Noncoding Deletions Expose a Novel Gene Critical for Intestinal Function

Danit Oz-Levi[1], Tsviya Olender[1*], Ifat Bar-Joseph[2,3*], Yiwen Zhu[4*], Dina Marek-Yagel[2], Iros Barozzi[4,22], Marco Osterwalder[4], Anna Alkelai[5], Elizabeth K. Ruzzo[6], Yujun Han[7], Erica Vos[8] , Haike Reznik-Wolf[2,3], Corina Hartman[3,9], Raanan Shamir[3,9], Batia Weiss[10], Rivka Shapiro[3,9], Ben Pode-Shakked[10], Pavlo Tatarskyy[1], Roni Milgrom[1], Michael Schvimer[11], Iris Barshack[3,11], Denise M. Imai[12], Devin Coleman-Derr[13], Diane E. Dickel[4], Alex S. Nord[4], Veena Afzal[4], Kelly Lammerts van Bueren[14], Ralston M. Barnes[14], Brian L. Black[14], Christopher N. Mayhew[15], Matthew F. Kuhar[15], Amy Pitstick[15], Mehmet Tekman[16], Horia C. Stanescu[16], James M.Wells[15,17], Robert Kleta[16], Wouter de Laat[8], David B. Goldstein[5], Elon Pras[2,3], Axel Visel[4,18,19], Doron Lancet[1,23], Yair Anikster[3,10,21,23*], and Len A. Pennacchio[4,19,20,23*]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel

[2]The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Ramat Gan, Israel

[3]The Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

[4]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

[5]Institute for Genomic Medicine, Columbia University Medical Center, New York, NY, USA

[6]Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles

[7]Center for Human Genome Variation, Duke University School of Medicine, Durham, North Carolina 27708, USA

[8]Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht, the Netherlands

[9]Schneider Children's Medical Center, Petach Tikva, Israel

[10]Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Ramat Gan, Israel

[11]Department of Pathology, Sheba Medical Center, Ramat Gan, Israel

[12]Comparative Pathology Laboratory University of California 1000 Old Davis Rd, Building R1 Davis, CA 95616, USA

[13]Plant Gene Expression Center, USDA-ARS, Albany, CA, 94710, USA

[14]Cardiovascular Research Institute, University of California, San Francisco, CA 94143-3120, USA

[15]Division of Developmental Biology at Cincinnati Children's Hospital Medical Center, Cincinnati OH 45229

[16]Centre for Nephrology, University College London, London, UK

[17]Division of Endocrinology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, 45229, USA

[18]School of Natural Sciences, University of California, Merced, CA 95343, USA

[19]U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA 94598, USA
[20]Comparative Biochemistry Program, University of California, Berkeley, CA 94720, USA
[21]Wohl Institute for Translational Medicine, Sheba Medical Center, Ramat Gan, Israel
[22]Current affiliation: Department of Surgery and Cancer, Imperial College London, London, UK
[23]Correspondence should be addressed to Y.A. (Yair.Anikster@sheba.health.gov.il), D.L. (Doron.Lancet@weizmann.ac.il), or L.A.P. (LAPennacchio@lbl.gov).
*These authors contributed equally

Large-scale genome sequencing is poised to provide an exponential increase in the discovery of disease-associated mutations, but the functional interpretation of such mutations remains challenging. Here we identify deletions of a sequence termed intestine-critical region (ICR) on chromosome 16 that cause intractable congenital diarrhea in infants. Transgenic mouse reporter assays show that the ICR contains a regulatory sequence that activates transcription during development of the gastrointestinal system. Targeted deletion of the ICR in mice caused symptoms recapitulating the human condition. Transcriptome analysis uncovered an unannotated open reading frame (*Percc1*) neighboring the regulatory sequence whose expression was lost in the developing gut of ICR knockout animals. Targeted deletion of the *Percc1* gene in mice caused phenotypes similar to those observed upon ICR deletion in mice and patients, whereas an ICR-driven *Percc1* transgene was sufficient to rescue the phenotypes found in ICR knockout mice. Taken together, our results identify a novel human gene critical for intestinal function and underscore the need for targeted *in vivo* studies for interpreting the growing number of clinical genetic findings that do not affect known protein-coding genes.

In contrast to whole exome sequencing (WES)[1], whole genome sequencing (WGS) can in principle identify mutations in non-coding sequences, as well as in genes that are not annotated in the reference genome. However, sequence variation affecting poorly annotated sequences outside of known genes is challenging to interpret because of the lack of structural and functional annotation of these regions. In the present study, we demonstrate how the identification of non-coding deletions in a small number of patients coupled to purpose-built mouse models can elucidate the regulatory and genic basis of an inherited severe disease **(Fig. 1)**.

Congenital diarrheal disorders are a heterogeneous group of inherited diseases of the digestive system and are frequently life-threatening if untreated[2,3,4] (see **Suppl. Text** for additional clinical background). We studied eight patients from seven unrelated families of common ethnogeographic origin with an autosomal recessive pattern of severe congenital malabsorptive diarrhea named IDIS (for **I**ntractable **D**iarrhea of **I**nfancy **S**yndrome)[4] (**Fig. 1a,b; Extended Data Fig. 1; Suppl. Text**). Initial WES analysis revealed no rare exonic sequence variants with the appropriate patient segregation. However, whole genome linkage analysis and haplotype reconstruction detected a single significant telomeric linkage interval on chromosome 16 (LOD = 4.26; **Extended Data**

Fig. S2, see **Suppl. Text**). We examined WES and WGS data from selected patients and observed a 7,013 bp deletion, termed ΔL, in the absence of other structural changes or coding mutations at the affected locus (**Fig. 1c, Extended Data Fig. 1** and **3, Suppl. Text**). Two of the patients (4.1 and 4.2) were compound heterozygous for ΔL, along with a second variant, termed ΔS, which contains a 3,101 bp deletion that partially overlaps ΔL, defining a minimal sequence termed intestine-critical region (ICR) of 1,528 bp (**Fig. 1c**). All eight patients in this study showed $ICR^{\Delta S/\Delta S}$, $ICR^{\Delta S/\Delta L}$ or $ICR^{\Delta L/\Delta L}$ genotypes, resulting in a homozygous deletion of the *ICR* that was not detected in any of the control groups examined (**Extended Data Fig. 1, Suppl. Text**). These data suggest that the deletion of the ICR causes the congenital diarrhea phenotype.

To explore possible non-coding functions of the ICR, we examined Encyclopedia of DNA Elements (ENCODE) data[5]. The ICR contains a 400 bp region with high evolutionary conservation across vertebrates, includes CpG island and DNase hypersensitivity signatures, and encompasses a cluster of multiple binding sites for transcription factors including FOXA1 and FOXA2[6,7] (**Fig. 1c**). To test the hypothesis that the ICR contains a regulatory sequence, we examined its' *in vivo* activity in a transgenic mouse reporter assay[8]. In transgenic embryos ranging from embryonic day (E) 11.5 to E14.5, we observed robust and reproducible reporter activity in the stomach, pancreas, and duodenum (**Fig. 2a,b**). These results support the notion that the ICR sequence lost in congenital diarrhea patients contains a gene regulatory sequence that is active in the developing digestive system.

To examine if loss of the ICR sequence is sufficient to cause the phenotypes observed in human patients, we deleted a 1,512 bp interval from the mouse genome that included the mouse orthologues of the human DNase hypersensitivity signature and predicted FOXA1/2 binding sites (**Fig. 1d**, **Extended Data Fig. 4**). Homozygous $chr17^{\Delta ICR/\Delta ICR}$ pups were born at the expected Mendelian frequency and showed no gross phenotypes at birth. However, starting within the first few days of life, $chr17^{\Delta ICR/\Delta ICR}$ mice displayed overall reduced size (**Figs. 1e, 2c**), low body weight (**Fig. 2d**), and substantially decreased survival (**Fig. 2e**). Examination of fecal pellets and internal organs revealed abnormal digestive tract function in $chr17^{\Delta ICR/\Delta ICR}$ mice (**Fig. 1f; Extended Data Fig. 5**), as well as changes in the composition of the intestinal microbiome in $chr17^{\Delta ICR/\Delta ICR}$ mice (**Extended Data Fig. 6; Suppl. Text**). Our combined results indicate that deletion of the

ICR in mice causes disruption of intestinal function, recapitulating the congenital diarrhea phenotype observed in human patients carrying homozygous ICR deletions.

Next, we sought to elucidate the molecular mechanisms through which deletion of the ICR causes deficiencies in gastrointestinal function. Targeted expression analysis of stomach and intestinal mouse tissue samples covering developmental stages ranging from E14.5 to P20 revealed an unannotated region flanking the ICR with very low levels of expression in wildtype prenatal stomach tissue (**Fig. 3a**). Expression of this sequence was completely lost in matched tissues from chr17$^{\Delta ICR/\Delta ICR}$ littermates, suggesting the presence of a gene previously unannotated in both the human and mouse genomes (**Extended Data Fig. 4c**). Sequence analysis identified an 897-bp open reading frame predicted to encode a protein that is **P**roline and glutamate (**E**) **R**ich and contains an N-terminal **C**oiled **C**oil domain (PERCC1, **Fig. 3b**). Despite overall strong evolutionary conservation across vertebrates (**Fig. 3c**), PERCC1 homology- and structure- based searches failed to identify similarities to known proteins. However, comparison of human and mouse non-synonymous and synonymous substitutions showed a dN/dS ratio of 0.17, further supporting that PERCC1 represents a *bona fide* protein-coding gene (**Fig. 3d**).

To compare the spatiotemporal expression of *Percc1* with the regulatory *in vivo* activity of the ICR, we performed mRNA *in situ* hybridization for *Percc1*. At E14.5, we observed a pattern of punctate *Percc1* expression in stomach, pancreas, and intestine highly reminiscent of ICR activity (**Extended Data Fig. 7, Fig. 2a,b**). To further establish a functional connection between the ICR regulatory sequence and the predicted open reading frame, we used genome editing to disrupt the *Percc1* open reading frame in mice. *Percc1*$^{-/-}$ mice mimicked the key phenotypes of chr17$^{\Delta ICR/\Delta ICR}$ mice, including low body weight (**Extended Data Fig. 8**) and abnormal appearance of intestinal content (n=11/12 [92%] in *Percc1*$^{-/-}$; n=20/21 [95%] in *chr17*$^{\Delta ICR/\Delta ICR}$; p=0.7, two-tailed T-test). Finally, to establish that Percc1 expression is sufficient to rescue the phenotypes resulting from deletion of the ICR regulatory sequence, we performed a complementation experiment in which we generated chr17$^{\Delta ICR/\Delta ICR}$ mice with an ICR-*Percc1* transgene. Upon complementation, we observed reversal of the reduced body weight, high lethality, and intestinal dysfunction found in chr17$^{\Delta ICR/\Delta ICR}$ mice (**Extended Data Fig. 9**). These results establish that lack of gastrointestinal expression of PERCC1, normally controlled by the ICR, causes the phenotypes observed in chr17$^{\Delta ICR/\Delta ICR}$ mice and likely in IDIS patients examined in this study.

To explore the cell type specificity and function of PERCC1, we generated a transgenic mouse line encoding a PERCC1-mCherry fusion protein (**Fig. 4a, Extended Data Fig. 10**). At postnatal day 8, some PERCC1-mCherry positive (PERCC1$^+$) cells were detected in the epithelium of the intestinal villi (**Extended Data Fig. 11a**), whereas distal stomach compartments displayed a high density of strongly positive cells, in particular in the epithelial layers of the pylorus, antrum, and corpus (**Fig. 4a, Extended Data Fig 11b**). Co-localization with the pan-endocrine marker Synaptophysin (SYP) revealed that a major fraction of PERCC1-positive cells in these compartments are endocrine cells (**Fig. 4a, Extended Data Fig. 11b**). The vast majority of PERCC1$^+$ cells detected in the antral epithelium were Gastrin-expressing G-cells (**Fig. 4a, antrum**), which are required for secretion of gastric acid and promote growth of the gastrointestinal tract. In contrast, the glandular epithelium at the entrance to the pyloric canal also contained clusters of PERCC1$^+$ cells without endocrine signatures located at the gland base, a region known to harbor gastric stem cells[9] (**Fig. 4a**). As the majority of PERCC1$^+$ cells in the distal stomach were G-cells, we next investigated whether loss of *Percc1* in mice affected the development of these cells. We observed that the number of Gastrin-expressing cells was reduced in the absence of *Percc1* (**Extended Data Fig. 11c**). Finally, analysis of 11,665 single cell transcriptomes from mouse intestinal epithelium[10] showed that *Percc1* is expressed in a small proportion of cells that are strongly enriched for enteroendocrine cells (EECs; p = 2.9e-20, chi-squared test, **Extended Data Fig. 12a**). *Percc1*-positive cells express *Sox4* and *Neurog3*, and their expression profiles are most consistent with an enteroendocrine progenitor identity (**Extended Data Fig. 12b,c**). These data indicate that expression of PERCC1 in gastrointestinal tissue is restricted primarily to gastric cells with an endocrine identity and G-cell signature, and suggest that disrupted development of these cells causes the observed phenotype.

To characterize the molecular consequences of loss of the ICR or *Percc1* in more detail, we examined RNA-seq data from relevant stages of mouse development. Among the 100 genes showing the greatest reduction in expression, seven encode gastrointestinal peptide hormones secreted by EECs[11] (**Figure 4b, Table 1, Suppl. Table S1**). Strikingly, reduction of *Gastrin* expression (*Gast*) appears as one of the most robust changes, along with changes in *Somatostatin* (*Sst*) and *Ghrelin* (*Ghrl*) expression (**Figure 4d**, **e**). We observed similar gene expression changes in duodenal and stomach biopsies obtained from an ICR$^{\Delta L/\Delta L}$ patient and an unaffected ICR$^{+/+}$ sibling (**Table 1; Suppl. Table S2; Extended Data Fig. 13**). Together, these results are consistent with major disruptions of normal gastrointestinal physiology in chr17$^{\Delta ICR/\Delta ICR}$ mice and human IDIS patients and highlight the close resemblance between the human disease condition and our mouse knockout models.

We also generated induced pluripotent stem cells (iPSC) from an ICR$^{\Delta L/\Delta L}$ patient and an unaffected ICR$^{+/+}$ sibling and differentiated them into human intestinal organoids (HIOs) (**Extended Data Fig. 14/15**)[12]. The gross morphology of ICR$^{\Delta L/\Delta L}$ and ICR$^{+/+}$ HIOs was similar, but at early stages (21 days) the ICR$^{\Delta L/\Delta L}$ HIOs showed more EECs than control HIOs. In contrast, by day 42 the ICR$^{\Delta L/\Delta L}$ HIOs showed a severe reduction in the number of EECs along with reduced expression of EEC markers Chromogranin A (CHGA) and Synaptophysin (SYP) (**Fig. 4c; Extended Data Fig. 11d**). These results suggest that disruption of the ICR is compatible with the initial formation of EECs, but interferes with their subsequent development.

Limited understanding of the *in vivo* function of human protein-coding genes and non-coding sequences continues to be a grand challenge preventing the systematic interpretation of disease-related data from WES and WGS studies. In the present work we establish that IDIS, a severe, recessively inherited gastrointestinal disease, is caused by microdeletions that disrupt a regulatory sequence required for normal intestinal expression of the previously unannotated *Percc1* gene. The molecular, cellular, and physiological phenotypes observed in human patients and engineered mice indicate that *Percc1* is required for normal development of EECs and, thereby, normal enteroendocrine hormone secretion. The phenotype of chr17$^{\Delta ICR/\Delta ICR}$ mice resembles that of mice with an intestinal-specific deletion of *Neurog3*, a proendocrine transcription factor required for development of EECs[13], further supporting abnormal EEC development as the cause of IDIS. The different etiologies of chronic diarrhea of infancy and enteropathies have recently been reviewed and classified into distinct categories[14]. Among these categories, disorders of enteroendocrine cell function, caused by mutations in *Neurog3*, *Arx*, *Pcsk1*, *Rfx6*, are described as a unique and separate entity that manifest with malabsorptive diarrhea. The phenotypes observed in human patients and engineered mice indicate that IDIS is most similar to this specific class of chronic diarrheal disorders. Beyond congenital diarrhea, our results serve as a reminder that, despite extensive annotation efforts, protein-coding genes associated with disease phenotypes remain to be discovered. As WGS is increasingly used for studies of rare diseases, our work underscores the importance of detailed experimental follow-up of such findings through *in vivo* models.

