Identification of master regulators in goblet cells and Paneth cells using transcriptomics profiling of gut organoids and multi-layered networks

Treveil A[1,2*], Sudhakar P[1,3*], Matthews Z J[4] *, Wrzesinski T[1], Jones E J[1,2], Brooks J[1,2,4,5], Olbei M[1,2], Hautefort I[1], Hall L J[2], Carding S R[2,4], Mayer U[6], Powell P P[4], Wileman T[2,4], Di Palma F[1], Haerty W[1,#], Korcsmáros T[1,2#]

[1] Earlham Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UZ, UK
[2] Quadram Institute, Norwich Research Park, Norwich, Norfolk, NR4 7UA, UK
[3] Department of Chronic Diseases, Metabolism and Ageing, KU Leuven, Belgium
[4] Norwich Medical School, University of East Anglia, Norwich, Norfolk, NR4 7TJ, UK
[5] Department of Gastroenterology, Norfolk and Norwich University Hospitals, Norwich, UK
[6] School of Biological Sciences, University of East Anglia, Norwich, Norfolk, NR4 7TJ, UK

* Equal contributions
[#] Joint corresponding authors

Corresponding author contact details:

Dr Tamas Korcsmaros
ORCID id : https://orcid.org/0000-0003-1717-996X
Tamas.Korcsmaros@earlham.ac.uk
Phone : 0044-1603450961
Fax : 0044-1603450021
Address: Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK

Summary

Dysfunction of intestinal epithelial cells is often associated with severe gut pathologies, such as inflammatory bowel disease (IBD). Although transcriptional signatures of intestinal cells have been identified, an integrated network analysis of specific cell types which identifies key regulators and disease relevance has been lacking. Here, we profiled the expression of mRNAs, microRNAs and long non-coding RNAs from mouse derived 3D intestinal organoids whose differentiation was directed towards Paneth cells or goblet cells. We generated regulatory networks specific to these cell types by integrating expression data into networks of transcriptional and post-transcriptional regulatory interactions. Analysis of the networks uncovered potential cell-type specific master regulators and highlighted possible regulatory rewiring between the cell types. Investigation of regulators identified links to cellular phenotypes associated with IBD. Application of our workflow can be used to disentangle multifactorial mechanisms of IBD and to propose regulators whose pharmacological targeting could be advantageous in treating Crohn's disease.

Keywords: Intestinal epithelium, Paneth cells, goblet cells, organoids, transcriptomics, regulatory networks, signalling, pathways, Crohn's disease

Introduction

Gut barrier integrity is critically important for efficient nutrient absorption and maintenance of intestinal homeostasis (Zhang et al., 2015) and is maintained by the combined action of the various cell types lining the intestinal epithelium (Okumura and Takeda, 2017). These intestinal epithelial cells serve to mediate signals between the gut microbiota and the host innate/adaptive immune systems (Duerkop et al., 2009; Gerbe and Jay, 2016). Disruption of epithelial homeostasis along with dysregulated immune responses are some of the underlying reasons behind the development of inflammatory bowel disease (IBD) such as Crohn's disease (CD) and ulcerative colitis (UC) (Mokry et al., 2014). Therefore, a greater understanding of the functions of intestinal cells will further our understanding of the aetiology of inflammatory gut conditions.

To date, various cell types have been identified in the intestinal epithelium based on specific functional and gene expression signatures. Paneth cells residing in the small intestinal crypts of Lieberkühn help to maintain the balance of the gut microbiota by secreting anti-microbial peptides, cytokines and other trophic factors (Bevins and Salzman, 2011). Located further up the intestinal crypts, goblet cells secrete mucin, which aggregates to form the mucus layer, which acts as a chemical and physical barrier between the intestinal lumen and the epithelial lining (Kim and Ho, 2010). Both of these cell populations have documented roles in gut-related diseases (Gersemann et al., 2011; Okamoto and Watanabe, 2016). Dysfunctional Paneth cells with reduced secretion of anti-microbial peptides have been shown to contribute to the pathogenesis of CD (Liu et al., 2016), while reduction in goblet cell numbers and defective goblet cell function has been associated with UC in humans (Gersemann et al., 2009).

To understand how changes in the function of intestinal epithelial cell types leads to disease, it is instrumental to first identify their molecular signatures and regulatory pathways in a healthy state. At least three separate studies have captured RNA signatures of healthy and/or diseased intestinal cell types by profiling messenger RNAs (mRNA), microRNAs (miRNA) and long non-coding RNAs (lncRNA) (Haber et al., 2017; Mirza et al., 2015; Peck et al., 2017). MiRNAs and lncRNAs, together with transcription factors (TFs) (proteins that regulate transcription), perform critical regulatory functions in maintaining the homeostasis of intestinal cells. Dysregulation of these regulatory functions has been associated with various gut pathologies (Chapman and Pekow, 2015; Mirza et al., 2015). However, a comprehensive analysis of miRNAs, lncRNAs and TFs in conjunction with the genes and proteins they regulate has not yet been performed on a systems-level in a standardized manner, including the connections between a regulator and its downstream effect. For example, biological interaction networks are a type of systems biology data representation, which aid the interpretation of -omics read-outs by identifying relevant signalling and regulatory pathways. In particular, the study of regulatory interactions using interaction networks, can be used to uncover how cells respond to changing environments at a transcriptional level and to investigate the downstream effects of gene mutations and knockouts (Luscombe et al., 2004).

Recent studies, particularly those focusing on diseases, have used biopsy samples to produce -omic read-outs (Balfe et al., 2018; Mirza et al., 2015). When the expression profile of diseased tissue is greatly different from that of a healthy control, these experiments produce interesting results. However, due to the cellular heterogeneity of the biopsies, the -omic readouts usually represent an average reading across different cell types (including semi-differentiated cells) thereby obfuscating the profiles of relatively rare yet fundamental cell types. Current understanding of many intestinal diseases such as CD and UC, implicates specific cell types in the dysregulation of homeostasis (Adolph et al., 2013). Whilst primary intestinal epithelial cells all originate from Lgr5+ stem cells, differentiation results in differences in gene expression as well as signalling and regulatory wiring (Crosnier et al., 2006; Vanuytsel et al., 2013). These differences can result in altered phenotypic functions, responses to stress and susceptibilities to specific dysregulations. Therefore, to understand the role of these cell types in diseases, it is necessary to study their molecular networks and pathways in a cell-type specific manner. Here, we reduced the heterogeneity often encountered with intestinal biopsies by employing the small intestinal organoid system (enteroids), which recapitulates many features of the *in vivo* intestinal tissue (Sato et al., 2009).

In this study, we identified Paneth cell and goblet cell specific mRNAs, miRNAs and lncRNAs, and applied this data to generate cell-type specific regulatory networks. To obtain the expression data, we used mouse derived 3D enteroids; either conventionally differentiated (control) or enriched in Paneth cells or goblet cells. In two recently published reports, we show that these enteroid models are useful for the investigation of health and disease related processes in intestinal cell types (Jones et al., 2019; Luu et al., 2018). We performed RNAseq transcriptomics analysis on the enteroids, and compared the readouts from Paneth cell and goblet cell enteroids to the conventionally differentiated enteroids. To the best of our knowledge, this is the first study profiling all the three regulator types (TFs, miRNAs, lncRNAs) from across different intestinal epithelial cell types. To interpret the data, we integrated the expression data with directed regulatory networks from resources containing transcriptional and post-transcriptional interactions. This integrative strategy led us to define regulatory network landscapes, master regulators of Paneth cells and goblet cells, and to highlight the contribution of regulatory rewiring in the diverse phenotypes of intestinal cells (Figure 1). Based on genetic and functional links between the networks we generated and IBD, we show that the workflow has potential for application to the study of cell-type specific mechanisms of IBD, as well as to propose targets for IBD pharmacological interventions.

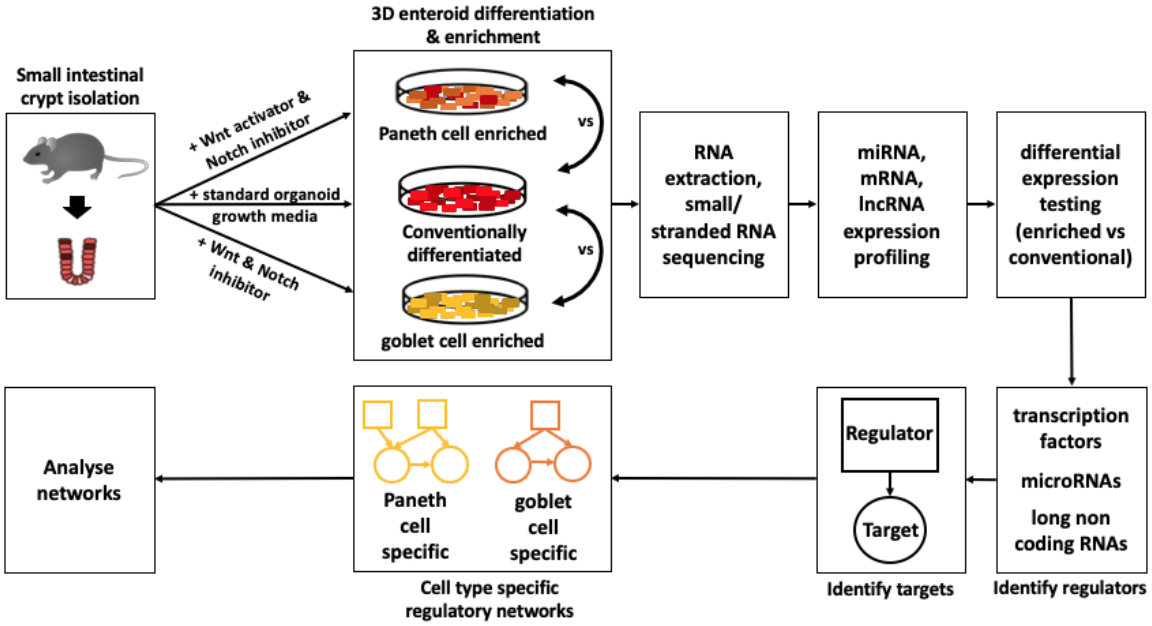

Figure 1: Schematic representation of the workflow to infer cell-type specific regulatory networks *from 3D enteroids differentiated into specific intestinal cell types. TF- transcription factor; lncRNA- long non-coding RNA; miRNA- microRNA; mRNA- messenger RNA.*

Results

Establishment of Paneth cell or goblet cell enriched *in vitro* models

We generated 3D self-organising enteroid cultures *in vitro* which recapitulated features of the *in vivo* small intestinal epithelium, including formation of an internal lumen and distinct crypt- and villus-like domains and presence of all major epithelial cell types (Figure 2A). To increase the technical feasibility of investigating Paneth cells and goblet cells, organoids were further enriched for these cells using a well-established and published protocol, presented in details in the Methods (Sato and Clevers, 2013; Yin et al., 2014). This method has been shown to produce organoids of which 70% cells displayed Paneth cell phenotype, for example (Yin et al., 2014). Indeed, by altering the stem cell culture medium, with addition of cell proliferation and secretory cell differentiation (Yin et al., 2014) we were able to enrich organoids for Paneth cells or goblet cells (Figure 2B).

