

1 **Title page**

2 **Comparison of human and murine enteroendocrine cells by transcriptomic and**
3 **peptidomic profiling**

4 **Short title: Enteroendocrine transcriptomics, peptidomics**

5 ‡Geoffrey P Roberts^{1,2}, ‡Pierre Larraufie¹, ‡Paul Richards^{1,3}, Richard G Kay¹, Sam G Galvin¹, Emily L
6 Miedzybrodzka¹, Andrew Leiter⁴, H. Joyce Li⁴, Leslie L Glass¹, Marcella KL Ma¹, Brian Lam¹, Giles SH Yeo¹,
7 Raphaël Scharfmann³, Davide Chiarugi¹, Richard H Hardwick², *Frank Reimann¹, *Fiona M Gribble¹

8 ‡ joint first authors

9 *joint senior and corresponding authors

- 10 1. Wellcome Trust – MRC Institute of Metabolic Science, University of Cambridge, UK
11 2. Cambridge Oesophago-gastric Centre, Addenbrooke's Hospital, Cambridge, UK
12 3. INSERM U1016, Institut Cochin, Université Paris-Descartes
13 4. Division of Gastroenterology, Department of Medicine, University of Massachusetts Medical School,
14 Worcester, Massachusetts, United States

15 *Correspondence to:*

16 Fiona Gribble – fmg23@cam.ac.uk

17 Frank Reimann – fr222@cam.ac.uk

18 Wellcome Trust – MRC Institute of Metabolic Science, Box 289, Addenbrooke's Hospital, Cambridge, CB2
19 0QQ

20 *Grant support:*

21 Metabolic Research Laboratories support was provided by the following core facilities: Disease Model Core,
22 Genomics and Transcriptomics Core, Histology Core, Imaging Core and NIHR BRC Core Biochemical Assay

1 Laboratory (supported by the MRC [MRC_MC_UU_12012/5] and Wellcome Trust [100574/Z/12/Z]). RNA-
2 sequencing was undertaken at the CRUK Cambridge Institute Genomics Core.

3 GR received an Addenbrooke's Charitable Trust / Evelyn Trust Cambridge Clinical Research Fellowship [16-
4 69], an EFSD project grant and a Royal College of Surgeons Research Fellowship. SG is supported by a grant
5 from the BBSRC. Research in the laboratory of Fiona Gribble and Frank Reimann is supported by the MRC
6 [MRC_MC_UU_12012/3], Wellcome Trust [106262/Z/14/Z, 106263/Z/14/Z] and research grants from
7 Medimmune. The MS instrument was funded by the MRC "Enhancing UK clinical research" grant
8 (MR/M009041/1). PR was supported by a postdoctoral grant from Agence Nationale de la Recherche
9 (Laboratoire d'Excellence Revive, Investissement d'Avenir; ANR-10-LABX-73). The RS lab is supported by ANR-
10 10-LABX-73 and the Bettencourt-Schueller foundation. This work was partially supported by the following
11 NIH grants to AL: R01 DK100223-02, R01 DK110614-01.

12 *Conflicts of interest*

13 FMG is a paid consultant for, and PR has (subsequent to his work on this project) taken up full time
14 employment with Kallyope, New York. The Gribble-Reimann lab hosts projects which receive funding from
15 Medimmune (FMG / FR). There are no other conflicts of interest to declare.

16 ***Data availability statement***

17 The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the
18 PRIDE partner repository with the dataset identifiers PXD009788 (10.6019/PXD009788) and PXD009796
19 (10.6019/PXD009796)(1). RNA-seq data has been deposited in the NCBI GEO repository (human –
20 GSE114853, mouse -GSE114913).

21

1 **Abstract**

2 Enteroendocrine cells (EECs) produce hormones that regulate food absorption, insulin secretion and
3 appetite. Both EECs and their peptide products are foci of drug discovery programmes for diabetes and
4 obesity. We compared the human and mouse EEC transcriptome and peptidome to validate mouse as a
5 model of the human enteroendocrine axis. We present the first RNA sequencing analysis of human EECs, and
6 demonstrate strong correlation with mouse, although with outliers including some low abundance G-protein
7 coupled receptors. Liquid chromatography mass spectrometry (LC-MS) identified peptide hormone gradients
8 along the human and mouse gut that should enhance progress in gut physiology and therapeutics.

9 **Keywords**

10 Gut hormones; GLP-1; enteroendocrine; PYY; GIP; bariatric surgery; transcriptome; peptidome; LC-MS;
11 RNAseq

12

1 **Introduction**

2 Enteroendocrine cells (EEC) are specialised hormone secreting cells in the intestinal epithelium which
3 monitor the quality and quantity of ingested foods. They produce at least 20 different hormones, mostly
4 peptides, that act in concert to coordinate digestion, peripheral nutrient disposal and appetite through
5 actions at local and distant target tissues. In the field of human metabolism, glucagon-like peptide-1 (GLP-1)
6 and peptideYY (PYY) have raised particular interest because of their central and pancreatic actions controlling
7 food intake and insulin secretion. GLP-1 based drugs are widely used for the treatment of diabetes and
8 obesity, and new gut hormone based therapeutics are under development, aiming to mimic the unrivalled
9 effectiveness of gastric bypass surgery on weight loss and type 2 diabetes resolution(2).

10 Our understanding of the physiology and pharmacology of human EECs is currently limited by a lack of
11 methods to identify and characterise this scattered cell population which only comprises ~1% of the intestinal
12 epithelium(3). By contrast, recent years have witnessed substantial progress in our understanding of murine
13 EEC physiology, facilitated by the generation of transgenic mice with fluorescently labelled EECs that enable
14 cell identification and functional characterisation through a range of approaches including fluorescence-
15 activated cell sorting (FACS), transcriptomics and live cell imaging(3-8). The lack of an evidence-based
16 platform for predicting the translatability of these findings from mouse to human represents, however, a
17 significant obstacle in drug discovery programmes. Whilst G-protein coupled receptors (GPCRs) that have
18 been identified and characterised in murine EECs represent promising candidates for therapeutic approaches
19 to enhance endogenous gut hormone secretion, early validation in a human model is strongly advisable(9).

20 Our objective was to compare mouse and human EECs at the transcriptomic and peptidomic levels. For
21 transcriptomics, we purified and RNA sequenced two subsets of human EECs distinguished by their
22 production (or not) of GLP-1, and compared the results with matching murine cell populations purified from
23 transgenic mouse models. Sequencing depth in both species was sufficient to enable the identification of low
24 abundance transcripts such as GPCRs and ion channels. We optimised protocols for peptide extraction and
25 liquid chromatography / tandem mass spectrometric (LC-MS/MS) analysis of mucosal tissue homogenates,

- 1 and identified exact sequences of the different gut peptides produced, and their relative quantities along the
- 2 length of the GI tract in human and mouse.
- 3

1 **Results**

2 *Cell and RNA collection for transcriptomics*

3 Tissue was obtained from 11 human participants recruited from sites in Cambridge and Paris (Supplementary
4 table 1). Jejunum samples were collected from all participants, and paired ileum and jejunum were obtained
5 from 2 transplant donors. Tissue pieces were digested, paraformaldehyde (PFA)-fixed and stained for
6 Chromogranin A (CHGA) and Secretogranin 2 (SCG2) as general markers for EECs, and for GLP-1 as a marker
7 of the EEC-subpopulation known as L-cells. By flow cytometry (Supplementary Figure 1A-C), we collected
8 pooled cell populations that were (i) positive for CHGA, SCG2 and GLP-1 (henceforth named GCG+), (ii)
9 positive for CHGA and SCG2 but negative for GLP-1 (henceforth named GCG-), and (iii) negative for all 3
10 markers (i.e. non-endocrine lineage cells, henceforth named negative). The GCG+ (L-cell) population
11 represented ~0.2% of all single cells examined, and the ratio of GCG+ to GCG- cells was ~1:5.

12 For the cross-species comparison, we collected unfixed murine EEC populations from the upper small
13 intestine of the mouse strain GLU-Venus (n=3) to identify *Gcg*-expressing L-cells (Supplementary Figure 1E-
14 G)(8), and of NeuroD1-Cre/EYFP mice(7) (n=3) to identify the total EEC population (Supplementary Figure 1I-
15 K). NeuroD1 is a well characterised EEC transcription factor found in the same cells as *Chga* in a previous
16 single cell analysis of murine EECs(10). GLU-Venus positive cells represented ~0.2% of singlets, and NeuroD1
17 positive cells ~0.6% of singlets.

18 RNA extracted from the purified fixed human cell populations had RIN values of only 2-3, with most RNA
19 fragments being 25-500 bases in length (Supplementary Figure 1D), but we hypothesised that the short
20 fragments might be sufficient for RNA sequencing using random primers. After RNA sequencing there were
21 high proportions of unmatched reads (up to 50% in some samples), likely due to the PFA modified input RNA,
22 but in all cases sequencing was achieved to a depth of at least 8.5×10^5 matched reads per sample. RNAs from
23 murine cell populations purified without fixation had RIN values of >7 (Supplementary Figure 1H,L) and were
24 sequenced to $>5 \times 10^6$ aligned reads per sample.

25 *Transcriptomics of human EEC populations*

1 In total we obtained individual RNA sequencing data from GCG+, GCG- and negative cells from each of 11
2 human jejunum samples and 2 human ileum samples. Principle component analysis (PCA) of the top 500
3 differentially expressed genes across all samples showed separation of the EECs (GCG+ and GCG-) from
4 negative cells on the first component, and of GCG+ cells from GCG- cells on the second component (Figure
5 1A). Pairwise analysis of key genes differentiating the cell populations from jejunum was performed using a
6 DESEQ2 model, and the normalised results for the 500 most differentially expressed genes are presented in
7 a heatmap (Figure 1B) which displays clear transcriptomic differences between the GCG+, GCG- and negative
8 cells(11). PCA did not demonstrate clustering of samples by BMI of the donor.

9 Both human EEC populations expressed a wide range of hormonal transcripts (Figure 2A,C,E). As expected
10 from the use of GLP-1 antibodies to purify GCG+ cells, the hormonal transcript showing the strongest
11 differential expression in GCG+ vs GCG- samples was *GCG* itself (Figure 2C). Consistent with previous findings
12 in mice(4, 12-14), we identified a range of additional hormonal transcripts in human GCG+ samples, including
13 *GIP* (glucose-dependent insulinotropic polypeptide), *CCK* (cholecystokinin), *NTS* (neurotensin), *PYY* and *SCT*
14 (secretin) as well as *MLN* (motilin) – a hormone produced by human but not mouse(4, 12-15). Compared with
15 GCG+ cells, GCG- cells had higher expression of *SCT*, *CCK*, *NTS*, *MLN*, *GHRL* (ghrelin) and *SST* (somatostatin),
16 together with *TPH1*, the enzyme responsible for serotonin biosynthesis in enterochromaffin cells. EECs also
17 expressed the putative gut hormones *UCN3* (urocortin 3), *PCSK1N* (ProSAAS) and *NPW* (neuropeptide W), as
18 well as lower levels of RNAs encoding several peptides not classically described as gut hormones such as *VGF*,
19 *GHRH* and *ADM*.

20 Transcripts of ~50 GPCRs were either enriched ($P < 0.1$) over the negative population, or expressed at > 100
21 CPM in one or both EEC populations (Figure 2B,D,F). Multiple GPCRs previously implicated in post-prandial
22 gut hormone secretion in mice were highly expressed in human EECs compared with negative cells, including
23 the fat-sensing receptors *FFAR1*, *FFAR2*, *FFAR3*, *FFAR4* and *GPR119*, the amino acid sensing receptors *CASR*
24 and *GPR142*, the butyrate and isovalerate sensing *OR51E1*(16), and the bile acid receptor, *GPBAR1*, as well
25 as the zinc receptor *GPR39*, and the alpha-ketoglutarate receptor *OXGR1*. Human EECs differentially
26 expressed a number of hormonal receptors compared with negative cells, including *SSTR1*, *SSTR2*, *SSTR5*,

1 *GIPR*, *AVPR1B*, *OXTR*, *CRHR2*, *RXFP4*, *GLP1R*, *NPY1R*, *CALCR*, *GALR1* and *SCTR*. Two opioid receptors, *OPRK1*
2 and *OPRD1*, not previously described as active in the gastrointestinal tract were identified in EECs, alongside
3 the oestrogen receptor *GPER1* and prostaglandin receptors *PTGDR2* and *PTGER3*. At least four orphan GPCRs
4 were differentially expressed in human EECs: *GPR148*, *GPR160*, *GPR173* and *GPR179*, hinting to as yet
5 undescribed pathways that may control gut hormone secretion.