Figure Legends

Figure 1. Overview of human and mouse locus and key findings. a/b, Selected family pedigrees and genotyping results for patients compound heterozygous for the two deletion alleles (**a**) and homozygous for one of the deletion alleles (**b**). **c/d,** Genomic map of the deletion alleles in human (**c; genome build GRCh37**) and mouse (**d**), indicating the location of ΔL and ΔS, as well as their minimal overlapping region ICR. Exome sequencing data is capped at up to 5 overlapping tags for visualization; vertebrate conservation is 100-vertebrate PhyloP; only selected transcription factor binding sites and DNase hypersensitivity clusters with signal in >20/125 ENCODE cell types shown. **e,** General appearance of wildtype (n=50) and chr17$^{\Delta ICR/\Delta ICR}$ (n=46) mice at 21 days after birth, showing overall significantly reduced size (see Fig 2d). **g,** Abnormal appearance of fecal pellets from chr17$^{\Delta ICR/\Delta ICR}$ mice (n=46).

Figure 2. Enhancer activity of the ICR and mouse deletion phenotypes. a-b, Enhancer reporter activity in transgenic mouse embryos. **a,** Mouse embryo cross-sections showing X-gal staining for β-galactosidase activity in E13.5 stomach, pancreas and duodenum as marked. **b,** E14.5 cross-section showing immunofluorescence with anti-β-galactosidase (ICR enhancer activity, red), anti-endomucin (endothelial cells, green), and DAPI (DNA, blue). **a/b,** Two embryos for each experiment and each condition were collected and a minimum of three sections from each embryo were examined. Representative sections are shown. **c,** Chr17$^{\Delta ICR/\Delta ICR}$ mice (n=46) are viable but show a reduction in size and weight compared to wild-type littermates (n=50). **d/e,** Reduction in body weight among surviving offspring (**d**) and increased mortality (**e**) of chr17$^{\Delta ICR/\Delta ICR}$ compared to wild-type. Body weight of female mice shown in (**d**); male wildtype and chr17$^{\Delta ICR/\Delta ICR}$ mice had similar genotype-dependent weight differences.

Figure 3. Discovery of a novel gene, *Perccl*, flanking the ICR. a, Gene expression levels of *Perccl* in gastrointestinal tissues from wildtype (wt) and chr17$^{\Delta ICR/\Delta ICR}$ mice. Highest levels of expression were detected in stomach at postnatal day 10 (P10). **b,** Mouse

genome view of *Percc1* gene localization and structure. Stranded RNA-seq data indicating expression loss of *Percc1* in knockout mice compared to controls (bottom panel). **c,** Detailed view of the *Percc1* gene and evolutionary conservation. **d,** *Percc1* codon position analysis illustrating the relaxation of constraint in the third codon position of the predicted Percc1 protein (n = 274). dN/dS ratio and corresponding P-value calculated using Phylogenetic Analysis using Maximum Likelihood (Chi-square distribution).

Figure 4: PERCC1 is abundant in G-cells and its genetic disruption impairs gastrointestinal peptide hormone expression and enteroendocrine cell development. **a,** Top: Generation of a reporter fusion transgene to track PERCC1 localization in murine gastrointestinal tissues. The genomic sequence spanning the ICR and Percc1 ORF were fused to mCherry. Lower left panels: PERCC1+ cells (red) in the pyloric antrum at postnatal day 8 (P8) show endocrine identity marked by Synaptophysin (Syp, green) and extensive overlap with the endocrine subset of Gastrin expressing G-cells (blue). Arrowheads indicate triple-positive cells. Nuclei are shown in gray. Lower right panels: PERCC1+ cells in the pyloric canal carry either endocrine G-cell identity (arrowheads) or appear frequently clustered at the gland base (arrows). **b,** RNA-seq from mouse stomach samples across different timepoints (n = 1 biological replicates) reveals reduced transcript levels of different gastric peptide hormones in ΔICR/ΔICR mice. *Sst*, *Somatostatin*; *Gast*, *Gastrin*; *Grhl*, *Ghrelin*; *Pyy*, *Peptide YY*; *Gcg*, *Glucagon*; *Fndc5*, *Fibronectin type III domain-containing protein 5*; *Adipoq*, *Adiponectin precursor*. **c,** Quantitative RT-PCR analysis of induced pluripotent stem cell (iPS) -derived human intestinal organoids (HIOs) from patients with disrupted PERCC1 (ΔL/ΔL) vs. control siblings. In contrast to Chromogranin A (CHGA), *Synaptophysin* (*SYP*) is significantly downregulated in patient-derived HIOs (p<0.05) (two-tailed, unpaired t-test). iPS, induced pluripotent stem cells. Box plot indicates median, interquartile values, range, outliers (circled dots) and individual technical replicates (from independent organoid preparations). n represents independent biological replicates with similar results. Scale bars, 50μm.

Fig 1

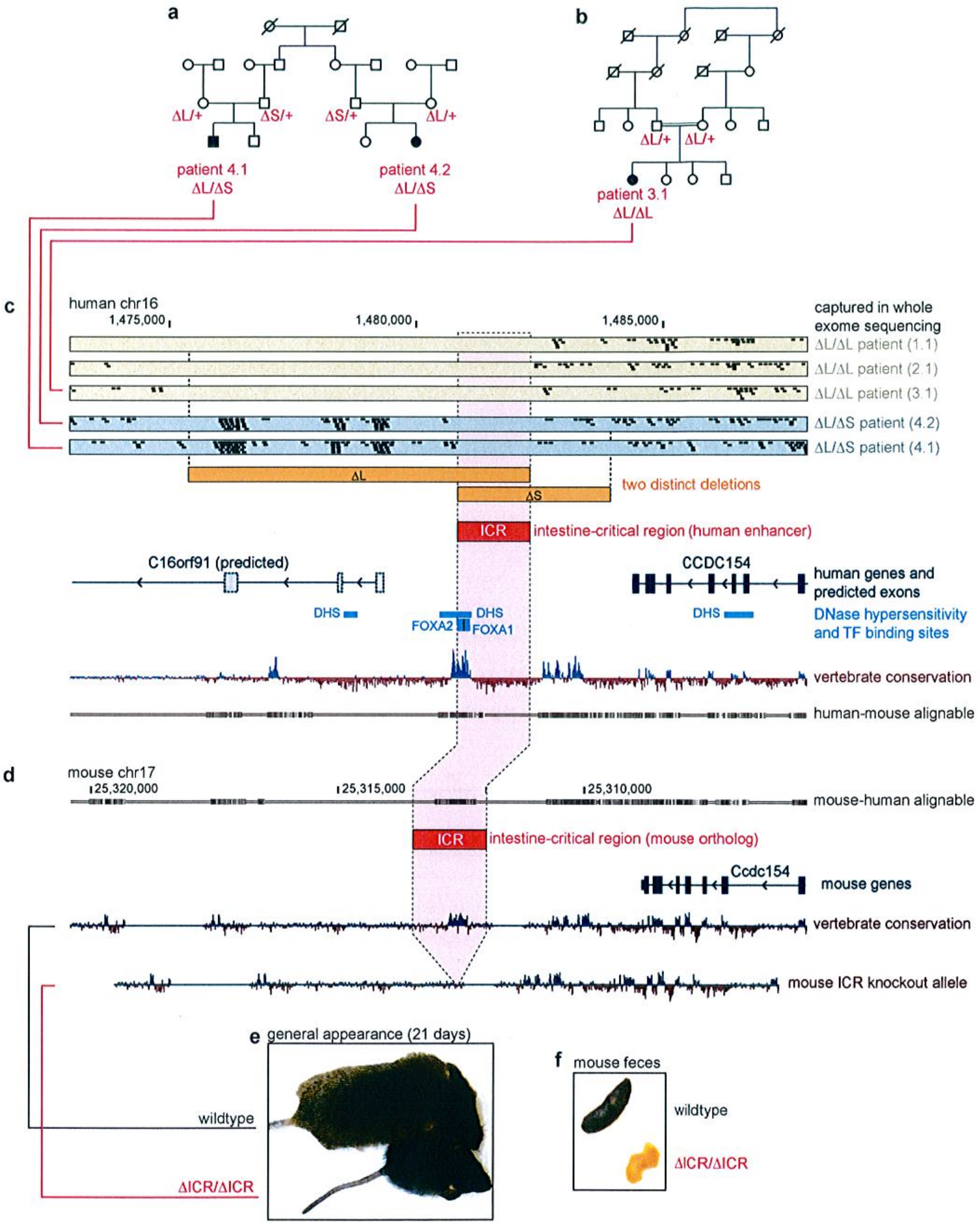

a
ΔL/+
ΔS/+
ΔS/+
ΔL/+
patient 4.1
ΔL/ΔS
patient 4.2
ΔL/ΔS
b
ΔL/+
ΔL/+
patient 3.1
ΔL/ΔL
c
human chr16
1,475,000
1,480,000
1,485,000
captured in whole exome sequencing
ΔL/ΔL patient (1.1)
ΔL/ΔL patient (2.1)
ΔL/ΔL patient (3.1)
ΔL/ΔS patient (4.2)
ΔL/ΔS patient (4.1)
ΔL
ΔS
two distinct deletions
ICR
intestine-critical region (human enhancer)
C16orf91 (predicted)
CCDC154
human genes and predicted exons
DHS
DHS
FOXA2
FOXA1
DHS
DNase hypersensitivity and TF binding sites
vertebrate conservation
human-mouse alignable
d
mouse chr17
25,320,000
25,315,000
25,310,000
mouse-human alignable
ICR
intestine-critical region (mouse ortholog)
Ccdc154
mouse genes
vertebrate conservation
mouse ICR knockout allele
e
general appearance (21 days)
wildtype
ΔICR/ΔICR
f
mouse feces
wildtype
ΔICR/ΔICR

Fig 2

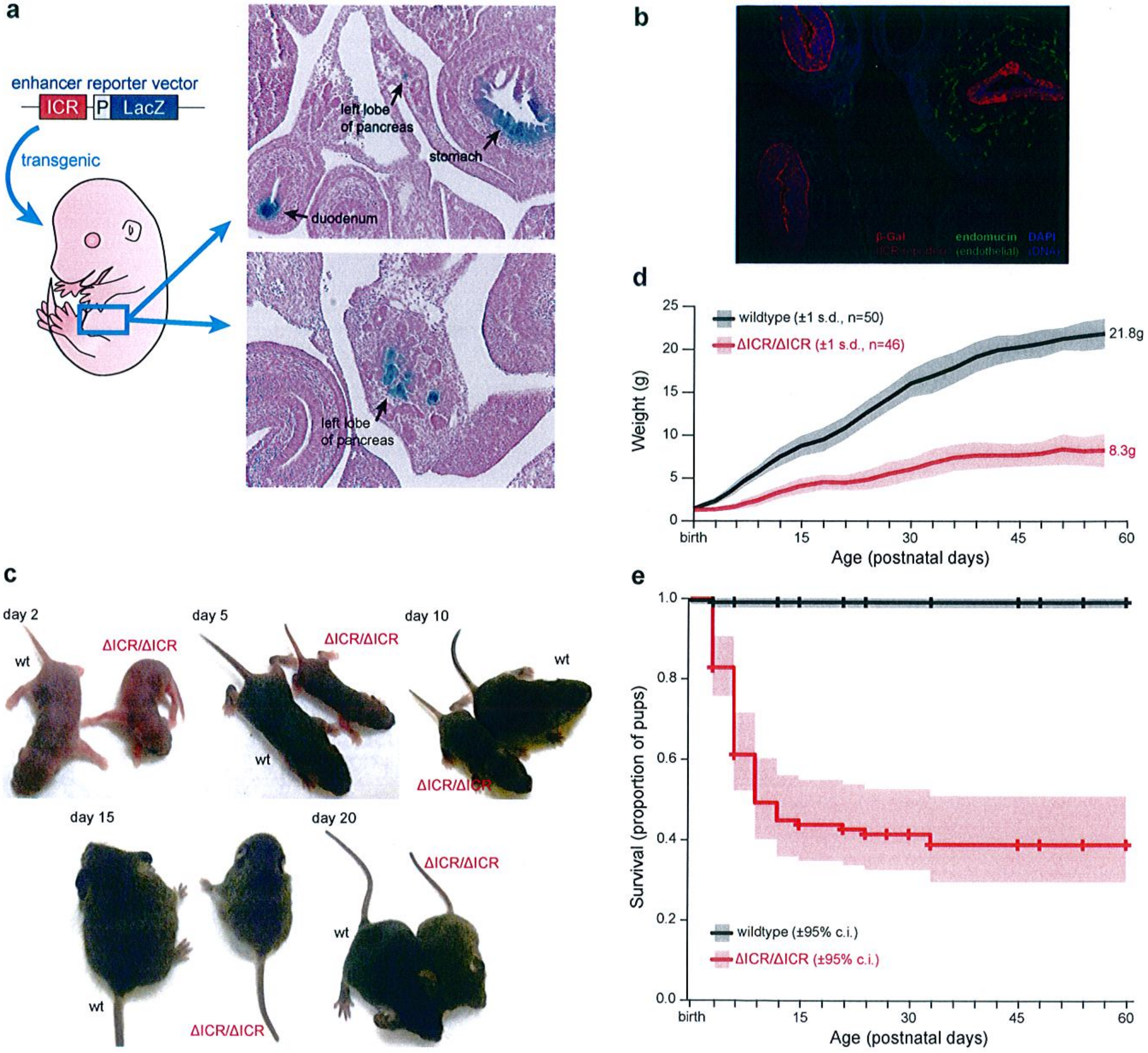

a
enhancer reporter vector
ICR
P
LacZ
transgenic
left lobe of pancreas
stomach
duodenum
left lobe of pancreas
b
β-Gal
ICR reporter
endomucin
(endothelial)
DAPI
(DNA)
c
day 2
wt
ΔICR/ΔICR
day 5
ΔICR/ΔICR
wt
day 10
wt
ΔICR/ΔICR
day 15
wt
ΔICR/ΔICR
day 20
ΔICR/ΔICR
wt
d
wildtype (±1 s.d., n=50)
ΔICR/ΔICR (±1 s.d., n=46)
Weight (g)
0
5
10
15
20
25
21.8g
8.3g
birth
15
30
45
60
Age (postnatal days)
e
Survival (proportion of pups)
0.0
0.2
0.4
0.6
0.8
1.0
wildtype (±95% c.i.)
ΔICR/ΔICR (±95% c.i.)
birth
15
30
45
60
Age (postnatal days)