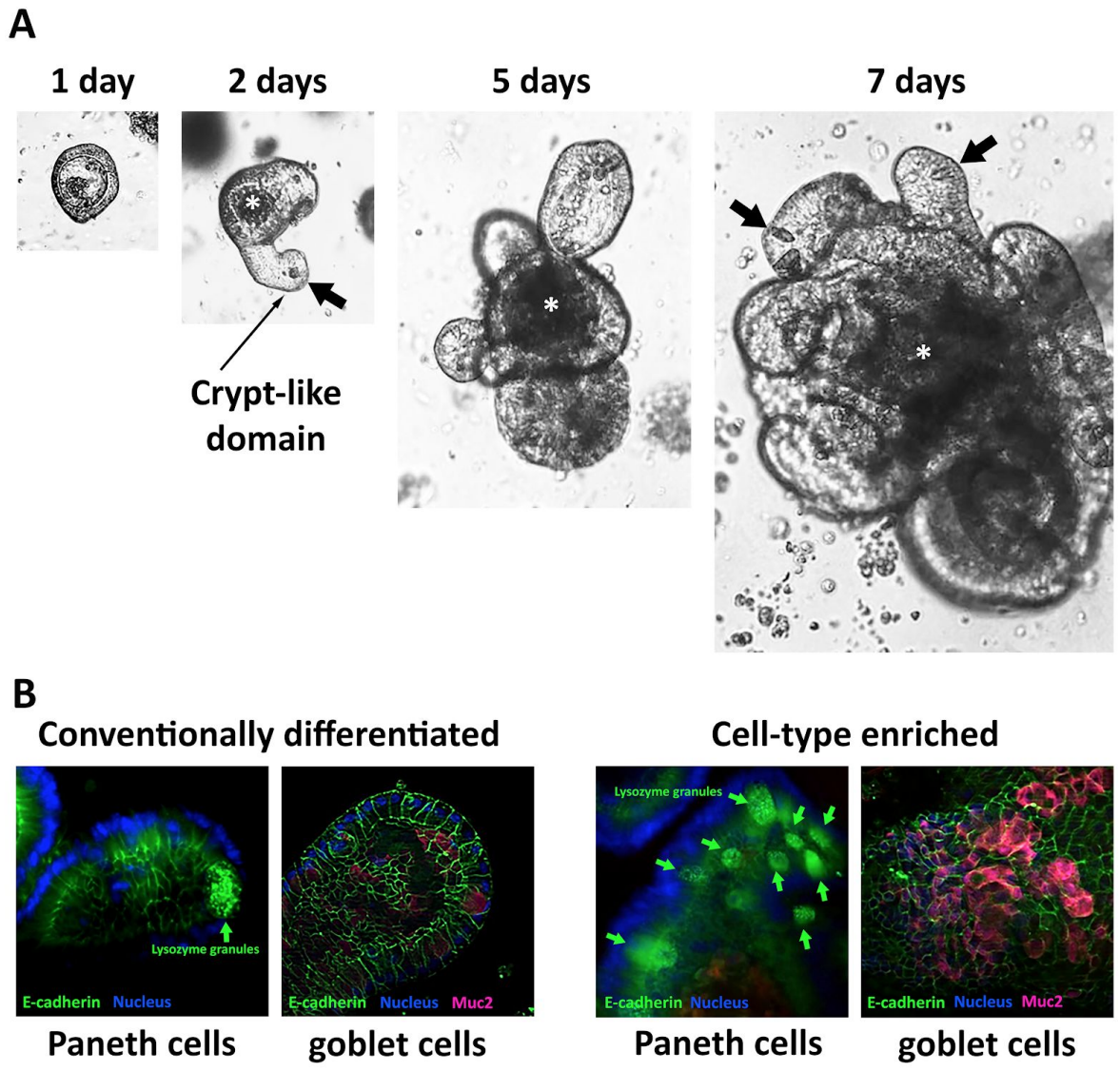

Figure 2: Small intestinal 3D organoid culture. *(A) Culture of isolated mouse small intestinal epithelial crypts in Matrigel matrix and ENR media (conventionally differentiated) for 7 days. Isolated crypts form 3D cysts which bud after 2 days of culture to form crypt- and villus-like domains. Paneth cells are clearly visible by light microscopy (Black arrows). Mucous and shedding cells accumulate in the central lumen of organoids (*). n = 3. (B) Cell-type specific enrichment illustrated by*

immunofluorescence labelling of cultured mouse 3D organoids, conventionally differentiated (left) and enriched for either Paneth cells or goblet cells (right). Lysozyme granules characteristic of Paneth cells were identified using a specific anti-lysozyme antibody and are indicated with a green arrow. Goblet cells were identified using a specific anti-Muc2 mucin antibody (pink).

Paneth cell and goblet cell marker genes are over represented in cell-type enriched enteroids compared to conventionally differentiated controls

Transcriptomics was used to determine gene expression dynamics in conventionally differentiated enteroids and enteroids enriched for Paneth cells and goblet cells. Differentially expressed genes were calculated by comparing the RNA expression levels (including protein coding genes, lncRNAs and miRNAs) of enteroids enriched for Paneth cells or goblet cells to those of conventionally differentiated enteroids. 4,135 genes were differentially expressed (at significantly higher or lower levels) in the Paneth cell dataset, and 2,889 were differentially expressed in the goblet cell dataset (Figure 3A, Table S1). The larger number of differentially expressed genes (DEGs) in the Paneth cell data could be attributed to the highly specialised nature of Paneth cells (Clevers and Bevins, 2013; Stappenbeck and McGovern, 2017). The majority of the DEGs were annotated as protein coding: 79% in the Paneth cell dataset and 84% in the goblet cell dataset. In addition, lncRNAs (Paneth cell, 11%; goblet cell, 9%), and miRNAs (Paneth cell, 4%; goblet cell, 2%) were identified among the DEGs and any gene not annotated as one of the above is classified as 'other', which includes pseudogenes and antisense genes (Figure 3A). Some of these DEGs were identified in both the Paneth cell and the goblet cell datasets, exhibiting the same direction of change compared to the conventionally differentiated enteroid data. In total, 1,363 genes were found upregulated in both the Paneth cell and goblet cells while 442 genes were found downregulated in both datasets. This result highlights considerable overlap between the cell types, and can be explained by the shared differentiation history and secretory function of both Paneth cells and goblet cells.

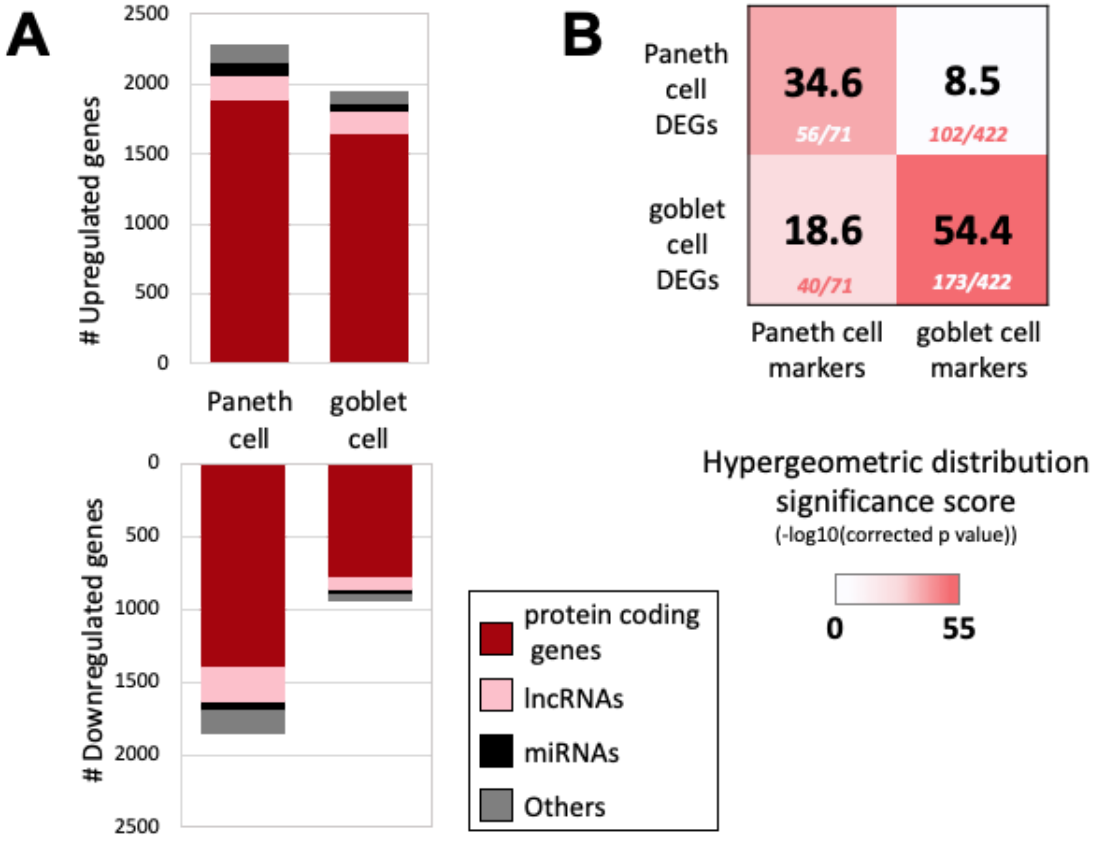

Figure 3: Differentially expressed genes in Paneth cells and goblet cells. *(**A**) Number of differentially expressed genes. Differential expression was defined as log2 fold change > |1| and false discovery rate <= 0.05. miRNA = microRNA; lncRNA = long non-coding RNA; Genes annotated as 'other' include pseudogenes and antisense genes. Goblet cell = goblet cell enriched organoids compared to normally differentiated; Paneth cell = Paneth cell enriched organoids compared to normally differentiated; downregulated = log2 fold change <= -1; upregulated = log2 fold change >= 1. (**B**) Enrichment of cell type specific marker genes in the differentially expressed gene (DEG) lists. Higher significance scores indicate greater enrichment. Number of markers in DEG list out of total number of markers shown below significance score. Also see Table S1,3.*

To validate the cell types present in the enteroids, the expression of five previously reported major cell-type specific markers were investigated across the enteroids using transcript abundances and RNA differential expression results (Table S2, Figure S2). The control enteroids and the cell-type enriched enteroids expressed all five markers: stem cells (*Lgr5*), enteroendocrine cells (*ChgA*), goblet cells (*Muc2*), Paneth cells (*Lyz1*) and epithelial cells (*Vil1*). We observed an upregulation of *Muc2, Lyz1* and *ChgA* in Paneth cell and goblet cell enriched enteroids compared to the control enteroids. *Lgr5* is downregulated in the cell-type enriched enteroids compared to the control enteroids, confirming the more pronounced differentiated status of the enteroids. Therefore, using primary cell-type specific markers we show that the enteroids contain all major cell types at quantities expected based on the culture mediums used.

To further investigate the cell-type specificity of the DEGs, we measured enrichment of Paneth cell and goblet cell marker genes in the upregulated DEG lists - given that DEGs also reflect the difference in cell-type proportions between the cell-type enriched and conventionally differentiated enteroids. While all tests were significant (hypergeometric model, q value ≤ 0.05), we identified greater enrichment of Paneth cell markers in the Paneth cell DEG list and goblet markers in the goblet cell DEG list (Figure 3B). This confirms that differential expression tests successfully discerned cell-type specific signatures from the RNAseq data, although the data also contains non-specific transcriptional noise from other cell types. We also noticed that both of the upregulated DEG lists are enriched with enteroendocrine markers, thus suggesting the presence, albeit at lower levels, of enteroendocrine cells in the enriched enteroids (Table S3). Confirming this observation, the enteroid differentiation protocols used have shown previously that in addition to Paneth cell and goblet cell enrichment, it results in enteroendocrine enrichment as well (Yin et al., 2014). In conclusion, we have used gene expression analysis to identify expression signatures of Paneth cells and goblet cells. In addition to image-based validation of the enteroid cell-type enrichment protocol (Figure 2B), application of extensive cell-type specific markers data shows that the Paneth cell and goblet cell expression datasets exhibit cell-type specific signatures, confirming successful cell-type enrichment of the enteroids.

Reconstruction of cell-type specific regulatory networks

To identify regulator-target relationships within the DEG lists and thus gain understanding of the biological context, we apply a network biology approach. We generate a large (universal network of non-specific molecular interactions known to occur in mice, by collating lists of published data (Table S4). The resulting network (termed the universal network) consists of 1,383,897 unique regulatory interactions connecting 23,801 molecular entities. All interactions within the network represent one of the following types of regulation: TF-TG, TF-lncRNA, TF-miRNA, miRNA-mRNA or lncRNA-miRNA. TF-TGs and TF-lncRNAs make up the majority of the network at 77% and 11% of all interactions, respectively. Due to its non-specific nature, this universal network contains many interactions not relevant for Paneth cells and goblet cells. In order to get a clearer and valid picture of cell-type specific regulatory interactions, we use the universal network to annotate the Paneth cell and goblet cell DEGs with regulatory connections. Combining these connections, we generate specific regulatory networks for Paneth cells and goblet cells.

In total, the Paneth cell network, generated using differential expression data from the Paneth cell enriched enteroids compared to the conventionally differentiated enteroids, contained 37,062 interactions connecting 208 unique regulators with 3,023 unique targets (Figure 4, Item S1). The

goblet cell network, generated using differential expression data from the goblet cell enriched enteroids compared to the conventionally differentiated enteroids, contained 19,171 interactions connecting 124 unique regulators with 2,095 unique targets (Figure 4, Item S1). 15.7% of all interactions (8,856 out of 56,234) were shared between the Paneth cell and goblet cell networks, however the interacting molecular entities (termed nodes) do not all exhibit the same direction of differential expression between the networks (comparing Paneth cell or goblet cell enriched enteroid data to the conventionally differentiated enteroid data). In each of the cell-type specific regulatory networks, a particular gene is represented (as a node) only once, but may be involved in multiple different interactions. In different interactions, a single node may act either as a regulator or as a target and in different molecular forms, for example, as a lncRNA in one interaction and as a target gene in another. In conclusion, we have generated specific regulatory interaction networks for Paneth cell and goblet cell types including transcriptional and post-transcriptional interactions.