6 Transcripts for a range of ion channel subunits were enriched in human EECs (Supplementary Figure 2),
7 consistent with previous reports that murine L-cells and enterochromaffin cells are electrically active. Human
8 GCG+ and GCG- EECs were particularly enriched for the voltage-gated calcium channel subunits *CACNA1A*
9 (P/Q-type), *CACNA1C* and *CACNA1D* (L-type), and *CACNA1H* (T-type)(17), and the voltage-gated sodium
10 channel subunits *SCN3A* and *SCN8A*. *TRPA1*, previously described to stimulate GLP-1 and 5-HT secretion in
11 mouse, was enriched in both EEC populations(18, 19), as was the amiloride/acid sensitive ion channel *ASIC5*,
12 and the K_{ATP} channel subunits *KCNJ11* and *ABCC8*(8, 20).

13 Transcription factor profiling of human GCG+ and GCG- cells is also shown in Supplementary Figure 2.

14 *Comparison of human jejunum vs ileum EEC transcriptomes*

15 PCA of the 3 cell populations from each of the 2 matched human jejunum and ileum samples separated EECs
16 from negative cells on the first component and the anatomical location of GCG+ and GCG- cells on the second
17 component (Figure 3A). This distinction was driven by a low number of differentially expressed genes, 279 in
18 GCG+ cells and 120 in GCG- cells, when analysed pairwise using a DESEQ2 model. Notable differentially
19 expressed transcripts that were higher in jejunal than ileal EECs (GCG+ or GCG-) included *GIP*, *CCK*, *SST*, *MLN*,
20 *SCT*, *GHRH*, *ASIC5*, and *TRPA1*; whereas transcripts higher in ileal than jejunal EECs included *GCG*, *NTS* and
21 *TAC1* (Figure 3,B,C).

22 *Comparison of human vs mouse EEC transcriptomes*

23 We compared human jejunal GCG+ cells with murine upper small intestinal GLU-Venus cells, representing
24 the GLP-1 secreting L-cell populations from each species (Figure 4). Only genes with 1:1 homology annotated
25 in Ensembl were included, and genes annotated as ribosomal, mitochondrial and small-nuclear were

1 excluded, giving a total of 15,507 genes for comparison. Log-log plots of normalised gene expression
2 indicated a strong correlation between the two species ($R^2=0.73$; Figure 4). Of the 354 genes lying outside
3 the 99% confidence interval of the linear model of human versus mouse, notable genes more highly
4 expressed in human than mouse GCG+ cells included *GIP*, *CHGA*, *ASIC5*, *GIPR*, *GPR142*, *SCTR*, *PTH2R*, *CHRNAS*
5 and *OPRK1*; whereas genes more highly expressed in mouse GCG+ cells included *Gpr174*, *Gpr171*, *Ghr*, *Grpr*,
6 *Ptger1*, *Cnr1*, *Insl5*, *Gpr22* and *Ghrl*.

7 Human jejunal GCG- cells, representing the wider murine EEC population (although depleted of L-cells) were
8 compared with murine NeuroD1-positive cells, revealing a strong correlation between these human and
9 murine EEC populations ($R^2=0.74$; Figure 5). Of the 380 genes lying outside the 99% confidence interval of
10 the linear model for this comparison, notable genes more highly expressed in human than murine EECs
11 included *GIPR*, *SCTR*, *GHRH*, *OPRK1*, *PTH2R* and *TRPV6*, and in murine compared with human cells included
12 *Ghrl*, *Tac1*, *lapp*, *Gpr22*, *Gast*, *Ptger1*, *Grpr* and *Ghr*.

13 Species-comparative data for subgroups of genes in L-cells and the total EEC population, separated by their
14 annotated roles as hormones, GPCRs, ion channels and transcription factors, are represented in Figures 5
15 and 6, in which the colour code additionally indicates whether the genes were more than four fold enriched
16 and significantly likely to be differentially expressed based on the DESEQ2 model (adjusted $P \leq 0.1$) in EECs
17 compared with negative cells in one or both species.

18 *Peptidomic analysis by LC-MS*

19 Peptide extraction and LC-MS protocols for analysis of fresh intestinal mucosal samples were optimised for
20 maximum peptide retrieval and identification (Supplementary Figure 3), and enabled reliable detection and
21 sequencing of peptides up to 65 amino acids in length. In the first instance, tissue samples were analysed by
22 nano LC-MS/MS and compared between human jejunum (n=4) and mouse mid-small intestine (n=4).
23 Peptides were assigned to their parental proteins by Peaks software, and were found to include known EEC
24 prohormones, granins and enteric neuronal signalling peptides as well as peptides derived from a variety of
25 house-keeping proteins that likely reflected the occurrence of some tissue damage/degradation prior to
26 homogenisation. Of the 463 and 705 different proteins matched in human and mouse respectively, 234 were

1 common between the two species, and showed good correlation ($R^2 = 0.54$, Figure 6A). To identify candidate
2 EEC-derived peptides, we restricted the analysis to peptides originating from genes that in the transcriptomic
3 analysis showed >4-fold higher expression in at least one EEC sample compared with the corresponding
4 negative cell population (Figure 6B). The known gut hormone genes and members of the chromogranin family
5 were mostly common to mouse and human, but motilin and ghrelin were found in human but not mouse
6 jejunum, which is interesting, as the transcriptomic data identified *GHRL* to be preferentially expressed in
7 mouse over human jejunum. A few peptides were derived from proteins not previously known to have
8 signalling roles, but further studies will be needed to determine whether these exhibit bioactivity or simply
9 reflect peptides released from EECs during tissue degradation. To search for novel candidate peptide
10 hormones, we also examined the transcriptomic data for unannotated transcripts that had a base mean value
11 >100 and were >16 fold more highly expressed in EECs than control cells. This analysis identified MIR7-
12 3HG(21), C1orf127 and C6orf141, but we were unable to detect corresponding peptides in the LC-MS data.

13 We then performed a longitudinal LC-MS analysis of known bioactive peptides along the length of the mouse
14 and human GI tract. Sequential samples were taken at 5cm intervals from the stomach to the rectum in mice
15 (n=4 for each location), and human biopsies were analysed from the stomach (n=5), duodenum (n=9),
16 jejunum (n=2), ileum (n=4), proximal colon (n=3), sigmoid colon (n=5) and rectum (n=3). Most EEC peptides
17 were identifiable in their known bioactive forms but fragments of pro-CCK mostly lacked the C-terminally
18 active 8 amino acids, perhaps reflecting an extraction artefact as CCK is known to be highly labile under
19 different conditions and our method was not optimized for very small peptides such as CCK-8. We therefore
20 used a robustly present peptide from the prohormone, CCK₂₁₋₄₄, as a surrogate of CCK production. Peptides
21 were depicted in separate heatmaps for mouse and human (Figure 7), divided into their origin from EEC
22 prohormones, granins and non-EECs (likely reflecting enteric neural peptides). The observed longitudinal
23 profiles broadly mirror historical immuno-staining patterns(22) but additionally provide details of the exact
24 peptide sequences and their post-translational modifications that cannot be deduced from antibody staining.
25 The human and mouse profiles were similar for most EEC peptides, with the notable exceptions that CCK,
26 NTS and SCT extended into the colon only in mice, and that Gastrin and Ghrelin could be detected in human
27 but not mouse proximal small intestine. Of the non-EEC peptides, NMU and galanin were found along the full

- 1 length of mouse small and large intestine but in human NMU was largely restricted to the small intestine and
- 2 galanin to the distal small intestine and colon/rectum.
- 3

1 **Discussion**

2 By RNA sequencing of fixed, FACS-purified cells from the human gut, we report here the transcriptome of
3 human L-cells and the wider EEC population, and a between-species comparison showing a strong similarity
4 with matching cells from the mouse. LC-MS based peptidomic analysis revealed longitudinal gradients of a
5 range of EEC peptides, including their sequences and post-translational modifications, along the length of
6 the human and mouse GI tract. The amalgamated transcriptomic and peptidomic picture provides a unique
7 insight into the sensory apparatus and hormonal secretory profile of human EECs, and the extent to which
8 the mouse is a valid model for studying the human enteroendocrine system.

9 The study is limited by many of the same challenges that have consistently hindered RNA and peptidomic
10 analyses. We relied on good quality internal epitopes to purify human EECs for transcriptomics, but this
11 required cell fixation and permeabilisation prior to cell separation by flow cytometry. The available RNA
12 extraction and sequencing kits were only just technically capable of dealing with PFA-modified RNA and
13 although the resulting high throughput sequencing consequently had a lower than desired mapping rate the
14 read depth was sufficient for detection of low abundance mRNAs such as those encoding receptors and ion
15 channels. The transcriptomes of human L-cells and GCG- EECs mapped robustly onto the homologous RNAseq
16 data from freshly purified matching murine cell populations. Genes without a 1:1 homology between species
17 or that were not protein encoding were excluded from the comparative analysis, but can be examined in the
18 raw deposited datasets. An alternative transcriptomic approach we could have used that would also not
19 depend on the availability of genetically labelled models is single cell RNA sequencing of large numbers of
20 dissociated cells, as recently applied to murine intestinal organoids(10). Although EECs accounted for only a
21 small percentage of the total cell count, they were readily identifiable from their transcriptomic
22 signatures(10). This approach could be extended to human organoids or fresh human tissue, but the potential
23 influence of prolonged organoid culture conditions on the cell transcriptome still needs to be established and
24 the outputs of most single cell RNA sequencing methods are currently limited by a read-depth that precludes
25 reliable quantification of low-abundance transcripts.

1 Although we separately RNA sequenced jejunal samples from 11 humans with varying BMI, we were unable
2 to detect any effect of BMI on the EEC transcriptome. Tissue samples, were, however, collected and sorted
3 in both France and the UK where pre-operative dietary intakes were likely quite different. As the high BMI
4 samples largely originated from the French site and we accounted for the donor location when constructing
5 the models of enriched transcripts, the study was insufficiently powered to detect small differences
6 attributable to obesity. We cannot therefore exclude the possibilities that altered EEC function might arise
7 as a result of obesity or variation in dietary intake, or contribute to the causation of obesity in a subset of
8 people.

9 GPCRs that are strongly and specifically expressed in EECs underlie some of the key sensory roles of these
10 cell types and deserve consideration as candidate drug targets for enhancing gut hormone secretion as a
11 therapy for type 2 diabetes and obesity. The GPCR repertoire of human EECs largely mirrored their well-
12 studied murine counterparts(9, 23-25), including expression of receptors for amino acids (*CASR*, *GPR142*),
13 triglyceride digestion products (*FFAR1*, *FFAR4*, *GPR119*) and bile acids (*GPBAR1*), as well as for hormones
14 such as somatostatin (*SSTR1*, *SSTR2*, *SSTR5*), GIP (*GIPR*) and arginine vasopressin (*AVPR1B*). Interestingly,
15 *GPR142* was highly expressed in human EECs, supporting current studies looking to exploit its ability to
16 stimulate both insulin and incretin hormone secretion(24).

