Regulatory network analysis of Paneth cell and goblet cell enriched gut organoids using transcriptomics approaches

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

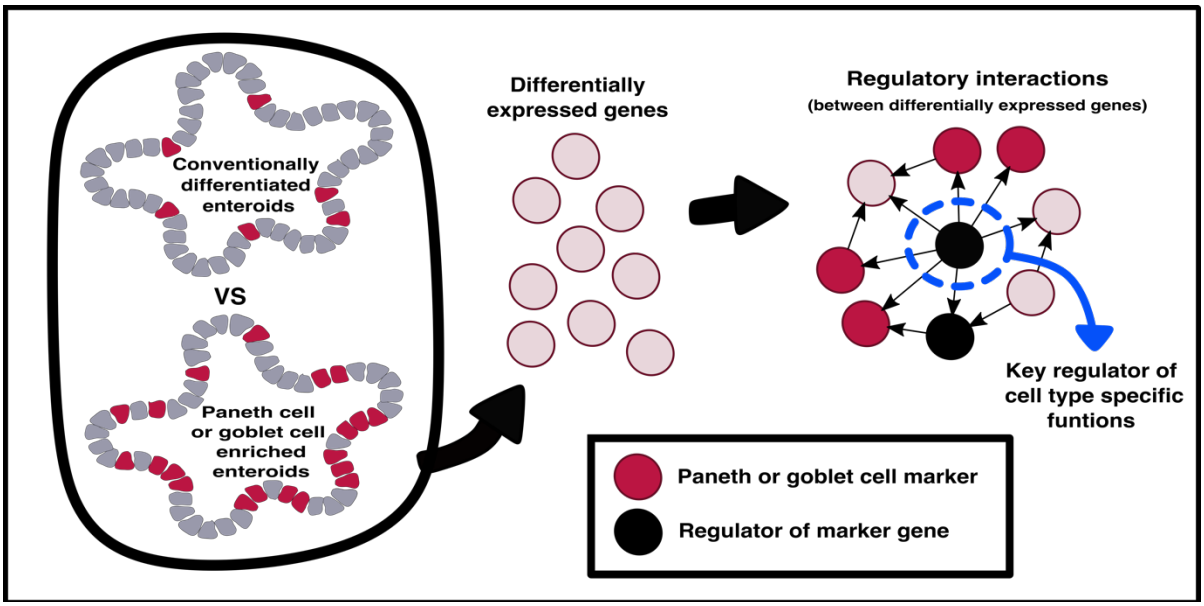
Abstract

The epithelial lining of the small intestine consists of multiple cell types, including Paneth cells and goblet cells, that work in cohort to maintain gut health. 3D *in vitro* cultures of human primary epithelial cells, called organoids, have become a key model to study the functions of Paneth cells and goblet cells in normal and diseased conditions. Advances in these models include the ability to skew differentiation to particular lineages, providing a useful tool to study cell type specific function/dysfunction in the context of the epithelium. Here, we use comprehensive profiling of mRNA, microRNA and long non-coding RNA expression to confirm that Paneth cell and goblet cell enrichment of murine small intestinal organoids (enteroids) establishes a physiologically accurate model. We employ network analysis to infer the regulatory landscape altered by skewing differentiation, and using knowledge of cell type specific markers, we predict key regulators of cell type specific functions: Cebpa, Jun, Nr1d1 and Rxra specific to Paneth cells, Gfi1b and Myc specific for goblet cells and Ets1, Nr3c1 and Vdr shared between them. Links identified between these regulators and cellular phenotypes of inflammatory bowel disease (IBD) suggest that global regulatory rewiring during or after differentiation of Paneth cells and goblet cells could contribute to IBD aetiology. Future application of cell type enriched enteroids combined with the presented computational workflow can be used to disentangle multifactorial mechanisms of these cell types and propose regulators whose pharmacological targeting could be advantageous in treating IBD patients with Crohn's disease or ulcerative colitis.

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

Table of contents

We demonstrate the application of network biology techniques to increase understanding of intestinal dysbiosis through studying transcriptomics data from Paneth and goblet cell enriched enteroids.



Introduction

Gut barrier integrity is critically important for efficient nutrient absorption and maintenance of intestinal homeostasis (1) and is maintained by the combined action of the various cell types lining the intestinal epithelium (2). These intestinal epithelial cells serve to mediate signals between the gut microbiota and the host innate/adaptive immune systems (3,4). 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) (5). Therefore, a greater insight into 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 (6). 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 (7). Both of these cell populations have documented roles in gut-related diseases (8,9). Dysfunctional Paneth cells with reduced secretion of anti-microbial peptides have been shown to contribute to the pathogenesis of CD (10), while reduction in goblet cell numbers and defective goblet cell function has been associated with UC in humans (11).

