1/50,000 (2% of enterocendocrine cells, which themselves are around 1% of the epithelial cells18), yielding an estimated total number of fewer than 10⁶ INS+ cells throughout the small intestine. This number is four orders of magnitude lower than the numbers of beta cells in the adult pancreas but only around 60-fold smaller than the estimated numbers of fetal beta cells (10⁶ × 0.420/70=6 × 10⁶). Because insulin cannot transfer through the placenta21, FIKL cells most likely do not affect the mother’s glucose homeostasis but might play a role in maintaining glucose homeostasis for the fetus, especially if there is a delay in fetal pancreatic beta cell maturation.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All data generated in this study are available at the Zenodo repository: https://doi.org/10.5281/zenodo.5457926. Immune cells from fetal samples 1100A and 1102 and the neonatal sample 1127 were presented in Olahoye et al.22. Data from Cao et al.23 are available at https://descartes.humenmatlab.org/bba/human-gene-expression-during-development/. Data from Baron et al.24 are available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GESE4133. Data from Elmentaite et al.25 are available at https://www.gutcellatlas.org/.

Code availability
All custom code used in this study is available at the Zenodo repository: https://doi.org/10.5281/zenodo.5457926.

References
21. Stras, S. F. et al. Maturation of the human intestinal immune system occurs early in fetal development. Dev. Cell 51, 357–373 (2019).
22. Konnikova, L. et al. High-dimensional immune phenotyping and transcriptional analyses reveal robust recovery of viable human immune and epithelial cells from frozen gastrointestinal tissue. Microsurg. Immunol. 11, 1684–1693 (2018).
23. Xin, H. et al. GMM-Demux: sample demultiplexing, multiplet detection, experiment planning, and novel cell-type verification in single cell sequencing. Genome Biol. 21, 188 (2020).
24. Luo, J., Erb, C. A. & Chen, K. Simultaneous measurement of surface proteins and gene expression from single cells. In: T-Cell Receptor Signaling Vol. 2111 35–46 (ed Liu, C.) (Springer, 2020).
Extended Data Fig. 1 | Analysis of differentially-expressed genes between the human fetal and neonatal small intestines. a–b, UMAP of small intestinal cells colored by age (a) or subject (b). c, MA plot showing differentially expressed genes between fetal and neonatal cells stratified by cell type. Genes in red have q-values below 0.1 and expression above 5e-5. INS, the most up-regulated gene in fetal enteroendocrine cells, is highlighted with a blue box.
Extended Data Fig. 2 | Characterization of the fetal enteroendocrine cell types. a, MA plot showing differentially expressed genes between fetal and neonatal enteroendocrine cells stratified by cell type. Genes in red have q-values below 0.05 and expression above 1e-4. INS, the most up-regulated gene in fetal K/L cells, is highlighted with a blue box. b, Spearman correlation distances between adult beta cells and the fetal endocrine cell types. Each dot represents a distance obtained from one of 100 bootstrap iterations, in which cells were sampled with replacement from the complete dataset. Distances from beta cells are significantly smaller for FIKL cells compared to all other enteroendocrine cell types (two-sided Wilcoxon rank-sum p = 1e-31 for the differences between FIKL cells and K/L INS- cells, the second closest enteroendocrine cell population). White circles are medians, gray boxes mark the 25–75 percentiles, gray lines extend from 1.5 times the interquartile range (IQR) above the 75 percentile to 1.5 times the IQR below the 25 percentile and truncated at the minimal or maximal measured values. c, RGS16 expression (log10 of the sum-normalized UMI counts).
Extended Data Fig. 3 | Analysis of gene regulatory networks. a, Clustergram of TF activities Z-score of mean AUC for each cell type. The list includes TFs with the most differential activities in the different cell types (Methods). b-g, Selected TFs from (a). In b-g, each dot is a cell, white circles are medians, gray boxes mark the 25–75 percentiles, gray lines extend from 1.5 times the interquartile range (IQR) above the 75 percentile to 1.5 times the IQR below the 25 percentile and truncated at the minimal or maximal measured values. Black lines connect the means.
Extended Data Fig. 4 | FIKL cells appear in earlier developmental time points. **a-b**, Combined UMAP of 1st trimester and 2nd trimester colored by enteroendocrine subtype (**a**) or INS expression (log10(normalized expression)), panel **b**. Dots with a black outline in **a-b** are 1st trimester cells, dots without an outline are 2nd trimester cells. **c**, Heatmap of INS+ cells enrichment in each enteroendocrine subtype. Colors are -log10(hypergeometric p-value for the enrichment of INS+ cells within each cell cluster).
Extended Data Fig. 5 | See next page for caption.
Extended Data Fig. 5 | In-situ validation of FIKL cells at different developmental stages. a, Intestinal section with a typical example of an INS+ cell marked with INS protein (green) from a 12 weeks GA fetus. Scale bar - 50 µm. b, Blowup of the INS+ cell from (a) stained with both INS-mRNA (magenta) and INS-protein (green). Scale bar - 5 µm. c-f, Representatives examples of INS+ cells in small intestine tissues from 12 weeks GA (c, 3 cells out of 20 found), 14 weeks GA (d, 2 cells out of 3 found), 18 weeks GA (e, 1 cell out of 1 found) and 21 weeks GA (f, 3 cells out of 23 found) stained with INS-mRNA (magenta), INS-protein (green). Scale bar - 5 µm. In all images nuclei are stained by DAPI (blue).
Extended Data Fig. 6 | See next page for caption.
Extended Data Fig. 6 | Insulin is not detected in embryonic mouse small intestine. **a**, Mouse embryo intestinal section (E16.5) stained with the pan-enteroendocrine cell marker Chga (gray) and blowups of selected Chga+ cells. Scale bar - 50 μm for big image and 10 μm for blowups. **b**, Mouse embryo pancreatic tissue stained with Ins2-mRNA (magenta) and INS-protein (green) shown as a positive control for the in-situ detection of mouse cells expressing insulin. Scale bar - 10 μm. In all images nuclei are stained by DAPI (blue). Images are representative of 6 E16.5 embryonic subjects and 5 P0 neonatal subjects.
Extended Data Fig. 7 | Expression of INSR in small intestinal cell types. INSR expression in all cell types of 1st trimester small intestinal data from^16 (a) and from 2nd trimester small intestine (b). Cell types are ranked by the mean INSR expression. Expression units are Seurat normalized (Methods).
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
- A description of any covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