Fig 3

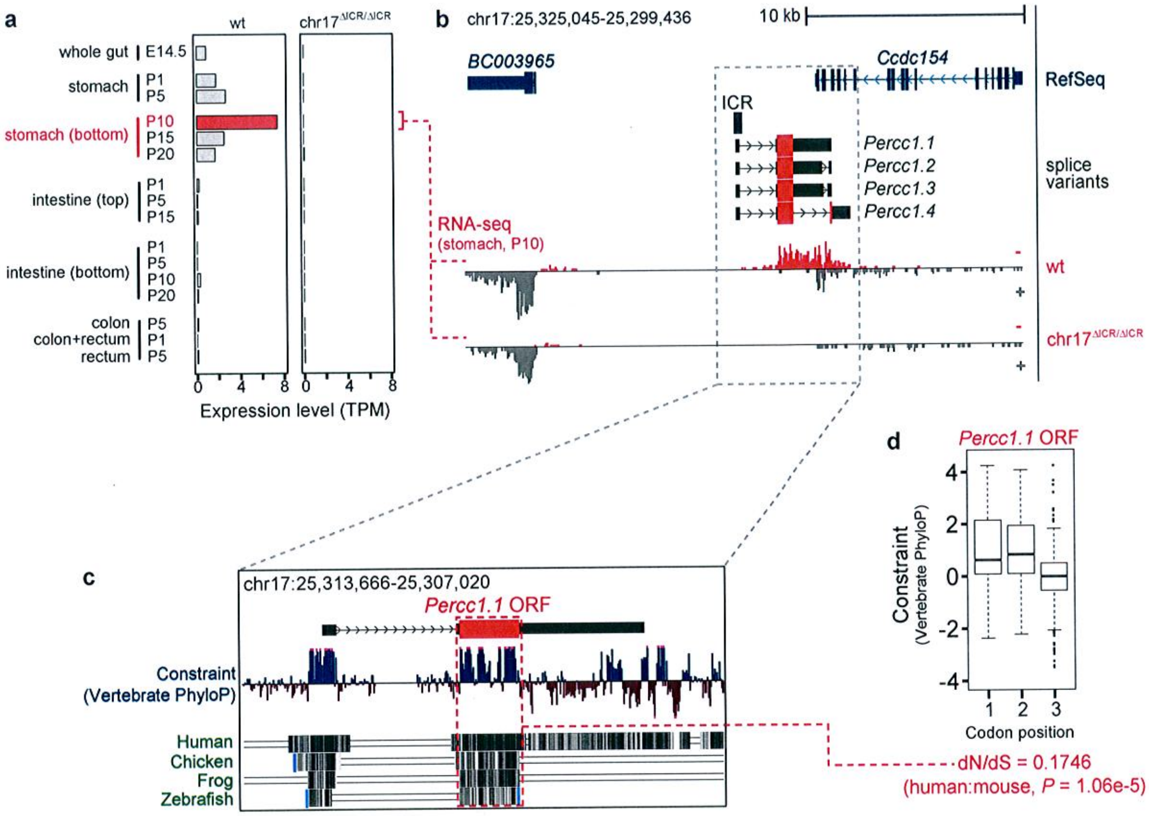
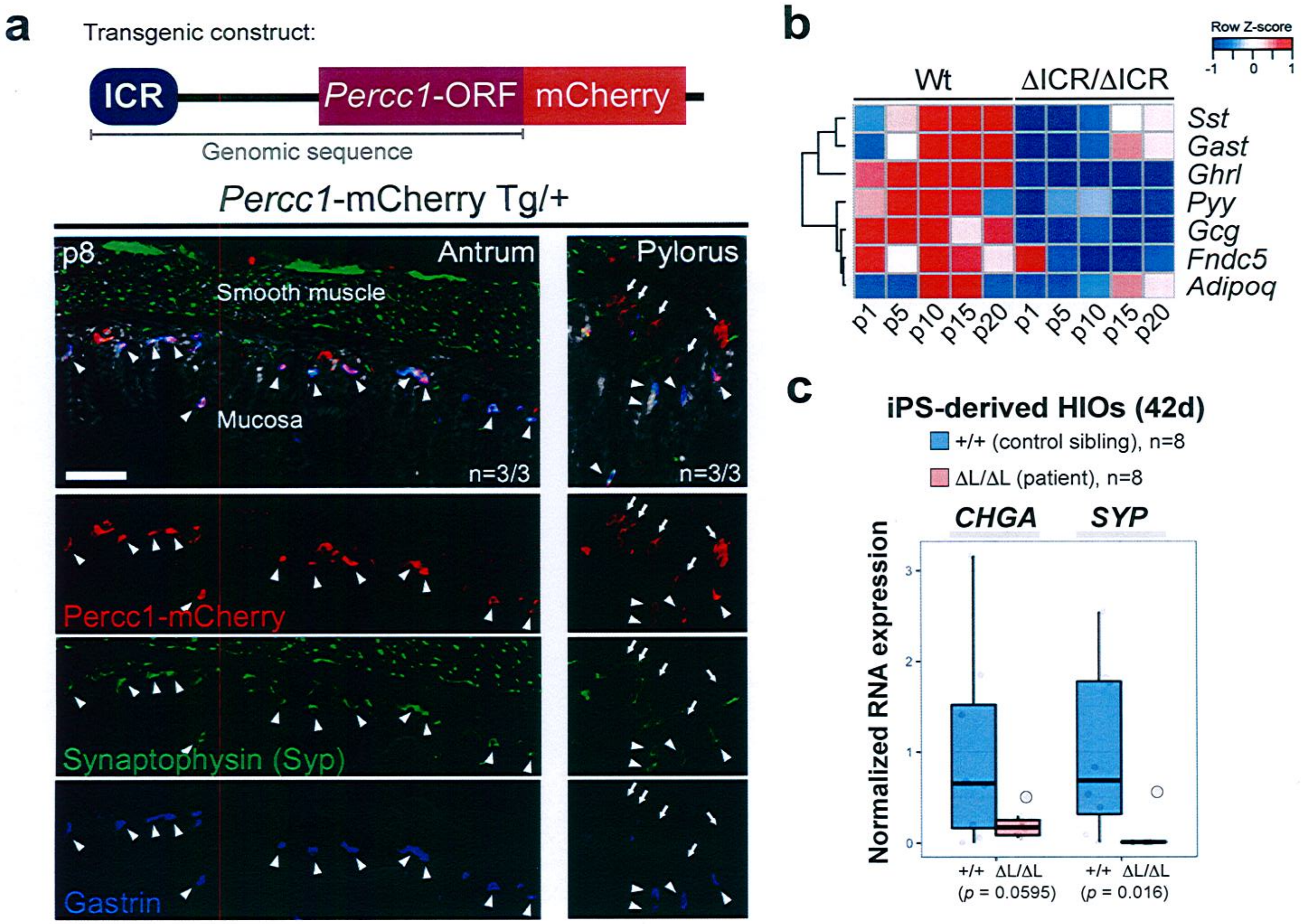

a
wt
chr17ΔICR/ΔICR
whole gut
E14.5
stomach
P1
P5
stomach (bottom)
P10
P15
P20
intestine (top)
P1
P5
P15
intestine (bottom)
P1
P5
P10
P20
colon
P5
colon+rectum
P1
rectum
P5
0
4
8
0
4
8
Expression level (TPM)
b
chr17:25,325,045-25,299,436
10 kb
BC003965
Ccdc154
RefSeq
ICR
Percc1.1
Percc1.2
Percc1.3
Percc1.4
splice variants
RNA-seq
(stomach, P10)
-
wt
+
-
chr17ΔICR/ΔICR
+
c
chr17:25,313,666-25,307,020
Percc1.1 ORF
Constraint
(Vertebrate PhyloP)
Human
Chicken
Frog
Zebrafish
dN/dS = 0.1746
(human:mouse, P = 1.06e-5)
d
Percc1.1 ORF
Constraint
(Vertebrate PhyloP)
4
2
0
-2
-4
1
2
3
Codon position

Fig 4


a
Transgenic construct:
ICR
Percc1-ORF
mCherry
Genomic sequence
Percc1-mCherry Tg/+
p8
Antrum
Pylorus
Smooth muscle
Mucosa
n=3/3
n=3/3
Percc1-mCherry
Synaptophysin (Syp)
Gastrin
b
Row Z-score
-1
0
1
Wt
ΔICR/ΔICR
Sst
Gast
Ghrl
Pyy
Gcg
Fndc5
Adipoq
p1
p5
p10
p15
p20
p1
p5
p10
p15
p20
c
iPS-derived HIOs (42d)
+/+ (control sibling), n=8
ΔL/ΔL (patient), n=8
CHGA
SYP
Normalized RNA expression
0
1
2
3
+/+
ΔL/ΔL
(p = 0.0595)
+/+
ΔL/ΔL
(p = 0.016)

Methods

Experimental Design

All animal work was reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare Committee. All mice used in this study were housed at the Animal Care Facility (the ACF) at LBNL. Mice were monitored daily for food and water intake, and animals were inspected weekly by the Chair of the Animal Welfare and Research Committee and the head of the animal facility in consultation with the veterinary staff. The LBNL ACF is accredited by the American Association for the Accreditation of Laboratory Animal Care (AAALAC). Generation of transgenic mice and manipulation of genomic sequence were performed in *Mus musculus* FVB strain mice, mice with C57/129S6 mixed background or W4/129S6 ES cells (see below for details). Animals of both sexes were used in the analysis. Sample size selection and randomization strategies were conducted as follows:

Transgenic mouse assays. Sample sizes were selected empirically based on our previous experience of performing transgenic mouse assays for >2,000 total putative enhancers (VISTA Enhancer Browser: https://enhancer.lbl.gov/). Mouse embryos or postnatal mice were excluded from further analysis if they did not contain the reporter transgene or if the developmental stage was not correct. All transgenic mice were treated with identical experimental conditions. Randomization and experimenter blinding were unnecessary and not performed.

Genomic knockouts. Sample sizes were selected empirically based on our previous studies[15]. All phenotypic characterization of knockout mice employed a matched littermate selection strategy. All phenotyped homozygous knockout mice described in the paper resulted from crossing heterozygous knockout mice together to allow for the comparison of matched littermates of different genotypes. Embryonic samples used for RNA-seq and immunofluorescence were dissected blind to genotype. RNA-seq libraries were prepared and sequenced in mixed batches (including both KO and WT samples).

Fixed

Subjects: IDIS patients were recruited at Schneider and Sheba medical centers in Israel. Clinical details of the subjects are provided in **Supp. Table S4**. All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants and/or their legal guardian involved in the study.

Exome sequencing and variants identification: Exome sequencing was performed using Agilent SureSelect Human All Exon technology (Agilent Technologies, Santa Clara, CA). The captured regions were sequenced using Genome Analyzer IIx (Illumina, Inc. San Diego, CA). The resulting reads were aligned to the reference genome (build 37) using the Burrows-Wheeler Alignment (BWA) tool[16]. We obtained 70× coverage, where a base was considered covered if ≥ 5 reads spanned the nucleotide. Genetic differences relative to the reference genome were identified by the SAMtools v0.1.7a variant calling program[17], which identifies both single nucleotide variants and small insertion-deletions (indels). Finally, the Sequence Variant Analyzer software (SVA v1.10)[18] was used to annotate all identified variants. For comparison to controls, we utilized 1000 samples subjected to exome or whole genome sequencing at the Center for Human Genome Variation (CHGV, Duke University, NC, USA), dbSNP, 1000 genomes, and NHLBI GO Exome-sequencing Project.

Whole genome sequencing: WGS of individual 2.1 was performed at CHGV, using the Illumina HiSeq platform (Illumina, Inc. San Diego, CA) and analyzed as described for exome data. 275 CHGV whole-genome sequenced, unrelated samples were used as controls. To detect copy number variants from WGS we used the Estimation by read depth with single-nucleotide variants (ERDS) tool[19].

Biopsy collection: Subjects underwent gastro-duodenoscopy following Institutional Review Board (IRB) approval (No. 9881-12-SMC) at Sheba Medical Center and written informed consent of the patients and family members.

RNA extraction from biopsies: RNA isolation from frozen biopsies was performed using TRI Reagent method (Sigma-Aldrich Inc.) according to the manufacturer's instructions or by Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA, USA). Integrity of the samples was measured for concentration and purity using a NanoDrop Spectrophotometer (Nanodrop Technologies, Wilmington, DE, USA).

RNA sequencing of human samples: Total RNA was prepared according to the Illumina RNA-seq protocol. Briefly, globin reduction, polyA enrichment, chemical fragmentation of the polyA RNA, cDNA synthesis, and size selection of 200bp cDNA fragments were performed. Next, the size-selected libraries were used for cluster generation on the flow cell and prepared flow cells were run on the Illumina HiSeq2000 (Illumina, Inc. San Diego, CA). We obtained a total of 74.18 million paired-end reads of a 100 bp for the affected sample and 72.53 million reads to the healthy sample. Reads were align to the human genome (NCBI37/hg19) using Tophat v2.0.4[20] with the default parameters. Gene expression quantification was performed with Cuffdiff v2.0.2[20] using the Illumina iGenome project UCSC annotation file as a reference. Differentially expressed genes were defined using the following thresholds: 1.5 linear fold change, p-value ≤ 0.05.

Quantitative Real-Time Reverse Transcriptase (RT) Polymerase Chain Reaction (qPCR): RNA extracted from biopsies was used for qPCR expression analyses. qPCR were performed using TaqMan Gene Expression Assays (Applied Biosystems, Foster City, CA, USA) using the Applied Biosystems StepOnePlus (Applied Biosystems). From 1 µg of biopsy RNA, cDNA was synthesized using the SuperScript® First-strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. A total of 20µl of cDNA was added with 30µl of water to 50µl of TaqMan universal PCR Master Mix (Applied Biosystems) and the resulting 100µl reaction mixtures were loaded onto a 96-well PCR plate. We used 14 different TaqMan Gene Expression Assays including three housekeeping genes with the following assays IDs: Hs00757713_m1 (MLN), Hs01074053_m1 (GHRL), Hs00175048_m1 (NTS), Hs00356144_m1 (SST), Hs00174945_m1 (PYY), Hs01062283_m1 (GAST), Hs00292465_m1 (ARX), Hs00174937_m1 (CCK), Hs00175030_m1 (GIP), Hs00219734_m1 (GKN1), Hs00699389_m1 (GKN2), The housekeeping genes we used were HMBS (Hs00609297_m1), ACTB (Hs99999903_m1) and GAPDH (Hs99999905_m1). Reference cDNA samples were synthesized using 200 ng of RNA from RNA extracted from stomach and duodenum tissues of two healthy controls (BioCat GmbH, Heidelberg, Germany) for use in the normalization calculations. qPCR for expression analysis on the missing exons in C16ORF91 was done using cDNA extracted from the Human Digestive System MTC Panel (Clontech Laboratories, Inc. Mountain View, CA).

Serum Collection: Whole blood was withdrawn into a Vacutainer serum tube without anti-coagulant. The blood was immediately treated with 1µM AEBSF (protease inhibitor) and remained at room temparature for 30 min to clot before centrifugation (15 min at 2500 rpm at 4°C).

ELISA: Serum hormone levels were determined using sandwich ELISA technique performed by the following commercial kits according to the manufacturer's instructions. Human Ghrelin (Total) ELISA COLD PACKS (Millipore, USA), Human PYY (Total) ELISA Kit (Millipore), and Human gastric inhibitory polypeptide (GIP) ELISA Kit (ENCO).

Linkage analysis and homozygosity mapping: Genome-wide SNP genotyping from DNA of 6 affected children and 22 relatives from families 1-5 was performed using the Illumina HumanCytoSNP-12v2-1_H, according to the manufacturer's recommendations (Illumina, Inc. San Diego, CA) in conjunction with SNP genotypes retrieved from whole exome data. For linkage studies 35,845 informative equally spaced SNP markers were chosen after filtering for Mendelian errors and unlikely genotypes. Genotypes were examined with the use of a multipoint parametric linkage analysis and haplotype reconstruction for an autosomal recessive model with complete penetrance and a disease allele frequency of 0.001 as previously described[21]. Homozygosity mapping was performed using PLINK v1.07[22] with the default parameters (length 1000 kb, SNP(N) 100, SNP density 50 kb/SNP, largest gap 1000 kb).

Deletion analysis: Boundaries for the two deletion alleles were determined by PCR using amplified DNA and Sanger sequencing. The specific primers used to amplify across both deletions and inside the overlapping region for the two deletions are reported in **Supp. Table S5**. In parallel, we used polymorphic markers that were identified by electronically screening genomic clones located on Chr16 0.86-2.8Mb. Primers were designed with the Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi/ from the Whitehead Institute, Massachusetts Institute of Technology, and Cambridge, MA). The specific primers used are reported in **Supp. Table S6**. Amplification of the polymorphic markers was performed in a 25µl reaction containing 50 ng of DNA, 13.4 ng of each primer, and 1.5 mM dNTPs in1.5 mM $MgCl_2$ PCR buffer with 1.2 U *Taq* polymerase (Bio-Line, London, UK). After an initial denaturation of 5 min at 95°C, 30 cycles were performed (94°C for 2 min, 56°C for 3 min, and 72°C for 1 min), followed by a final step of 7 min at 72°C. PCR products were electrophoresed on an automated genetic analyzer (Prism 3100; Applied Biosystems, Inc. [ABI], Foster City, CA). The breakpoints coordinates were : ΔL- chr16: 1475365-1482378, ΔS- chr16:1480850-1483951, with an overlapping region at chr16: 1480850-1482378 (ICR).

Mouse transgenic assays: The candidate gene regulatory sequence (chr 16: 1479875 – 1480992) was PCR amplified from human genomic DNA and was cloned into the *hsp68*-lacZ transgenic vector containing a minimal *hsp68* promoter coupled to a *lacZ* reporter gene. The purified transgene construct was microinjected into fertilized FVB/N mouse oocytes, which were implanted into pseudopregnant foster females and embryos were collected at E11.5 through E14.5. Transgenic activity was determined by X-gal staining to detect β-galactosidase activity. Only patterns observed in at least three different embryos resulting from independent transgenic events were considered reproducible positive enhancers. All animal work was reviewed and approved by the Lawrence Berkeley National Laboratory Animal Welfare and Research Committee.

Generation of ICR knockout mice (chr17$^{\Delta ICR/\Delta ICR}$): Homologous arms were generated by PCR (see **Suppl. Table S7** for primers) and cloned into the ploxPN2T vector, which contains a neomycin (G418) resistant cassette flanked by loxP sites for positive selection, and an HSV-tk cassette for negative selection. Constructs were linearized and electroporated (20μg) into W4/129S6 mouse embryonic stem cells (Taconic). The electroporated cells were selected under G418 (150μg/ml) and 0.2μM FIAU for a week. Surviving colonies were picked and expanded on 96-well plates, screened by PCR and sequencing with primers outside but flanking the homology arm. Clones that were correctly targeted were electroporated with 20μg of the Cre recombinase-expressing plasmid TURBO-Cre. TURBO-Cre was provided by Dr. Timothy Ley of the Embryonic Stem Cell Core of the Siteman Cancer Center, Washington University Medical School.