Recent studies have employed single cell transcriptomics sequencing of tissue samples to characterise the proportion and signatures of different epithelial cell types in the intestines of healthy and IBD patients (12–14). However, to provide deeper insights into the role of specific cell populations (such as Paneth cells and goblet cells) in IBD, *in vitro* models are required for in depth testing and manipulation. Such models can be used to study specific mechanisms of action, host-microbe interactions, intercellular communication, patient specific therapeutic responses and to develop new diagnostic approaches. Due to ease of manipulation, observation and analysis, organoid models, including small intestinal models (enteroids), are increasingly used in the IBD field (15–17). Enteroid cell culture systems employ growth factors to expand and differentiate Lgr5+ stem cells into spherical models of the small intestinal epithelia which recapitulate features of the *in vivo* intestinal tissue (18–20). It has been shown that these enteroids contain all the major cell types of the intestinal epithelium, and exhibit normal *in vivo* functions (21). These models, generated using intestinal tissue from mice, from human patient biopsies or from induced pluripotent stem cells (iPSCs), have proven particularly valuable for the study of complex diseases which lack other realistic models and exhibit large patient variability, such as IBD (22,23). Small molecule treatments have been developed that skew the differentiation of enteroids towards Paneth cell or goblet cell lineages, improving representation of these cells within the enteroid cell population (24,25). Specifically, differentiation can be directed towards the Paneth cell lineage through the addition of DAPT, which inhibits notch signaling, and CHIR99021, which inhibits GSK3β-mediated β-catenin degradation. Enteroid cultures enriched in goblet cells can be generated through the addition of DAPT and IWP-2, an inhibitor of Wnt signaling (25). Whilst these methods do not present single cell type resolution, they provide useful tools to study Paneth cell and goblet cell populations in the context of the other major epithelial cell types (26). A recent study by Mead et al. found that Paneth cells from enriched enteroids more closely represent their *in vivo* counterparts than those isolated from conventionally differentiated enteroids, based on transcriptomics, proteomics and morphologic data (27). Furthermore, we have shown that enteroids enriched for Paneth cells and goblet cells recapitulate *in vivo* characteristics on the proteomics level (28,29) and that they are a useful tool for the investigation of health and disease related processes in specific intestinal cell types (29).

Nevertheless, the effect of Paneth cell and goblet cell enrichment of enteroids on key regulatory landscapes has not been extensively characterised. In this study, we comprehensively profiled mRNAs, miRNAs and lncRNAs from mouse derived, 3D conventionally differentiated enteroids (control), Paneth cell enriched enteroids (PCeE) and goblet cell enriched enteroids (GCeE) to determine the extent to which these enteroids display increased Paneth cell and goblet cell signatures. We applied a systems-level analysis of regulatory interactions within the PCeEs and GCeEs to further characterise the effect of cell type enrichment and to predict key molecular regulators involved with Paneth cell and goblet cell specific functions. This analysis was carried out using interactions networks, which are a primary method to collate, visualise and analyse biological systems. These networks are a type of systems biology data representation, which aids the interpretation of -omics read-outs by contextualising genes/molecules of interest and identifying relevant signalling and regulatory pathways (30,31). In the

presented analysis, the nodes of the interaction networks represent genes/molecules of interest from the transcriptomics data and the edges represent regulatory connections (molecular interactions) between the nodes inferred from databases. Studying regulatory interactions using interaction networks has been proved useful to uncover how cells respond to changing environments at a transcriptional level, to prioritise drug targets and to investigate the downstream effects of gene mutations and knockouts (32–34).

We used network approaches to interpret the PCeE and GCeE transcriptomic data by integrating directed regulatory connections from resources containing transcriptional and post-transcriptional interactions. This integrative strategy led us to define regulatory network landscapes altered by Paneth cell and goblet cell enrichment of enteroids, termed as PCeE network and GCeE network, respectively (Figure 1). By incorporating known Paneth cell and goblet cell markers, we used these networks to predict master regulators of Paneth cell and goblet cell differentiation and/or maintenance in the enriched enteroids. Furthermore, we highlighted varying downstream actions of shared regulators between the cell types. This phenomenon, called regulatory rewiring, highlights the importance of changes in regulatory connections in the function and differentiation of specific cell types. Taken together, we show that cell type enriched enteroids combined with the presented network biology workflow have potential for application to the study of epithelial dysfunction and mechanisms of action of multifactorial diseases such as IBD.

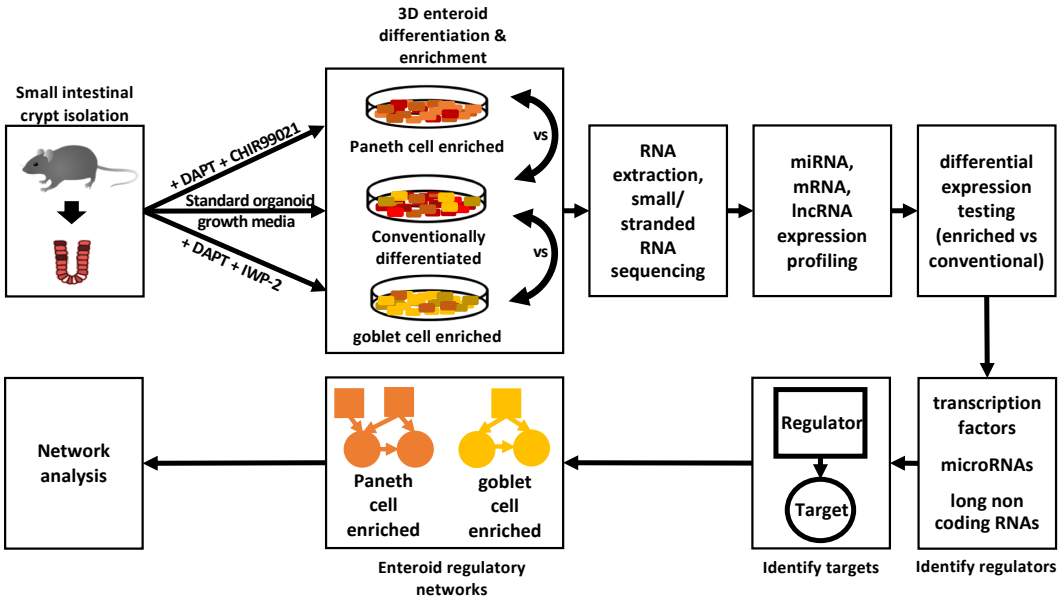


Figure 1: Schematic representation of the workflow to infer regulatory network landscapes altered by Paneth cell and goblet cell enrichment of enteroids *PCeE/GCeE network - Paneth cell enriched enteroid / goblet cell enriched enteroid network; TF- transcription factor; lncRNA- long non-coding RNA; miRNA- microRNA; mRNA- messenger RNA.*