Give P values as exact values whenever suitable.

- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

Software for data collection used: Cell Ranger 4.0.0

Data analysis

Single cell data analysis was performed using R version 4.0.2, Seurat package version 3.2.2, SCENIC version 0.11.2. Image analysis was performed using Fiji software.

Additional custom codes are available at the Zenodo repository under the following: https://doi.org/10.5281/zenodo.5457926

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All data generated in this study is available at the Zenodo repository under the following https://doi.org/10.5281/zenodo.5457926. Immune cells from fetal samples 1100A, 1102 and the neonatal sample 1127 have been presented in Olaloye et al. Data from Cao, J. et al. is available at https://descartes.brotmanbary.org/bbi/human-gene-expression-during-development/. Data from Baron, M. et al. is available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84133. Data from Elmentale, R. et al. is available at https://www.gutcellatlas.org/.
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](http://nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | Sample size were determined by availability of donors within the sampling ages. No statistical methods were used to calculate appropriate sample size. We followed standards in the field and Human Cell Atlas criteria. |
| Data exclusions | Single cells used in the final analysis were filtered based on criteria detailed in the Methods section (number of genes, U/Ms and mitochondrial content). Non excluded raw data is available at the Zenodo repository https://doi.org/10.5281/zenodo.5457925. |
| Replication | Our findings were replicated in multiple independent human samples, both for the scRNAseq identification of the intestinal INS+ cells and for the in-situ validations. When displaying representative images (Extended Data Figure 5) we have reported the number of replicates. |
| Randomization | Sample collection was based on availability of donors. We allocated donor samples into developmental groups based on age. |
| Blinding | Blinding was not relevant in this study, since the aim was the identification of the existence of a rare sub-population in the intestine. All experiments included a positive control pancreatic sample to ensure smFISH libraries and IF antibodies were working robustly. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---|---|
| n/a | n/a |
| □ X Antibodies | □ X Involved in the study |
| □ | □ ChIP-seq |
| □ | □ Flow cytometry |
| □ | □ MRI-based neuroimaging |
| □ Eukaryotic cell lines | □ Palaeontology and archaeology |
| □ □ □ □ □ | □ Palaeontology and archaeology |
| □ Palaeontology and archaeology | □ Palaeontology and archaeology |
| □ □ □ □ □ | □ Palaeontology and archaeology |
| □ Human research participants | □ Human research participants |
| □ □ □ □ □ | □ Human research participants |
| □ Clinical data | □ Clinical data |
| □ □ □ □ □ | □ Clinical data |
| □ Dual use research of concern | □ Dual use research of concern |