Clones positive for Neo removal were screened by PCR and checked for G418 sensitivity. PCR products covering the deleted region and part of homologous arms were gel purified and sequenced to confirm the deletion of the ICR.

Correctly targeted clones were subsequently injected into C57BL/6J blastocyst stage embryos. Chimeric mice were then crossed to C57BL/6J mice (Charles River) as well as

129S6/SvEvTac (Taconic) to generate heterozygous ICR-null mice, followed by breeding of heterozygous littermates to generate homozygous null mice.

Genotyping of chr17$^{\Delta ICR}$ mice: Genomic DNA was extracted from a 0.2 to 0.3-cm section of tail that was incubated overnight in lysis buffer (containing 100mM Tris-HCl pH 8.5, 5mM EDTA, 0.2% SDS, 200mM NaCl and 50µg Proteinase K) at 55°C. Genotyping was carried out using standard PCR techniques (see **Suppl. Table S7** for primers). 1-2µl of 50- to 100-fold diluted tail lysate was used in a 20µl PCR containing 200 µM dNTP, 1.5 mM $MgCl_2$, 5 pmole of each forward and reverse primer and 0.5U of Taq polymerase.

RNA sequencing of mouse tissues: Total RNA was extracted from different intestinal regions and stomach of mice at E11.5, P1, P5, P10, P15 and P20 using TRIzol Reagent (Invitrogen). RNAseq libraries were then constructed using Illumina TruSeq Stranded Total RNA Sample Preparation Kit following manufacturer's recommendation. The libraries were sequenced using a 50bp single end strategy with four samples per lane on an Illumina HiSeq instrument, and data were analyzed using the same protocols as described for human, though with the mm9 mouse reference and Illumina iGenome project mouse genome annotation data. The RNA-seq data can be accessed with GEO number GSE94245. DAVID v6.7[23] was run separately on the differentially expressed genes in the P10 stomach (167 down-regulated, 187 up-regulated, **Supp. Table S1**) or in the P10 intestine (326 down-regulated, 327 up-regulated, **Supp. Table S1**). The resulting lists of clustered terms were inspected for highly significant enrichment (q-value ≤0.05) for biological process (GO), molecular function (MF), or pathway (KEGG) (see **Supp. Table S3**). The most significant term for each cluster was retained. In total, 59 genes showing any of these annotations were also found in the top 100 deregulated genes (either up- or down- regulated, separately for each tissue). A selection of genes from these 59 is highlighted in **Table 1**.

16S amplicon analysis (iTags) of microbial community diversity: Feces and gut content samples were collected from chr17$^{\Delta ICR/\Delta ICR}$ mice and wild type littermates. DNA was extracted from these samples using PowerFecal DNA Isolation Kit (MO Bio Laboratories). V4 16S regions were amplified from the DNA samples using barcoded primers and 5 PRIME™ HotMasterMix (Fisher Scientific) as previously described[24]. Amplicons were pooled in equal amount, purified with AMPureXP magnetic beads (Beckman Coulter), and sequenced.

***Perccl* RT-PCR, RT-qPCR and 5', 3'-RACE:** Total RNA was extracted from intestine and stomach of mice using TRIzol Reagent (ThermoFisher Scientific) and total RNA was DNase (Promega) treated. First strand cDNA was generated from the DNase-treated total RNA using SuperScript First-Strand Synthesis System (ThermoFisher Scientific). RT-qPCR was performed using KAPA SYBR FAST Roche LightCycler 480 2X qPCR Master Mix (KAPA Biosystems). To identify the expression boundaries of the *Perccl* transcript(s), regular RT-PCR, 5'-RACE and 3'-RACE were performed (primers used can be found in **Supp. Table S8**). Standard RT-PCR used Platinum Taq DNA Polymerase High fidelity (ThermoFisher Scientific) while 5'- and 3'- RACE used the SMARTer RACE 5'/3' Kit (Clontech). PCR products were gel purified and Sanger sequenced. The cDNA and predicted protein sequence are available in Genbank record KY964488. It should be noted that since our original identification of *perccl*, the Havana gene annotation project manually curated this open reading frame as uncharacterized LOC105371045.

***Perccl* knockout mice by CRISPR/Cas9:** *Perccl* knockout mice were generated by CRISPR/Cas9 editing as previously described[25],[26]. Single guide RNAs (sgRNAs) were constructed using 60-mer oligonucleotides and an sgRNA cloning vector (Addgene plasmid 41824)[27] according to the following protocol: http://www.addgene.org/static/cms/files/hCRISPR_gRNA_Synthesis.pdf. The sgRNA target site sequences are provided in **Supp. Table S9.** Cas9 mRNA was generated using a human codon optimized Cas9 gene from plasmid pDD921[28]. T7promoter-Cas9-polyA and T7promoter-sgRNA amplicons were PCR amplified from pDD921 and sgRNA clones, respectively, using Phusion polymerase (New England Biolabs). Cas9 RNA was generated by *in vitro* transcription from the T7promoter-Cas9-polyA amplicon using mMESSAGE mMACHINE T7 Kit (ThermoFisher Scientific), following manufacturer's instruction. sgRNA RNA was generated by *in vitro* transcription from the T7promoter-sgRNA amplicon using MEGAshortscript Kit (ThermoFisher Scientific), following manufacturer's instruction. *In vitro* transcribed RNA was cleaned using MEGAclear Kit (ThermoFisher Scientific), following manufacturer's instruction. RNA was eluted into RNase-free Microinjection Buffer (10 mM Tris, pH7.5; 0.1 mM EDTA). The RNA was then assessed by electrophoresis on a 10% TBE Urea PAGE gel.

Microinjection and generation of genetically-modified mice: A mixture of 100 ng/μl Cas9 RNA and 50 ng/μl total sgRNA RNAs in Microinjection Buffer was injected into cytoplasm of fertilized mouse strain FVB/N mouse oocytes. Pups generated (F0) were screened by PCR (primers listed in **Supp. Table S9**) and resulting PCR products were sequenced to identify deletion breakpoints.

Complementation of chr17$^{\Delta ICR/\Delta ICR}$ with a *Percc1* mouse transgene: An 8530bp region was PCR amplified from W4/129S6 mouse genomic DNA (primer sequences can be found in **Suppl. Table S10)**. The region amplified includes all putative *Percc1* exons, a possible promoter, and the 3'-UTR. The resulting PCR product was cloned into pCR2.1 vector (Invitrogen TOPO® TA Cloning® Kit) and verified by sequencing. To generate transgenic mice, a purified, linear transgene fragment was injected into the pronucleus of fertilized eggs from chr17$^{\Delta ICR/+}$ intercrosses. Offspring were PCR genotyped for existence of the *percc1* transgene and examined for phenotypes found in chr17$^{\Delta ICR/\Delta ICR}$ mice.

Generation of a PERCC1-mCherry fusion protein transgenic mice: A 3308bp fragment covering the *Percc1* promoter and last amino acid of the *Percc1* coding region was PCR amplified and cloned into the pcDNA3 mCherry LIC cloning vector 6B (Addgene, #30125) 5' of the mCherry cDNA to generate a mouse Percc1-mCherry fusion construct (primers can be found in **Suppl. Table S11)**. The construct was linearized with NotI and injected into the pronucleus of fertilized FVB eggs to generate transgenic mice.

Western blotting: Tissues (stomach and intestine) were lysed and homogenized in T-PER™ Tissue Protein Extraction (ThermoFisher Scientific, cat # 78510) plus Protease Inhibitor (Sigma, cat# P8340). Protein concentration was determined using Pierce™ Coomassie (Bradford) Protein Assay Kit (ThermoFisher cat# 23200). 100μg of lysate per well of each sample was loaded on Bolt™ 4-12% Bis-Tris Plus Gel (ThermoFisher, cat# NW04120BOX), along with SeeBlue® Plus2 Pre-Stained Standard (ThermoFisher Scientific, LC5925) and 0.1 μg mCherry protein (VWR cat# 10190-818) as a positive control. After electrophoresis at 200V for 35 minutes, the gel was blotted on 0.45 μm pore size PVDF membrane using XCell II™ Blot Module (ThermoFisher Scientific). The membrane was then blocked with 3% nonfat milk in TBS (Sigma, T8793). Primary antibody, mCherry Antibody (ThermoFisher Scientific, cat# PA5-34974), was added to the blot at 1:1000 dilution in 3% nonfat milk in TBS, and incubated overnight at 4 °C.

Following several washes with TBST (Sigma, T9039-10PAK), HRP-conjugated goat anti-rabbit antibody (ThermoFisher cat# 31460) was added at 1:3000 dilution in TBST and incubated at room temperature for 1 h. mCherry protein was detected with 1-Step Ultra TMB-Blotting Solution (ThermoFisher Scientific, cat# 37574).

Tissue embedding and cryosectioning: The stomach was dissected out from Perccl-mCherry transgenic mice at age of P8 and fixed in 4% PFA for 2-3 hours at 4 °C. After several washes with PBS, the tissue was soaked in 10%, 20% and 30% sucrose in PBS, each for 1 h at 4°C and then embedded in a 1:1 mixture of Tissue-Tek O.C.T. Compound (VWR, cat# 25608-930) and 30% sucrose in PBS. Embedded tissue was sectioned using a cryostat (10μm sections).

Immunofluorescence and cell counting: Sections were washed with PBS 3× for 5 min each wash, incubated in 0.2% TritonX-100 in PBS for 20 min and washed again with 3× in PBS for 5 min each wash at room temperature. Sections were blocked with 1% BSA in 0.1% PBT for 1 h at room temperature. Primary antibodies against mCherry (1:500, ThermoFisher Scientific, cat# PA5-34974), Synaptophysin (1:1000, Synaptic Systems, cat#101004), Gastrin (1:500, C-20, Santa Cruz Biotechnology, sc-7783), anti-E-Cadherin (1:500, BD Biosciences, cat#610181), beta-galactosidase (1:500, Abcam, cat#ab9361) and Endomucin (1:500, eBiosciences cat#14-5851-82) were used and incubated overnight at 4°C. After washing sections with 3× with PBS for 5 min each wash, sections were incubated for 1h at room temperature with combinations of Alexa Fluor 488 and 594 (for double fluorescence) or Alexa Fluor 488, 568 and 647 (for triple fluorescence) conjugate secondary antibodies (ThermoFisher Scientific) each at a 1:1000 dilution in 1% BSA in PBS. Sections were then washed 2× with PBS, treated with Hoechst for counterstaining of nuclei and mounted in Mowiol or VECTASHIELD HardFSet Mounting Medium (Vector Laboratories, H-1500). Gastrin positive cells were counted on consecutive stomach sections of wild-type (N=3) and ICR knockout (N=2) mice at P8. The fraction of gastrin expressing cells was calculated by normalization to the total number of nuclei determined via Cell Profiler (http://cellprofiler.org). Cell counts from each antral side (as shown in the schematic of **Extended Data Figure 11c**) were averaged.

Mouse Intestine Single Cell Analysis: Single-cell RNA-seq profiles[29] were retrieved from the Gene Expression Omnibus[30] (GEO, GSE92332). SAM alignment files[17] were downloaded from the Short Read Archive using the *sam-dump* utility (part of the SRA Toolkit v2.8.1, http://ncbi.github.io/sra-tools/). Coordinates of the reads mapping to *Perccl* were extracted using SAMtools v1.3.1[17], along with the corresponding cell barcodes and UMIs (CB:Z and UB:Z, respectively). These allowed the assignment of each unique transcript (via UMI) to the parental cell (via cell barcode). Cell-type classification as described in *Haber et al*[29]. were parsed from GSE92332_Regional_UMIcounts.txt.gz (available on the GEO, GSE92332). All the further data processing steps, plots and *p*-values calculations were performed in the statistical computing environment R v.3.4.2 (www.r-project.org).

Histological analysis of human biopsies: FFPE blocks were sectioned at a thickness of 4μm and a positive control was added on the right side of the slides. All immunostainings were fully calibrated on a Benchmark XT staining module (Ventana Medical Systems Inc., USA). Briefly, after sections were dewaxed and rehydrated, a CC1 Standard Benchmark XT pretreatment for antigen retrieval (Ventana Medical Systems) was selected for all immunostainings: Chromogranin A (1:500, Dako, Denmark), Synaptophysin, (1:200, Life Technologies, Invitrogen, USA), GAST, CGC and STT. Detection was performed with iView DAB Detection Kit (Ventana Medical Systems Inc., USA) and counterstained with hematoxylin (Ventana Medical Systems Inc., USA). After the run on the automated stainer was completed, slides were dehydrated in ethanol solutions (70%, 96%, and 100%) for one min each. Sections were then cleared in xylene for 2 min, mounted with Entellan and cover slips were added.

Generation of induced pluripotent stem cells (iPSCs) from patient lymphocytes: Whole blood was isolated by routine venipuncture from patient 2.1 and two healthy siblings (2.3- heterozygous carrier, 2.4- unaffected WT) at Sheba Medical Center in Israel, in preservative-free 0.9% sodium chloride containing 100U/ml heparin. Blood was then shipped overnight to Cincinnati Children's Hospital Medical Center for iPS cell generation. Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood by Ficoll centrifugation as previously described[28] and were used to derive iPSCs. Briefly, PBMCs were cultured for 4 d in DMEM containing 10% FCS, 100ng/ml SCF, 100ng/ml TPO, 100ng/ml IL3, 20ng/ml IL6, 100ng/ml Flt3L, 100ng/ml GM-CSF, and 50ng/ml M-CSF (Peprotech). Transduction using a polycistronic lentivirus expressing Oct4, Sox2, Klf4, cMyc and dTomato was performed[31] following the second day of culture in this media.

Transduced cells were then cultured for an additional 4 d in DMEM containing 10% FCS, 100ng/ml SCF, 100ng/ml TPO, 100ng/ml IL3, 20ng/ml IL6, and 100ng/ml Flt3L. Media was changed every other day. PBMCs were then plated on 0.1% gelatin-coated dishes containing 2 × 10^4 irradiated MEFs/cm^2 (GlobalStem, Rockville, MD), and cultured in hESC media containing 20% knockout serum replacement, 1mM L-glutamine, 0.1mM β-mercaptoethanol, 1× non-essential amino acids, and 4ng/ml bFGF until iPSC colony formation. Putative iPSC colonies were then manually excised and re-plated in feeder free culture conditions consisting of matrigel (BD BioSciences, San Jose, CA) and mTeSR1 (STEMCELL Technologies, Vancouver, BC). Lines exhibiting robust proliferation and maintenance of stereotypical human pluripotent stem cell morphology were then expanded and cryopreserved before use in experiments. Standard metaphase spreads and G-banded karyotypes were determined by the CCHMC Cytogenetics Laboratory.

Differentiation of iPSCs into intestinal organoids: The differentiation of induced human pluripotent stem cells was performed as previously described[17,32,33] with minor modifications. Briefly, two clonal iPSC lines from each donor were dispase passaged into a matrigel-coated 24-well tissue culture plate and cultured for 3 d in mTeSR1. Following definitive endoderm differentiation, the monolayers were treated for 4 d with RPMI medium 1640 (Gibco) containing 2% defined fetal calf serum, 1× non-essential amino acids, 3μM CHIR99021 (Stemgent) and 500ng/ml rhFGF4 (R&D Systems) to induce hindgut spheroid morphogenesis. After the 4th day, "day 0" HIOs were collected and embedded in matrigel matrix and cultured in Advanced DMEM/F12 (Gibco) containing 100 U/ml penicillin/streptomycin (Gibco), 2mM L-Glutamine (Gibco), 15mM HEPES (Gibco), N2 Supplement (Gibco), B27 Supplement (Gibco), and 100ng/mL rhEGF (R&D Systems) for up to 42 days, splitting, passaging, and changing the media periodically.

HIOs collected for immunofluorescence analysis were fixed in 4% paraformaldehyde for 1-2 h at room temperature, washed overnight at 4°C in PBS, and embedded in O.C.T. Compound (Sakura). 8-10μm thick sections were incubated with primary antibodies overnight at 4°C in 10% normal donkey serum, 0.05% Triton X-100-PBS solution and subsequently incubated with secondary antibodies for 1 h at room temperature. The primary antibodies used were: FoxA2 (1:500; Novus), E-Cadherin (1:500; R&D Systems), Synaptophysin (1:1000; Synaptic Systems), CDX2 (1:500; Biogenex), Pdx1 (1:5000;

Abcam). All secondary antibodies (AlexaFluor; Invitrogen) were used at 1:500 dilution. Confocal microscopy images were captured with a 20× plan apo objective on a Nikon A1Rsi Inverted, using settings of 0.5 pixel dwell time, 1024 resolution, 2× line averaging, and 2.0× A1 plus scan.