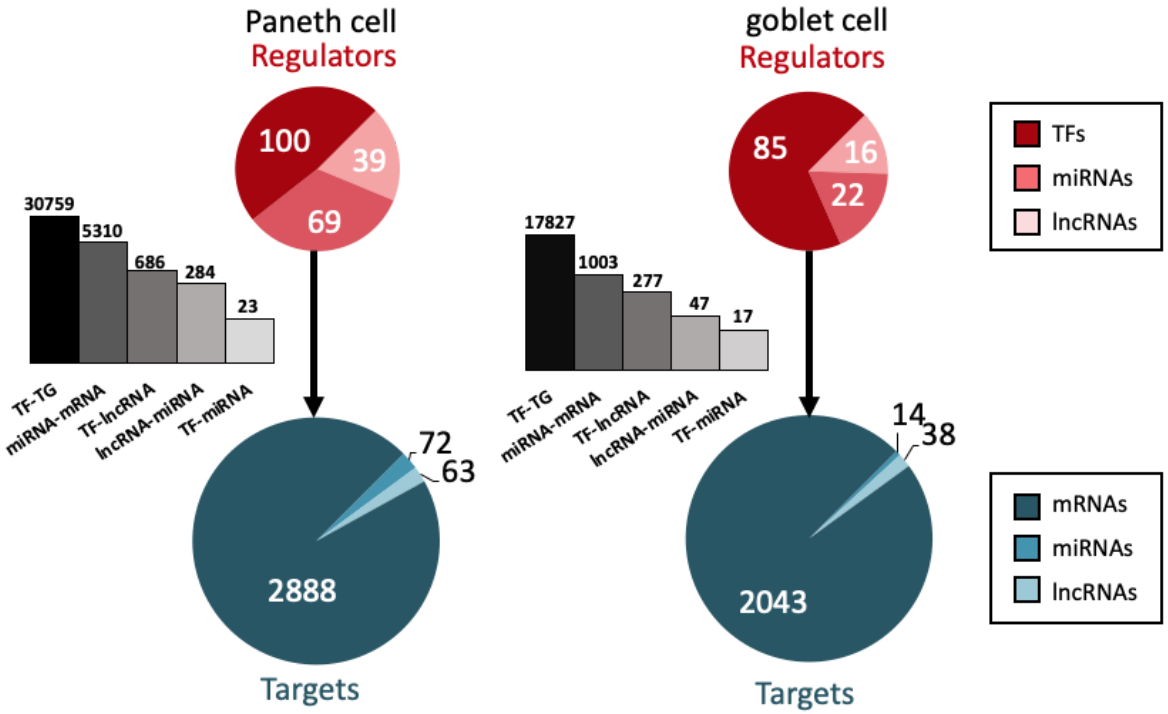

Figure 4: Cell-type specific regulatory network summary for Paneth cell (left) and goblet cell (right) datasets. *TF- transcription factor; miRNA- microRNA; lncRNA- long non-coding RNA. Total number of each regulator type shown in red, number of each target type shown in blue. In the targets pie-chart, mRNAs also represent protein coding genes and proteins, miRNAs also represent miRNAs genes and lncRNAs also represent lncRNA genes. Size of circles represents log10(total unique regulators/targets). Bar chart represents the distribution of interaction types in the networks (log10 scale). Also see Item S1.*

Identification of potential cell-type specific master regulators

Cell specific differentiation markers represent genes performing functions specific to a particular cell-type. Therefore, we expect that the regulators of these marker genes will have an important role in determining the function of different cell types. To identify these regulators, we extracted subnetworks from the Paneth cell and goblet cell regulatory networks consisting of all relevant cell-type specific markers and their direct regulators. 33 markers specific for Paneth cells were identified in the Paneth cell regulatory network with 62 possible regulators. In the goblet cell regulatory network, 150 markers were identified with 63 possible regulators (Figure 5, Table S5). Observing the ratio of regulators and markers in the cell-type specific marker networks (subnetworks), the Paneth cell dataset has, on average, 1.88 regulators for each marker. On the other hand, the goblet cell

dataset exhibits only 0.42 regulators for each marker. The quantity of markers identified in each sub-network (33 in the Paneth network and 150 in the goblet network) correlates with the number of marker genes identified by Haber *et al.* (Haber et al., 2017). However, far fewer regulators were identified in the goblet cell subnetwork per marker than for the Paneth cell subnetwork. Whilst the underlying reason for this discrepancy is unknown, it could potentially be evidence of the complex regulatory environment required to integrate and respond to the arsenal of signals recognised by Paneth cells in comparison to goblet cells (Stappenbeck and McGovern, 2017).

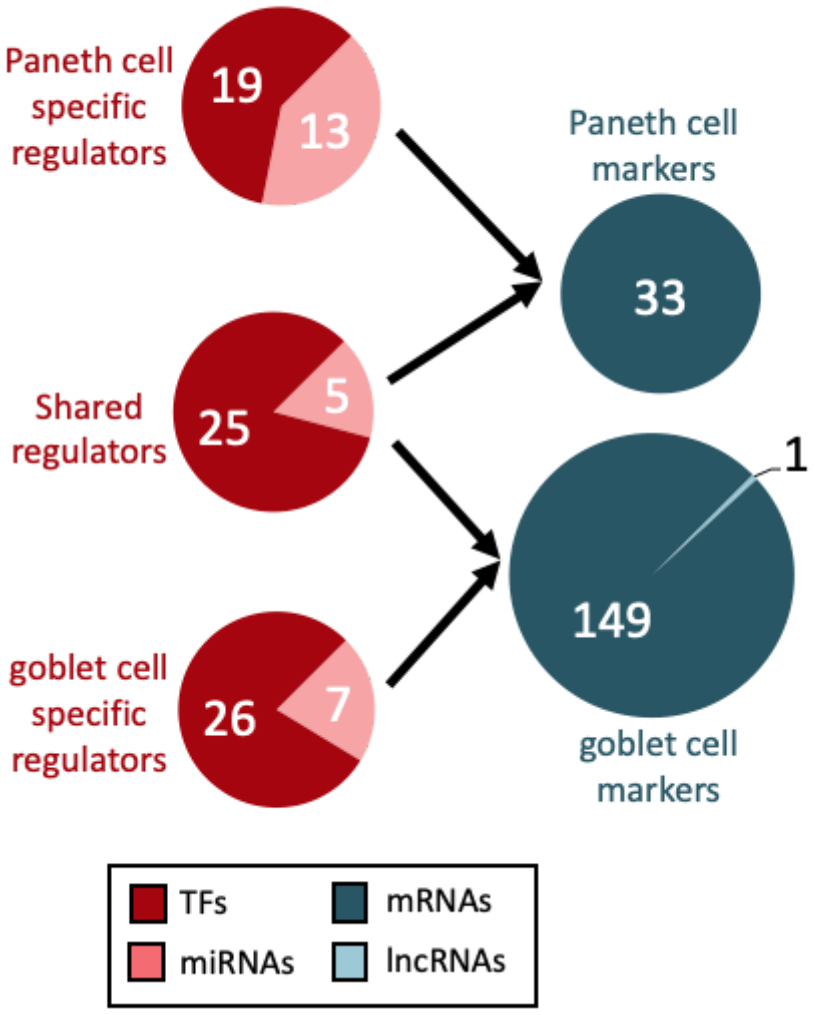

Figure 5: Cell-type specific regulator-marker subnetwork summary for Paneth cell (above) and goblet cell (below) datasets. *TF- transcription factor; miRNA- microRNA; lncRNA- long non-coding RNA. Total number of each regulator type shown in red, number of each target type shown in blue. Regulators have been categorised based on their membership in the two networks - shared regulators are present in both networks. In the targets pie-chart, mRNAs also represent protein coding genes. Size of circles represents log10 (total unique regulators/targets). See also Table S5-7.*

Of the 95 cell-type specific marker regulators, we identified approximately one-third (30/95) as present in both subnetworks (Figure 5). Given that the markers are different between the cell types, a regulator shared between the Paneth cell and goblet cell marker sub-networks must show an altered pattern of regulatory targeting in the two cell types. This phenomenon, referred to as regulatory rewiring, often results in functional differences of shared regulators in different environments (Han et al., 2017) - for example, in this case, between the Paneth cells and goblet cells.

Further investigation of the distinct regulator-marker interactions highlights a gradient of regulator specificity. We generated matrices to visualise the markers controlled by each regulator in the goblet cell (Figure 6A) and the Paneth cell (Figure 6B,C) marker subnetworks. Each coloured square indicates that a marker (shown on the y-axis) is regulated by the corresponding regulator (shown on the x-axis). Squares are coloured green if the associated regulator is shared between the Paneth cell and goblet cell subnetworks and orange if they are specific to one marker subnetworks. A collection of regulators (both cell-type specific and shared) appear to regulate large proportions of the markers in the networks. For example, Ets1, Nr3c1 and Vdr regulate >50% of the markers in both the Paneth cell and the goblet cell subnetworks. Specific to the Paneth cell subnetwork, Cebpa, Jun, Nr1d1 and Rxra regulate >50% of the markers. Specific to the goblet cell subnetwork, Gfi1b and Myc regulate >50% of the markers. These regulators are potential master regulators of the given cell types. In contrast, regulators such as Mafk in the Paneth cell subnetwork and Spdef in the goblet cell subnetwork regulate only one marker. These regulators likely have more functionally specific roles. Together, these results highlight potential regulators which likely play key roles in specification and maintenance of Paneth and goblet cells and their functions.

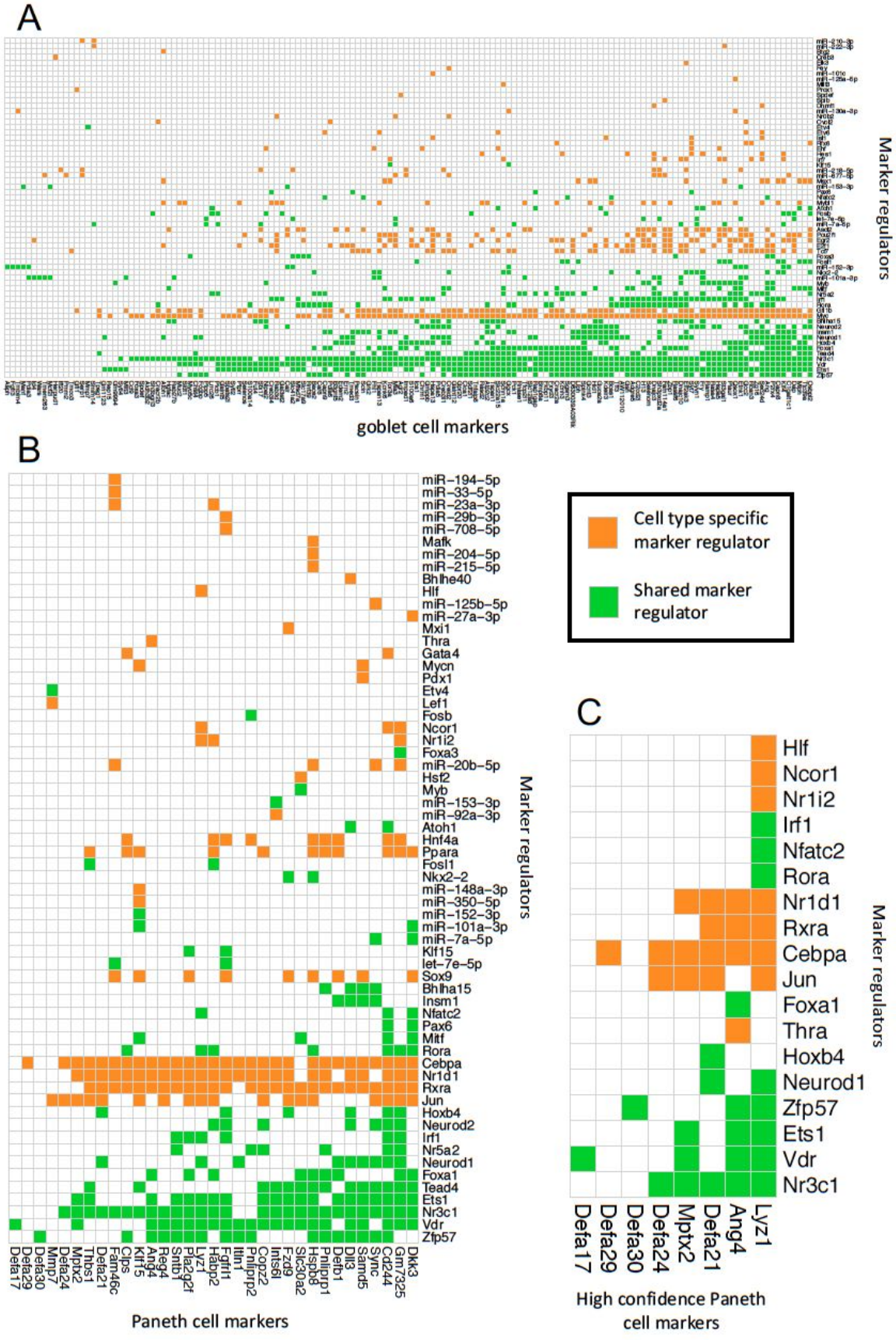

Figure 6: Matrices of regulator-marker interactions. *Regulators on y-axis, markers (regulator targets) on y-axis. Orange boxes indicate regulator-marker interactions where the regulator is only*

found in one cell-type marker subnetwork. Green boxes signify that the regulator is found in both the Paneth and the goblet marker subnetworks. ***A:*** *All goblet cell markers from (Haber et al., 2017) and their regulators.* ***B:*** *All Paneth cell markers from (Haber et al., 2017) and their regulators.* ***C:*** *High quality Paneth cell markers from (Haber et al., 2017) and their regulators. Table of interactions in Table S5.*