17 RNAseq analysis revealed expression in human and mouse EECs of genes encoding known gut peptides, as
18 well as several putative hormones previously described in the gut or elsewhere. Of the hormone-encoding
19 genes with a 1:1 homology between species, there was a good correlation between the expression profiles
20 of mouse and human. Motilin was expressed in human but not murine EECs, consistent with previous
21 literature(15). Our optimised peptide extraction protocol combined with nano-LC-MS analysis enabled
22 identification of the exact peptide sequences biosynthesised in human and mouse intestinal mucosa,
23 including post-translational modifications, for peptides from ~8-10 to a maximum length of 65 amino acids.
24 From the proglucagon gene, for example, we detected multiple processed and pre-processed products,
25 including GRPP, oxyntomodulin, GLP-1₇₋₃₆ amide, GLP-1₇₋₃₇, GLP-1₁₋₃₇, IP₁₃₁₋₁₄₂, IP-GLP2 and GLP-2. Intact
26 (pancreatic-type) glucagon was detected in samples from the mouse stomach, but was undetectable in the

1 remainder of the intestine and colon from both species, conflicting with recent suggestions that the small
2 intestine can secrete intact glucagon(26), but consistent with our own recent findings that post-prandial
3 glucagon concentrations were not altered following gastric bypass surgery in lean subjects despite dramatic
4 increases in GLP-1(27). Whether the gut might adapt to produce glucagon in pathological conditions such as
5 obesity or type 2 diabetes cannot be concluded from our data. In addition to the set of well-recognised gut
6 hormones, LC-MS also identified some additional peptides encoded by EEC-enriched genes, including
7 peptides derived from PCSK1N, chromogranins and secretogranins. Whether any of these have specific
8 physiological roles or are simply inactive by-products of enzymatic processing of the contents of secretory
9 vesicles, requires further evaluation.

10 Mapping of gut hormone production along the GI tract length has previously been performed using immuno-
11 staining methods and extraction/immuno-assays for specific peptides(22, 28). Immuno-staining relies on the
12 detection of all antigenic sequences binding to a primary antibody and is a reliable method for localising cells
13 producing any given prohormone but cannot generally distinguish whether the prohormone was processed
14 or unprocessed, or post-translationally modified. Nevertheless, immunostaining methods have resulted in
15 longitudinal maps of EEC sub-types producing different prohormones(22). Our LC-MS method provides a
16 surprisingly robust mirror of these immuno-staining maps, whilst assigning an exact peptide sequence to
17 each identified peptide, clearly distinguishing e.g. oxyntomodulin from glucagon, and PYY₁₋₃₆ from PYY₃₋₃₆,
18 despite these peptide pairs differing only by extensions of a few amino acids at the C and N-terminus,
19 respectively. The method also benefits from being unbiased, providing MS-based identification of all peaks
20 triggered from the liquid chromatography separation, rather than analysing only a sub-group of specified
21 peptide sequences. Interestingly we identified acylated as well as non-acylated ghrelin from the human
22 jejunum despite our previous finding that plasma acylated ghrelin levels were undetectable in humans after
23 total gastrectomy(27). We were surprised to find high levels of PYY₃₋₃₆ as well as PYY₁₋₃₆ in tissue
24 homogenates, suggesting that dipeptidyl-peptidases (DPP) are active within L-cells themselves, whereas GLP-
25 1(7-36amide) was much more abundant than GLP-1(9-36amide), indicating they had not been subject to DPP
26 cleavage. Why GLP-1 but not PYY seems protected from DPP-mediated processing in L-cells, despite both
27 peptides being located in the same vesicular pool (manuscript under preparation), remains unclear.

1

2 In summary, we have developed techniques for RNA sequencing of purified human EECs, and for peptidomic
3 analysis of mucosal homogenates using LC-MS. Comparison of the human and mouse EEC transcriptomes
4 revealed strong global similarities between the two species. Variation at the level of a few key individual
5 genes could, however, have profound implications for the use of mouse as a model species in drug
6 development programmes targeting specific receptors. Our mouse/human comparative datasets provide
7 tools for confirming the validity of using mouse as a model for investigating given signalling pathways and
8 receptors or peptide processing in humans, and the human EEC GPCR-ome can be used independently as a
9 potential source of drug targets in the human enteroendocrine system. Longitudinal peptide mapping of the
10 GI tract by LC-MS is a key step towards understanding the metabolic benefits of gastric bypass surgery, as
11 the surgical procedures result in a shift in the location of nutrient digestion and absorption to more distal
12 regions of the small intestine, with consequent stimulation of the more distal EEC population and release of
13 their distinct profiles of peptide hormones. Devising strategies to mimic gastric bypass surgery using
14 injectable peptide mimetics or through stimulation of the endogenous EEC population is a key but highly
15 topical challenge for industry and academia alike.

16

1 **Methods**

2 *Ethics*

3 This study was conducted in accordance with the principles of the Declaration of Helsinki and good clinical
4 practice. Human ethical approvals were given by Cambridge Central and South Research Ethics Committees
5 (ref: 09/H0308/24, 16/EE/0338, 15/EE/0152) and the Inserm ethics committee and Agence de la
6 biomédecine (ref: PFS16-004). Animal work was regulated under the Animals (Scientific Procedures) Act 1986
7 Amendment Regulations 2012 and conducted following ethical review by the University of Cambridge Animal
8 Welfare and Ethical Review Body.

9 ***Human tissue transcriptome***

10 *Sample collection*

11 Samples of human jejunum discarded during surgery were collected during total gastrectomy for treatment
12 or prophylaxis of gastric cancer, or Roux-en-Y gastric bypass for obesity. All were from the point of entero-
13 enterostomy 50cm distal to the ligament of Treitz. Two matched samples of jejunum and terminal ileum were
14 collected during organ procurement from transplant donors. Data were collected on age, gender and BMI,
15 and participants stratified as lean vs obese (BMI>30kg/m²). Tissue samples from different regions of the
16 human GI tract for LC-MS were obtained from Addenbrooke's Human Research Tissue Bank and the
17 Cambridge Biorepository for Translational Medicine.

18 Samples were immediately placed in cold Leibovitz's L-15 media (Thermo Scientific, Waltham, MA, USA) and
19 processed to the point of fixation or homogenised and stored at -70°C within 6 hours.

20 *Tissue preparation for FACS*

21 FACS and RNA extraction from fixed human cells followed a modified version of the MARIS protocol(29).
22 Intestine was rinsed in cold phosphate buffered saline (PBS) and the muscular coat removed. Diced mucosa
23 was digested twice in 0.1% w/v collagenase XI (Sigma-Aldrich, MO, USA) in Hanks' Buffered Saline solution
24 (HBSS) #9394 (Sigma-Aldrich) for 30 minutes each, shaking vigorously every 10 minutes. Supernatants were
25 triturated, passed through a 50µm filter and centrifuged at 300g. Pellets were resuspended in PBS and fixed

1 in 4% w/v paraformaldehyde (PFA) at 4°C for 20 minutes. PFA-fixed cells were washed twice in nuclease free
2 1% w/v bovine serum albumin (BSA) in PBS, and if a FACS facility was not immediately available, were
3 suspended in 1% w/v BSA and 4% v/v RNAsin plus RNase inhibitor (Promega, WI, USA) in PBS at 4°C overnight.
4 Cells were permeabilised with either a single 30 minute incubation with 0.1% v/v Triton x100 (Sigma-Aldrich)
5 in 1% w/v BSA in PBS prior to antibody staining, or by the addition of 0.1% w/v Saponin (Sigma-Aldrich) to
6 solutions in all steps from this point until after the first wash post-secondary antibody staining, with identical
7 results.

8 Primary antibody staining was for one hour in 4% v/v RNAsin, 1% w/v BSA, 1% v/v goat anti-GLP-1 (Santa
9 Cruz, Dallas, TX, USA; sc7782), 2% v/v rabbit anti-CHGA (Abcam, Cambridge, UK; Ab15160), 0.25% v/v rabbit
10 anti-SCG2 (Abcam, Ab12241) in PBS at 4°C. Cells were then washed twice in 1% w/v BSA, 1% v/v RNAsin, and
11 secondary antibody staining was for 30 minutes in 4% v/v RNAsin, 1% w/v BSA, 0.2% v/v donkey anti-goat
12 Alexa 555, 0.2% v/v donkey anti-rabbit Alex 647 in PBS at 4°C. Cells were washed twice then suspended in
13 4% v/v RNAsin, 1% w/v BSA in PBS on ice for FACS.

14 *FACS*

15 Cell populations were sorted on a BD FACS ARIA III in the Cambridge NIHR BRC cell phenotyping hub or at
16 Institut Cochin, Paris. Single cells positive for Alexa 647 but not Alexa 555 (i.e. CHGA/SCG2 +ve / GLP-1 -ve)
17 were classified as GCG- enteroendocrine cells. Single cells positive for both Alexa 647 and Alexa 555 were
18 classified as GCG+ enteroendocrine cells. At least 5000 cells were collected for each positive population.
19 Twenty thousand double negative cells were collected as the negative (i.e. non-enteroendocrine) cell
20 population. Cells were sorted into 2% v/v RNAsin in PBS at 4°C.

21 *RNA extraction*

22 RNA was extracted using the Ambion Recoverall Total nucleic acid isolation kit for FFPE (Ambion, CA, USA)
23 with modifications to the protocol as below. The FACS sorted cell suspension was centrifuged at 3000g for 5
24 minutes at 4°C and the pellet resuspended in 200µl digestion buffer with 4µl protease and incubated at 50°C
25 for 3 hours. The solution was then stored at -70°C for at least 12 hours prior to further extraction. After

1 thawing, RNA was extracted using the manufacturer's protocol with the exception of performing 2x 60µl
2 elutions from the filter column in the final step.

3 The RNA solution was concentrated using a RNEasy Minelute cleanup kit (Qiagen, Hilden, Germany). RNA
4 aliquots were diluted to 200µl with nuclease free water. The standard manufacturer's protocol was followed
5 with the exception that 700µl, not 500µl, of 100% ethanol was added to the solution in step two, to generate
6 optimum binding conditions for the PFA fragmented RNA. RNA concentration and quality was analysed using
7 an Agilent 2100 Bioanalyser (Agilent, CA, USA).

8 *Sequencing*

9 cDNA libraries were created using the Clontech SMARTer Stranded Total RNA-Seq Kit – Pico Input Mammalian
10 v1 (Takara Bio, USA). RNA input quantity was 5ng and the non-fragmentation protocol was used. The
11 standard manufacturer's protocol was followed with the exception that 175µl of AMPure beads were used
12 for the final bead purification to ensure recovery of the small fragments of RNA arising from PFA fixation.
13 Sixteen PCR cycles were used for amplification.

14 50 base single-end sequencing was performed using an Illumina HiSEQ 4000 at the CRUK Cambridge Institute
15 Genomics Core.

16 ***Mouse transcriptome***

17 *Sample collection and preparation for FACS*

18 Female NeuroD1-Cre/EYFP (mixed background: 3-10 generations back-crossed with C57BL6) and GLU-Venus
19 mice (C57BL6)(7, 8)aged 8-10 weeks were killed by cervical dislocation (n=3 each). Diced mucosa from the
20 proximal 10cm of small intestine was digested twice in 0.1% w/v collagenase XI in HBSS at 37°C for 30 minutes
21 each. Cells were pelleted by centrifugation at 100g for 1 minute, triturated and passed through a 50µm filter.
22 Cells were stained with DAPI (1µg/ml) for 5 minutes at room temperature, washed twice and sorted in HBSS
23 on a FACSJazz sorter at the Cambridge NIHR BRC cell phenotyping hub.

24 *FACS and RNA extraction*

1 All positive cells, and 20,000 negative cells were collected separately into aliquots of 500µl of buffer RLT+
2 (Qiagen), with 143mM β-mercaptoethanol. RNA was extracted using a RNeasy Micro plus kit (Qiagen) and
3 quantified using an Agilent 2100 Bioanalyser

4 *Sequencing*

5 2ng of each RNA was used for cDNA amplification by SPIA amplification using the Ovation RNAseq system V2
6 kit (Nugen, CA, USA). 1µg of cDNA was then fragmented to ~200bp by sonication (Diagenode, Liege, Belgium)
7 and adaptors for the indexing were added using the Ovation Rapid DR multiplex 1-96 kit. Samples were
8 pooled and concentrated together with a MinElute column (Qiagen) to reach a concentration of 10nM. Single
9 end 50 base sequencing was performed at the CRUK Cambridge Institute Genomics Core with an Illumina
10 HiSeq4000.

11 ***RNAseq pipeline***

12 Quality control and trimming of adaptors was performed using FastQC(30). Sequenced transcripts were
13 mapped to the human (GRCh37) and mouse (GRCm38) genomes using TopHat 2.1.0 and raw counts
14 generated using Cufflinks 2.2.1(31-33). Differential gene expression analysis was performed in RStudio using
15 DESEQ2(11). Gene annotation was pulled from the Ensembl dataset held in BioMart(31). Receptor and ion
16 channel lists were generated from the IUPHAR “targets and families” list(34). Graphical output used ggplot2
17 and pheatmap in RStudio(35).