Total RNA was extracted from HIOs using a NucleoSpin RNA II kit (Macherey-Nagel), and cDNA was synthesized with SuperScript VILO (Invitrogen) using 300ng RNA. qPCR analysis was performed with TaqMan Fast Advanced Master Mix and custom designed TaqMan Array 96-Well FAST Plates (Applied Biosystems) consisting of the following targets: 18S-Hs99999901_s1; GAPDH-Hs99999905_m1; ARX-Hs00292465_m1; CHGA-Hs00900370_m1; SYP-Hs00300531_m1; NTS-Hs00175048_m1.

Supplementary Information is available in the online version of the paper.

Acknowledgments

The authors wish to thank the patients and their families for their cooperation and support. This work was supported by grants to D.L. from the SysKid EU FP7 project (241544), the Wolfson Family Charitable Trust and the Crown Human Genome Center at the Weizmann Institute of Science. A.V. and L.A.P. were supported by NHLBI grant R24HL123879, and NHGRI grants R01HG003988, and U54HG006997, and UM1HG009421 where research was conducted at the E.O. Lawrence Berkeley National Laboratory and performed under Department of Energy Contract DE-AC02-05CH11231, University of California. J.M.W. and Y.A. were supported by the Cincinnati Children's - Sheba Medical Center's Joint Research Fund. J.M.W. and M.F.K. were supported by NIH grants (1R01DK092456 and 1U18NS080815) as well as a digestive disease center grant (P30 DK0789392). iPSC lines were generated in collaboration with the Cincinnati Children's Pluripotent Stem Cell Facility. R.K. was supported by the the David and Elaine Potter Charitable Foundation. B.L.B. was supported by NIH grants HL089707, HL064658, and HL136182. I.B. was funded through an Imperial College Research Fellowship. This work was performed in partial fulfillment of the requirements for a Ph.D. degree of D.O.L., Weizmann Institute of Science, Rehovot, Israel and IBJ, The Sackler Faculty of Medicine, Tel Aviv University, Israel.

Author contributions

CH, RS, RS, BW, BPS, IB and YA recruited patients, provided patient care, and characterized the symptoms. BW, BPS, and YA obtained biopsies. DOL, DBG, EKR managed the sequencing, DOL, TO, AA, performed the bioinformatic analyses and discovered the mutation. MT, HCS, and RK provided SNP genotyping and linkage analysis. IBJ, DMY, HRF, IB, MO, and RM did experimental work, including mutation characterization. YZ, MO, HW, ASN, DED, KLVB, RMB, BLB, AV and LAP did the transgenic characterization and knockout generation. JMW, MFK, AP and CNM did the iPSC generation and human intestinal organoid studies. DOL, AV, LP and DL wrote the manuscript. RS, DBJ, EP, LAP, DL and YA provided leadership to the project. All authors contributed to the final manuscript.

Author Information

Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. Correspondence and requests for materials should be addressed to YA (Yair.Anikster@sheba.health.gov.il), DL (Doron.Lancet@weizmann.ac.il), or LAP (LAPennacchio@lbl.gov).

References

1. Bamshad, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745–755 (2011).

2. Avery, G. B., Villavicencio, O., Lilly, J. R. & Randolph, J. G. Intractable diarrhea in early infancy. *Pediatrics* **41**, 712–722 (1968).

3. Straussberg, R. *et al.* Congenital intractable diarrhea of infancy in Iraqi Jews. *Clin. Genet.* **51**, 98–101 (1997).

4. Canani, R. B. & Terrin, G. Recent progress in congenital diarrheal disorders. *Curr. Gastroenterol. Rep.* **13**, 257–264 (2011).

5. Qu, H. & Fang, X. A brief review on the Human Encyclopedia of DNA Elements (ENCODE) project. *Genomics Proteomics Bioinformatics* **11**, 135–141 (2013).

6. Calo, E. & Wysocka, J. Modification of enhancer chromatin: what, how, and why? *Mol. Cell* **49**, 825–837 (2013).

7. Eeckhoute, J. *et al.* Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res.* **19**, 372–380 (2009).

8. Pennacchio, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. *Nature* **444**, 499–502 (2006).

9. Barker, N. *et al.* Lgr5(+ve) stem cells drive self-renewal in the stomach and build long-lived gastric units in vitro. *Cell Stem Cell* **6**, 25–36 (2010).

10. Caron, T. J., Scott, K. E., Fox, J. G. & Hagen, S. J. Tight junction disruption: Helicobacter pylori and dysregulation of the gastric mucosal barrier. *World J. Gastroenterol.* **21**, 11411–11427 (2015).

11. Helander, H. F. & Fändriks, L. The enteroendocrine 'letter cells' - time for a new nomenclature? *Scand. J. Gastroenterol.* **47**, 3–12 (2012).

12. Spence, J. R. *et al.* Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature* **470**, 105–109 (2011).

13. Mellitzer, G. *et al.* Loss of enteroendocrine cells in mice alters lipid absorption and glucose homeostasis and impairs postnatal survival. *J. Clin. Invest.* **120**, 1708–1721 (2010).

14. Thiagarajah, J. R. *et al.* Advances in Evaluation of Chronic Diarrhea in Infants. *Gastroenterology* **154**, 2045-2059.e6 (2018).

15. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243 (2018).

16. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* **25**, 1754–1760 (2009).

17. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).

18. Ge, D. *et al.* SVA: software for annotating and visualizing sequenced human genomes. *Bioinforma. Oxf. Engl.* **27**, 1998–2000 (2011).

19. Zhu, M. *et al.* Using ERDS to infer copy-number variants in high-coverage genomes. *Am. J. Hum. Genet.* **91**, 408–421 (2012).

20. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* **7**, 562–578 (2012).

21. Bockenhauer, D. *et al.* Epilepsy, ataxia, sensorineural deafness, tubulopathy, and KCNJ10 mutations. *N. Engl. J. Med.* **360**, 1960–1970 (2009).

22. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

23. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).

24. Lindemann, S. R. *et al.* The epsomitic phototrophic microbial mat of Hot Lake, Washington: community structural responses to seasonal cycling. *Front. Microbiol.* **4**, 323 (2013).

25. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).

26. Yang, H. *et al.* One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370–1379 (2013).

27. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).

28. Kvon, E. Z. *et al.* Progressive Loss of Function in a Limb Enhancer during Snake Evolution. *Cell* **167**, 633-642.e11 (2016).

29. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333–339 (2017).

30. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* **41**, D991-995 (2013).

31. Warlich, E. *et al.* Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. *Mol. Ther. J. Am. Soc. Gene Ther.* **19**, 782–789 (2011).

32. McCracken, K. W., Howell, J. C., Wells, J. M. & Spence, J. R. Generating human intestinal tissue from pluripotent stem cells in vitro. *Nat. Protoc.* **6**, 1920–1928 (2011).

33. Glusman, G., Caballero, J., Mauldin, D. E., Hood, L. & Roach, J. C. Kaviar: an accessible system for testing SNV novelty. *Bioinforma. Oxf. Engl.* **27**, 3216–3217 (2011).

Extended Data Figure 1. Family pedigrees. Filled black symbols are affected, and deletion genotypes are indicated in red. Exome sequencing was done for individuals 1.1, 2.1, 3.1, 4.1, 4.2; whole genome sequencing was done for individual 2.1. Transcriptome analysis done for 2.1, 2.4. Patient 1.1 (*) was found to have uniparental disomy (UPD).

Extended Data Figure 2: Whole genome linkage analysis. Analysis of SNP genotyping performed on six of the patients in families 1-5 and their 22 relatives detected a single significant telomeric linkage interval on chr16 with a max LOD score of 4.26. Haplotype reconstruction confirmed this interval with flanking marker rs207435 (chr16: 2,984,868) and showed two distinct disease haplotypes in an either homozygous setting in affected individuals for disease allele 1 (i.e. ΔL) in families 2, 3, 5, or a compound heterozygous setting for disease alleles 1 and 2 (i.e. ΔS) in family 4. All affected individuals carrying disease allele 1 showed an identical disease haplotype from rs533184 (chr16: 1,155,025) to rs397435 (chr16: 2,010,138).

Extended Data Figure 3: Schematic of reads covering exons in the *C16orf91* gene, for the five exome-sequenced patients and for three controls sequenced under identical conditions. The first three patients with a ΔL/ΔL genotype had zero-coverage in the three upstream exons (right). The last two patients with a ΔL/ΔS genotype had non-zero coverage in these exons, but significantly lower than controls. The downstream exons (left) had high coverage in all subjects. Numbers indicate scale in sequencing reads per base.

Extended Data Figure 4: Targeted deletion of the ICR non-coding sequence in mice. a, Overview of targeting approach. See **Methods** for details. **b,** Genotyping results obtained from genomic DNA (n=554) isolated from the tails of homozygous and heterozygous ICR deletion mice, compared to a wild type control. See **Methods** for primers and details. **c,** *Perccl* expression derived from mouse RNA-Seq from control littermates (left) and knockout mice (right). Tissues and timepoints are indicated to the left of each plot.

Extended Data Figure 5. Modified intestinal content in wild type (left) compared to chr17$^{\Delta ICR/\Delta ICR}$ mice at postnatal day 5 (n=45) (right).

Extended Data Figure 6. IRS deletion causes changes in intestinal and fecal microbiome composition. Microbial communities in different intestinal compartments and feces were profiled by 16S rRNA-based sequence profiling. **a,** Family-level relative abundance profiles of the top fifteen most abundant prokaryotic families for wild type (n=22) and chr17$^{\Delta ICR/\Delta ICR}$ (n=21) intestinal and fecal samples, organized by sample type. The most pronounced changes were observed in colon and fecal samples. **b,** Heatmap of log-transformed read counts for those genera exhibiting the greatest variance (top 60%) across all fecal samples. The abundance profiles exhibit perfect clustering of the fecal samples (rows) into wild type (n=6) and chr17$^{\Delta ICR/\Delta ICR}$ (n=7) groups. **c,** Bar plots of Shannon's diversity for all fecal samples from panel **b** grouped into wild type and chr17$^{\Delta ICR/\Delta ICR}$ sample types.

Extended Data Figure 7. Gastrointestinal X-gal (ß-gal) staining of ICR-reporter transgenic embryos compared to *Perccl* mRNA *in situ* hybridization . **a,b** E14.5 sections from a beta-galactosidase ICR-driven transgene. **c,d** *Perccl* mRNA *in situ* hybridization analysis on E14.5 wild-type sections. For X-gal staining and *in situ* hybridization experiments, two embryos for each experiment and each condition were collected at E14.5 and a minimum of three sections from each embryo were examined. Representative sections are shown.

Extended Data Figure 8. *Perccl* knockout mice have reduced body weight. *Perccl* knockout mice (n=38) weight comparison to littermate controls (n=25) (green boxes perccl knockouts, black diamonds littermate controls). Center point of lines represent the mean. Shaded areas represent +/- 1SD. *Perccl* knockout mice were generated in an FVB/N genetic background.

Extended Data Figure 9. *Perccl* transgenic rescue of the severe body weight phenotype found in chr17$^{\Delta ICR/\Delta ICR}$ mice. An 8.5kb *Perccl* mini gene was constructed (Supp. Table 10) and used to generate a

Perccl mouse line over-expressing PERCC1. Through introduction of this trangene into the chr17$^{\Delta ICR/\Delta ICR}$ mouse genetic background, we observed rescue of all the phenotypes found in chr17$^{\Delta ICR/\Delta ICR}$ mice including severe body weight reduction. Chr17$^{\Delta ICR/\Delta ICR}$ mice were generated in a mixed 129/C57Bl6 background. Center point of lines represent the mean. Shaded areas represent +/- 1SD. P values were determined using a 2 tailed T-test. n.s.= p value 0.8-1.0.

Extended Data Figure 10. Western blot analysis of PERCC1:mCherry fusion protein. Two stable transgenic lines (B3269 and B3309) were established through standard pronuclear microinjection of fertilized mouse eggs. Protein extracts from juvenile mice (P13/14) were run by SDS-PAGE and transferred for western hybridization. Lanes: 1, Molecular weight marker. 2/3, Line B3269. 4/5, Line B3309. 6, wild type control. 7, mCherry positive control protein. 8, Molecular weight marker. mCherry is predicted to be 28.8kD and the PERCC1:mCherry fusion protein 59kD with both proteins running ~5kD larger. Line B3309 does not express the fusion protein in contrast to Line B3269 (likely due to a position effect). These experiments were performed four times.

Extended Data Figure 11. Identification of cells with PERCC1+ identity in duodenum and stomach and impact of PERCC1 ablation on G-cells. Co-localization of Percc1-mCherry immunofluorescence with the pan-endocrine marker Synaptophysin at postnatal day 8 reveals PERCC1-positive cells in the duodenum and stomach corpus. **a,** Dispersed PERCC1+ cells (red) are observed in the villi of the duodenum. Upper panel: Cross-section through villi illustrates absence of endocrine identity (green) in these cells. Lower panel: Sagittal section showing distribution of PERCC1+ cells in the epithelium of villi (CDH1, green). **b,** A subpopulation of PERCC1+ cells (red) near the gland base of the corpus epithelium (mucosa) expresses Synaptophysin (SYP, green). Arrowheads mark double positive cells. Arrows mark a minor fraction of PERCC1+ cells detected in longitudinal smooth muscle cells (lSM) but not in circular smooth muscle cells (cSM). DAPI-stained nuclei are shown blue. **c,** Upper panel: Schematic depicting the anatomical compartments of the distal stomach and location of sections used for cell counting. Lower panel: Reduction of the fraction of G-cells observed predominantly in the pyloric antrum of *Perccl*-deficient (ΔICR/ΔICR) mice at P8. Box plots indicate median, interquartile values, range and individual biological replicates. Outliers are shown as circled data points. *p*, unpaired, two-tailed t-test. **d,** Comparative immunofluorescence analysis shows reduced numbers of Gastrin-expressing cells (red) in the absence of *Perccl* (ΔICR/ΔICR) in the pyloric antrum at P8. Synaptophysin-expressing endocrine cells are colored green and nuclei gray. *n* represents independent biological replicates with similar results. Scale bars, 50μm.

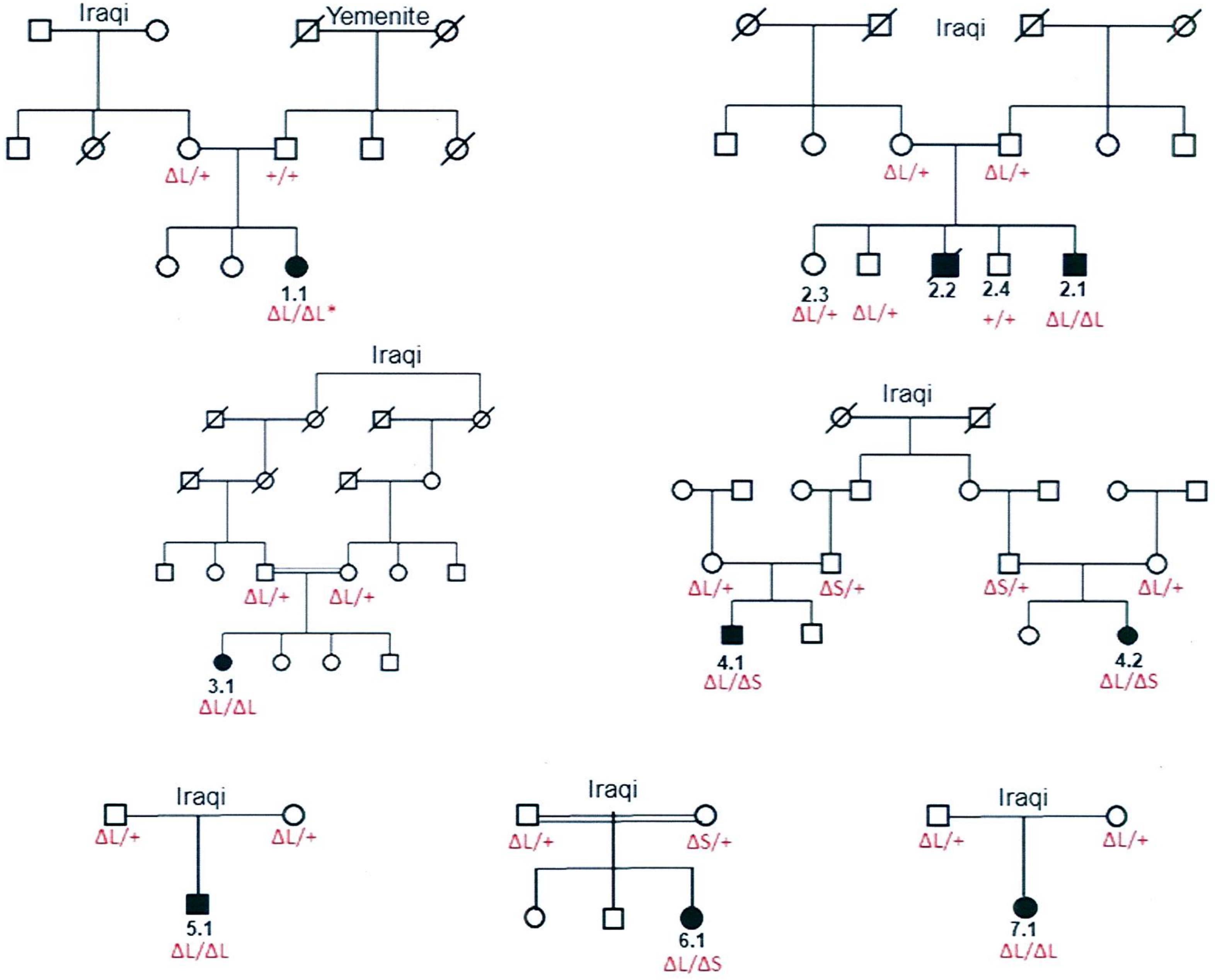
Extended Data Figure 12. a, Bar chart on the left shows the fraction of the total cells profiled in Haber *et al.* 2017 (n=11,665) assigned to each one of the major cell types identified. Bar chart on the right shows the same information but limited to those cells expressing Percc1 (n=8). **b,** Same as **(a)** but limited to Enteroendocrine Cells. **(a-b)** p-values were calculated using a Chi-squared test, using data from the corresponding left panel as reference. **c,** Box plots showing the distributions of the normalized expression values for known enteroendocrine-cell-associated transcription factors and hormones in the eight Percc1-positive cells from **(a)**. In boxplots, the middle line denotes median; box denotes interquartile range (IQR); and whiskers denote 1.5 x IQR.