Results

Secretory lineages are over-represented in cell type enriched enteroids compared to conventionally differentiated controls

We generated 3D self-organising enteroid cultures *in vitro* from murine small intestinal crypts (18–20). In addition to conventionally differentiated enteroids, we generated enteroids enriched for Paneth cells and goblet cells using well-established and published protocols, presented in detail in the Methods (20,25). Bulk transcriptomics data was obtained from each set of enteroids to determine genes with differential expression resulting from enteroid skewing protocols. 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 (absolute log2 fold change ≥ 1 and false discovery rate ≤ 0.05)) in the PCeE dataset, and 2,889 were differentially expressed in the GCeE dataset (Figure 2A, Table S1). The larger number of differentially expressed genes (DEGs) in the PCeE data could be attributed to the highly specialised nature of Paneth cells (35,36). The majority of the DEGs were annotated as protein coding: 79% in the PCeE dataset and 84% in the GCeE dataset. In addition, we identified lncRNAs (PCeE, 11%; GCeE, 9%) and miRNAs (PCeE, 4%; GCeE, 2%) among the DEGs (Figure 2A). Some of these DEGs were identified in both the PCeE and the GCeE 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 PCeE and the GCeE data, while 442 genes were found downregulated in both datasets. This result highlights considerable overlap between the results of skewing enteroids towards Paneth cells and goblet cells, and can be explained by the shared differentiation history and secretory function of both Paneth cells and goblet cells.

Pathway analysis was employed to study functional associations of the DEGs (Figure 2B). The PCeE-specific DEGs were associated with a number of metabolic pathways, including Metabolism of vitamins and cofactors, Pyruvate metabolism and Citric Acid (TCA) cycle and Cholesterol biosynthesis. On the other hand, GCeE-specific DEGs were associated with the cell cycle through pathways such as Cell Cycle Checkpoints, DNA replication and G1/S transition. Pathways associated with the shared DEGs included Transmission across Chemical Synapses, Integration of energy metabolism and a number of pathways linked to hormones. As hormone functions are characteristic of enteroendocrine cells, this analysis suggests that enteroendocrine cells are enriched in both the PCeEs and the GCeEs.