18 ***Comparative transcriptomics***

19 Mouse and human data sets were compared using only the 15,507 genes present in both datasets, not
20 annotated as ribosomal, mitochondrial or small-nuclear and described with one-to-one homology according
21 to the Ensembl mouse-human homology dataset(31). Normalised CPM (counts per million) were generated
22 from the respective DESEQ2 models and compared for the human GCG+ population versus the murine GLU-
23 Venus population and the human GCG- population versus the murine NeuroD1 population. Linear models
24 were generated of the log₁₀ CPM of the human vs murine datasets by a total least squares strategy, 99%
25 confidence intervals calculated and the outliers hand searched for relevant genes.

1 Examining just the human samples, DESEQ2 models were generated for the following sets of jejunum
2 samples: GCG+ versus negative; GCG- versus negative; GCG+ versus GCG-; GCG+ lean versus obese; GCG-
3 lean versus obese. Participant paired DESEQ2 analyses were also performed comparing GCG+ and GCG-
4 populations from the jejunum and ileum of the two transplant donor participants, for whom there were
5 matched jejunum and ileum samples. An adjusted P value of $\leq .1$ was defined as the cut off for statistical
6 significance when examining for differential gene expression between populations.

7 ***Tissue homogenates***

8 Small pieces (~2x2mm, 15-45 mg) of mucosa from different regions of the human GI tract or from every 5cm
9 of the GI tract from C57BL6 mice were homogenised in 250 μ l of 6M guanidine hydrochloride (Sigma-Aldrich)
10 with Lyzing MatrixD (MPbio), in a FastPrep-24 for 4x40s at 6ms⁻¹. Samples were stored at -70C before further
11 processing. Proteins were precipitated by adding 80% acetonitrile in water then centrifuged at 12000g at 4°C
12 for 5 minutes. The aqueous phase containing the peptides was collected, dried on a centrifugal vacuum
13 concentrator and stored at -70°C before analysis

14 ***Mass spectrometry***

15 Samples were extracted using a Waters HLB μ Elution solid-phase extraction (SPE) plate (Waters, MA, USA)
16 after being resuspended in 500 μ L 0.1% v/v formic acid in water as described previously and analysed after
17 reduction/alkylation(36). Human homogenates and mouse homogenates for interspecies comparison were
18 analysed using nano-flow based separation and electrospray approaches on a Thermo Fisher Ultimate 3000
19 nano LC system coupled to a Q Exactive Plus Orbitrap mass spectrometer (ThermoScientific). High-flow
20 separation for the longitudinal mouse analysis was as previously described(36). Downstream analysis was
21 performed using Peaks 8.0 software (Waterloo, ON, Canada) against the human and the mouse Swissprot
22 databases (downloaded 26th October 2017)(37), with a fixed cysteine carbamidomethylation and variable
23 methionine oxidation, N-terminal acetylation and pyro-glutamate and C-terminal amidation modifications.
24 Manual searches were performed for other modifications. Peptides of interest were quantified by measuring
25 peak areas for selected m/z ranges and retention times corresponding to the peptide sequences and
26 normalised by tissue weight.

1 *Contributions*

2 GR, PR, PL, RS, RH, FR and FG designed the study. Tissue handling and FACS were performed by GR, PR, PL
3 and LG. cDNA library production and bioinformatics analysis were performed by GR, PR, PL, MM and BL with
4 support from GY and DC. Mass spectrometry was by PL, RK, EM, SG and GR. NeuroD1 mice were generated
5 by AL and HJL. The manuscript was written by GR, FR and FG. All authors have had editorial oversight of the
6 final manuscript version.

7

8

9

1 **Acknowledgements**

2 We would like to thank the surgical team in the Cambridge Oesophago-gastric centre, Sébastien Czernichow,
3 Cyrille Ramond, The Cambridge Biorepository for Translational Medicine and the NIHR Cambridge Human
4 Research Tissue Bank for assistance with collecting donor tissue. FACS was supported by the Institut Cochin
5 in Paris and the Cambridge NIHR BRC Cell Phenotyping Hub. Metabolic Research Laboratories support was
6 provided by the following core facilities: Disease Model Core, Genomics and Transcriptomics Core, Histology
7 Core, Imaging Core and NIHR BRC Core Biochemical Assay Laboratory (supported by the MRC
8 [MRC_MC_UU_12012/5] and Wellcome Trust [100574/Z/12/Z]). RNA-sequencing was undertaken at the
9 CRUK Cambridge Institute Genomics Core.

10 GR received an Addenbrooke's Charitable Trust / Evelyn Trust Cambridge Clinical Research Fellowship [16-
11 69], an EFSD project grant and a Royal College of Surgeons Research Fellowship. SG is supported by a grant
12 from the BBSRC. Research in the laboratory of Fiona Gribble and Frank Reimann is supported by the MRC
13 [MRC_MC_UU_12012/3], Wellcome Trust [106262/Z/14/Z, 106263/Z/14/Z] and research grants from
14 Medimmune. The MS instrument was funded by the MRC "Enhancing UK clinical research" grant
15 (MR/M009041/1). PR was supported by a postdoctoral grant from Agence Nationale de la Recherche
16 (Laboratoire d'Excellence Revive, Investissement d'Avenir; ANR-10-LABX-73). The RS lab is supported by ANR-
17 10-LABX-73 and the Bettencourt-Schueller foundation. This work was partially supported by the following
18 NIH grants to AL: R01 DK100223-02, R01 DK110614-01.

19

1 References

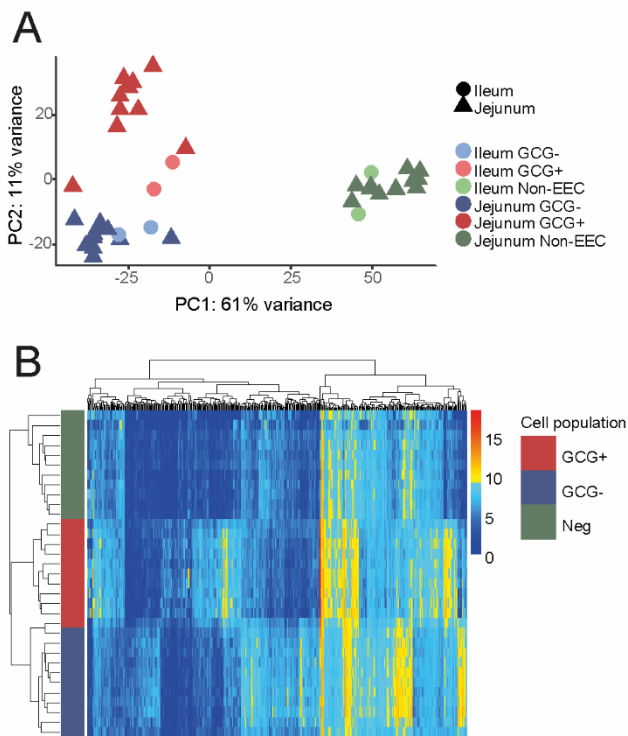
- 2 1. Vizcaino JA, Csordas A, del-Toro N, Dianas JA, Griss J, Lavidas I, Mayer G, Perez-Riverol Y, Reisinger
3 F, Ternent T, et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.*
4 2016;44(D1):D447-56.
- 5 2. Holst JJ, and Madsbad S. Mechanisms of surgical control of type 2 diabetes: GLP-1 is key factor.
6 *Surgery for obesity and related diseases : official journal of the American Society for Bariatric*
7 *Surgery.* 2016;12(6):1236-42.
- 8 3. Gribble FM, and Reimann F. Enteroendocrine Cells: Chemosensors in the Intestinal Epithelium.
9 *Annu Rev Physiol.* 2016;78(277-99).
- 10 4. Glass LL, Calero-Nieto FJ, Jawaid W, Larraufie P, Kay RG, Gottgens B, Reimann F, and Gribble FM.
11 Single-cell RNA-sequencing reveals a distinct population of proglucagon-expressing cells specific to
12 the mouse upper small intestine. *Mol Metab.* 2017;6(10):1296-303.
- 13 5. Svendsen B, Pais R, Engelstoft MS, Milev NB, Richards P, Christiansen CB, Egerod KL, Jensen SM,
14 Habib AM, Gribble FM, et al. GLP1- and GIP-producing cells rarely overlap and differ by bombesin
15 receptor-2 expression and responsiveness. *J Endocrinol.* 2016;228(1):39-48.
- 16 6. Adriaenssens A, Lam BY, Billing L, Skeffington K, Sewing S, Reimann F, and Gribble F. A
17 Transcriptome-Led Exploration of Molecular Mechanisms Regulating Somatostatin-Producing D-
18 Cells in the Gastric Epithelium. *Endocrinology.* 2015;156(11):3924-36.
- 19 7. Li HJ, Kapoor A, Giel-Moloney M, Rindi G, and Leiter AB. Notch signaling differentially regulates the
20 cell fate of early endocrine precursor cells and their maturing descendants in the mouse pancreas
21 and intestine. *Developmental biology.* 2012;371(2):156-69.
- 22 8. Reimann F, Habib AM, Tolhurst G, Parker HE, Rogers GJ, and Gribble FM. Glucose sensing in L cells:
23 a primary cell study. *Cell metabolism.* 2008;8(6):532-9.
- 24 9. Reimann F, Tolhurst G, and Gribble FM. G-protein-coupled receptors in intestinal chemosensation.
25 *Cell metabolism.* 2012;15(4):421-31.
- 26 10. Grun D, Lyubimova A, Kester L, Wiebrands K, Basak O, Sasaki N, Clevers H, and van Oudenaarden A.
27 Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature.*
28 2015;525(7568):251-5.
- 29 11. Love MI, Huber W, and Anders S. Moderated estimation of fold change and dispersion for RNA-seq
30 data with DESeq2. *Genome biology.* 2014;15(12):550.
- 31 12. Habib AM, Richards P, Cairns LS, Rogers GJ, Bannon CA, Parker HE, Morley TC, Yeo GS, Reimann F,
32 and Gribble FM. Overlap of endocrine hormone expression in the mouse intestine revealed by
33 transcriptional profiling and flow cytometry. *Endocrinology.* 2012;153(7):3054-65.
- 34 13. Egerod KL, Engelstoft MS, Grunddal KV, Nohr MK, Secher A, Sakata I, Pedersen J, Windelov JA,
35 Fuchtbauer EM, Olsen J, et al. A major lineage of enteroendocrine cells coexpress CCK, secretin,
36 GIP, GLP-1, PYY, and neurotensin but not somatostatin. *Endocrinology.* 2012;153(12):5782-95.
- 37 14. Grunddal KV, Ratner CF, Svendsen B, Sommer F, Engelstoft MS, Madsen AN, Pedersen J, Nohr MK,
38 Egerod KL, Nawrocki AR, et al. Neurotensin Is Coexpressed, Coreleased, and Acts Together With
39 GLP-1 and PYY in Enteroendocrine Control of Metabolism. *Endocrinology.* 2016;157(1):176-94.
- 40 15. Sanger GJ, Wang Y, Hobson A, and Broad J. Motilin: towards a new understanding of the
41 gastrointestinal neuropharmacology and therapeutic use of motilin receptor agonists. *British*
42 *journal of pharmacology.* 2013;170(7):1323-32.
- 43 16. Bellono NW, Bayrer JR, Leitch DB, Castro J, Zhang C, O'Donnell TA, Brierley SM, Ingraham HA, and
44 Julius D. Enterochromaffin Cells Are Gut Chemosensors that Couple to Sensory Neural Pathways.
45 *Cell.* 2017;170(1):185-98.e16.
- 46 17. Rogers GJ, Tolhurst G, Ramzan A, Habib AM, Parker HE, Gribble FM, and Reimann F. Electrical
47 activity-triggered glucagon-like peptide-1 secretion from primary murine L-cells. *The Journal of*
48 *physiology.* 2011;589(Pt 5):1081-93.
- 49 18. Emery EC, Diakogiannaki E, Gentry C, Psichas A, Habib AM, Bevan S, Fischer MJ, Reimann F, and
50 Gribble FM. Stimulation of GLP-1 secretion downstream of the ligand-gated ion channel TRPA1.
51 *Diabetes.* 2015;64(4):1202-10.