Extended Data Figure 13. Validation of human RNA-Seq data by quantitative RT-PCR in duodenum tissue from 2 different patients and control tissue. Six peptide hormones relative expression levels are presented compared with normal duodenum tissue (control), which is represented as 1. Relative expression levels for patients represent the average between two patients (patient 1.1 and 5.1). Gene symbols are provided beneath each pairwise comparison. NTS, neurotensin. GCG, glucagon. CCK, choleocystokinin. GAST, gastrin. SST, somatostatin. GIP, gastric inhibitory polypeptide.

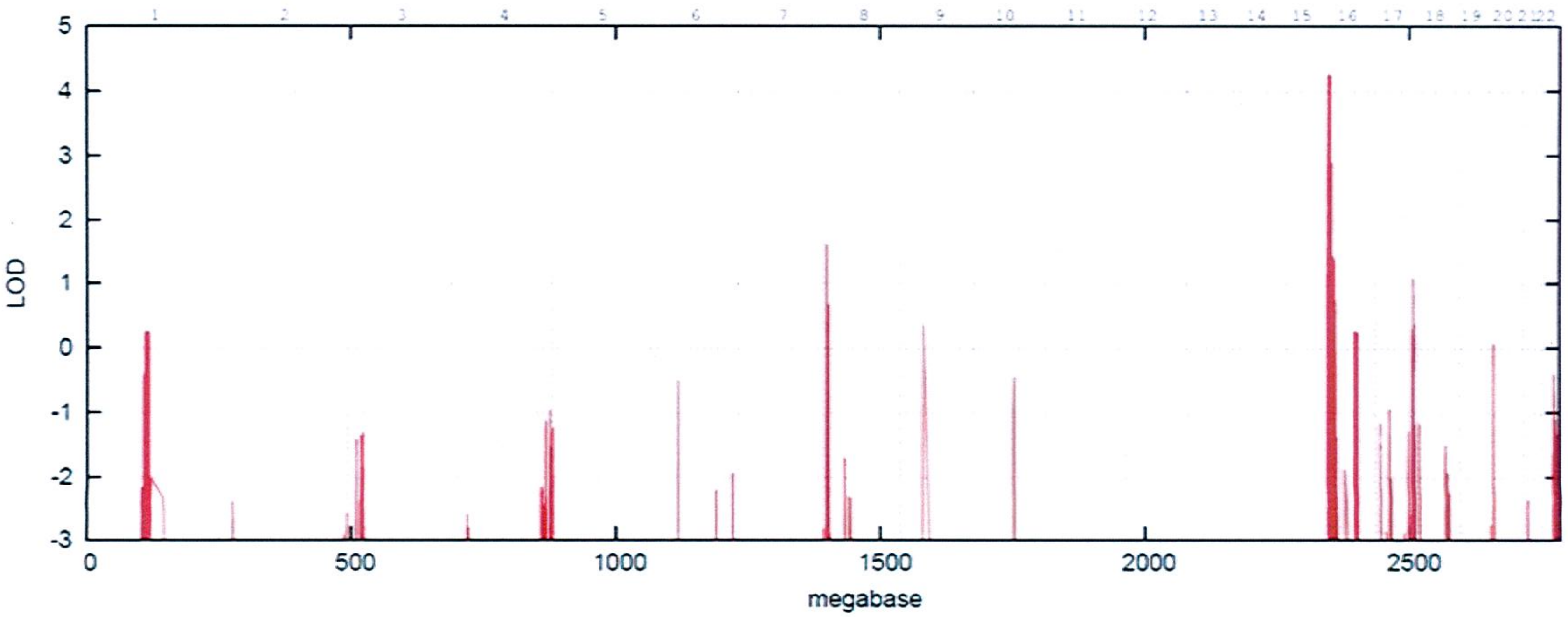
Extended Data Figure 14. Human intestinal organoids (HIOs) generated from affected patient, carrier and wild-type control all showing normal morphology.

Extended Data Figure 15. Affected patient, carrier and wild-type control-iPSC lines showing normal karyotype.

ED Fig 1

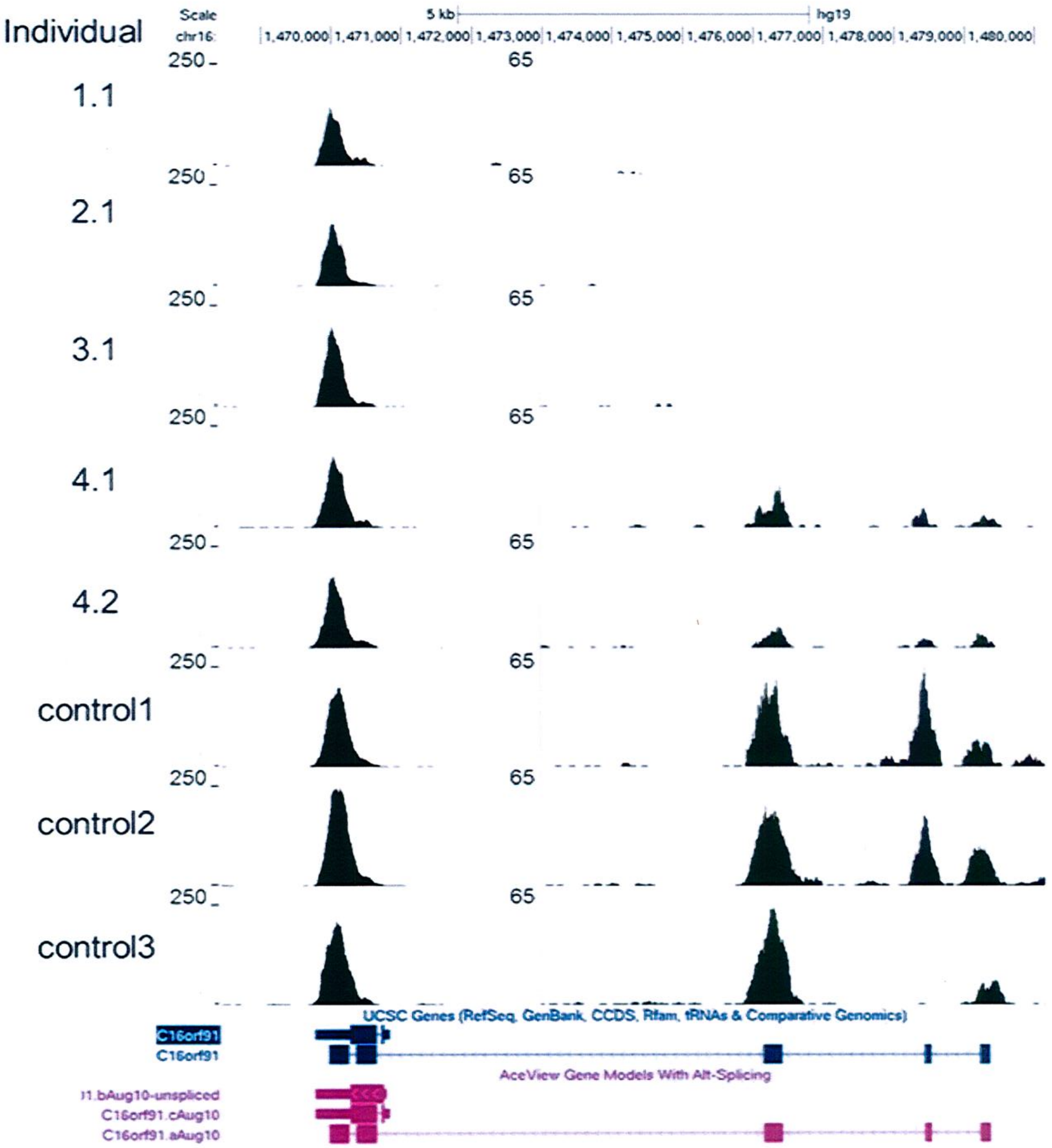

Iraqi
Yemenite
ΔL/+
+/+
1.1
ΔL/ΔL*
Iraqi
ΔL/+
ΔL/+
2.3
ΔL/+
ΔL/+
2.2
2.4
+/+
2.1
ΔL/ΔL
Iraqi
ΔL/+
ΔL/+
3.1
ΔL/ΔL
Iraqi
ΔL/+
ΔS/+
ΔS/+
ΔL/+
4.1
ΔL/ΔS
4.2
ΔL/ΔS
Iraqi
ΔL/+
ΔL/+
5.1
ΔL/ΔL
Iraqi
ΔL/+
ΔS/+
6.1
ΔL/ΔS
Iraqi
ΔL/+
ΔL/+
7.1
ΔL/ΔL

ED Fig 2

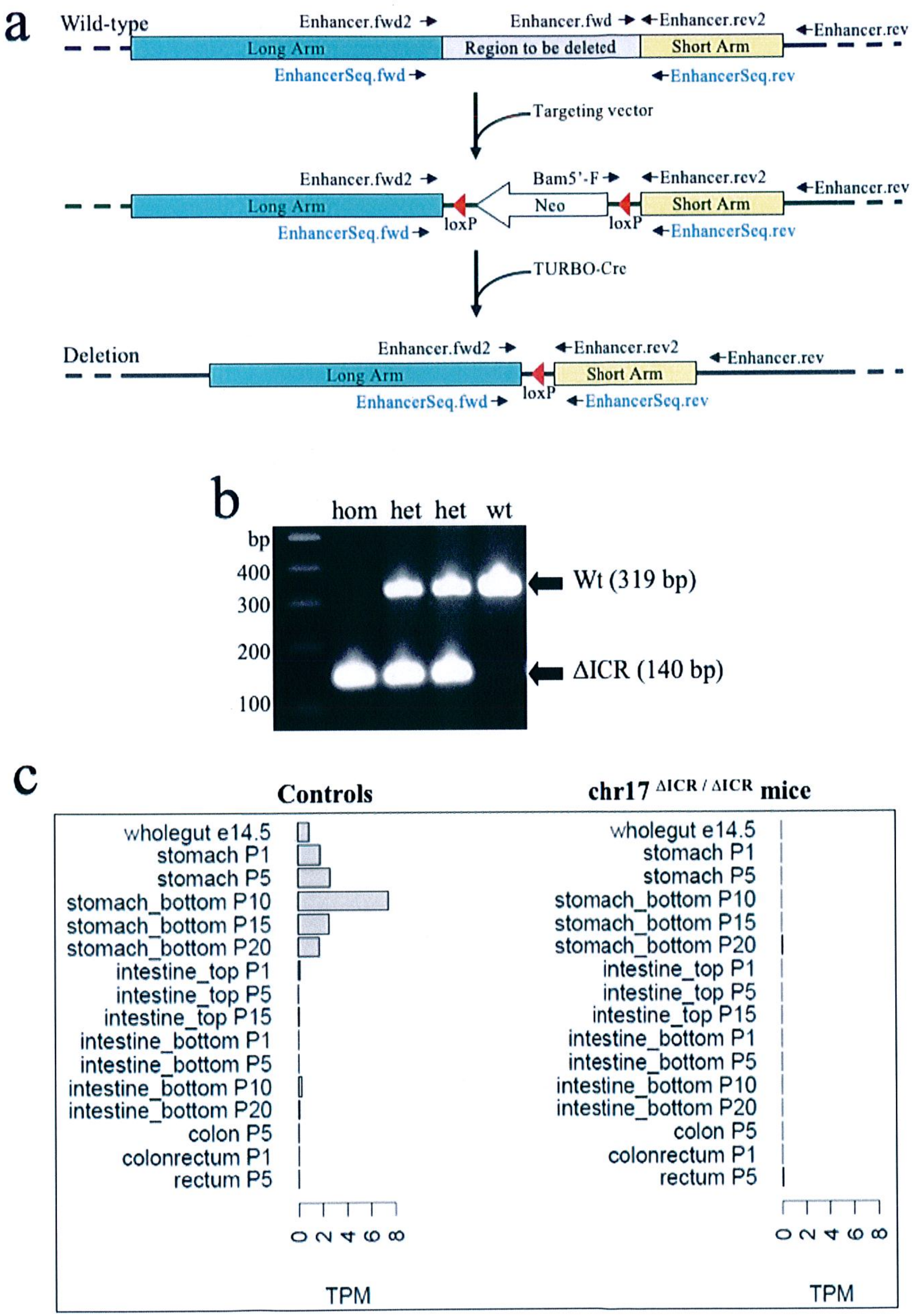


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
5
4
3
2
1
0
-1
-2
-3
LOD
0
500
1000
1500
2000
2500
megabase

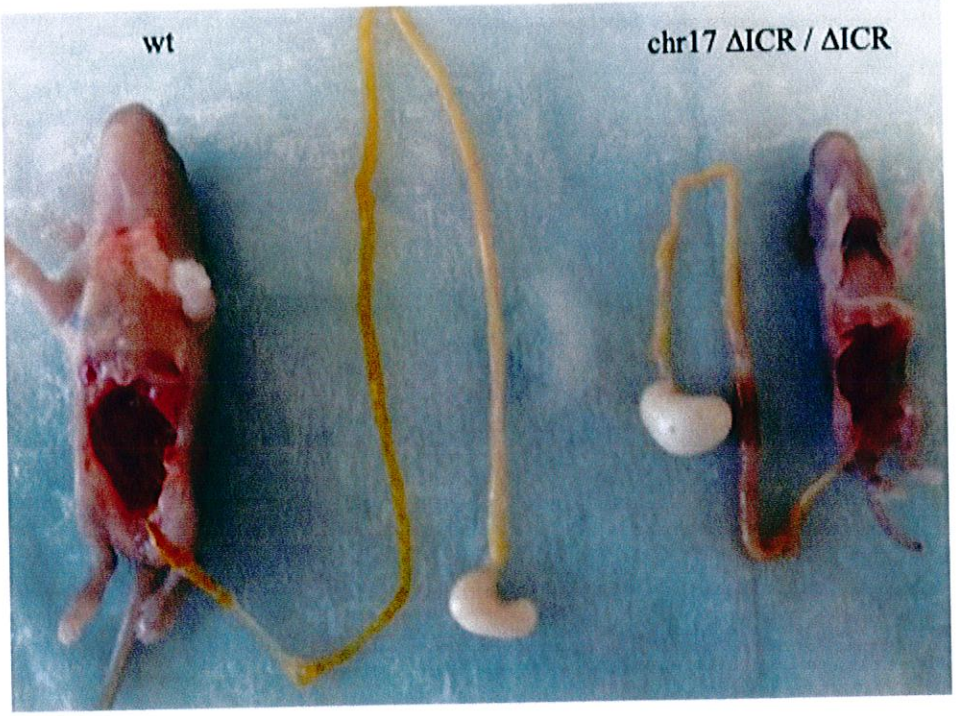
ED Fig 3


Individual
Scale
chr16:
5 kb
hg19
1,470,000
1,471,000
1,472,000
1,473,000
1,474,000
1,475,000
1,476,000
1,477,000
1,478,000
1,479,000
1,480,000
250
65
1.1
250
65
2.1
250
65
3.1
250
65
4.1
250
65
4.2
250
65
control1
250
65
control2
250
65
control3
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
C16orf91
C16orf91
AceView Gene Models With Alt-Splicing
)1.bAug10-unspliced
C16orf91.cAug10
C16orf91.aAug10