Putative cell-type specific regulators exhibit rewiring between Paneth cells and goblet cells

Cell-type marker genes, which carry out cell-type specific functions, are different between Paneth cell and goblet cell marker subnetworks. Therefore, the shared regulators we observed in both marker subnetworks must target different marker genes. To do this, the regulators must have different regulatory connections in the different cell types and therefore be rewired. We aimed to investigate whether any of the 30 identified shared marker regulators are rewired between the whole regulatory networks and thus have different functions in the two different cell types. To quantify rewiring of each of these regulators, we observed their targets in the main cell-type specific regulatory networks using the Cytoscape application, DyNet. Each regulator is assigned a rewiring score depending on how different their targets are between the two regulatory networks (Table S6) Using these rewiring scores, we identified the five most rewired regulators (of 30) as Etv4, let-7e-5p, miR-151-3p, Myb and Rora. Functional enrichment analysis was carried out on the targets of these regulators to test whether the targets specific to the Paneth cell and goblet cell networks have different functions (hypergeometric model, q value ≤ 0.1) (Table S7). Across all five regulators the general trend indicated that regulator targets specific to the Paneth cell regulatory network are associated with metabolism; regulator targets specific to the goblet cell regulatory network are associated with cell cycle and DNA repair. This suggests these functions are key features of the different cell types. Looking at the regulators in more detail, the goblet cell specific targets of miR-151-3p, for example, are significantly enriched in functions relating to antigen presentation, cell junction organisation, Notch signalling and the calnexin/calreticulin cycle. None of these functions are enriched in the shared or Paneth specific targets. Of particular interest is the calnexin/calreticulin cycle, which is known to play an important role in ensuring proteins in the endoplasmic reticulum are correctly folded and assembled (Leach and Williams, 2013). Dysfunction of protein folding and the presence of endoplasmic reticulum stress are both associated with IBD (Kaser and Blumberg, 2009; Kaser et al., 2008, 2011). Therefore, we assume that miR-151-3p plays a role in the secretory pathway of goblet cells and could be an interesting target for IBD research.

In addition, different functional profiles were also observed for the targets of Rora in the Paneth cell and goblet cell regulatory networks: targets present in both networks are significantly associated with mitosis, whereas those specific to the Paneth cell network are associated with metabolism, protein localisation, nuclear receptor transcription pathway, circadian clock and hypoxia induced signalling. Goblet cell specific targets of Rora are connected to Notch signalling, cell cycle and signalling by Rho GTPases (associated with cell migration, adhesion and membrane trafficking) and interferon. Altogether these observations show that some regulators that are shared between the Paneth cell and goblet cell regulatory networks have different targets in the two cell types. This suggests that regulatory rewiring has occurred between the Paneth and goblet cell types. The functions of these altered regulatory targets are different between the two cell types, resulting in functional differences in the actions of the same regulators.

Evaluating the disease relevance of the cell type specific master regulators

To investigate the relevance of the identified regulators in human gastrointestinal diseases, such as IBD, we carried out literature searches of the three groups of predicted master regulators: those specific to the Paneth cell marker regulatory subnetwork (Cebpa, Jun, Nr1d1 and Rxra), those specific to the goblet cell marker regulatory subnetwork (Gfi1b and Myc) and those which appear to regulate many of the markers in both the Paneth and the goblet subnetworks (Ets1, Nr3c1 and Vdr). We identified five genes (*Ets1, Nr1d1, Rxra, Nr3c1 and Vdr*), corresponding to 71% (5/7) of the Paneth cell associated master regulators and 60% (3/5) of the goblet cell associated master regulators, which

have connections to inflammation, autophagy and/or inflammatory bowel disease (IBD), as shown in Table 1. Interestingly, four of these genes (all apart from *Ets1*), encode nuclear hormone receptors.

Putative master regulator	Autophagy/inflammation/IBD associations	References
NR1D1 (REV-ERBa)	Modulates autophagy and lysosome biogenesis in macrophages leading to antimycobacterial effects	(Chandra et al., 2015)
	SNP rs12946510 which has associations to IBD, acts as a cis-eQTL for NR1D1	(Mirza et al., 2015)
NR3C1 (glucocorticoid receptor)	Associations with cellular proliferation and anti-inflammatory responses	(Oakley and Cidlowski, 2013)
	Exogenous glucocorticoids are heavily used as anti-inflammatory therapy for IBD	(Prantera and Marconi, 2013; Rutgeerts, 1998))
	ATG16L1, an autophagy related gene, was down-regulated in patients who do not respond to glucocorticoid treatment	(De Iudicibus et al., 2011; Dubois-Camacho et al., 2017)
	Transcriptionally regulates NFKβ1, a SNP affected gene in ulcerative colitis	(Dinkel et al., 2016; Yemelyanov et al., 2007)
VDR (Vitamin D Receptor)	Regulates autophagy in Paneth cells through ATG16L1 – dysfunction of autophagy in Paneth cells has been linked to Crohn's disease	(Bakke et al., 2018; Wu et al., 2015)
	Induces antimicrobial gene expression in other cell lines	(Gombart et al., 2005; Wang et al., 2004)
	Specific polymorphisms in the VDR genes have been connected to increased susceptibility to IDB	(Pei et al., 2011)
	A study looking at colonic biopsies of IBD patients observed reduced VDR expression compared to healthy biopsies	(Abreu et al., 2004)
	Interacts with five SNP affected UC genes	(Bovolenta et al., 2012; Lesurf et al., 2016)
RXRa (Retinoid X Receptor Alpha)	Heterodimerizes with VDR (see above)	
ETS1 (ETS Proto-Oncogene 1)	Important role in the development of hematopoietic cells and Th1 inflammatory responses	(Grenningloh et al., 2005; Mouly et al., 2010)
	Angiogenic factors in the VEGF-Ets-1 cascades are upregulated in UC and downregulated in CD	(Konno et al., 2004)

	IBD susceptibility gene	(Li et al., 2018)

Table 1: *Literature associations relating to autophagy, inflammation and IBD for putative master regulators.*

Additionally, given the possible relationships between the identified master regulators and IBD, we tested the potential of our cell-type specific regulatory networks for the study of CD. To do this, we checked for the presence of known CD susceptibility genes in the Paneth cell and goblet cell whole regulatory networks, using lists obtained from two studies (Farh et al., 2015; Jostins et al., 2012). Of the 81 CD susceptibility genes, 22 (27%) were identified in the Paneth cell network and 12 (15%) in the goblet cell network (Supplementary table 8). Of these, one gene acts as a regulator in each network. In the Paneth cell network, the CD associated gene *Dbp* regulates *Bik,* which encodes the BCL2 interacting killer, a pro-apoptotic, death promoting protein. In the goblet cell network, the CD associated gene *Notch2* regulates *Notch3* and *Hes1.* Ultimately, by integrating functional annotations obtained through literature searches, we show that the Paneth cell and goblet cell regulatory networks contain genes with direct and indirect associations with IBD. Consequently, these networks and the workflow to reconstruct and analyse them have great potential for the study of IBD pathomechanisms in specific intestinal cell types.

Discussion

By generating and integrating enteroid RNAseq datasets with regulatory networks, we identified master regulators of specific marker genes of related secretory cells, Paneth cells and goblet cells. Furthermore, we highlighted the role of regulatory rewiring with shared regulators having different target genes between the two cell types, raising the possibility that these regulatory networks are responsible for regulating functions relevant in intestinal diseases, such as IBD.

We utilised enteroids (small intestinal organoid cultures) to measure transcriptional signatures of Paneth cells and goblet cells. The use of this model was favoured over the heterogeneous biopsy samples because it provided epithelial cells from one genetic background that could be enriched for Paneth cells or goblet cells (Yin et al., 2014). Recent publications confirmed that within cell-type enriched enteroids, transcriptomic changes are well correlated to *in vivo* gene expression (Mead et al., 2018) and showed that these models are useful for the investigation of health and disease related processes (Jones et al., 2019; Luu et al., 2018; Mead and Karp, 2019). Furthermore, in the future, the described workflow can be extended to test the response of enteroids to external stimuli, such as bacteria, or to grow enteroids from human-derived biopsies, enabling patient-specific experiments.

We showed that the differential expression datasets generated from the Paneth cell and goblet cell enteroids were enriched for marker genes associated with these cell types identified by (Haber et al., 2017). Other epithelial cell types may also be present, for example we found evidence for enteroendocrine cells, but the enteroids still provided greater cell-type specific resolution than using whole tissue biopsy samples. In future, we hope to significantly reduce noise by applying flow cytometry on the enteroids to generate pools of cell types for subsequent sequencing - but this approach is technically more challenging (Jung and Jung, 2016). Using pools of cell types is preferable to single cell sequencing as it mitigates effects of intercellular variation.

By comparing the transcriptional signature of small and stranded RNAs in the conventionally differentiated enteroids to the cell-type enriched enteroids, we identified genes with differential expression relevant to the enriched cell-type. We hypothesise that the larger quantity of DEGs observed in the Paneth cell data represents the more diverse selection of environmental signals integrated by Paneth cells and the lesser role of the goblet cell in the small intestine, given that they are primarily located in the colon (Clevers and Bevins, 2013; Stappenbeck and McGovern, 2017). Using a universal network of non-specific molecular interactions we annotated the DEGs with transcriptional and post-transcriptional regulatory connections. MiRNAs and lncRNAs are included in these networks as they have been shown to perform critical regulatory and mediatory functions in maintaining intestinal homeostasis (Chapman and Pekow, 2015; Farh et al., 2015; Mirza et al., 2015). Nevertheless, the addition of further ‘omics data-types to the described approach could generate a

more holistic view of cellular molecular mechanisms, including the ability to observe post-translational regulation. These networks will not contain every possible regulatory interaction within the cell type of interest but will contain interactions which are likely relevant to cell-type specific functions. For example, whilst regulators do not necessarily show strong co-expression with their targets, where co-expression exists, there is a greater chance that the association is functionally interesting. Therefore, we can use these networks to represent and analyse current biological knowledge as well as to generate hypotheses and guide further research.

Within these regulatory networks we identified cell-type specific markers and their regulators. Given that cell-type specific marker genes should be involved with cell-type specific functions, we predict that regulators of these markers will regulate cell-type specific functions. During this analysis, we identified a collection of regulators which target Paneth cell specific markers in the Paneth cell regulatory networks as well as goblet cell specific markers in the goblet cell regulatory network. This highlighted the presence of regulatory rewiring between the Paneth cell and goblet cell regulatory networks - where a regulator has altered regulatory connections/targets between two conditions, in this case two different types of cell. Downstream rewiring analysis yielded Etv4, let-7e-5p, miR-151-3p, Myb and Rora as the regulators with the most altered targets between the Paneth cell and goblet cell regulatory networks. Functional analysis on the targets of these regulators highlighted an overrepresentation of metabolism associated targets in the Paneth network. This result supports current understanding that Paneth cells rely on high levels of protein and lipid biosynthesis for secretory functions (Cadwell et al., 2008). In addition, recent evidence suggests that Paneth cells play an important role in metabolically supporting stem cells (Rodríguez-Colman et al., 2017). Functional analysis revealed an overrepresentation of cell cycle associated targets in the goblet cell network. As terminally differentiated cells do not usually undergo mitosis, the cell cycle connection could be explained by the presence of many semi-differentiated goblet-like cells, which has previously been observed (Paulus et al., 1993). This analysis highlights apparent redundancy and/or cooperation of regulators which control similar cell-type specific functions and shows the potential importance of regulatory rewiring in the evolution of cell-type specific pathways and functions.