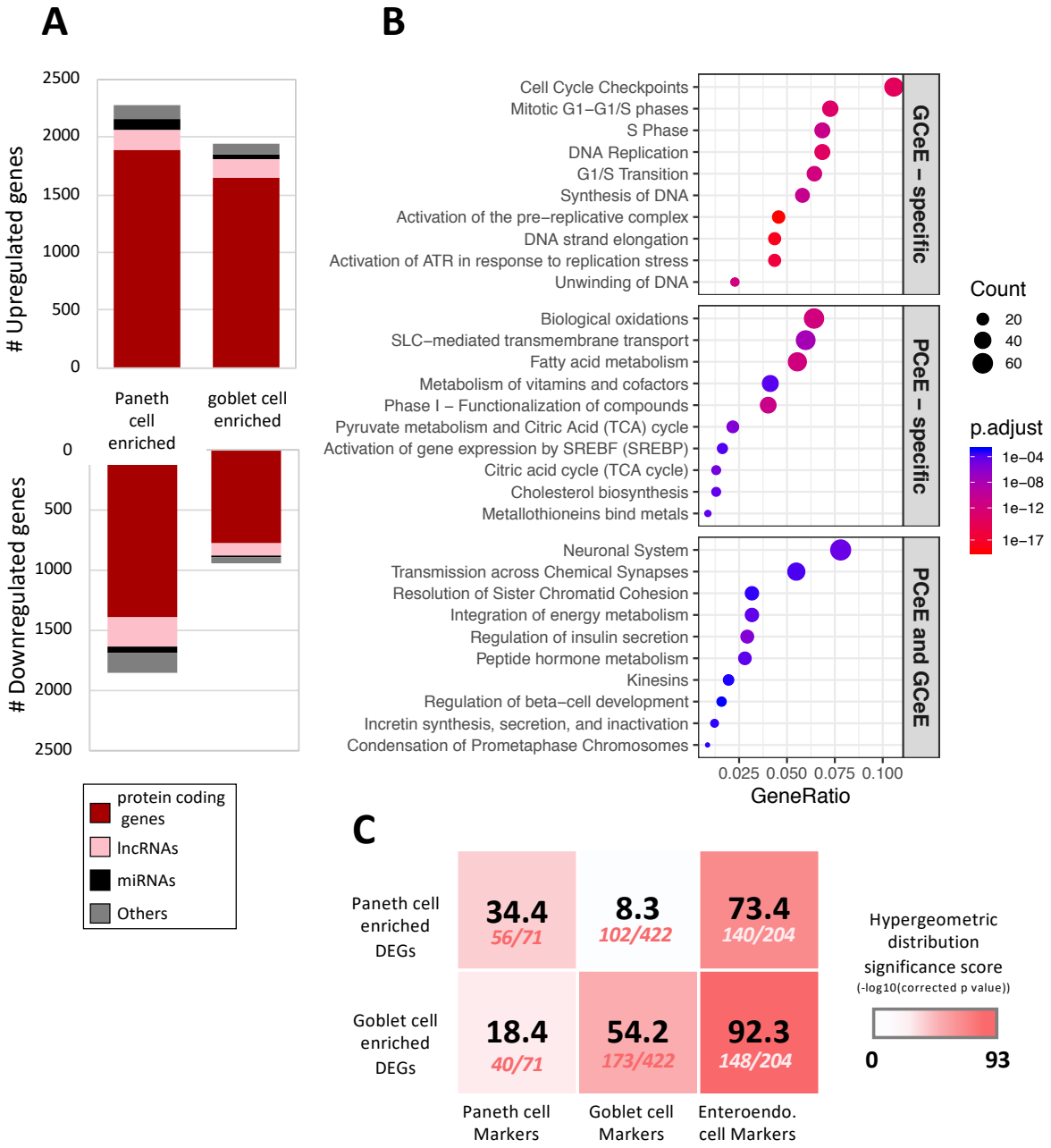


Figure 2: Differentially expressed genes in Paneth cell enriched enteroids and goblet cell enriched enteroids (compared to conventionally differentiated enteroids). ***(A)*** *Number of differentially expressed genes (DEGs). miRNA = microRNA; lncRNA = long non-coding RNA; Genes annotated as 'other' include pseudogenes and antisense genes.* ***(B)*** *Top 10 Reactome pathways of the 50 most significantly differentially expressed genes (by FDR).* ***(C)*** *Enrichment of cell type specific marker genes in the DEG lists. Higher significance scores indicate greater enrichment. Number of markers in DEG list out of the total number of markers shown below significance score. Also see Table S1,4.*

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 investigated markers: *Lgr5* (stem cells), *ChgA* (enteroendocrine cells), *Muc2* (goblet cells), *Lyz1* (Paneth cells) and *Vil1* (epithelial cells). We observed an upregulation of *Muc2, Lyz1* and *ChgA* and a downregulation of *Lgr5* in PCeEs and GCeEs compared to the control enteroids, confirming the more pronounced differentiated status of the enteroids. In addition a number

of Paneth cell specific antimicrobial peptide genes were differentially expressed in the PCeE dataset, including *Ang4, Reg3γ, Pla2g2a and Defa2* (Table S3). Some of these genes were also differentially expressed in the GCeE dataset but with smaller log fold change values, e.g. *Lyz1* and *Ang4*. Conversely, a number of goblet cell mucin related genes (including *Muc2* and *Tff3*) were differentially expressed in both datasets although all genes exhibited a smaller increase in the PCeEs (Table S3). Therefore, using primary cell type specific markers, antimicrobial peptide genes and mucin-related genes, we show that the enteroids contain all major cell types, and that Paneth cell are most upregulated in the PCeEs, while goblet cells are most upregulated in the GCeEs. We also note that both differentiation methods resulted in increases of other secretory cell types as well.

To further investigate secretory cell type specific signatures of the enteroids, we measured enrichment of secretory lineages in the upregulated DEG lists using additional marker genes of Paneth cells, goblet cells and enteroendocrine cells. While all tests were significant (hypergeometric model, q value ≤ 0.05), we identified greater enrichment of Paneth cell markers in the PCeE DEG list and goblet markers in the GCeE DEG list (Figure 2B). This confirms that both enteroid enrichment protocols were successful in increasing the proportion of their target cell type, but also increased proportions of other secretory lineages, albeit to a lesser extent (Table S4). This observation confirms previous studies that these enteroid differentiation protocols result in enteroendocrine enrichment in addition to Paneth cell and goblet cell enrichment (25,28).

In conclusion, we have used image-based validation, pathway analysis and marker gene investigation to show successful enrichment of target cell types in the PCeE's and GCeE's. We also highlighted an additional increase in other secretory lineages, particularly enteroendocrine cells, as a result of both enrichment protocols.

Reconstruction of regulatory networks altered by enteroid differentiation skewing

To gain an understanding of the regulatory changes occuring when enteroids development is skewed, we applied a network biology approach, and identified regulator-target relationships within the DEG lists. We generated a large 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) consisted 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 the current biological context. In order to get a clearer and valid picture of regulatory interactions occurring in our enteroids, we used the universal network to annotate the PCeE DEGs and GCeE DEGs with regulatory connections. Combining these connections, we generated specific regulatory networks for PCeEs and GCeEs.

In total, the PCeE network, generated using differential expression data from the PCeEs compared to the conventionally differentiated enteroids, contained 37,062 interactions connecting 208 unique regulators with 3,023 unique targets (Figure 3, Item S1). The GCeE network, generated using differential expression data from the GCeEs compared to the conventionally differentiated enteroids, contained 19,171 interactions connecting 124 unique regulators with 2,095 unique targets (Figure 3, Item S1). 15.7% of all interactions (8,856 out of 56,234) were shared between the PCeE and GCeE networks, however the interacting molecular entities in these interactions (termed nodes) did not all exhibit the same direction of differential expression between the networks (comparing PCeE or the GCeE data to the conventionally differentiated enteroid data). In each of the enriched enteroid regulatory networks, a particular gene was represented (as a node in the network) only once, but may have been involved in multiple different interactions. In different interactions, a single node could 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 regulatory interaction networks, including transcriptional and post-transcriptional interactions, which illustrate the effect of skewing enteroid differentiation towards Paneth cell and goblet cell lineages.