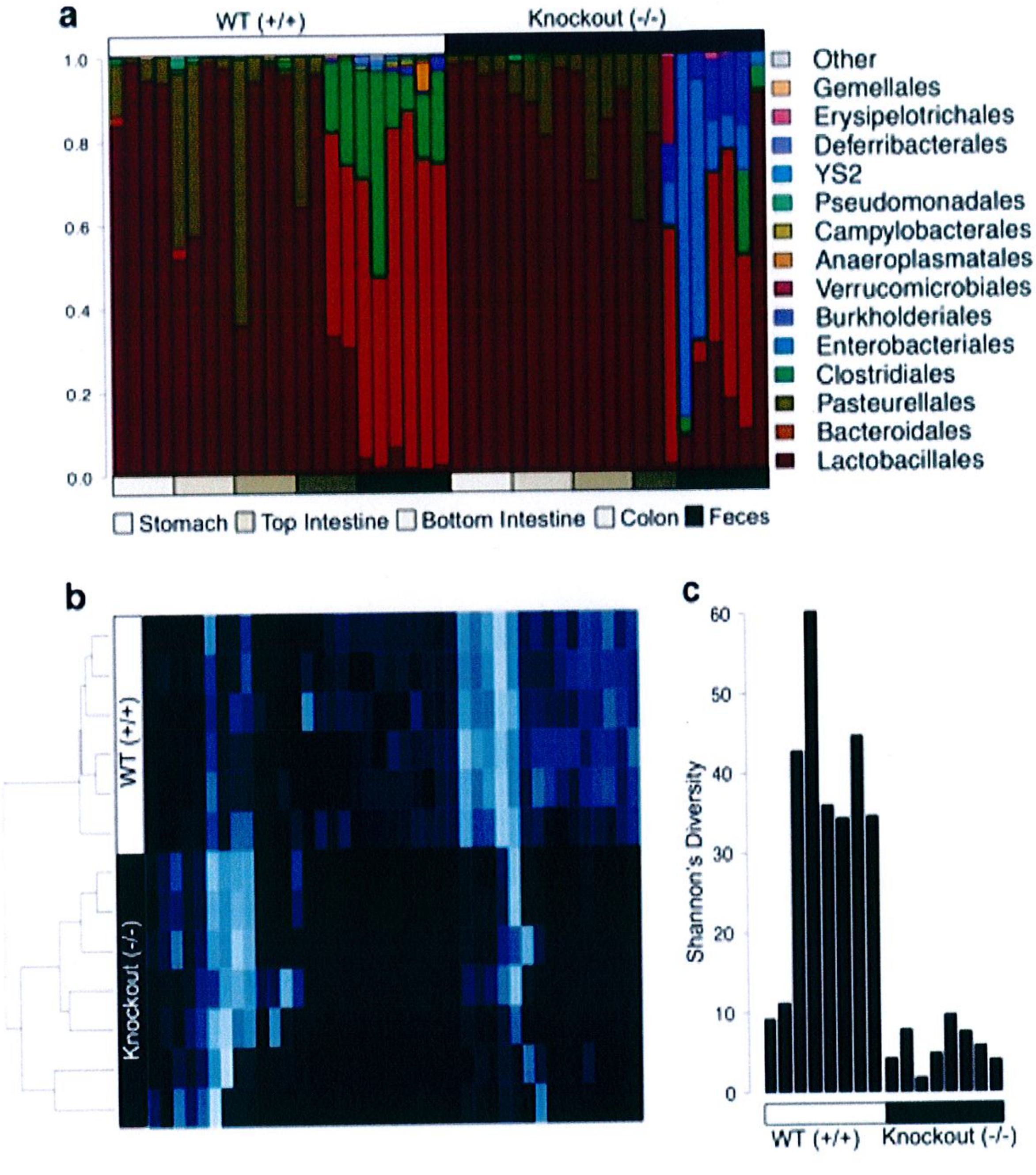
ED Fig 4


a
Wild-type
Enhancer.fwd2
Enhancer.fwd
Enhancer.rev2
Enhancer.rev
Long Arm
Region to be deleted
Short Arm
EnhancerSeq.fwd
EnhancerSeq.rev
Targeting vector
Enhancer.fwd2
Bam5'-F
Enhancer.rev2
Enhancer.rev
Long Arm
Neo
Short Arm
loxP
loxP
EnhancerSeq.fwd
EnhancerSeq.rev
TURBO-Cre
Deletion
Enhancer.fwd2
Enhancer.rev2
Enhancer.rev
Long Arm
Short Arm
loxP
EnhancerSeq.fwd
EnhancerSeq.rev
b
hom
het
het
wt
bp
400
300
200
100
Wt (319 bp)
ΔICR (140 bp)
c
Controls
chr17 ΔICR / ΔICR mice
wholegut e14.5
stomach P1
stomach P5
stomach_bottom P10
stomach_bottom P15
stomach_bottom P20
intestine_top P1
intestine_top P5
intestine_top P15
intestine_bottom P1
intestine_bottom P5
intestine_bottom P10
intestine_bottom P20
colon P5
colonrectum P1
rectum P5
0
2
4
6
8
TPM
0
2
4
6
8
TPM

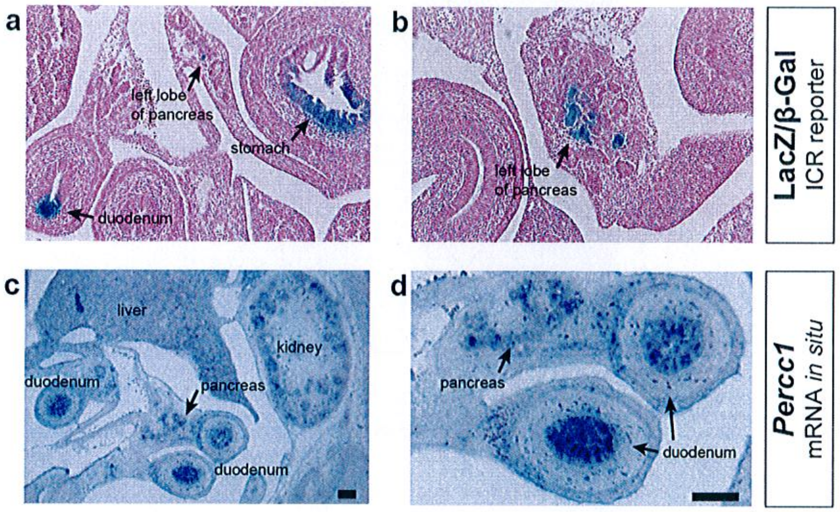
ED Fig 5



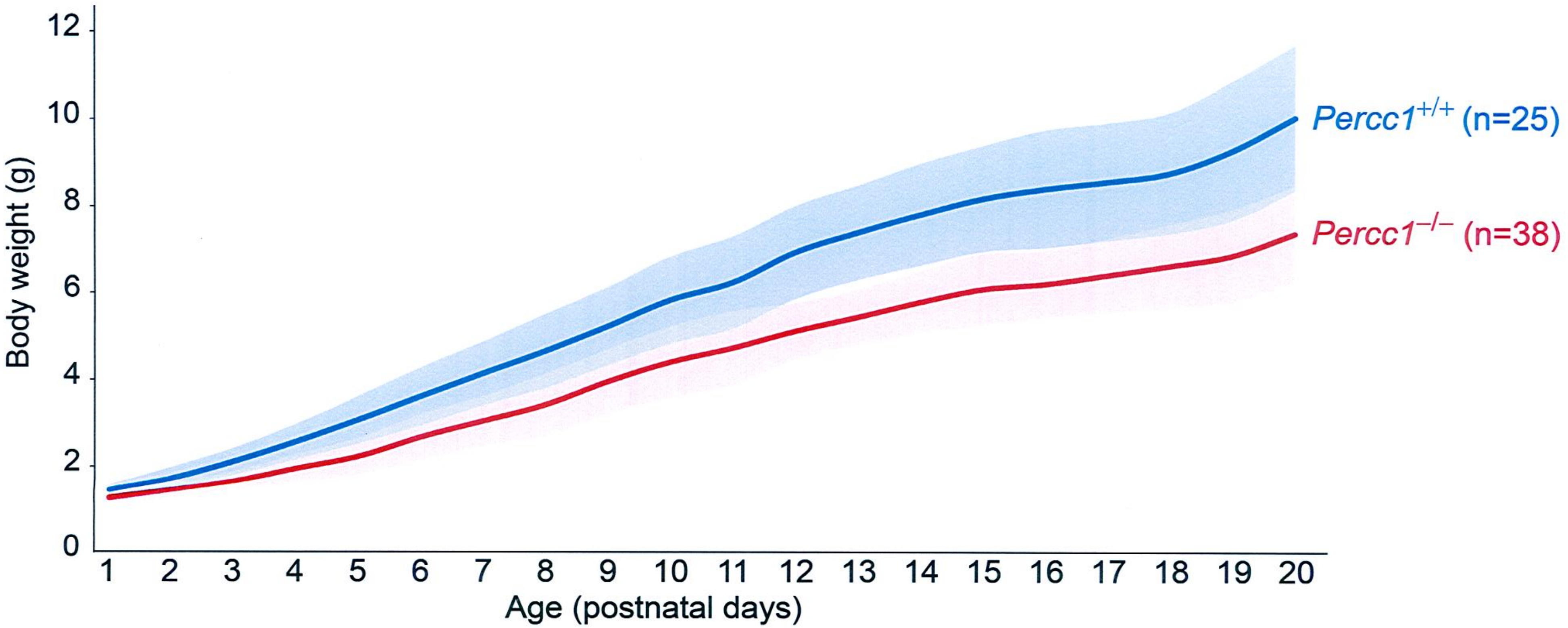
ED Fig 6


a
WT (+/+)
Knockout (-/-)
1.0
0.8
0.6
0.4
0.2
0.0
Other
Gemellales
Erysipelotrichales
Deferribacterales
YS2
Pseudomonadales
Campylobacterales
Anaeroplasmatales
Verrucomicrobiales
Burkholderiales
Enterobacteriales
Clostridiales
Pasteurellales
Bacteroidales
Lactobacillales
Stomach
Top Intestine
Bottom Intestine
Colon
Feces
b
WT (+/+)
Knockout (-/-)
c
60
50
40
30
20
10
0
Shannon's Diversity
WT (+/+)
Knockout (-/-)

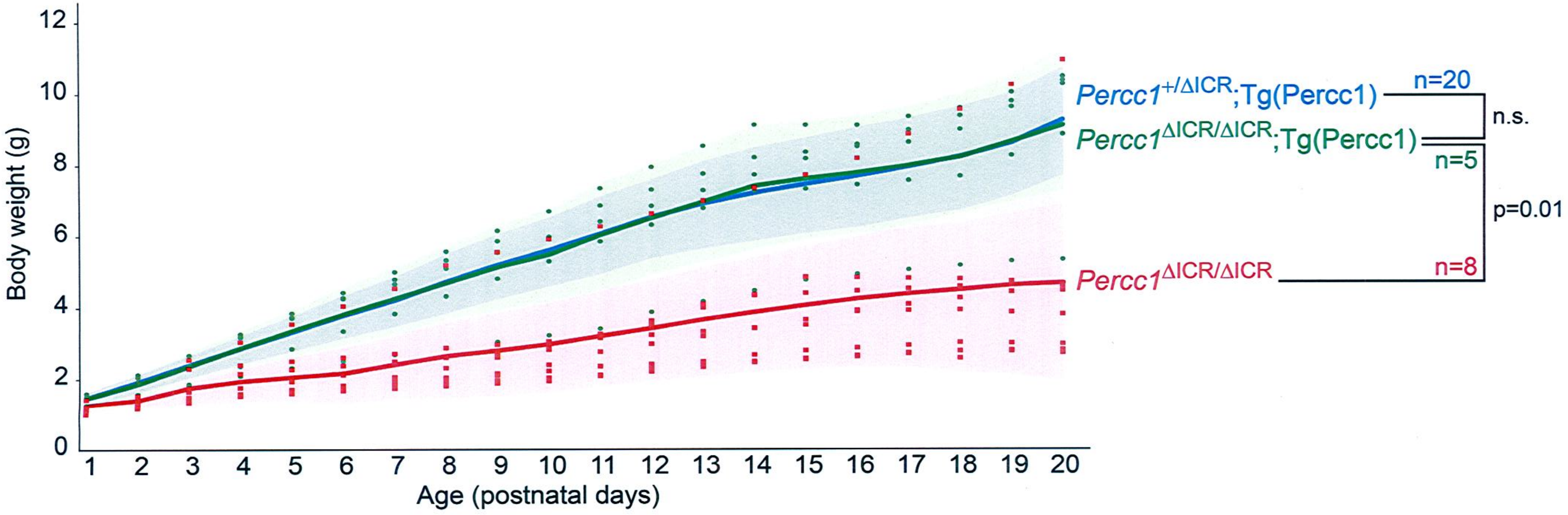
ED Fig 7


a
left lobe
of pancreas
stomach
duodenum
b
left lobe
of pancreas
LacZ/β-Gal
ICR reporter
c
liver
kidney
duodenum
pancreas
duodenum
d
pancreas
duodenum
Percc1
mRNA in situ

ED Fig 8

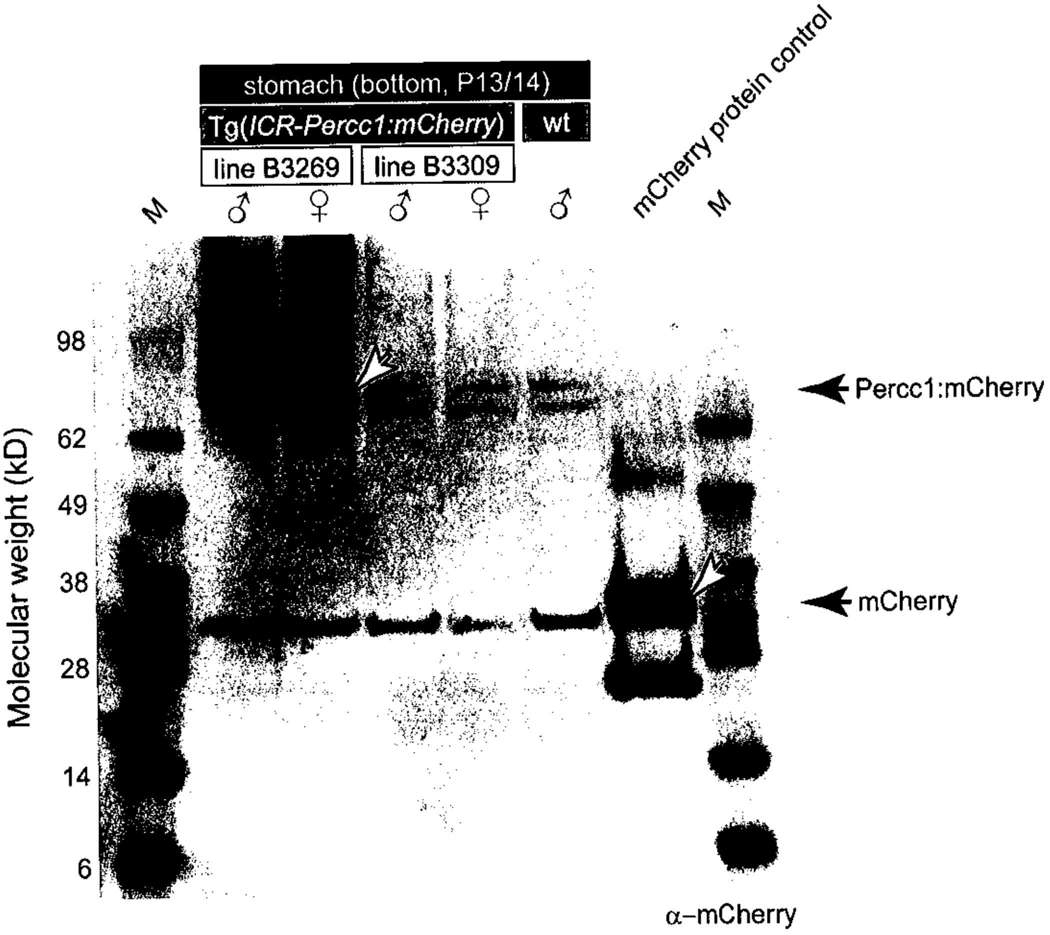

12
10
8
6
4
2
0
Body weight (g)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Age (postnatal days)
Percc1$^{+/+}$ (n=25)
Percc1$^{-/-}$ (n=38)