As an extension of the cell-type specific marker analysis, we identified putative cell-type master regulators which control at least 50% of the cell-type specific markers in the Paneth cell and goblet cell regulatory networks. Literature investigation highlighted that many of these regulators, particularly those associated with Paneth cells, have connections to autophagy, inflammation and IBD (Table 1). It is known that Paneth and goblet cell immune-associated secretory functions, which play a major role in gut homeostasis, are highly dependent on the cellular process autophagy (Cadwell et al., 2008; Jones et al., 2019; Patel et al., 2013; Stappenbeck, 2010). For example, it has recently been shown that the release of lysozyme from Paneth cells requires a form of autophagy called secretory autophagy, in which LC3+ vesicles containing lysozyme are routed to the apical surface of the cell for secretion (Bel et al., 2017). This secretory autophagy was shown to be disrupted in mice harbouring a Crohn's disease risk allele in the *ATG16L1* gene, thus providing a link between secretion, inflammation, autophagy and IBD. In addition, dysfunction of secretion and autophagy in Paneth cells and goblet cells has been associated IBD through mechanisms such as reduced intracellular bacterial killing, increased endoplasmic reticulum stress and impaired secretion of mucus from goblet cells (Cadwell et al., 2009; Gersemann et al., 2009; Kaser and Blumberg, 2009; Patel et al., 2013). Furthermore, many IBD susceptibility genes are associated with autophagy function, further highlighting the link between autophagy, secretion and IBD (Lassen and Xavier, 2017). Ultimately, the literature associations with the predicted cell-type specific master regulators, particularly those relating to the Paneth cell, highlight the importance of autophagy, secretion and inflammation in Paneth cell and goblet cell function and suggest that dysregulation of key cell master regulators could lead to IBD.

In addition, we identified CD susceptibility genes in the Paneth cell and goblet cell networks. Two of the identified susceptibility genes act as regulators in the networks: in the Paneth cell network, the protein encoded by *Dbp,* and in the goblet cell network, the protein encoded by *Notch2* (Table S8).In the Paneth cell network *Dbp,* encoding the D site binding protein, regulates *Bik,* which encodes the BCL2 interacting killer, a pro-apoptotic, death promoting protein. Interestingly, rate of apoptosis has been implicated in IBD disease mechanisms (Nunes et al., 2014) and has been associated with IBD drug response (Aghdaei et al., 2018). Therefore this finding highlights a possible regulatory connection between CD susceptibility genes and IBD pathology on a Paneth cell specific level. In the

goblet cell, *Notch2* regulates *Notch3* and *Hes1.* Real *et al*. demonstrated that through Notch signalling, HES1 can inhibit the expression of *NR3C1* (which we predict to be a master regulator in Paneth cells and goblet cells), thus blocking glucocorticoid resistance in T-cell acute lymphoblastic leukaemia (Real et al., 2009). This is relevant to IBD given that glucocorticoids are a common treatment for IBD patients (Prantera and Marconi, 2013) Additionally, the repression of *HATH1* by HES1 via Notch signalling has been previously associated with goblet cell depletion in humans (Zheng et al., 2011). Goblet cell depletion is commonly observed in IBD patients, therefore this observation shows a possible regulatory connection between *Notch2* and IBD. Ultimately, the observation of IBD susceptibility genes in the regulatory networks of these enteroids highlights possible application of this model system to the study disease regulation in specific intestinal cell types, through understanding specific mechanistic pathways. In addition, combined with patient genetic profiles, this approach could help to understand patient specific drug responses and identify new targets for drug actions.

We believe this study is the first comprehensive analysis of miRNA, lncRNA and TF regulatory connections within small intestinal epithelial cells. The intestinal cell-type specific approach, focusing on key secretory cells often implicated in IBD, enables identification of key regulatory mechanisms relevant in gut homeostasis and related diseases. This approach provides a new level of resolution compared to many previous studies utilising heterogeneous cell populations, such as those studying human biopsies. This is particularly pertinent as specific cell types are currently implicated in the pathogenesis of many intestinal diseases, such as IBD (Adolph et al., 2013). Whilst this study has its limitations, the presented workflow has great potential for application in future studies. Such studies can overcome current issues through application of fluorescence-activated cell sorting of the enteroids (as opposed to cell-type enrichment) and through the addition of further -omics datasets such as proteomics. The workflow can also be applied to other conditions, for example for the study of different cell types or on patient-derived organoids. Specifically, the Paneth cell and goblet cell specific data presented here provide potential targets for further study of IBD mechanisms.

In conclusion, we describe an integrative systems biology study to combine enteroid transcriptomics data into cell-type specific regulatory networks, incorporating information on transcriptional and post-transcriptional regulation. We applied this approach, using Paneth cell and goblet cell enriched enteroids, to reconstruct regulatory landscapes and identify probable master regulators of these cell types. Adaptation of this workflow to patient derived organoids, genetic knockout and/or microbially challenged enteroids, alongside appropriate validation, will aid discovery of key regulators and signalling pathways of healthy and disease associated intestinal cell types.

Acknowledgement

The authors are grateful for the helpful discussions to the past and present members and visitors of the Korcsmaros and Wileman groups.

Competing interests

None

Author contributions

EJ and ZM performed the experimental work with the organoids. AT, PS, TW and MO carried out the bioinformatic analysis. PP, LH, TW, JB, FDP, WH and TK designed and supervised the experiments. AT, PS, ZM, EJ, IH and TK wrote the manuscript.

Funding

This work was supported by a fellowship to TK in computational biology at the Earlham Institute (Norwich, UK) in partnership with the Quadram Institute (Norwich, UK), and strategically supported by the Biotechnological and Biosciences Research Council, UK grants (BB/J004529/1, BB/P016774/1 and BB/CSP17270/1). ZM was supported by a PhD studentship from Norwich Medical School. PP was supported by the BBSRC grant BB/J01754X/1. AT and MO were supported by the BBSRC Norwich Research Park Biosciences Doctoral Training Partnership (grant BB/M011216/1). TW and WH were supported by an MRC award (MR/P026028/1). Next-generation sequencing and library construction was delivered via the BBSRC National Capability in Genomics and Single Cell (BB/CCG1720/1) at Earlham Institute by members of the Genomics Pipelines Group.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents and resources may be directed to, and will be fulfilled by the corresponding author Tamas Korcsmaros (tamas.korcsmaros@earlham.ac.uk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animal handling

C57BL/6J mice of both sexes were used for enteroid generation. All animals were maintained in accordance with the Animals (Scientific Procedures) Act 1986 (ASPA).

Small intestinal organoid culture

Murine enteroids were generated as described previously (Jones et al., 2019; Sato and Clevers, 2013; Sato et al., 2009). Briefly, the entire small intestine was opened longitudinally, washed in cold PBS then cut into ~5mm pieces. The intestinal fragments were incubated in 30mM EDTA/PBS for 5 minutes, transferred to PBS for shaking, then returned to EDTA for 5 minutes. This process was repeated until five fractions had been generated. The PBS supernatant fractions were inspected for released crypts. The crypt suspensions were passed through a 70µm filter to remove any villus fragments, then centrifuged at 300xg for 5 minutes. Pellets were resuspended in 200µl phenol-red free Matrigel (Corning), seeded in small domes in 24-well plates and incubated at 37°C for 20 minutes to allow Matrigel to polymerise. Enteroid media containing EGF, Noggin and R-spondin (ENR; (Sato et al., 2009)) was then overlaid. Enteroids were generated from three separate animals for each condition, generating three biological replicates.

To chemically induce differentiation, on days two, five and seven post-crypt isolation, ENR media was changed to include additional factors for each cell-type specific condition: 3µM CHIR99021 (Tocris) and 10µM DAPT (Tocris) [Paneth cells]; 2µM IWP-2 (Tocris) and 10µM DAPT [goblet and enteroendocrine cells] (Yin et al., 2014).

Small intestinal organoid immunofluorescence

On day eight post-crypt isolation, enteroids were fixed with 4% paraformaldehyde (PFA; Sigma-Aldrich) for 1 hour at 4°C prior to permeabilization with 0.1% triton X-100 (Sigma-Aldrich) and incubation in blocking buffer containing 10% goat serum (Sigma-Aldrich). Immunostaining was performed overnight at 4°C using primary antibodies: mouse anti-E-cadherin (BD Transduction Laboratories), rabbit anti-muc2 (Santa Cruz) and rabbit anti-lysozyme (Dako), followed by Alexa Fluor-488 and -594 conjugated secondary antibodies (ThermoFisher Scientific). DNA was stained with DAPI (Molecular Probes). Images were acquired using a fluorescence microscope (Axioimager.M2, equipped with a Plan-Apochromat 63x/1.4 oil immersion objective) and analysed using ImageJ/FIJI V1.51.

METHOD DETAILS

RNA extraction

On day eight post-crypt isolation (allowing optimal cell type-enrichment as previously shown; Yin et al., 2014), enteroids were extracted from Matrigel using Cell Recovery Solution (BD Bioscience),

rinsed in PBS and RNA was extracted using miRCURY RNA Isolation Tissue Kit (Exiqon) according to the manufacturer's protocol.

Stranded RNA library preparation

The enteroid transcriptomics libraries were constructed using the NEXTflex™ Rapid Directional RNA-Seq Kit (5138-07) using the polyA pull down beads from Illumina TruSeq RNA v2 library construction kit (RS-122-2001) with the NEXTflex™ DNA Barcodes – 48 (514104) diluted to 6µM. The library preparation involved an initial QC of the RNA using Qubit DNA (Life technologies Q32854) and RNA (Life technologies Q32852) assays as well as a quality check using the PerkinElmer GX with the RNA assay (PN:CLS960010). In details, 1µg of RNA was purified to extract mRNA with a poly- A pull down using biotin beads, fragmented and first strand cDNA was synthesised. This process reverse transcribes the cleaved RNA fragments primed with random hexamers into first strand cDNA using reverse transcriptase and random primers. The second strand synthesis process removes the RNA template and synthesizes a replacement strand to generate ds cDNA. Directionality is retained by adding dUTP during the second strand synthesis step and subsequent cleavage of the uridine containing strand using Uracil DNA Glycosylase. The ends of the samples were repaired using the 3' to 5' exonuclease activity to remove the 3' overhangs and the polymerase activity to fill in the 5' overhangs creating blunt ends. A single 'A' nucleotide was added to T1S1the 3' ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provided a complementary overhang for ligating the adapter to the fragment. This strategy ensured a low rate of chimera formation. The ligation of a number indexing adapters to the ends of the DNA fragments prepared them for hybridisation onto a flow cell. The ligated products were subjected to a bead-based size selection using Beckman Coulter XP beads (A63880). As well as performing a size selection this process removed the majority of unligated adapters. Prior to hybridisation to the flow cell the samples were amplified to enrich for DNA fragments with adapter molecules on both ends and to amplify the amount of DNA in the library. The strand that was sequenced is the cDNA strand. The insert size of the libraries was verified by running an aliquot of the DNA library on a PerkinElmer GX using the High Sensitivity DNA chip (PerkinElmer CLS760672) and the concentration was determined by using a High Sensitivity Qubit assay and q-PCR. Libraries were then equimolar pooled and checked by qPCR to ensure the libraries had the necessary sequencing adapters ligated.

Small RNA library preparation

The small RNA libraries were made using the TruSeq Small RNA Library Prep Kits, six-base indexes distinguish samples and allow multiplexed sequencing and analysis using 48 unique indexes ((Set A: indexes 1 -12 (RS-200-0012), Set B: indexes 13–24(RS-200-0024), Set C: indexes 25–36 (RS-200-0036), Set D: indices 37–48 (RS-200-0048)) (TruSeq Small RNA Library Prep Kit Reference Guide (15004197 Rev.G)). The TruSeq Small RNA Library Prep Kit protocol is optimised for an input of 1µg of total RNA in 5 µl nuclease-free water or previously isolated microRNA may be used as starting material (minimum of 10–50 ng of purified small RNA). Total RNA is quantified using the Qubit RNA HS Assay kit (ThermoFisher Q32852) and quality of the RNA is established using the Bioanalyzer RNA Nano kit (Agilent Technologies 5067-1511), it is recommended that RNA with an RNA Integrity Number (RIN) value ≥ 8 is used for these libraries as samples with degraded mRNA are also likely to contain degraded small RNA.