- 1 19. Nozawa K, Kawabata-Shoda E, Doihara H, Kojima R, Okada H, Mochizuki S, Sano Y, Inamura K,
2 Matsushime H, Koizumi T, et al. TRPA1 regulates gastrointestinal motility through serotonin release
3 from enterochromaffin cells. *Proc Natl Acad Sci U S A*. 2009;106(9):3408-13.
- 4 20. Parker HE, Habib AM, Rogers GJ, Gribble FM, and Reimann F. Nutrient-dependent secretion of
5 glucose-dependent insulinotropic polypeptide from primary murine K cells. *Diabetologia*.
6 2009;52(2):289-98.
- 7 21. Tanaka S, Tatsumi K, Okubo K, Itoh K, Kawamoto S, Matsubara K, and Amino N. Expression profile
8 of active genes in the human pituitary gland. *Journal of molecular endocrinology*. 2002;28(1):33-44.
- 9 22. Sjolund K, Sanden G, Hakanson R, and Sundler F. Endocrine cells in human intestine: an
10 immunocytochemical study. *Gastroenterology*. 1983;85(5):1120-30.
- 11 23. Diakogiannaki E, Pais R, Tolhurst G, Parker HE, Horscroft J, Rauscher B, Zietek T, Daniel H, Gribble
12 FM, and Reimann F. Oligopeptides stimulate glucagon-like peptide-1 secretion in mice through
13 proton-coupled uptake and the calcium-sensing receptor. *Diabetologia*. 2013;56(12):2688-96.
- 14 24. Lin HV, Efanov AM, Fang X, Beavers LS, Wang X, Wang J, Gonzalez Valcarcel IC, and Ma T. GPR142
15 Controls Tryptophan-Induced Insulin and Incretin Hormone Secretion to Improve Glucose
16 Metabolism. *PLoS one*. 2016;11(6):e0157298.
- 17 25. Psichas A, Larraufie PF, Goldspink DA, Gribble FM, and Reimann F. Chylomicrons stimulate incretin
18 secretion in mouse and human cells. *Diabetologia*. 2017;60(12):2475-85.
- 19 26. Knop FK. EJE PRIZE 2018: A gut feeling about glucagon. *Eur J Endocrinol*. 2018;178(6):R267-r80.
- 20 27. Roberts GP, Kay RG, Howard J, Hardwick RH, Reimann F, and Gribble FM. Gastrectomy with Roux-
21 en-Y reconstruction as a lean model of bariatric surgery. *Surgery for obesity and related diseases :
22 official journal of the American Society for Bariatric Surgery*. 2018.
- 23 28. Baldissera FG, Holst JJ, Jensen SL, and Krarup T. Distribution and molecular forms of peptides
24 containing somatostatin immunodeterminants in extracts from the entire gastrointestinal tract of
25 man and pig. *Biochim Biophys Acta*. 1985;838(1):132-43.
- 26 29. Hrvatin S. DF, O'Donnell C.W., Gifford D.K., Melton D.A. MARIS: Method for Analyzing RNA
27 following Intracellular Sorting. *PLoS one*. 2016;9(3):e89459.
- 28 30. Andrews S. FastQC: a quality control tool for high throughput sequence data.
29 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
- 30 31. . BioMart. <https://www.ensembl.org/biomart/martview/c68208e4252974c97a2be0f8947c62b7>.
31 2017.
- 32 32. Trapnell C, Pachter L, and Salzberg SL. TopHat: discovering splice junctions with RNA-Seq.
33 *Bioinformatics (Oxford, England)*. 2009;25(9):1105-11.
- 34 33. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, and
35 Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and
36 isoform switching during cell differentiation. *Nature biotechnology*. 2010;28(5):511-5.
- 37 34. . Targets and families list. <http://www.guidetopharmacology.org/download.jsp>. 2018.
- 38 35. Wickham H. New York: Springer-Verlag; 2009.
- 39 36. Kay RG, Galvin S, Larraufie P, Reimann F, and Gribble FM. LC/MS based detection and semi-
40 quantitative analysis of INSL5 in human and murine tissues. *Rapid Commun Mass Spectrom*. 2017.
- 41 37. . Swissprot / Uniprot database. www.uniprot.org. 2016.

42

43

1 **Figures**

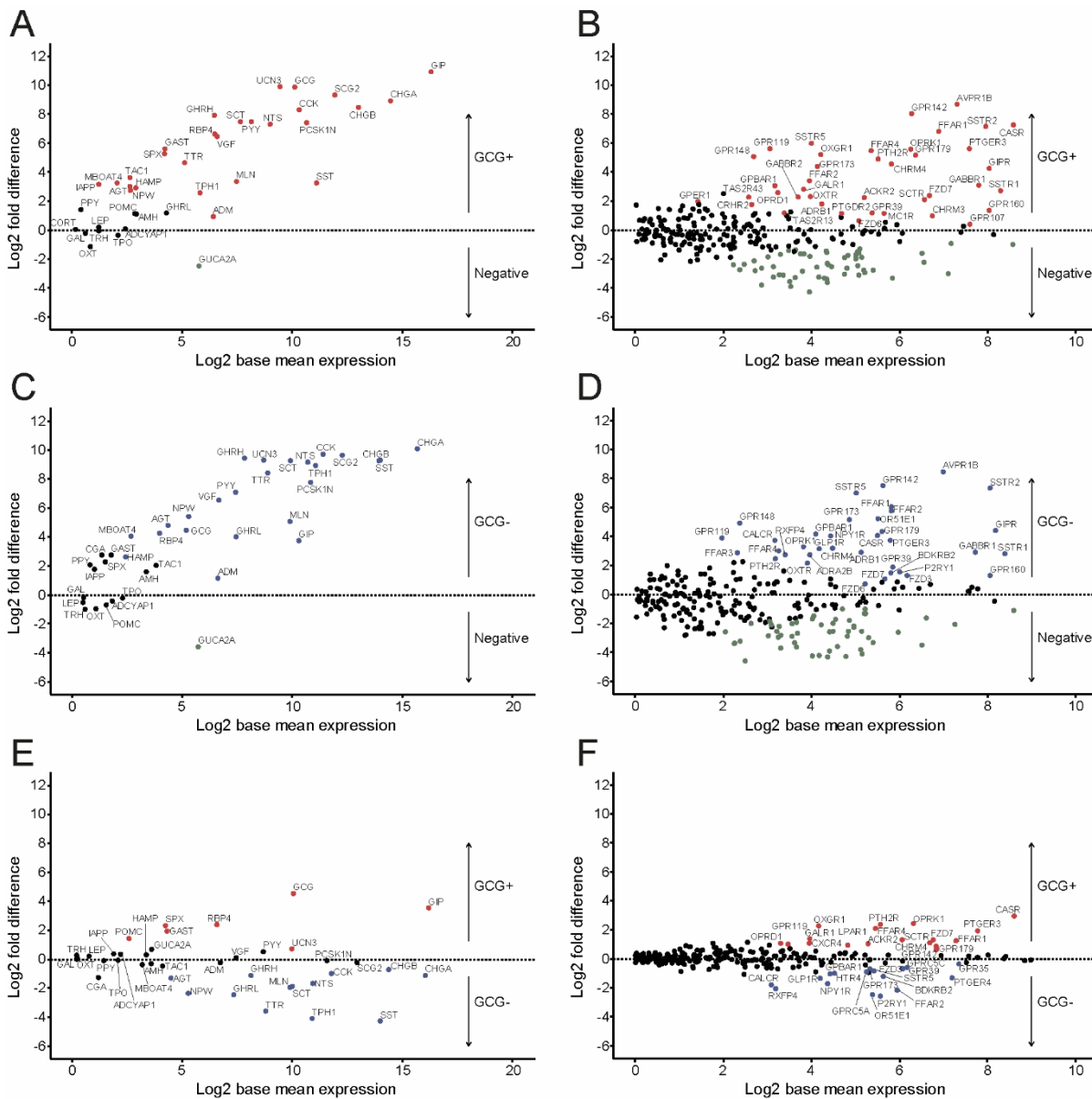


2

3 *Figure 1. Transcriptomic distinction between cell populations from human small intestine*

4 (A) Principal component analysis plot of all human samples (n=3 cell populations from each of 11 jejunal and
5 2 ileal tissue samples), differentiated by cell population and anatomical site. (B) Heatmap of 500 most
6 differentially expressed genes in human jejunum samples (n=3 cell populations per each of 11 participants),
7 y axis – cell population identified by coloured bar, x axis – genes.

8

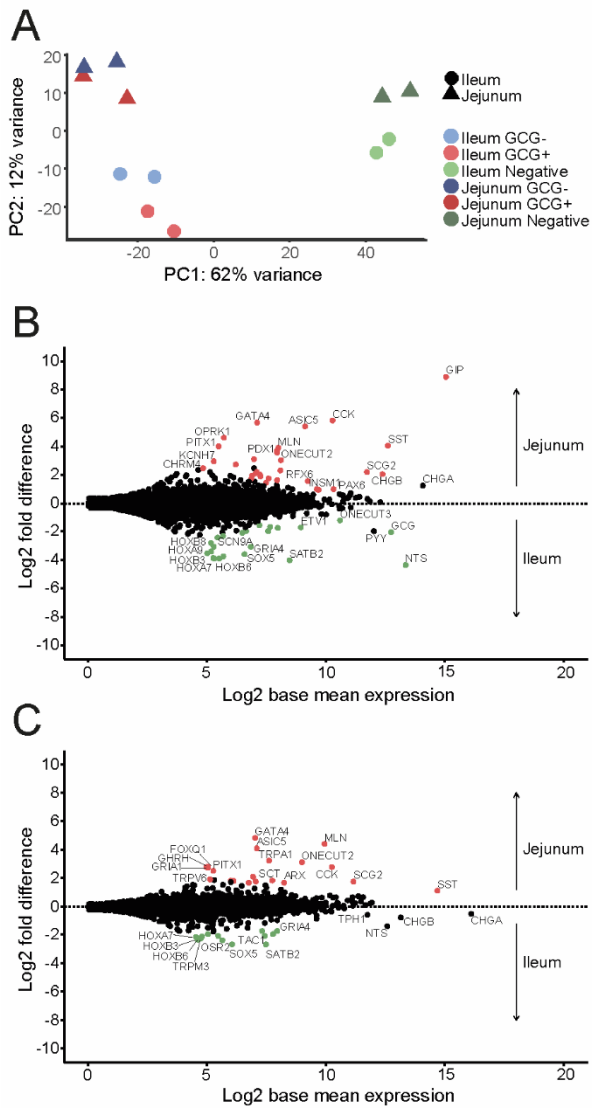


1

2 *Figure 2. Transcripts enriched in human EECs*

3 Enrichment vs expression plots for human jejenum EEC populations. Enrichment is presented as the log2 fold
 4 difference between the cell populations indicated, and expression is presented as the log2 base mean
 5 normalised expression extracted from the DESEQ2 model. (A,B) GCG+ vs negative, (C,D) GCG- vs negative,
 6 (E,F) GCG+ vs GCG-. Hormones and granins are shown in A,C,E; receptors and ion channels are shown in B,D,F.
 7 Red – enriched (adjusted $P \leq .1$ in DESEQ2 model) in GCG+; Blue – enriched in GCG-; Green – enriched in
 8 negative cells.

9



1

2 *Figure 3. Comparison between EECs from human jejunum and ileum*

3 (A) Principal component analysis plot of matched human jejunum vs ileum samples, labelled by cell
 4 population and anatomical location (n=3 cell populations from both anatomical regions from each of 2
 5 participants). (B,C) Enrichment vs expression plots for GCG+ (B), and GCG- (C), cell populations. Enrichment
 6 is presented as the log2 fold difference between the cell populations indicated, and expression is presented
 7 as the log2 base mean normalised expression extracted from the DESEQ2 model. Red – enriched (adjusted
 8 $P < .1$ in DESEQ2 model) in jejunum; Green – enriched in ileum.