ED Fig 9

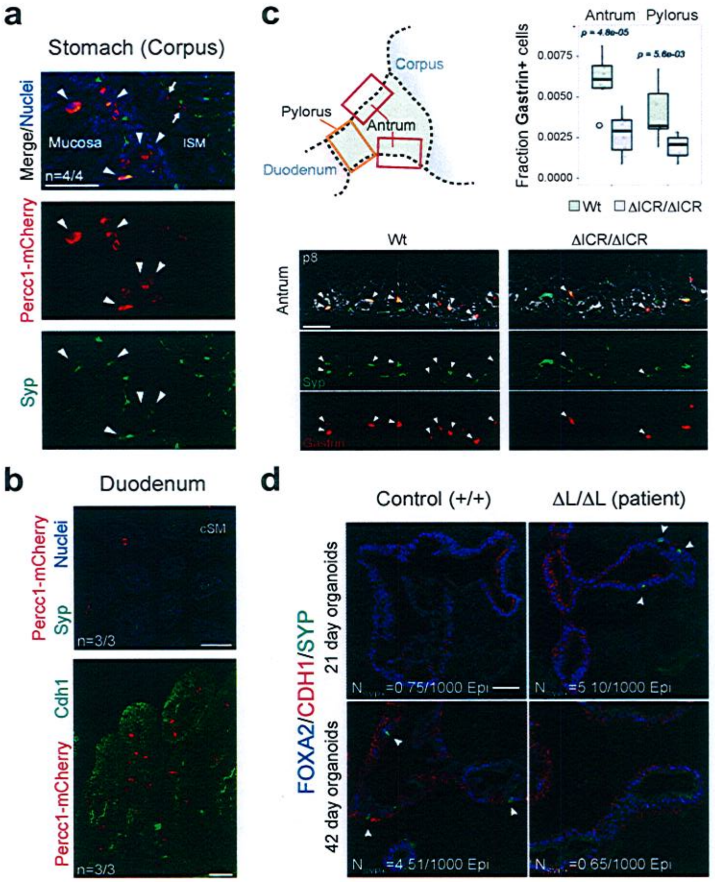


Body weight (g)
0
2
4
6
8
10
12
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Age (postnatal days)
Percc1+/ΔICR;Tg(Percc1)
n=20
n.s.
Percc1ΔICR/ΔICR;Tg(Percc1)
n=5
p=0.01
Percc1ΔICR/ΔICR
n=8

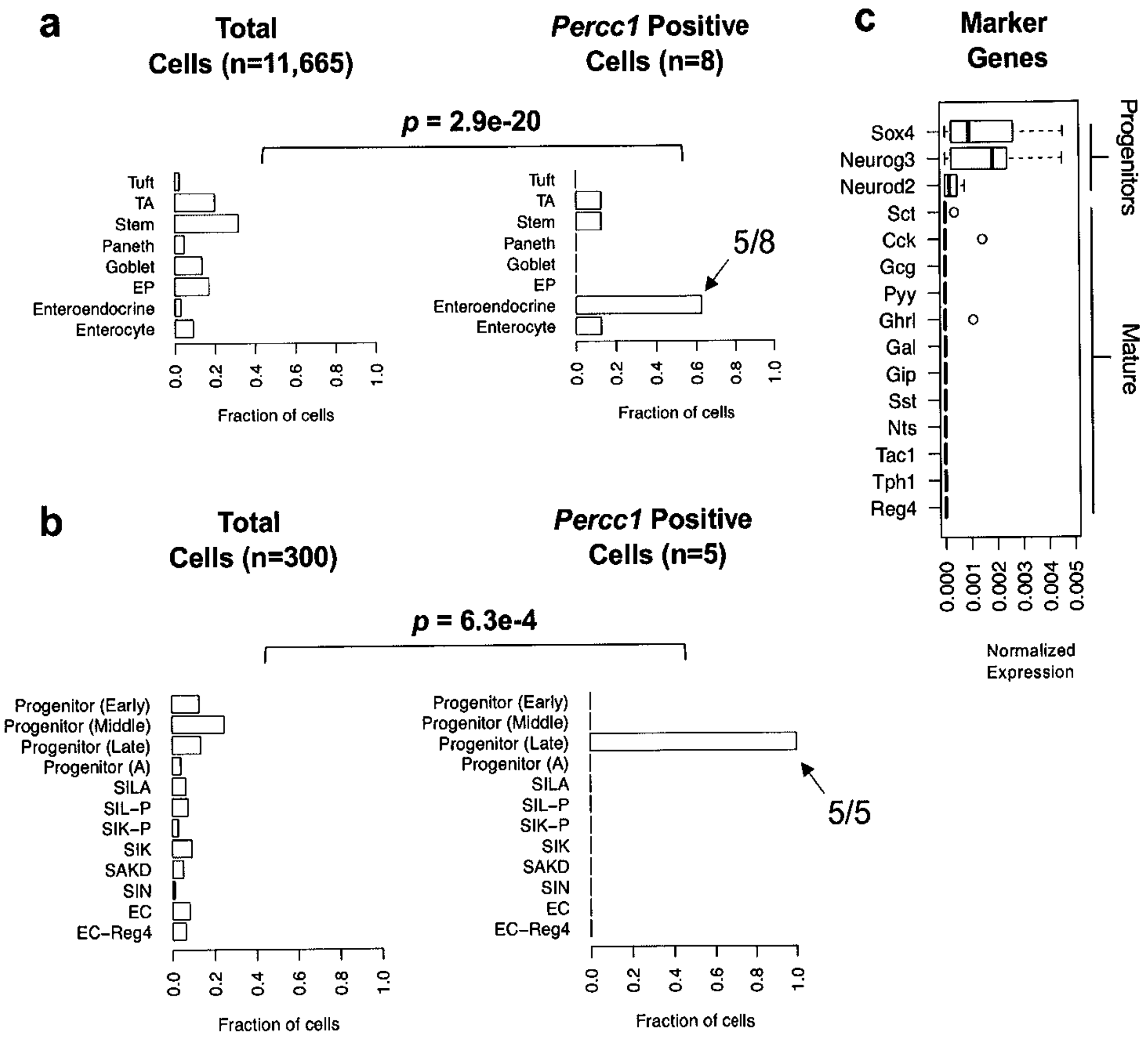
ED Fig 10


stomach (bottom, P13/14)
Tg(*ICR-Percc1:mCherry*)
wt
line B3269
line B3309
M
♂
♀
♂
♀
♂
mCherry protein control
M
Molecular weight (kD)
98
62
49
38
28
14
6
Percc1:mCherry
mCherry
α–mCherry

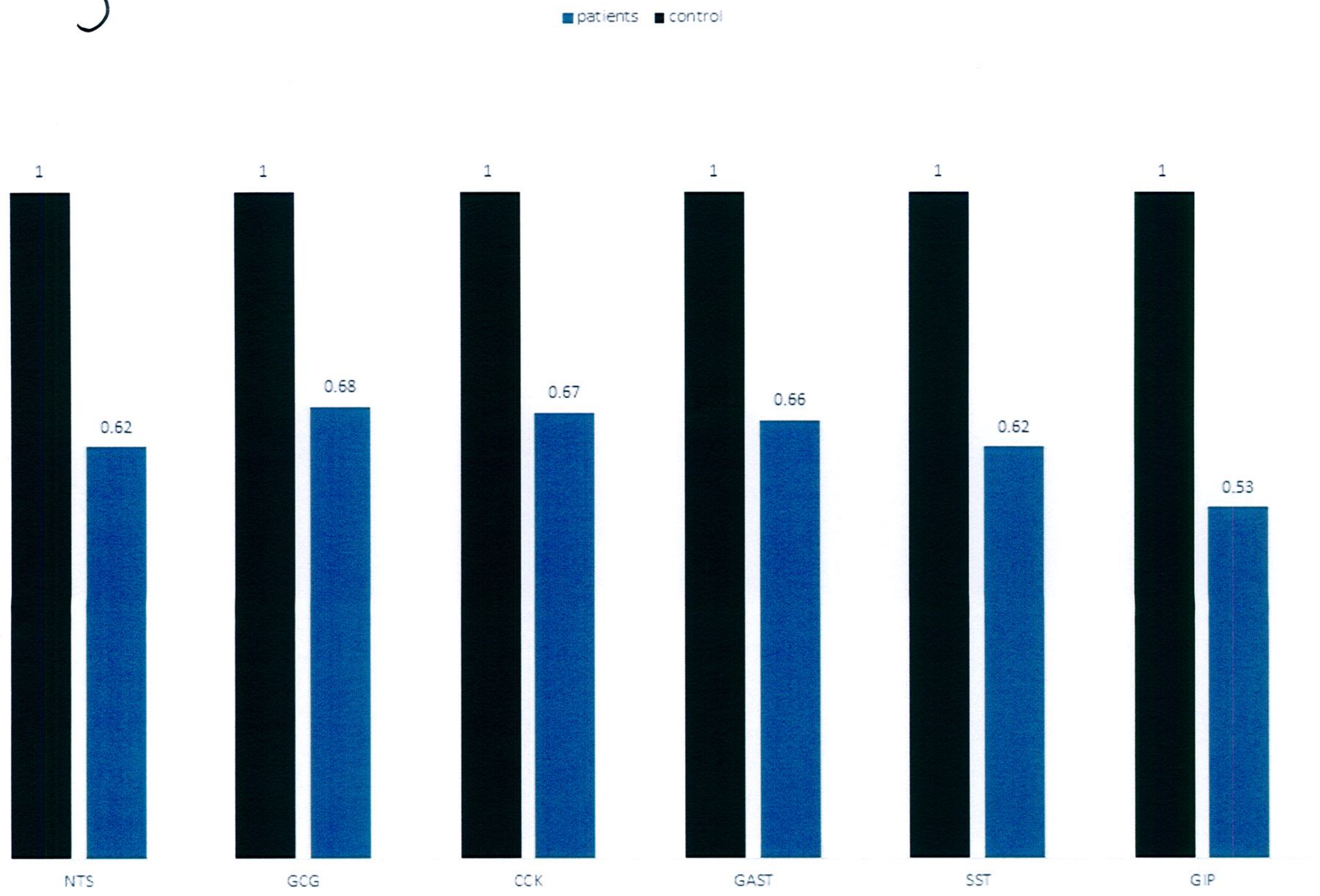
ED Fig 11


a
Stomach (Corpus)
Merge/Nuclei
Mucosa
ISM
n=4/4
Percc1-mCherry
Syp
b
Duodenum
Percc1-mCherry
Syp
Nuclei
cSM
n=3/3
Cdh1
Percc1-mCherry
n=3/3
c
Corpus
Pylorus
Antrum
Duodenum
Antrum Pylorus
p = 4.8e-05
p = 5.6e-03
Fraction Gastrin+ cells
0.0075
0.0050
0.0025
0.0000
🟩Wt □ΔICR/ΔICR
Wt
ΔICR/ΔICR
Antrum
p8
Syp
Gastrin
d
Control (+/+)
ΔL/ΔL (patient)
FOXA2/CDH1/SYP
21 day organoids
42 day organoids
N=0 75/1000 Epi
N=5 10/1000 Epi
N=4 51/1000 Epi
N=0 65/1000 Epi

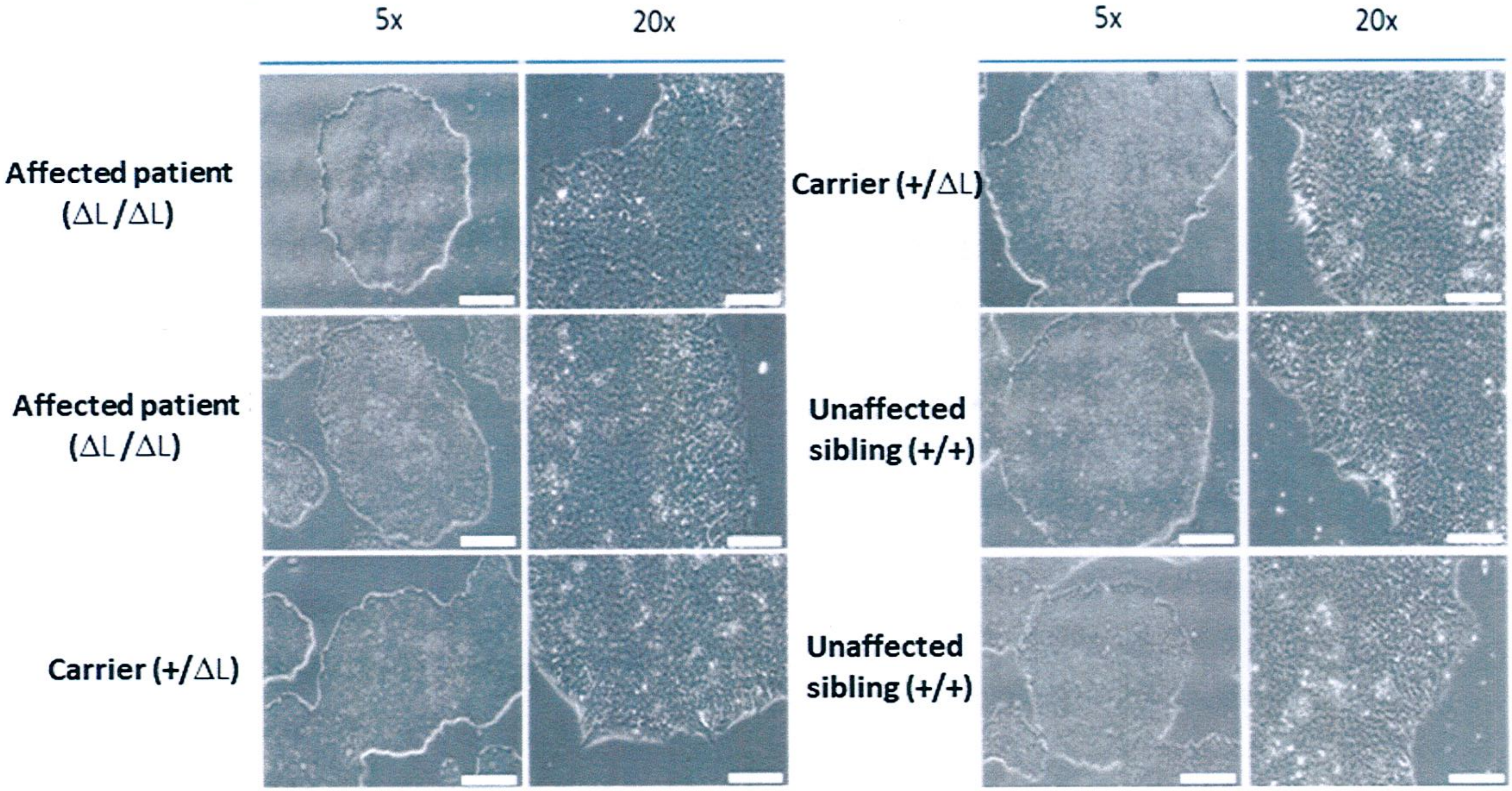
ED Fig 12


a
Total
Cells (n=11,665)
Percc1 Positive
Cells (n=8)
p = 2.9e-20
Tuft
TA
Stem
Paneth
Goblet
EP
Enteroendocrine
Enterocyte
0.0
0.2
0.4
0.6
0.8
1.0
Fraction of cells
5/8
b
Total
Cells (n=300)
Percc1 Positive
Cells (n=5)
p = 6.3e-4
Progenitor (Early)
Progenitor (Middle)
Progenitor (Late)
Progenitor (A)
SILA
SIL-P
SIK-P
SIK
SAKD
SIN
EC
EC-Reg4
Fraction of cells
5/5
c
Marker
Genes
Sox4
Neurog3
Neurod2
Sct
Cck
Gcg
Pyy
Ghrl
Gal
Gip
Sst
Nts
Tac1
Tph1
Reg4
Progenitors
Mature
0.000
0.001
0.002
0.003
0.004
0.005
Normalized
Expression

ED Fig 13


■ patients ■ control
1
0.62
NTS
1
0.68
GCG
1
0.67
CCK
1
0.66
GAST
1
0.62
SST
1
0.53
GIP

ED Fig 14


5x
20x
5x
20x
Affected patient (ΔL/ΔL)
Carrier (+/ΔL)
Affected patient (ΔL/ΔL)
Unaffected sibling (+/+)
Carrier (+/ΔL)
Unaffected sibling (+/+)