This protocol generates small RNA libraries directly from RNA by ligating adapters to each end of the RNA molecule, the 3' adapter is modified to target microRNAs and other small RNAs that have a 3' hydroxyl group resulting from enzymatic cleavage by Dicer or other RNA processing enzymes, whilst most mature microRNAs have a 5' phosphate group. Reverse transcription is used to create cDNA, and subsequent PCR amplification of the cDNA (standard protocol is 14 cycles of PCR) is used to generate libraries. Library purification combines using BluePippin cassettes (Sage Science Pippin Prep 3% Cassettes Dye-Free (CDF3010), set to collection mode range 125-160bp) to extract the library molecules followed by a concentration step (Qiagen MinElute PCR Purification (cat. no. 28004)) to produce libraries ready for sequencing. Library concentration and size are established using HS DNA Qubit and HS DNA Bioanalyser. The resulting libraries were then equimolar pooled and qPCR was performed on the pool prior to clustering.

Stranded RNA sequencing on HiSeq 100PE

The final pool was quantified using a KAPA Library Quant Kit, denatured in NaOH and combined with HT1 plus a 1% PhiX spike at a final running concentration of 10pM.The flow-cell was clustered using HiSeq PE Cluster Kit v3 (Illumina PE-401-3001) utilising the Illumina PE HiSeq Cluster Kit V3 cBot recipe V8.0 method on the Illumina cBot. Following the clustering procedure, the flow-cell was loaded onto the Illumina HiSeq2000 instrument following the manufacturer's instructions with a 101 cycle paired reads and a 7 cycle index read. The sequencing chemistry used was HiSeq SBS Kit v3 (Illumina FC-401-3001) with HiSeq Control Software 2.2.68 and RTA 1.18.66.3. Reads in bcl format were demultiplexed based on the 6bp Illumina index by CASAVA 1.8, allowing for a one base-pair mismatch per library, and converted to FASTQ format by bcl2fastq.

Small RNA sequencing on HiSeq Rapid 50SE

The final pool was quantified using a KAPA Library Quant Kit, denatured in NaOH and combined with HT1 plus a PhiX spike at a final running concentration of 20pM. The libraries were hybridized to the flow-cell using TruSeq Rapid Duo cBot Sample Loading Kit (Illumina CT-403-2001), utilising the Illumina RR_TemplateHyb_FirstExt_VR method on the Illumina cBot. The flow-cell was loaded onto the Illumina HiSeq2500 instrument following the manufacturer's instructions with a 51 cycle single read and a 7 cycle index read. The sequencing chemistry used was HiSeq SBS Rapid Kit v2 (Illumina FC-402-4022) with a single read cluster kit (Illumina GD-402-4002), HiSeq Control Software 2.2.68 and RTA 1.18.66.3. Reads in bcl format were demultiplexed based on the 6bp Illumina index by CASAVA 1.8, allowing for a one base-pair mismatch per library, and converted to FASTQ format by bcl2fastq.

Differentially expressed transcripts

The quality of stranded reads was assessed by FastQC software (version 0.11.4) (Andrews, 2010). All reads coming from technical repeats were concatenated together and aligned (in stranded mode, i.e. with '--rna-strandness RF' flag) using HISAT aligner (version 2.0.5) (Kim et al., 2015). Subsequently, a reference-based de novo transcriptome assembly was carried out for each biological repeat and merged together using StringTie (version 1.3.2) with following parameters: minimum transcript length of 200 nucleotides, minimum FPKM of 0.1 and minimum TPM of 0.1 (Pertea et al., 2015, 2016). Coding potential of each novel transcript was determined with CPC (version 0.9.2) and CPAT (version 1.2.2) (Kong et al., 2007; Wang et al., 2013). From the novel transcripts, only non-coding transcripts (as predicted by both tools) were included in final GTF file. Gene and transcript abundances were estimated with kallisto (version 0.43.0) (Bray et al., 2016). Sleuth (version 0.28.1) R library was used to perform differential gene expression (Pimentel et al., 2017). mRNAs and lncRNAs with an absolute log2 fold change of 1 and q value ≤ 0.05 were considered to be differentially expressed.

The small RNA reads were analysed using the sRNAbench tool within the sRNAtoolbox suite of tools (Rueda et al., 2015) . The barcodes from the 5' end and adapter sequences from the 3' end were removed respectively. Zero mismatches were allowed in detecting the adapter sequences with a minimum adapter length set at 10. Only reads with a minimum length of 16 and a read-count of 2 were considered for further analysis. The mice miRNA collection was downloaded from miRBase version 21 (Kozomara and Griffiths-Jones, 2014). The trimmed and length filtered reads were then mapped to the mature version of the miRBase miRNAs in addition to the annotated version of the mouse genome (version mm10). No mismatches were allowed for the mapping. A seed length of 20 was used for the alignment with a maximum number of multiple mappings set at 10. Read-counts normalised after multiple-mapping were calculated for all the libraries. The multiple-mapping normalised read-counts from the corresponding cell-type enriched enteroids were compared against the conventionally differentiated enteroids to identify differentially expressed miRNAs in a pair-wise manner using edgeR (Robinson et al., 2010). miRNAs with an absolute log2 fold change of 1 and false discovery rate ≤ 0.05 were considered to be differentially expressed.

Enrichment of marker genes

Cell-type specific marker gene lists were obtained a mouse single cell sequencing survey (Haber et al., 2017). The cell-type specific signature genes for Paneth, goblet and enteroendocrine cell types were obtained from the droplet-based and the plate-based methods. Gene symbols were converted to Ensembl gene IDs using bioDBnet db2db (Mudunuri et al., 2009).

Hypergeometric distribution testing was carried out using a custom R script to measure enrichment of cell-type specific marker genes in the differentially upregulated gene sets. To standardise the universal dataset, only markers which are present in the output of the Wald test (genes with variance greater than zero among samples) were used. Similarly, to enable fair comparisons, only differentially expressed protein coding genes and documented lncRNAs were used from the DEG lists, as was surveyed in the cell-type specific marker paper. Bonferroni correction was applied to the hypergeometric distribution p values to account for multiple testing and significance scores were calculated using -log10(corrected p value). For the mapping of marker genes to the interaction networks, no filters were applied.

Reconstruction of molecular networks

Mice regulatory networks containing directed regulatory layers were retrieved from multiple databases (Table S4): miRNA-mRNA (ie., miRNAs regulating mRNAs) and lncRNA-miRNA (ie., lncRNAs regulating miRNAs) interactions were downloaded from TarBase v7.0 (Vlachos et al., 2015) and LncBase v2.0 (Paraskevopoulou et al., 2016), respectively. Only miRNA-mRNA and lncRNA-miRNA interactions determined using HITS-CLIP (Chi et al., 2009) experiments were considered. Regulatory interactions between transcription factors (TFs) and miRNAs (ie. TFs regulating miRNAs) were retrieved from TransmiR v1.2 (Wang et al., 2010), GTRD (Yevshin et al., 2017) and TRRUST v2 (Han et al., 2015, 2018). Co-expression based inferences were ignored from all the above resources. Transcriptional regulatory interactions (ie., TFs regulating target genes) were inferred using data from ORegAnno 3.0 (Lesurf et al., 2016), GTRD and TRUSST. In cases where transcriptional regulatory interactions are derived from high-throughput datasets such as ChIP-seq, we attributed the regulatory interaction elicited by the bound transcription factor to genes which lie within a 10kb window on either side of the ChIP-seq peak (ORegAnno) or meta-cluster (in the case of GTRD). TF-lncRNA interactions (ie., TFs regulating lncRNAs) were also inferred based on the ChIP-seq binding profiles represented by meta-clusters in GTRD. We used only TF-lncRNA interactions within intergenic lncRNAs to avoid assigning false regulatory interactions due to the high number of instances where the lncRNAs overlap with protein-coding genes. In addition, no overlaps were allowed between the coordinates of the ChIP-seq peaks / meta-clusters and any gene annotation. Only if the first annotation feature within a 10kb genomic window downstream to the ChIP-seq peak / meta-cluster was designated as an intergenic lncRNA, a regulatory interaction between the TF and the lncRNA was assigned. Bedtools (Quinlan and Hall, 2010) was used for the custom analyses to look for overlaps between coordinates. All the nodes in the collected interactions were represented by their Ensembl gene IDs for standardization. A summary of the interactions collected from each resource and the quality control criteria applied is given in Table S4.

To generate cell-type specific regulatory networks, interactions in this collated universal network were filtered using the differential expression data (Figure 1). The assumption was made that if both nodes of a particular interaction were expressed in the RNAseq data, the interaction is possible within the cell-type of interest. Furthermore, to filter for the interactions of prime interest, only nodes which were differentially expressed and their associated interactors were included in the cell-type specific networks.

Master regulator analysis

To identify potential master regulators of the Paneth cell and the goblet cell types, the upstream regulators of cell-type specific markers (from (Haber et al., 2017)) were investigated. To do this, all markers were mapped to the relevant networks then subnetworks were extracted consisting of markers and their regulators (Supplementary table 5).

Regulatory rewiring analysis

To calculate rewiring scores for regulators, sub-networks were extracted (from the Paneth cell and goblet cell regulatory networks) containing just the regulator of interest and its downstream targets. For each regulator of interest, the subnetworks from the Paneth cell and goblet cell datasets were compared using the Cytoscape app DyNet (Goenawan et al., 2016; Shannon et al., 2003). The degree corrected D_n score was extracted for each regulator and used to quantify rewiring of the regulator's downstream targets between the Paneth cell and the goblet cell regulatory networks. Functional analysis was carried out on the targets of the top five most rewired regulators. For each regulator, the targets were classified based on whether they are present in only the Paneth cell regulatory network, only the goblet cell regulatory network or in both networks. Each group of targets was tested for functional enrichment (hypergeometric model, q value ≤ 0.1) based on Reactome and KEGG annotations using the ReactomePA R package (Fabregat et al., 2018; Kanehisa et al., 2017; Ogata et al., 1999; Yu and He, 2016) following conversion from mouse to human identifiers using Inparanoid (v8) (O'Brien et al., 2005; Sonnhammer and Östlund, 2015).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters including the exact value of n and statistical significance are reported in the Figures and Figure Legends. n represents the number of enteroid biological replicates generated. Where relevant, data is judged to be statistically significant when Bonferroni corrected p value ≤ 0.01. Genes with absolute log2 fold change of ≥ |1| and false discovery rate ≤ 0.05 were considered to be differentially expressed. Based on principal component analysis of transcript expression, one biological replicate from the Paneth cell enriched enteroids was identified as an outlier and removed (Figure S1).