9

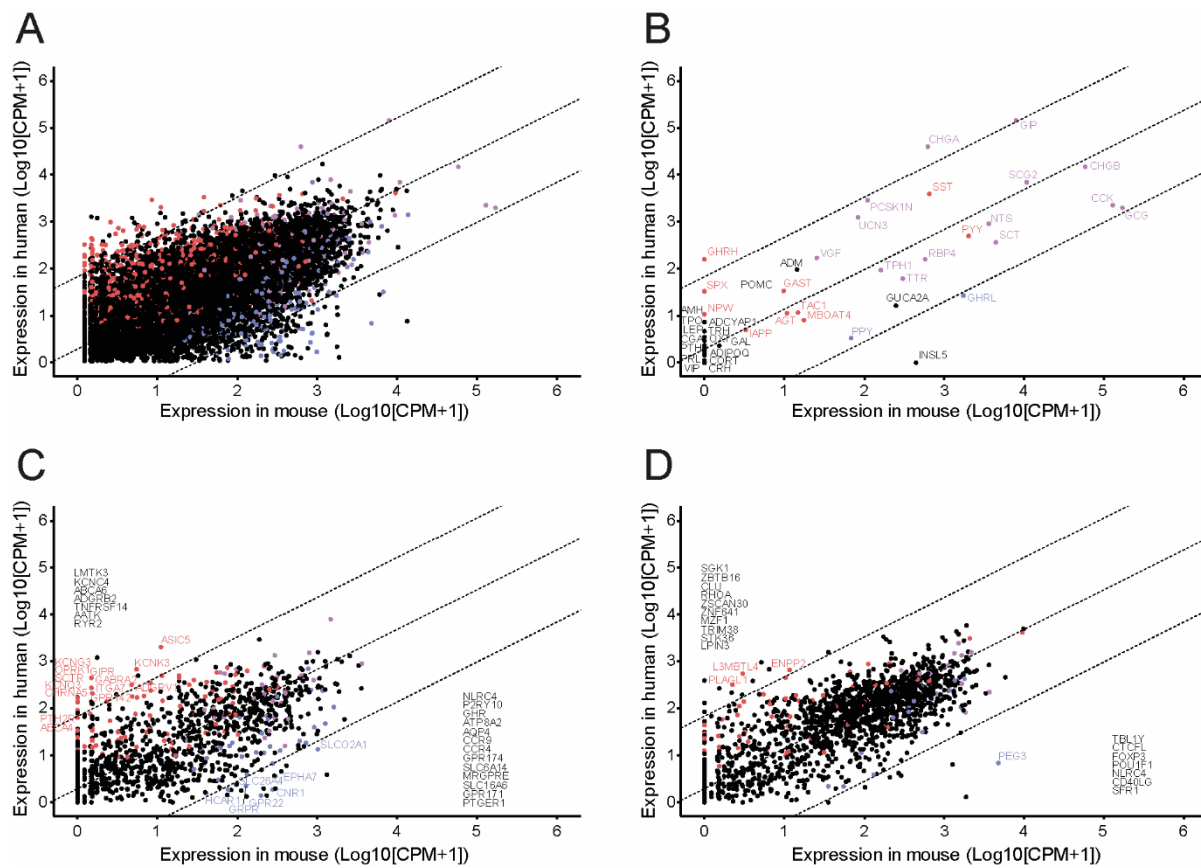
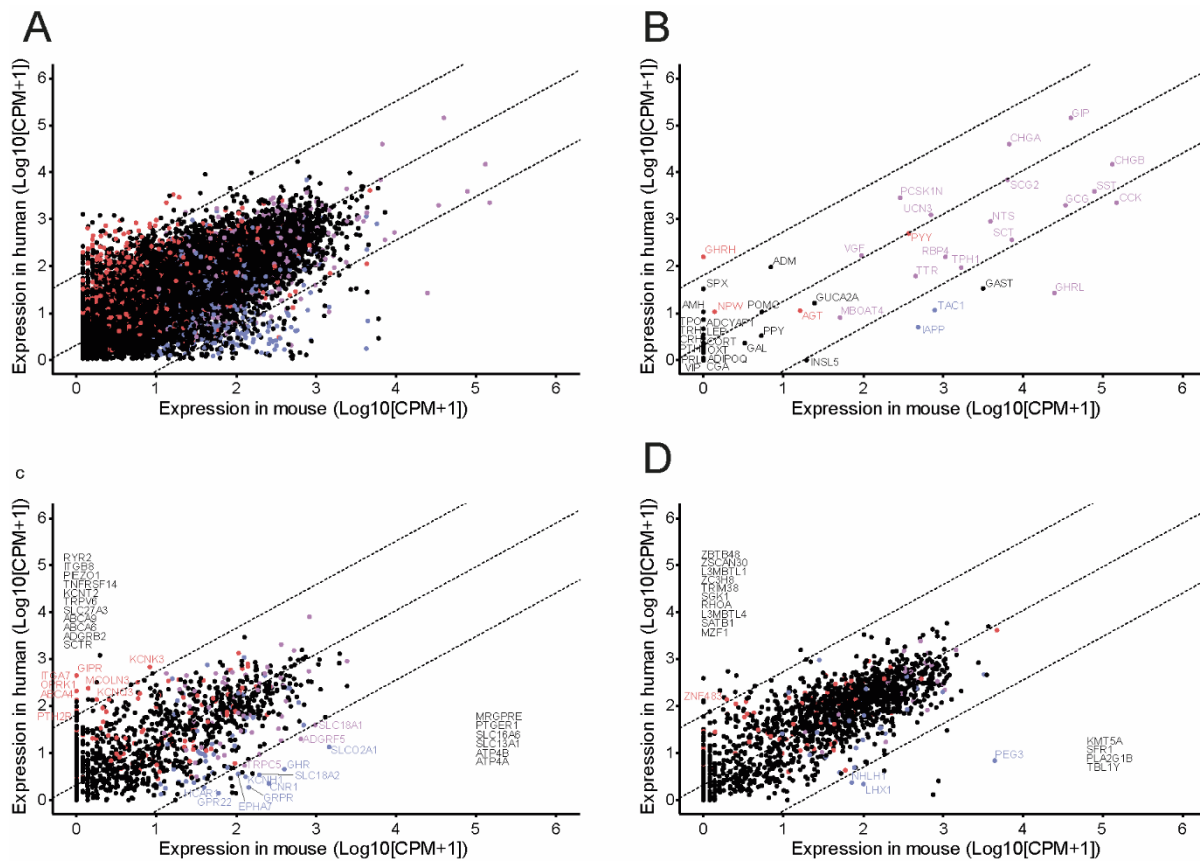


Figure 4. Comparison between human and mouse L-cells

Human versus mouse jejunal L cell gene expression (log₁₀ normalised counts per million +1; n = 11 humans, n = 3 mice). (A) all genes with 1:1 homology between species, excluding mitochondrial, ribosomal and small nuclear transcripts (n=15,507). (B) Hormones. (C) GPCRs and ion channels. (D) Transcription factors. Dashed lines are linear regression and 99% confidence interval. Each dot represents normalised CPM+1 for one gene. Red – enriched (>4x fold change) and differentially expressed (adjusted P<=.1) for human GCG+ vs Negative cell populations, but not murine GLU-Venus vs Negative cell populations in relevant DESEQ2 model. Blue – enriched and differentially expressed for murine GLU-Venus vs Negative cell populations, but not human GCG+ vs Negative cell populations. Purple – enriched and differentially expressed in both murine GLU-Venus and human GCG+ cells versus relevant negative cell populations. Black – not enriched and differentially expressed in either human GCG+ or murine GLU-Venus cells versus relevant negative cell populations. All genes are labelled in B, and genes outside the 99% CI are labelled in C and D, with those not differentially expressed or enriched in either population listed along the axis.



1

2 *Figure 5. Comparison between human GCG- and mouse NeuroD1 cells*

3 Human versus mouse jejunal EEC (GCG-) gene expression (log10 normalised counts per million +1; n = 11

4 humans, n = 3 mice). (A) all genes with 1:1 homology between species, excluding mitochondrial, ribosomal

5 and small nuclear transcripts (n=15,507). (B) Hormones. (C) GPCRs and ion channels. (D) Transcription factors.

6 Dashed lines are linear regression and 99% confidence interval. Each dot represents normalised CPM+1 for

7 one gene. Red – enriched (>4x fold change) and differentially expressed (adjusted P<=.1) for human GCG- vs

8 Negative cell populations, but not murine NeuroD1 vs Negative cell populations in relevant DESEQ2 model.

9 Blue – enriched and differentially expressed for murine NeuroD1 vs Negative cell populations, but not human

10 GCG- vs Negative cell populations. Purple – enriched and differentially expressed in both murine NeuroD1

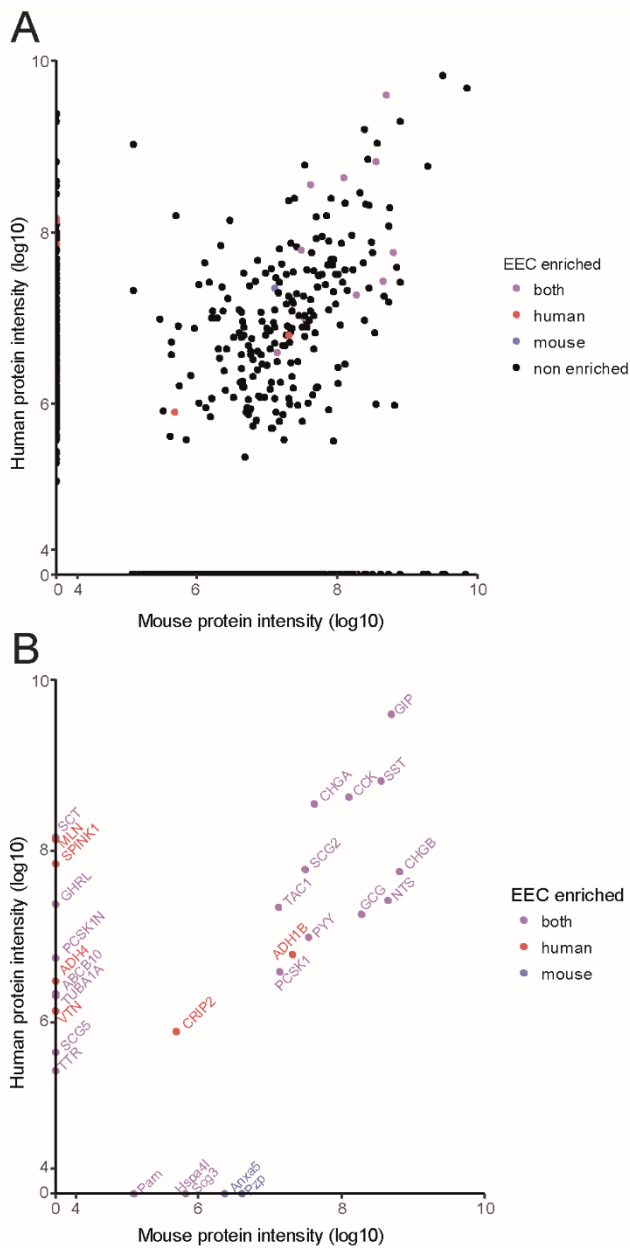
11 and human GCG- cells versus relevant negative cell populations. Black – not enriched and differentially

12 expressed in either human GCG- or murine NeuroD1 cells versus relevant negative cell populations. All genes

13 are labelled in B, and genes outside the 99% CI are labelled in C and D, with those not differentially expressed