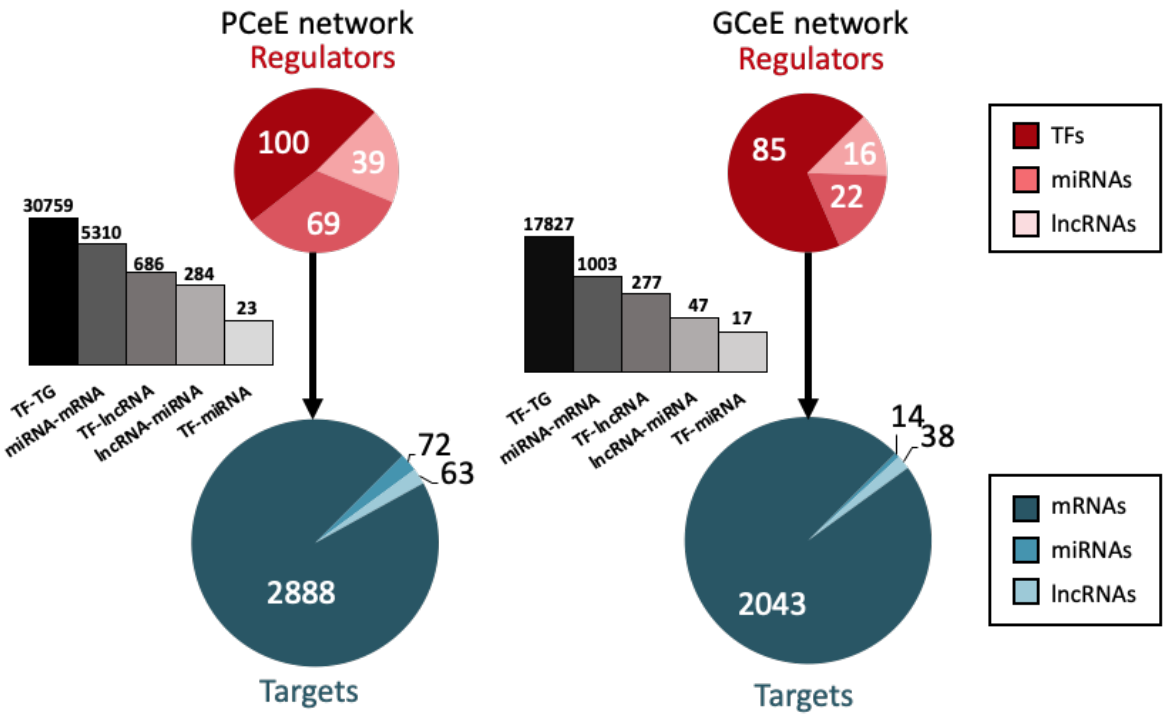


Figure 3: Regulatory network summary for Paneth cell enriched enteroid (left) and goblet cell enriched enteroid (right) datasets. *TF- transcription factor; miRNA- microRNA; lncRNA- long non-coding RNA. Total number of each regulator type shown in red, number of each target type shown in blue. In the 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

Through pathway and marker analysis we predicted that our PCeE and GCeE datasets (i.e. DEG lists), and consequently our regulatory networks, contain signatures from the cell type of interest as well as additional noise from other secretory lineages. To focus specifically on the cell type specific elements of the networks, we used previously identified cell type specific markers to extract predicted Paneth cell and goblet cell regulators from our PCeE and GCeE networks. As cell type specific markers represent genes performing functions specific to a particular cell type, we expected that the regulators of these marker genes will have an important role in determining the function of said cell type. To identify these regulators, we extracted from the PCeE and GCeE networks, all relevant cell type specific markers and their direct regulators. These new networks were termed the Paneth cell subnetwork and goblet cell subnetwork respectively. The Paneth cell subnetwork contained 33 markers specific for Paneth cells with 62 possible regulators. The goblet cell subnework contained 150 markers with 63 possible regulators (Figure 4, Table S6). Observing the ratio of regulators and markers, the Paneth cell subnetwork had, on average, 1.88 regulators for each marker. On the other hand, the goblet cell subnetwork exhibited only 0.42 regulators for each marker. The quantity of markers identified in each subnetwork (33 in the Paneth network and 150 in the goblet network) correlates with the number of marker genes identified by Haber *et al.* (12). 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 (35).