DATA AND SOFTWARE AVAILABILITY

Small and stranded RNA-seq data has been deposited in the European Nucleotide Archive (ENA) with accession numbers PRJEB32354 and PRJEB32366 respectively.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Matrigel phenol-red free	Corning	356237
DMEM / F12 GlutaMAX (ENR media)	Invitrogen	21885-025
P/S (ENR media)	Fisher	1052882
L-glutamine (ENR media)	Invitrogen	25030-024
N2 (ENR media)	Invitrogen	17502-048
B27 (ENR media)	Invitrogen	12587-010
N-acetylcysteine (ENR media)	Sigma-Aldrich	A9165

Recombinant human R-spondin 1 (ENR media)	VWR	4645-RS
Murine recombinant epidermal growth factor (ENR media)	Invitrogen	PMG8043
Murine recombinant noggin (ENR media)	PeproTech	250-38
CHIR99021	Tocris	4423
DAPT	Tocris	2634/10
IWP-2	Tocris	3533/10
Cell Recovery Solution	BD Bioscience	354253
Paraformaldehyde (4%)	Sigma-Aldrich	P6148
Triton X-100 (1%)	Sigma-Aldrich	X100
Goat serum (10%)	Sigma-Aldrich	G9023
Mouse anti-E-cadherin	BD Transduction Laboratories	610404
Rabbit anti-muc2	Santa Cruz	sc-15334
Rabbit anti-lysozyme	Dako	F 0372
Alexa Fluor-488 conjugated secondary antibodies	ThermoFisher Scientific	A20181
Alexa Fluor-594 conjugated secondary antibodies	ThermoFisher Scientific	A20185
DAPI	Molecular Probes	D1306
NEXTflex™ DNA Barcodes - 48	Perkin Elmer	514104
Beckman Coulter XP beads	Beckman Coulter	A63880
TruSeq Small RNA Library Prep Kits	Illumina	15004197
Critical Commercial Assays		
miRCURY RNA Isolation Tissue Kit	Exiqon	300115
NEXTflex™ Rapid Directional RNA-Seq Kit	Perkin Elmer	5138-07
TruSeq RNA v2 library construction kit	Illumina	RS-122-2001

Qubit RNA HS Assay kit	ThermoFisher	Q32852
Bioanalyzer RNA Nano kit	Agilent Technologies	5067-1511
Pippin Prep 3% Cassettes Dye-Free	Sage Science	CDF3010
MinElute PCR Purification	Qiagen	28004
KAPA Library Quant Kit	Roche	07960140001
HiSeq PE Cluster Kit v3	Illumina	PE-401-3001
HiSeq SBS Kit v3	Illumina	FC-401-3001
TruSeq Rapid Duo cBot Sample Loading Kit	Illumina	CT-403-2001
HiSeq SBS Rapid Kit v2	Illumina	FC-402-4022
HiSeq Rapid Cluster Kit v2	Illumina	GD-402-4002
Experimental Models: Organisms/Strains		
Mouse: C57BL/6J	Charles River	632
Software and Algorithms		
FIJI (1.51)	ImageJ	https://fiji.sc/
HiSeq Control Software (2.2.68)	Illumina	https://emea.illumina.com/systems/sequencing-platforms/hiseq-2500/products-services/hiseq-control-software.html
Real-Time Analysis (RTA) (1.18.66.3)	Illumina	http://emea.support.illumina.com/sequencing/sequencing_software/real-time_analysis_rta/downloads.html
CASAVA (1.8)	Illumina	http://emea.support.illumina.com/downloads/casava_software_version_18_user_guide_(15011196_b).html?langsel=/ch/
FastQC (0.11.4)	(Andrews, 2010)	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
HISAT (2.0.5)	(Kim et al., 2015)	https://ccb.jhu.edu/software/hisat2/index.shtml
StringTie (1.3.2)	(Pertea et al., 2015)	https://ccb.jhu.edu/software/stringtie/
CPC (0.9.2)	(Kong et al., 2007)	http://cpc.cbi.pku.edu.cn
CPAT (1.2.2)	(Wang et al., 2013)	http://code.google.com/p/cpat/

kallisto (0.43.0)	(Bray et al., 2016)	https://pachterlab.github.io/kallisto/download
Sleuth (0.28.1)	(Pimentel et al., 2017)	https://pachterlab.github.io/sleuth/download
sRNAbench	(Rueda et al., 2015)	https://bioinfo2.ugr.es/ceUGR/srnabench/
edgeR	(Robinson et al., 2010)	https://www.bioconductor.org/packages/release/bioc/html/edgeR.html
bioDBnet db2db	(Mudunuri et al., 2009)	https://biodbnet-abcc.ncifcrf.gov/db/db2db.php
Bedtools	(Quinlan and Hall, 2010)	https://bedtools.readthedocs.io/en/latest/
Cytoscape (3.6.1)	(Shannon et al., 2003)	https://cytoscape.org/download.html
DyNet Cytoscape app (1.0.0)	(Goenawan et al., 2016)	http://apps.cytoscape.org/apps/dynet
ReactomePA R package	(Yu and He, 2016)	https://bioconductor.org/packages/release/bioc/html/ReactomePA.html
Inparanoid (v8)	(Sonnhammer and Östlund, 2015)	http://inparanoid.sbc.su.se/cgi-bin/index.cgi
Other		
miRBase (v21)	(Kozomara and Griffiths-Jones, 2014)	http://www.mirbase.org/
Mouse genome (mm10 /Genome Reference Consortium Mouse Build 38)	Genome Reference Consortium	https://www.ensembl.org/Mus_musculus/Info/Annotation
TarBase (v7.0)	(Vlachos et al., 2015)	http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index
LncBase (v2.0)	(Paraskevopoulou et al., 2016)	http://carolina.imis.athena-innovation.gr/diana_tools/web/index.php?r=lncbasev2%2Findex-experimental
TransmiR (v1.2)	(Wang et al., 2010)	http://www.cuilab.cn/transmir
GTRD	(Yevshin et al., 2017)	http://gtrd.biouml.org/
TRRUST (v2)	(Han et al., 2015, 2018)	https://www.grnpedia.org/trrust/
ORegAnno (3.0)	(Lesurf et al., 2016)	http://www.oreganno.org/dump/

Supplemental items

Table S1: Differentially expressed genes from cell-type enriched enteroids vs conventionally differentiated organoids (q value ≤ 0.05 and log2 fold change ≥ |1|). Gene annotations included. Related to Figure 3.

Table S2: Expression and differential expression values for primary cell-type markers. Expression given as mean transcripts per million (TPM) for each enteroid type. Log2 fold changes values given where differential expression criteria passed (q value ≤ 0.05 and log2 fold change ≥ |1|). Control- normally differentiated enteroids; Paneth- Paneth cell enriched enteroids; goblet- goblet cell enriched enteroids.

Table S3: Hypergeometric distribution testing of cell-type specific marker enrichment in upregulated differentially expressed gene lists. Related to Figure 3.

Table S4: A summary of the physical interactions compiled to generate the universal network. Related to reconstruction of molecular networks.

Table S5: Regulator-marker interactions in the Paneth and the goblet subnetworks, including regulator specificity. Related to Figure 5.

Table S6: Rewiring analysis results for the marker regulators present in the Paneth and the goblet marker subnetworks. D_n score generated using Cytoscape app DyNet. Related to Figure 5.

Table S7: Functional enrichment analysis of the top five most rewired (shared) marker regulators (p ≤ 0.1). Related to Figure 5.

Table S8: Crohn's disease associated genes in the cell-type specific regulatory networks.

Figure S1: Principal component analysis of Paneth cell enriched enteroid transcriptomics data from each biological replicate. Related to mapping and identification of differentially expressed transcripts.

Figure S2: Transcript abundances and differential expression of five major cell-type markers. **A:** Mean transcript abundances in the conventionally differentiated, goblet cell enriched and Paneth cell enriched enteroids. **B:** Log2 fold change in the goblet cell enriched enteroid vs conventional enteroid analysis and the Paneth cell enriched enteroid vs conventional enteroid analysis. Data only presented where the differential expression criteria passed (q value ≤ 0.05 and log2 fold change ≥|1|).

Item S1: Cytoscape file for the Paneth cell and goblet cell regulatory networks. Related to Figure 4.

References

Abreu, M.T., Kantorovich, V., Vasiliauskas, E.A., Gruntmanis, U., Matuk, R., Daigle, K., Chen, S., Zehnder, D., Lin, Y.C., Yang, H., et al. (2004). Measurement of vitamin D levels in inflammatory bowel disease patients reveals a subset of Crohn's disease patients with elevated 1,25-dihydroxyvitamin D and low bone mineral density. Gut *53*, 1129–1136.

Adolph, T.E., Tomczak, M.F., Niederreiter, L., Ko, H.-J., Böck, J., Martinez-Naves, E., Glickman, J.N., Tschurtschenthaler, M., Hartwig, J., Hosomi, S., et al. (2013). Paneth cells as a site of origin for intestinal inflammation. Nature *503*, 272–276.

Aghdaei, H.A., Kadijani, A.A., Sorrentino, D., Mirzaei, A., Shahrokh, S., Balaii, H., Geraci, M., and Zali, M.R. (2018). An increased Bax/Bcl-2 ratio in circulating inflammatory cells predicts primary response to infliximab in inflammatory bowel disease patients. United European Gastroenterol. J. *6*, 1074–1081.

Andrews, S. (2010). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

Bakke, D., Lu, R., Agrawal, A., Zhang, Y., and Sun, J. (2018). 18 myeloid vitamin d receptor signaling regulates paneth cell function and intestinal homeostasis. Gastroenterology *154*, S41.

Balfe, A., Lennon, G., Lavelle, A., Docherty, N.G., Coffey, J.C., Sheahan, K., Winter, D.C., and O'Connell, P.R. (2018). Isolation and gene expression profiling of intestinal epithelial cells: crypt isolation by calcium chelation from in vivo samples. Clin. Exp. Gastroenterol. *11*, 29–37.

Bel, S., Pendse, M., Wang, Y., Li, Y., Ruhn, K.A., Hassell, B., Leal, T., Winter, S.E., Xavier, R.J., and Hooper, L.V. (2017). Paneth cells secrete lysozyme via secretory autophagy during bacterial infection of the intestine. Science *357*, 1047–1052.

Bevins, C.L., and Salzman, N.H. (2011). Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. Nat. Rev. Microbiol. *9*, 356–368.

Bovolenta, L.A., Acencio, M.L., and Lemke, N. (2012). HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. BMC Genomics *13*, 405.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Cadwell, K., Liu, J.Y., Brown, S.L., Miyoshi, H., Loh, J., Lennerz, J.K., Kishi, C., Kc, W., Carrero, J.A., Hunt, S., et al. (2008). A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. Nature *456*, 259–263.

Cadwell, K., Patel, K.K., Komatsu, M., Virgin, H.W., and Stappenbeck, T.S. (2009). A common role for Atg16L1, Atg5 and Atg7 in small intestinal Paneth cells and Crohn disease. Autophagy *5*, 250–252.

Chandra, V., Bhagyaraj, E., Nanduri, R., Ahuja, N., and Gupta, P. (2015). NR1D1 ameliorates Mycobacterium tuberculosis clearance through regulation of autophagy. Autophagy *11*, 1987–1997.

Chapman, C.G., and Pekow, J. (2015). The emerging role of miRNAs in inflammatory bowel disease: a review. Therap. Adv. Gastroenterol. *8*, 4–22.

Chi, S.W., Zang, J.B., Mele, A., and Darnell, R.B. (2009). Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature *460*, 479–486.

Clevers, H.C., and Bevins, C.L. (2013). Paneth cells: maestros of the small intestinal crypts. Annu. Rev. Physiol. *75*, 289–311.

Crosnier, C., Stamataki, D., and Lewis, J. (2006). Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. Nat. Rev. Genet. *7*, 349–359.

De Iudicibus, S., Franca, R., Martelossi, S., Ventura, A., and Decorti, G. (2011). Molecular mechanism of glucocorticoid resistance in inflammatory bowel disease. World J. Gastroenterol. *17*, 1095–1108.

Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kühn, H., Behrendt, A., et al. (2016). ELM 2016--data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res. *44*, D294-300.

Dubois-Camacho, K., Ottum, P.A., Franco-Muñoz, D., De la Fuente, M., Torres-Riquelme, A., Díaz-Jiménez, D., Olivares-Morales, M., Astudillo, G., Quera, R., and Hermoso, M.A. (2017). Glucocorticosteroid therapy in inflammatory bowel diseases: From clinical practice to molecular biology. World J. Gastroenterol. *23*, 6628–6638.

Duerkop, B.A., Vaishnava, S., and Hooper, L.V. (2009). Immune responses to the microbiota at the intestinal mucosal surface. Immunity *31*, 368–376.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The reactome pathway knowledgebase. Nucleic Acids Res. *46*, D649–D655.

Farh, K.K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J.H., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

Gerbe, F., and Jay, P. (2016). Intestinal tuft cells: epithelial sentinels linking luminal cues to the immune system. Mucosal Immunol. *9*, 1353–1359.

Gersemann, M., Becker, S., Kübler, I., Koslowski, M., Wang, G., Herrlinger, K.R., Griger, J., Fritz, P., Fellermann, K., Schwab, M., et al. (2009). Differences in goblet cell differentiation between Crohn's disease and ulcerative colitis. Differentiation. *77*, 84–94.

Gersemann, M., Stange, E.F., and Wehkamp, J. (2011). From intestinal stem cells to inflammatory bowel diseases. World J. Gastroenterol. *17*, 3198–3203.

Goenawan, I.H., Bryan, K., and Lynn, D.J. (2016). DyNet: visualization and analysis of dynamic molecular interaction networks. Bioinformatics *32*, 2713–2715.

Gombart, A.F., Borregaard, N., and Koeffler, H.P. (2005). Human cathelicidin antimicrobial peptide (CAMP) gene is a direct target of the vitamin D receptor and is strongly up-regulated in myeloid cells by 1,25-dihydroxyvitamin D3. FASEB J. *19*, 1067–1077.

Grenningloh, R., Kang, B.Y., and Ho, I.-C. (2005). Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses. J. Exp. Med. *201*, 615–626.

Haber, A.L., Biton, M., Rogel, N., Herbst, R.H., Shekhar, K., Smillie, C., Burgin, G., Delorey, T.M., Howitt, M.R., Katz, Y., et al. (2017). A single-cell survey of the small intestinal epithelium. Nature *551*, 333–339.