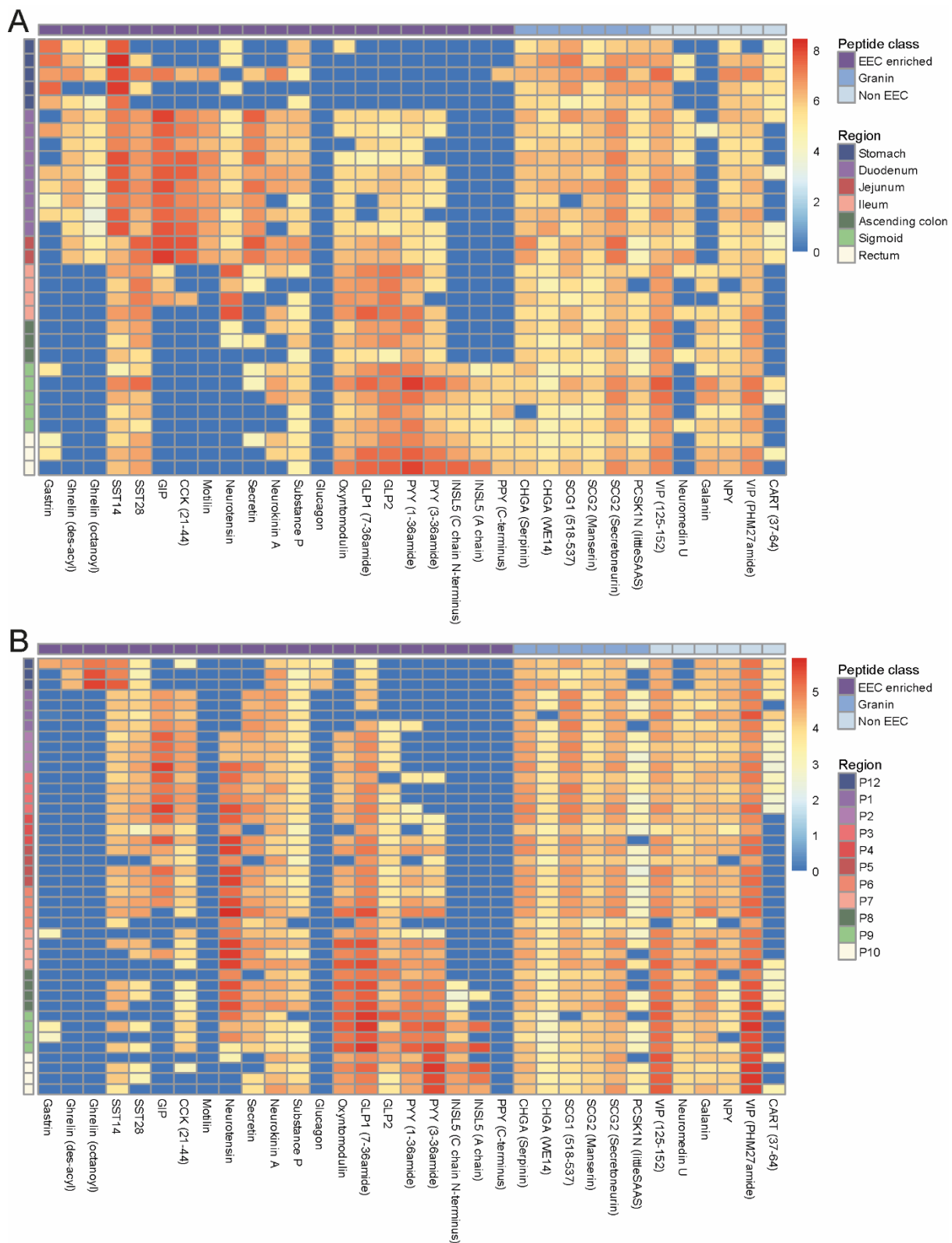
14 or enriched in either population listed along the axis.

15



2 *Figure 6. Comparison between human and mouse jejunum peptidome*

3 Protein intensity as calculated by Peaks v8.0 for all proteins detected corresponding to genes with 1:1
4 homology between human and mouse for jejunum mucosal homogenates. (A) All proteins, (B) proteins from
5 genes enriched in at least one of the species from the human and the mouse transcriptome datasets.
6 Enrichment was defined as adjusted $P \leq .1$, fold difference >4 and base mean expression >50 from DESEQ2
7 model. Colours indicate in which species the mRNA for the genes was enriched.



1
2 *Figure 7. Longitudinal profiles of gut peptides along the human and mouse gut*

3 Heatmap of gut peptide quantification normalised by tissue weight (log10 scale) along human (A) and mouse
4 (B) GI tract. Blue indicates not detected in the sample. Rows: Samples ordered from proximal to distal and
5 colour coded by their region of origin. Mouse regions: P12: stomach lesser curvature, P1 to P7: small intestine

- 1 from proximal to distal, sampling every 5 cm, P8 to P10: large intestine (proximal, mid and distal). Columns:
- 2 peptide (using the human name if different between human and mouse), classified by origin (purple: classical
- 3 EEC hormone peptides; medium blue: peptides from granins; light blue: enteric neuron bioactive peptides).
- 4

1

2

1 **Supplementary material**

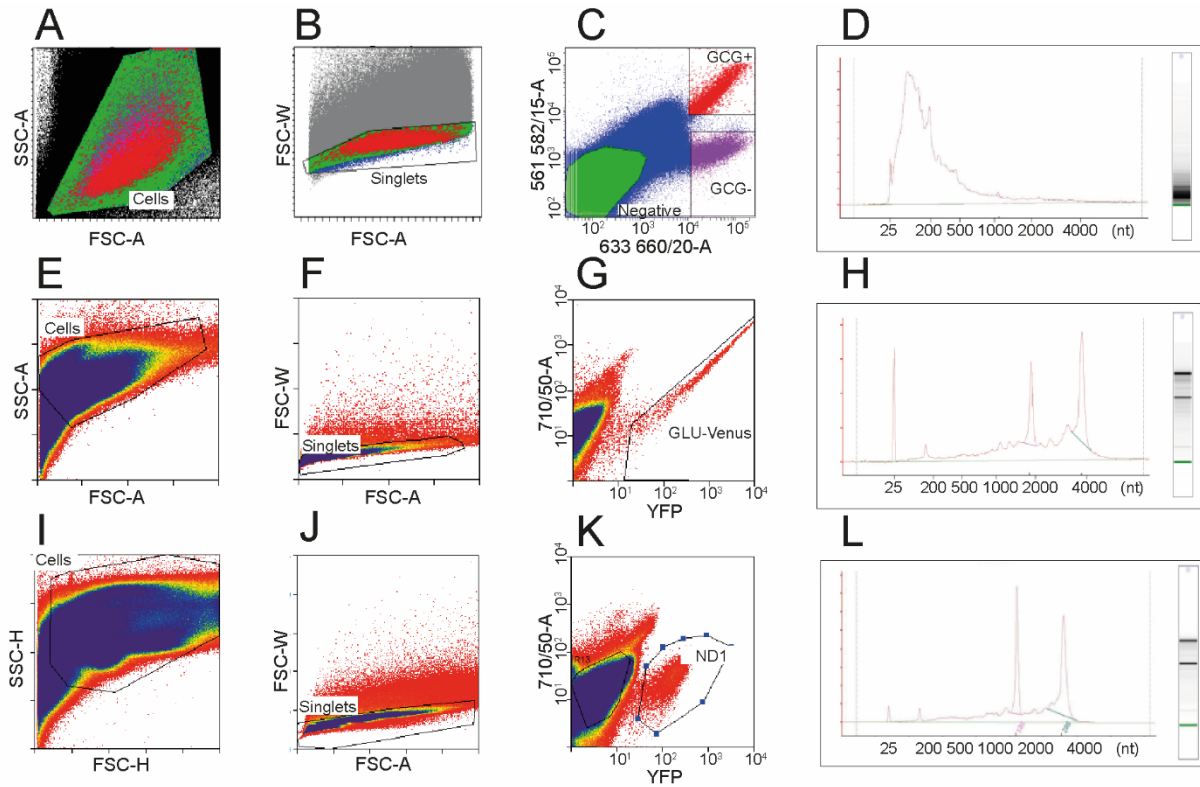
ID	Source	Age	Gender	BMI
R3	France – RYGB	N/A	N/A	>40
R4	France – RYGB	24	F	47
R5	France – RYGB	31	F	43
R6	France – RYGB	46	M	41
D1	France – donor	27	M	24
D2	France – donor	26	F	21
HGH006	UK – TG	57	M	34
HGH007	UK – TG	22	M	29
HGH008	UK – TG	23	M	28
HGH010	UK – TG	64	M	27
HGH018	UK – TG	61	M	37

Supplementary table 1. Demographics of human participants from transcriptomic study. RYGB = Roux-en-Y Gastric Bypass; donor = organ transplant donor; TG = Total gastrectomy for prevention / treatment of gastric cancer

2

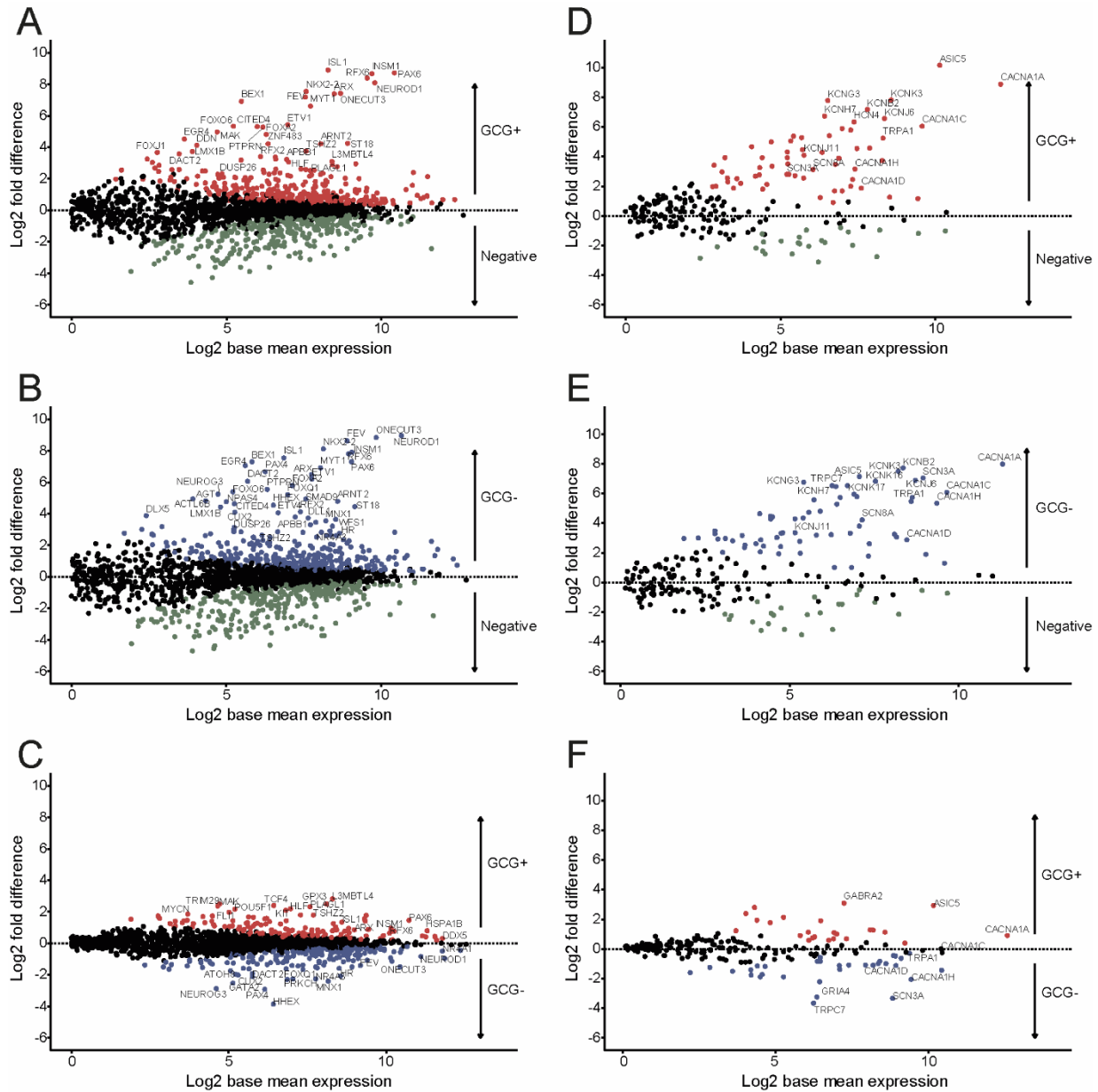
3

1



2

3



Supplementary figure 2.