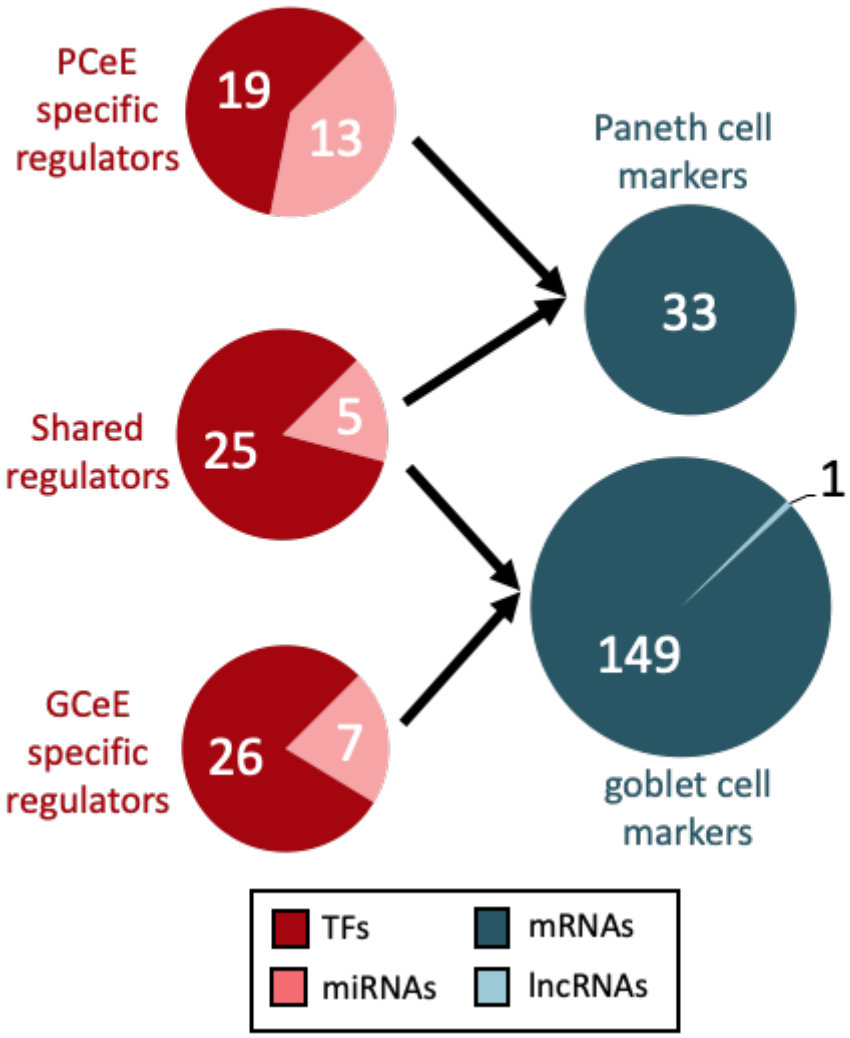


Figure 4: Regulator-marker subnetwork summary for Paneth cell (upper) and goblet cell (lower) datasets. *TF- transcription factor; miRNA- microRNA; lncRNA- long non-coding RNA. Total number of each regulator type shown in red, number of each target type shown in blue. Regulators have been categorised 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).*

Of the 95 marker regulators, we identified approximately one-third (30/95) as present in both subnetworks (Figure 4). Given that the markers are different between the cell types, a regulator shared between the Paneth cell and goblet cell subnetworks 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 (37) - for example, in this case, between the Paneth cells and goblet cells.

Further investigation of the distinct regulator-marker interactions highlighted a gradient of regulator specificity. We generated matrices to visualise the markers controlled by each regulator in the goblet cell (Figure 5A) and the Paneth cell (Figure 5B,C) 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 subnetwork. A collection of regulators (both

subnetwork specific and shared) appear to regulate large proportions of the markers. 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 represent potential master regulators of differentiation or maintenance of the given cell types in the enriched enteroids. 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 in cell type enriched enteroids.

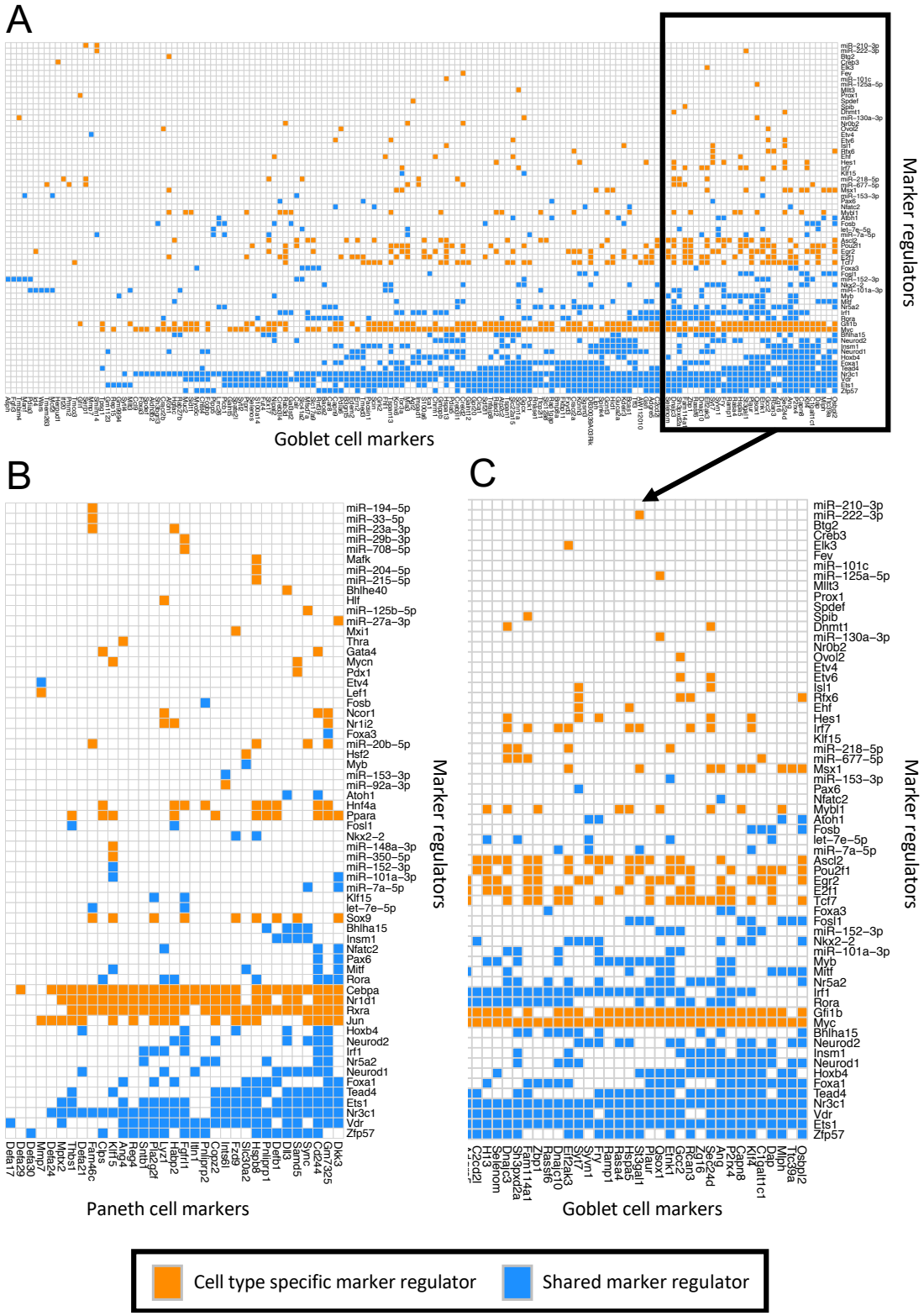


Figure 5: Matrices of interactions between markers and their regulators in the Paneth cell and goblet cell subnetworks. *Regulators on y-axis, markers (regulator targets) on y-axis. Orange boxes indicate the interaction of a regulator and a marker where the regulator is only found in one of the two subnetworks. Blue boxes signify that the regulator is found in both the Paneth cell and the goblet cell subnetworks.* ***A:*** *All goblet cell markers (12) and their regulators in the goblet cell subnetwork.* ***B:*** *All Paneth cell markers (12)from (12) and their regulators in the Paneth cell subnetwork.* ***C:*** *Sub-section of A showing the markers (and their regulators) which have the most regulatory connections. Interactions in Table S6.*