Antibodies

| Antibodies | guinea pig anti-insulin antibody (1:900, polyclonal, Dako, cat: A0564, lot: 10109333), donkey anti-guinea pig Cy3 (1:300, Polyclonal, Jackson Laboratories cat:706-165-148, lot: 154462), goat anti-guinea pig FITC (1:300, Polyclonal, Abcam cat: ab6904, lot: GR3345928-1). |

Validation

All antibodies were validated on published papers.

Guinea pig anti-insulin antibody (Dako, A0564) was used in: Massumi, Mohammad, et al. "An abbreviated protocol for in vitro generation of functional human embryonic stem cell-derived beta-like cells." PloS One 11.10 (2016): e0164457. Used for IF and FC analysis.

Donkey anti-guinea pig Cy3 (Jackson Laboratories 706-165-148) antibody was used in: Belle, Morgane, et al. "3Dimensional visualization and analysis of early human development." Cell 169.1 (2017): 161-173.

Goat anti-guinea pig FITC (Abcam ab6904) antibody was used in: Farack, Lydia, et al. "Transcriptional heterogeneity of beta cells in the intact pancreas." Developmental cell 48.1 (2019): 115-125.

Animals and other organisms

Policy information about studies involving animals, ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | All mice experiments were approved by the Institutional Animal Care and Use Committee of the Weizmann Institute of Science (protocol #06530821-2) and performed in accordance with institutional guidelines. Pregnant C57BL/6 female mice (age 8-10 weeks) were used in order to work on embryos at E16.5 (6 subjects) and pups at P0 (5 subjects). Mice were sacrificed at the same day of |
arrival from Envigo. Before arrival, mice were housed at 5 per cage, maintained at a constant temperature of 20-24°C and humidity of 50-80%, exposed at all times to a 12h light/12h dark cycle and had access to food and water ad libitum.

Wild animals
The study did not involve wild animals.

Field-collected samples
the study did not involve samples collected from the field.

Ethics oversight
All mice experiments were approved by the Institutional Animal Care and Use Committee of the Weizmann Institute of Science and performed in accordance with institutional guidelines.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics
Our study analyzed 2 fetal samples at 21 weeks GA (one female one unknown), 2 fetal samples at 23 weeks GA [one female one unknown] and 2 neonatal samples [both females] for single cell RNA sequencing. In-situ validations were performed on 1 sample at 12 weeks GA (unknown gender), 1 sample at 14 weeks GA (male), 1 sample at 15 weeks GA (unknown gender), 1 sample at 17 weeks GA (male), 2 sample at 18 weeks GA (male), 1 sample at 19 weeks GA (unknown gender), 4 sample at 21-22 weeks GA [2 females, 2 unknown], 3 sample at 23 weeks GA [2 females, 1 unknown], 3 Neonatal samples [females].

Recruitment
Fresh small intestinal (SI) tissue from human neonatal samples were obtained from surgical resections in infants with IRB approval (IRB# PRO17070226). Human fetal SI was obtained through University of Pittsburgh Biospecimen core after IRB approval (IRB# PRO18010491). Fetal samples: All subjects undergoing therapeutic termination procedures at the University of Pittsburgh Medical School were approached by the biobank intermediary about participating in research. Informed consent was obtained from those that expressed interest. Reasons for elective terminations were not disclosed to the researchers and thus no overt bias was identified. A signed maternal informed consent was obtained for all fetal samples. Neonatal samples: Samples were collected from all subjects from birth to two years of age undergoing intestinal surgery from the department of pathology prior to fixation in formalin. All samples were provided in a deidentified fashion. No overt bias was identified. No consent was obtained for neonatal samples as they were collected without any identifying information under a discarded specimen protocol that was deemed non-human research by the University of Pittsburgh IRB.

Ethics oversight
All experiments were approved by that University of Pittsburgh Institutional Review Board, IRB# PRO18010491 (Mucosal Immune Development in Fetal Tissue) and IRB# PRO17070226.

Note that full information on the approval of the study protocol must also be provided in the manuscript.