Differential vs mean expression plots for human jejunum EEC populations. Differential expression is presented as the log(2) fold difference between the cell populations indicated, and expression is presented as the log(2) base mean normalised expression extracted from the DESEQ2 model. (A,D) GCG+ vs negative, (B,E) GCG- vs negative (C,F), GCG+ vs GCG-. Transcription factors are shown in A,B,C; ion channels are shown in D,E,F. Red – enriched (i.e. adjusted P<0.1) in GCG+; Blue – enriched in GCG-; Green – enriched in negative cells. Key genes are labelled

1

2

1 **Supplementary figure 3 – Peptidomics optimisation studies**

2 Peptidomics is a relatively new and under-developed field within proteomics, despite the varied and
3 significant roles of peptides as signalling molecules in many organisms. Peptides, due to their small size in
4 comparison to proteins, have specific properties that can be used to extract and analyse them intact by mass-
5 spectrometry, giving information about their exact processing and post-translational modifications. Several
6 methods to extract peptides from different tissues or fluids have been published, but none was developed
7 to study intestinal tissues and the large variety of hormone peptides they produce.

8 Peptide extraction and purification for LC-MS/MS studies initially requires the complete disruption of the
9 tissue of interest while inhibiting all protease activity. This is to avoid contamination with degradation
10 products of larger proteins as well as degradation of the peptides of interest. The low-molecular weight
11 proteins and peptides must then be separated from the rest of the proteins and other molecules. We
12 compared several methods adapted from the literature to develop a robust and reliable method to quantify
13 peptide levels from intestinal tissues. Methods tested for tissue lysis and protein extraction included tissue
14 homogenization in 0.25% acetic acid, 8M urea, 90% methanol with 1% acetic acid, guanidine hydrochloride
15 or 80% acetonitrile, or using Trizol to extract both proteins and peptides from the same sample¹⁻⁴ Peptides
16 were separated from proteins using either 10kDa molecular weight cut-off filters or 80% acetonitrile. Samples
17 were finally purified using SPE columns, dried and reduced alkylated before being analysed by mass-
18 spectrometry.

19 The different methods allowed the detection of many unique peptides by LC-MS/MS with an important
20 variability between the different methods, from 1778 unique peptides for the extraction using 8M urea to
21 25 with 90% methanol/1% acetic acid. Many corresponded to fragments of longer peptides or proteins and
22 could represent degradation products, coming either from the protein turn-over of the tissue or degradation
23 happening during sample preparation. However, most of the peptides were unique to the extraction method
24 (Supp Fig 2a), due to their different chemical properties and also due to different degradation levels. Focusing
25 on the analysis of the classical gut peptide hormones significantly reduced the number of hits, and clearly
26 showed that most of the expected intact peptides were detected using the guanidine HCl extraction method
27 (Supp Fig2b). In addition to improving the number of detected hormone peptides, the guanidine HCl
28 extraction method showed the highest levels of these peptides when quantified (Supp Fig 2c). Urea and Trizol
29 extraction were the second best methods for peptide levels and variety of peptide hormones extracted, and
30 tissue homogenization using Trizol can be a reliable technique to extract both RNA and peptides from a same
31 sample. We also compared peptide extraction from the proteins using either protein precipitation with 80%
32 ACN or using 10kDa filters, the latter resulting in the loss of many peptides and decreased levels (Supp Fig
33 2c).

1 We then confirmed the reproducibility of the guanidine HCl +80% ACN extraction method, as the %CV from
2 5 replicates repeated twice from different mice was around 20% for most of the peptides, with higher %CV
3 for peptides detected with lower abundance (Supp Fig 2d), and we could also show that peptide
4 quantification was linear with the amount of starting material in a range from 2 to 40 mg (Supp Fig 2e),
5 indicating that tissue weight can be a satisfactory normalisation factor to compare different samples.

6 Tissue homogenization in guanidine HCl followed by protein precipitation with 80% acetonitrile is therefore
7 a promising method to extract peptides from intestinal tissue for mass-spectrometry analysis, allowing the
8 detection a large range of peptides and their robust quantification.

9

10 **Homogenization optimization**

11 Small pieces of mouse jejunal mucosa of similar size (~40mg) were homogenized in different conditions in
12 duplicates with lysing Matrix D using a FastPrep-24 for 4 x 40s at 6ms⁻¹, and peptide extracted before being
13 dried on a centrifugal concentrator. Samples were resuspended in 500µL 0.1% aqueous formic acid and
14 purified by solid-phase extraction using a Waters HLB µElution solid-phase extraction (SPE) plate as
15 described. Samples were analysed using high-flow rate based LC-MS method as described before.

16 For the acetic acid, guanidine HCl and urea extractions, tissues were directly incubated in 300µL of the lysing
17 solution (0.25% acetic acid aqueous, guanidine HCl 6M or urea 8M) and homogenized. 200µL of the
18 homogenate was collected into a new tube after centrifugation to remove tissue debris and lysing matrix
19 (5min, 2000g, 4°C). 80% acetonitrile (800µL) was added to precipitate the proteins and peptides in solution
20 were collected following centrifugation (10min, 12,000g, 4°C). For the guanidine HCl homogenates, addition
21 of 80% acetonitrile induced a separation between an organic phase (top) and aqueous phase (bottom) in
22 addition to the protein precipitation, with the peptides in the aqueous phase (data not shown). Peptides
23 from guanidine HCl homogenization were also extracted using a Vivaspin 500 MCWO 10kDa (GE-healthcare)
24 instead of using 80% ACN, collecting the flow-through after a 30 minutes spin at 16,000g at 4°C. Flow-through
25 was then acidified adding 10% formic acid solution and extracted by SPE.

26 For the 80% ACN and methanol/acetic acid extraction, tissues were directly homogenized in 80% acetonitrile
27 aqueous or 90% methanol 1% aqueous acetic acid, 200 µL of homogenate was recovered from the lysing D
28 matrix and spun at 12,000g at 4°C for 10 minutes to separate the precipitated proteins. Peptides in
29 supernatant were transferred to a new tube and dried before SPE extraction as previously.

30 For the Trizol extraction, tissue was homogenized in 500 µL Trizol and the homogenates transferred to a new
31 tube. 100µL chloroform was added and mixed vigorously and sample centrifuged after 10min incubation at
32 RT (15min, 12,000g, 4°C). The upper phase (aqueous) was then transferred to a new tube (from which RNA
33 can be extracted using standard methods), 800uL 80% ACN was added to precipitate any remaining proteins

1 and sample centrifuged (10min, 12,000g, 4°C) and supernatant dried before SPE extraction. The lower
2 (phenol) phase was treated with 100% ethanol (1:3 ethanol:phenol phase by volume) to precipitate DNA and
3 the mixture centrifuged (10min, 2,000g, 4°C) then transferred to a new tube. Proteins and peptides were
4 precipitated by addition of 750 µL cold acetone and centrifugation at 7500g for 5min at 4°C. Pellets were
5 washed once with 1mL 0.3M guanidine HCl in ethanol and air dried before being resuspended in 200µL 8M
6 urea. Finally, proteins were precipitated with 80% ACN, centrifuged (10min, 12,000g, 4°C) and the
7 supernatant (containing the peptides) was transferred to a fresh tube, dried and SPE extracted as previously.

8

- 9 1. Basak, O. *et al.* Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables
10 Differentiation of Hormone-Producing Enteroendocrine Cells. *Cell Stem Cell* **0**, R106-188 (2016).
- 11 2. Van Camp, K. A., Baggerman, G., Blust, R. & Husson, S. J. Peptidomics of the zebrafish *Danio rerio*: In
12 search for neuropeptides. *J. Proteomics* **150**, 290–296 (2017).
- 13 3. Van Dijck, A. *et al.* Comparison of extraction methods for peptidomics analysis of mouse brain
14 tissue. *J. Neurosci. Methods* **197**, 231–237 (2011).
- 15 4. Ziros, P. G., Chartoumpakis, D. V. & Sykiotis, G. P. A Simple Protocol for High Efficiency Protein
16 Isolation After RNA Isolation from Mouse Thyroid and Other Very Small Tissue Samples. in 383–393
17 (Humana Press, New York, NY, 2016). doi:10.1007/978-1-4939-3756-1_25

18

19

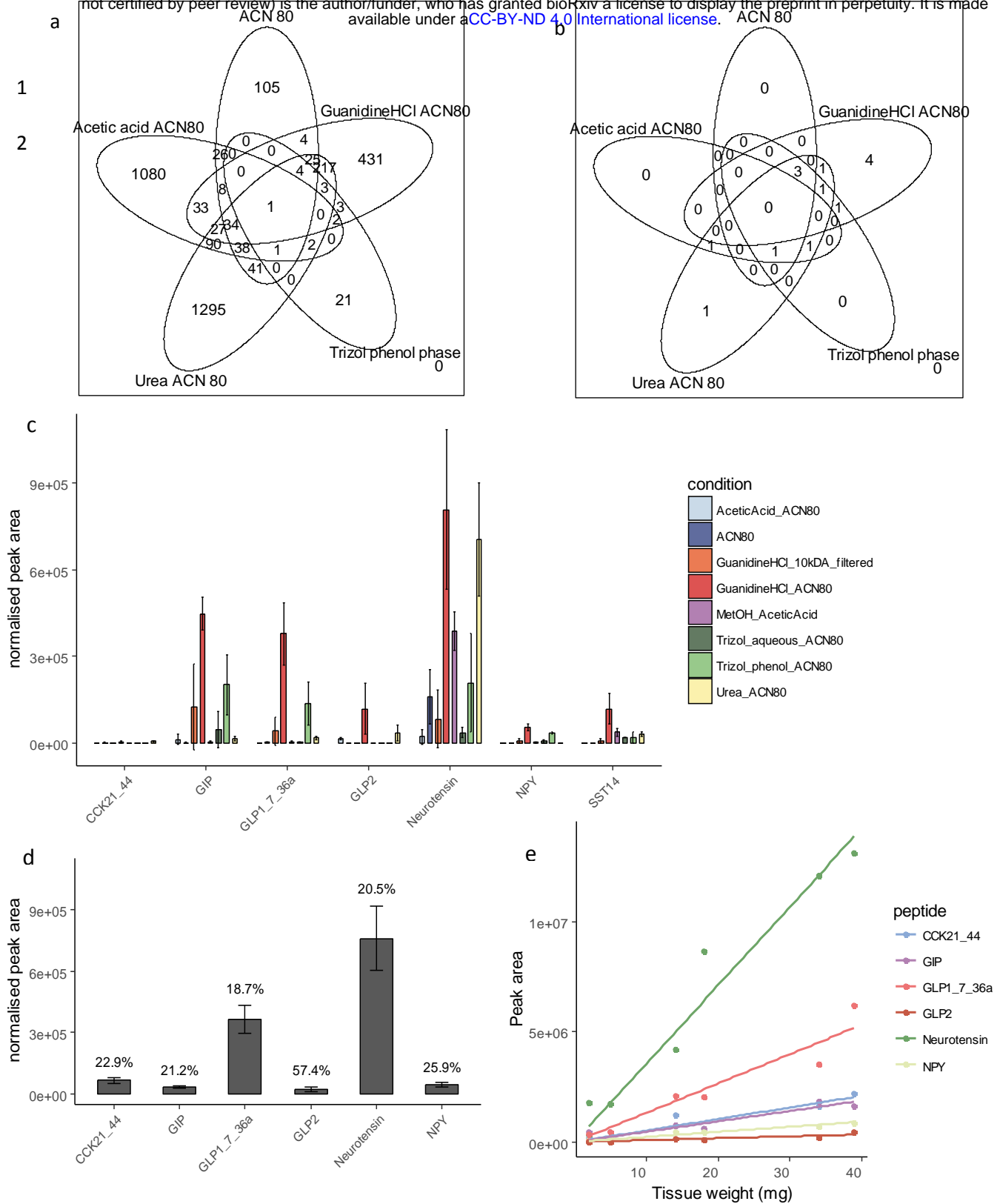


Figure legend (Suppl fig 3).

a and b: Venn diagrams showing the total number of unique peptides (a) and the number of unique peptides matching classical hormone peptides (b) found in the different methods for peptide extraction. c: Quantification of gut hormone peptides using different extraction methods, normalised by tissue weight. Data is represented as mean +/- SD from 2 replicates. d: Quantification variability for different peptides measured from 10 homogenates from 2 different mice. Data is the mean value +/- SD and values indicated above are the %CV. e: Raw peak area of peptide quantification for different samples spanning different weights from 2 to 40 mg plotted against the tissue weight. Linear regression is plotted, showing a good correlation between tissue weight and peptide quantification.