Regulators of cell type specific markers exhibit rewiring between Paneth cells and goblet cells

Cell type specific markers, which carry out cell type specific functions, are inherently different between the Paneth cell and goblet cell subnetworks (mutually exclusive). Therefore, the regulators observed in both Paneth cell and goblet cell subnetworks (shared regulators) are expected to target different marker genes. To do this, the regulators must have different regulatory connections in the different cell types, a phenomenon termed 'rewiring' (38). We extended the analysis to the original regulatory networks (PCeE and GCeE networks) to investigate whether any of the 30 identified shared regulators are rewired between the whole PCeE and the GCeE networks, and thus are highly likely to have different functions in the two types of enriched enteroids as well as between Paneth and goblet cells. To quantify rewiring of each of these regulators, we observed their targets in the PCeE and GCeE networks using the Cytoscape application, DyNet. DyNet assigns each regulator a rewiring score depending on how different their targets are between the two regulatory networks (Table S7). 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 PCeE and GCeE networks have different functions (hypergeometric model, q value ≤ 0.1) (Table S8). Across all five regulators the general trend indicated that targets specific to the PCeE network are associated with metabolism; targets specific to the GCeE network are associated with cell cycle and DNA repair. As pathway analysis carried out on the enteroid DEGs (Figure 2B) identified the same phenomenon, this suggests that the rewired regulators could be key drivers of transcriptional changes during the skewing of enteroid differentiation towards Paneth cell or goblet cell lineages. In addition, given that the strongest signal of enriched enteroids represents their enriched cell type, we predict that these functions are key features of Paneth cells and goblet cells in the enteroids, and that the rewired regulators are important drivers of cell type specific functions.

Looking at the regulators in more detail, the GCeE 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 PCeE 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 (39). Dysfunction of protein folding and the presence of endoplasmic reticulum stress are both associated with IBD (40–42). Therefore, we predict 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 PCeE and GCeE regulatory networks: targets present in both networks are significantly associated with mitosis, whereas those specific to the PCeE network are associated with metabolism, protein localisation, nuclear receptor transcription pathway, circadian clock and hypoxia induced signalling. GCeE 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 of the regulators of both Paneth cell and goblet cell marker genes have different targets (with different associated functions) between the PCeE and the GCeE networks. This suggests that regulatory rewiring occurs between Paneth cell and goblet cell types.

Evaluating the disease relevance of the subnetwork 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 markers (Cebpa, Jun, Nr1d1 and Rxra), those specific to the goblet cell markers (Gfi1b and Myc) and those which appear to regulate many of the markers of both cell types (Ets1, Nr3c1 and Vdr). We identified five genes (*Ets1, Nr1d1, Rxra, Nr3c1 and Vdr*), which correspond to 71% (5/7) of the Paneth cell associated master regulators and 60% (3/5) of the goblet cell associated master regulators to have associations with 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	(43)
	SNP rs12946510 which has associations to IBD, acts as a cis-eQTL for NR1D1	(44)
NR3C1 (glucocorticoid receptor)	Associations with cellular proliferation and anti-inflammatory responses	(45)
	Exogenous glucocorticoids are heavily used as anti-inflammatory therapy for IBD	(46,47))
	ATG16L1, an autophagy related gene, was down-regulated in patients who do not respond to glucocorticoid treatment	(48,49)
	Transcriptionally regulates NFKβ1, a SNP affected gene in ulcerative colitis	(50,51)
VDR (Vitamin D Receptor)	Regulates autophagy in Paneth cells through ATG16L1 – dysfunction of autophagy in Paneth cells has been linked to Crohn's disease	(52,53)
	Induces antimicrobial gene expression in other cell lines	(54,55)
	Specific polymorphisms in the VDR genes have been connected to increased susceptibility to IDB	(56)
	A study looking at colonic biopsies of IBD patients observed reduced VDR expression compared to healthy biopsies	(57)
	Interacts with five SNP affected UC genes	(58,59)
RXRa (Retinoid X Receptor Alpha)	Heterodimerizes with VDR (see above)	(60)
ETS1 (ETS Proto-Oncogene 1)	Important role in the development of hematopoietic cells and Th1 inflammatory responses	(61,62)
	Angiogenic factors in the VEGF-Ets-1 cascades are upregulated in UC and downregulated in CD	(63)
	IBD susceptibility gene	(64)

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 the regulatory networks we reconstructed in this work for the study of CD. To do

this, we checked for the presence of 97 known CD-associated genes in the PCeE and GCeE networks, using lists obtained from two studies (65,66). We found that the PCeE network was significantly enriched in CD associated genes (22/97, hypergeometric model, q value ≤ 0.01) while the GCeE network was not significantly enriched (12/97, hypergeometric model, q value = 0.11) (Table S9). Of the CD-associated genes, one 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 cell type enriched enteroid RNAseq datasets with regulatory networks, we characterised the regulatory environment altered by differentiation skewing. Through focusing on cell type specific markers, we used the regulatory networks to predict master regulators of Paneth cells and goblet cells and to highlight the role of regulatory rewiring in cell differentiation. Given the relevance of Paneth cell and goblet cell dysfunction in IBD, future application of cell type enriched enteroids combined with our network analysis workflow can be used to disentangle multifactorial mechanisms of IBD.