Han, H., Shim, H., Shin, D., Shim, J.E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., et al. (2015). TRRUST: a reference database of human transcriptional regulatory interactions. Sci. Rep. *5*, 11432.

Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., et al. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res. *46*, D380–D386.

Han, P., Gopalakrishnan, C., Yu, H., and Wang, E. (2017). Gene Regulatory Network Rewiring in the Immune Cells Associated with Cancer. Genes (Basel) *8*.

Jones, E.J., Matthews, Z.J., Gul, L., Sudhakar, P., Treveil, A., Divekar, D., Buck, J., Wrzesinski, T., Jefferson, M., Armstrong, S.D., et al. (2019). Integrative analysis of Paneth cell proteomic and transcriptomic data from intestinal organoids reveals functional processes dependent on autophagy.

Dis. Model. Mech.

Jostins, L., Ripke, S., Weersma, R.K., Duerr, R.H., McGovern, D.P., Hui, K.Y., Lee, J.C., Schumm, L.P., Sharma, Y., Anderson, C.A., et al. (2012). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature *491*, 119–124.

Jung, J., and Jung, H. (2016). Methods to analyze cell type-specific gene expression profiles from heterogeneous cell populations. Animal Cells Syst (Seoul) *20*, 113–117.

Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361.

Kaser, A., and Blumberg, R.S. (2009). Endoplasmic reticulum stress in the intestinal epithelium and inflammatory bowel disease. Semin. Immunol. *21*, 156–163.

Kaser, A., Lee, A.-H., Franke, A., Glickman, J.N., Zeissig, S., Tilg, H., Nieuwenhuis, E.E.S., Higgins, D.E., Schreiber, S., Glimcher, L.H., et al. (2008). XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. Cell *134*, 743–756.

Kaser, A., Flak, M.B., Tomczak, M.F., and Blumberg, R.S. (2011). The unfolded protein response and its role in intestinal homeostasis and inflammation. Exp. Cell Res. *317*, 2772–2779.

Kim, Y.S., and Ho, S.B. (2010). Intestinal goblet cells and mucins in health and disease: recent insights and progress. Curr. Gastroenterol. Rep. *12*, 319–330.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12*, 357–360.

Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. *35*, W345-9.

Konno, S., Iizuka, M., Yukawa, M., Sasaki, K., Sato, A., Horie, Y., Nanjo, H., Fukushima, T., and Watanabe, S. (2004). Altered expression of angiogenic factors in the VEGF-Ets-1 cascades in inflammatory bowel disease. J. Gastroenterol. *39*, 931–939.

Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. *42*, D68-73.

Lassen, K.G., and Xavier, R.J. (2017). Genetic control of autophagy underlies pathogenesis of inflammatory bowel disease. Mucosal Immunol. *10*, 589–597.

Leach, M.R., and Williams, D.B. (2013). Calnexin and Calreticulin, Molecular Chaperones of the Endoplasmic Reticulum - Madame Curie Bioscience Database - NCBI Bookshelf.

Lesurf, R., Cotto, K.C., Wang, G., Griffith, M., Kasaian, K., Jones, S.J.M., Montgomery, S.B., Griffith, O.L., and Open Regulatory Annotation Consortium (2016). ORegAnno 3.0: a community-driven resource for curated regulatory annotation. Nucleic Acids Res. *44*, D126-32.

Li, D., Haritunians, T., Potdar, A., and McGovern, D.P.B. (2018). 16 Genetic analysis identified novel loci associated with IBD. Inflamm. Bowel Dis. *24*, S14–S14.

Liu, T.-C., Gurram, B., Baldridge, M.T., Head, R., Lam, V., Luo, C., Cao, Y., Simpson, P., Hayward, M., Holtz, M.L., et al. (2016). Paneth cell defects in Crohn's disease patients promote dysbiosis. JCI Insight *1*, e86907.

Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A., and Gerstein, M. (2004). Genomic analysis of regulatory network dynamics reveals large topological changes. Nature *431*, 308–312.

Luu, L., Matthews, Z.J., Armstrong, S.D., Powell, P.P., Wileman, T., Wastling, J.M., and Coombes, J.L. (2018). Proteomic Profiling of Enteroid Cultures Skewed toward Development of Specific

Epithelial Lineages. Proteomics *18*, e1800132.

Mead, B.E., and Karp, J.M. (2019). All models are wrong, but some organoids may be useful. Genome Biol. *20*, 66.

Mead, B.E., Ordovas-Montanes, J., Braun, A.P., Levy, L.E., Bhargava, P., Szucs, M.J., Ammendolia, D.A., MacMullan, M.A., Yin, X., Hughes, T.K., et al. (2018). Harnessing single-cell genomics to improve the physiological fidelity of organoid-derived cell types. BMC Biol. *16*, 62.

Mirza, A.H., Berthelsen, C.H., Seemann, S.E., Pan, X., Frederiksen, K.S., Vilien, M., Gorodkin, J., and Pociot, F. (2015). Transcriptomic landscape of lncRNAs in inflammatory bowel disease. Genome Med. *7*, 39.

Mokry, M., Middendorp, S., Wiegerinck, C.L., Witte, M., Teunissen, H., Meddens, C.A., Cuppen, E., Clevers, H., and Nieuwenhuis, E.E.S. (2014). Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium. Gastroenterology *146*, 1040–1047.

Mouly, E., Chemin, K., Nguyen, H.V., Chopin, M., Mesnard, L., Leite-de-Moraes, M., Burlen-defranoux, O., Bandeira, A., and Bories, J.-C. (2010). The Ets-1 transcription factor controls the development and function of natural regulatory T cells. J. Exp. Med. *207*, 2113–2125.

Mudunuri, U., Che, A., Yi, M., and Stephens, R.M. (2009). bioDBnet: the biological database network. Bioinformatics *25*, 555–556.

Nunes, T., Bernardazzi, C., and de Souza, H.S. (2014). Cell death and inflammatory bowel diseases: apoptosis, necrosis, and autophagy in the intestinal epithelium. Biomed Res. Int. *2014*, 218493.

Oakley, R.H., and Cidlowski, J.A. (2013). The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. J. Allergy Clin. Immunol. *132*, 1033–1044.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., and Kanehisa, M. (1999). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *27*, 29–34.

Okamoto, R., and Watanabe, M. (2016). Role of epithelial cells in the pathogenesis and treatment of inflammatory bowel disease. J. Gastroenterol. *51*, 11–21.

Okumura, R., and Takeda, K. (2017). Roles of intestinal epithelial cells in the maintenance of gut homeostasis. Exp. Mol. Med. *49*, e338.

O'Brien, K.P., Remm, M., and Sonnhammer, E.L.L. (2005). Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. *33*, D476-80.

Paraskevopoulou, M.D., Vlachos, I.S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., Zagganas, K., Tsanakas, P., Floros, E., Dalamagas, T., et al. (2016). DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. Nucleic Acids Res. *44*, D231-8.

Patel, K.K., Miyoshi, H., Beatty, W.L., Head, R.D., Malvin, N.P., Cadwell, K., Guan, J.-L., Saitoh, T., Akira, S., Seglen, P.O., et al. (2013). Autophagy proteins control goblet cell function by potentiating reactive oxygen species production. EMBO J. *32*, 3130–3144.

Paulus, U., Loeffler, M., Zeidler, J., Owen, G., and Potten, C.S. (1993). The differentiation and lineage development of goblet cells in the murine small intestinal crypt: experimental and modelling studies. J. Cell Sci. *106 (Pt 2)*, 473–483.

Peck, B.C.E., Mah, A.T., Pitman, W.A., Ding, S., Lund, P.K., and Sethupathy, P. (2017). Functional transcriptomics in diverse intestinal epithelial cell types reveals robust microrna sensitivity in intestinal stem cells to microbial status. J. Biol. Chem. *292*, 2586–2600.

Pei, F.H., Wang, Y.J., Gao, S.L., Liu, B.R., Du, Y.J., Liu, W., Yu, H.Y., Zhao, L.X., and Chi, B.R. (2011). Vitamin D receptor gene polymorphism and ulcerative colitis susceptibility in Han Chinese. J.

Dig. Dis. *12*, 90–98.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. *11*, 1650–1667.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. Nat. Methods *14*, 687–690.

Prantera, C., and Marconi, S. (2013). Glucocorticosteroids in the treatment of inflammatory bowel disease and approaches to minimizing systemic activity. Therap. Adv. Gastroenterol. *6*, 137–156.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Real, P.J., Tosello, V., Palomero, T., Castillo, M., Hernando, E., de Stanchina, E., Sulis, M.L., Barnes, K., Sawai, C., Homminga, I., et al. (2009). Gamma-secretase inhibitors reverse glucocorticoid resistance in T cell acute lymphoblastic leukemia. Nat. Med. *15*, 50–58.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rodríguez-Colman, M.J., Schewe, M., Meerlo, M., Stigter, E., Gerrits, J., Pras-Raves, M., Sacchetti, A., Hornsveld, M., Oost, K.C., Snippert, H.J., et al. (2017). Interplay between metabolic identities in the intestinal crypt supports stem cell function. Nature *543*, 424–427.

Rueda, A., Barturen, G., Lebrón, R., Gómez-Martín, C., Alganza, Á., Oliver, J.L., and Hackenberg, M. (2015). sRNAtoolbox: an integrated collection of small RNA research tools. Nucleic Acids Res. *43*, W467-73.

Rutgeerts, P. (1998). The use of oral topically acting glucocorticosteroids in the treatment of inflammatory bowel disease. Mediators Inflamm. *7*, 137–140.

Sato, T., and Clevers, H. (2013). Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. Science *340*, 1190–1194.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., et al. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Sonnhammer, E.L.L., and Östlund, G. (2015). InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. *43*, D234-9.

Stappenbeck, T.S. (2010). The role of autophagy in Paneth cell differentiation and secretion. Mucosal Immunol. *3*, 8–10.

Stappenbeck, T.S., and McGovern, D.P.B. (2017). Paneth Cell Alterations in the Development and Phenotype of Crohn's Disease. Gastroenterology *152*, 322–326.

Vanuytsel, T., Senger, S., Fasano, A., and Shea-Donohue, T. (2013). Major signaling pathways in intestinal stem cells. Biochim. Biophys. Acta *1830*, 2410–2426.

Vlachos, I.S., Paraskevopoulou, M.D., Karagkouni, D., Georgakilas, G., Vergoulis, T., Kanellos, I., Anastasopoulos, I.-L., Maniou, S., Karathanou, K., Kalfakakou, D., et al. (2015). DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. Nucleic Acids

Res. *43*, D153-9.

Wang, J., Lu, M., Qiu, C., and Cui, Q. (2010). TransmiR: a transcription factor-microRNA regulation database. Nucleic Acids Res. *38*, D119-22.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. *41*, e74.

Wang, T.-T., Nestel, F.P., Bourdeau, V., Nagai, Y., Wang, Q., Liao, J., Tavera-Mendoza, L., Lin, R., Hanrahan, J.W., Mader, S., et al. (2004). Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression. J. Immunol. *173*, 2909–2912.

Wu, S., Zhang, Y.-G., Lu, R., Xia, Y., Zhou, D., Petrof, E.O., Claud, E.C., Chen, D., Chang, E.B., Carmeliet, G., et al. (2015). Intestinal epithelial vitamin D receptor deletion leads to defective autophagy in colitis. Gut *64*, 1082–1094.

Yemelyanov, A., Czwornog, J., Chebotaev, D., Karseladze, A., Kulevitch, E., Yang, X., and Budunova, I. (2007). Tumor suppressor activity of glucocorticoid receptor in the prostate. Oncogene *26*, 1885–1896.

Yevshin, I., Sharipov, R., Valeev, T., Kel, A., and Kolpakov, F. (2017). GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. Nucleic Acids Res. *45*, D61–D67.

Yin, X., Farin, H.F., van Es, J.H., Clevers, H., Langer, R., and Karp, J.M. (2014). Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny. Nat. Methods *11*, 106–112.

Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol. Biosyst. *12*, 477–479.

Zhang, K., Hornef, M.W., and Dupont, A. (2015). The intestinal epithelium as guardian of gut barrier integrity. Cell. Microbiol. *17*, 1561–1569.

Zheng, X., Tsuchiya, K., Okamoto, R., Iwasaki, M., Kano, Y., Sakamoto, N., Nakamura, T., and Watanabe, M. (2011). Suppression of hath1 gene expression directly regulated by hes1 via notch signaling is associated with goblet cell depletion in ulcerative colitis. Inflamm. Bowel Dis. *17*, 2251–2260.