Analysis using known cell type specific markers confirmed that skewing differentiation of enteroids towards Paneth cells or goblet cells results in an increase in the target cell signatures at the transcriptomics level. In addition, signatures of enteroendocrine cells were increased in both differentiation protocols, albeit at a lower amount than the target cell type. This finding correlates with previous investigations at both the transcriptomic and proteomic levels (25,27–29) and is likely due to the shared differentiation pathways of these secretory cells. Nevertheless, as enteroids contain a mixed population of cell types by nature and because intercellular communication is key to a functioning epithelium (19,67), the increased proportion of non-targeted secretory lineages should not be an issue for the application of these models to research. In fact, the enrichment of specific cell types is beneficial for enteroid-based research to increase the signal originating from a specific population of cells and to provide a larger population of cells of interest for downstream single cell analysis of enteroids, which is particularly beneficial when studying rare populations such as Paneth cells. A number of previous studies have shown that these models, which offer a simplified and manipulatable version of the intestinal environment, are useful for the investigation of health and disease related processes (26,28,29).

Given the observed additional increase in secretory lineages in the enriched organoids, we chose to use cell type specific markers to extract interactions specific to Paneth cells and goblet cells from the generated regulatory networks. Using this approach, we identified possible regulators of cell type specific functions in Paneth cells and goblet cells. Some of these regulators were predicted to be important in both cell types, but exhibited differential targeting patterns between the PCeE and the GCeE networks, indicating rewiring of regulators between the cell types. This 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, something which has been shown previously to occur (68,69). Functional analysis of the targets of the most rewired regulators (Etv4, let-7e-5p, miR-151-3p, Myb and Rora) highlights an overrepresentation of metabolism associated targets in the PCeE network and cell cycle associated targets in the GCeE network. A similar result was observed when functional analysis was carried out on genes with significantly different expression levels between the cell type enriched enteroids and the conventionally differentiated enteroids (Figure 2B). This suggests that transcriptional changes during the skewing of enteroid differentiation could be driven by rewired regulators and that these functions are key features of Paneth cells and goblet cells in the enteroids. The latter is supported by current understanding that Paneth cells rely on high levels of protein and lipid biosynthesis for secretory functions (70)important role in metabolically supporting stem cells (71). Additionally, as terminally-differentiated cells do not undergo cell division, this result suggests that enteroid goblet cell signatures are derived from a large population of semi-differentiated goblet-like cells, a phenomenon previously observed in tissue sample based studies (13,72). Extension of our workflow to single cell sequencing of enteroid cells could validate these findings by providing greater cell type specificity. However, these techniques pose further technical and economic challenges (73,74). Specifically, a large number of organoids must be sequenced to mitigate cellular complexity and batch heterogeneity and powerful, reproducible and accurate computational pipelines are required to analyse such data (75).

We predicted key regulators involved in differentiation or maintenance of Paneth cells and goblet cells in the enteroids: Cebpa, Jun, Nr1d1 and Rxra specific to Paneth cells, Gfi1b and Myc specific for goblet cells and Ets1, Nr3c1 and Vdr shared between them. Validation of these regulators poses significant challenges due to their wide expression and broad function range. If the master regulators are controlling differentiation as opposed to cell function maintenance, evaluating lineage arrest or delay can be carried out using a gene knockout or knock down. However, the effects of pleiotropy will

significantly hamper the results and such a study would require significant follow-up studies. On the other hand, if key regulators were predicted by applying the presented computational workflow to condition-specific organoids compared to control organoids (eg. drug treated organoids vs non-treated organoids), the validation would be much simpler. Literature investigation highlighted that many of the predicted master regulators, particularly those associated with Paneth cells, have connections to autophagy, inflammation and IBD (Table 1). While some of these associations are related to different cell types, we assume that if a gene can contribute to a specific function in one cell type it may also contribute in another. Given that IBD is linked to dysfunction of Paneth cells and goblet cells and associated with autophagy via immune-associated secretory functions (29,70,76,77), these results suggest that dysregulation of key cell master regulators could contribute to IBD.

In addition, we identified Crohn's disease (CD) susceptibility genes in the PCeE and GCeE networks, finding a significant enrichment in the PCeE network only. Given that Paneth cell dysfunction is specifically associated with CD, this finding highlights the relevance of the generated networks to the *in vivo* situation. Two of the identified susceptibility genes act as regulators in the networks: in the PCeE network, the protein encoded by *Dbp,* and in the GCeE network, the protein encoded by *Notch2* (Table S9). *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 (78) and has been associated with IBD drug response (79). Therefore this finding highlights a possible regulatory connection between CD susceptibility genes and IBD pathology on a Paneth cell specific level. In the GCeE network, *Notch2* regulates *Notch3* and *Hes1.* It has been previously demonstrated that this pathway can block glucocorticoid resistance in T-cell acute lymphoblastic leukaemia via NR3C1 (predicted master regulator) (80)lucocorticoids are a common treatment for IBD patients (47) Furthermore, this pathway has been previously associated with goblet cell depletion in humans (81)commonly observed in IBD patients. Ultimately, the observation of IBD susceptibility genes in the regulatory networks of these enteroids highlights possible application of this model system to study disease regulation in specific intestinal cell types, through understanding specific mechanistic pathways.

We have shown how network biology techniques can be applied to generate interaction networks representing the change in regulatory environments between two sets of enteroids. Here we presented this workflow, and by using transcriptomics data we characterised the effect of Paneth cell and goblet cell differentiation skewing protocols on enteroids., However the described workflow could be applied to a variety of 'omics datasets and enteroid conditions. For example, to test the response of enteroids to external stimuli, such as bacteria, and on enteroids grown from human-derived biopsies, enabling patient-specific experiments. The application 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. In this study, we integrated miRNA and lncRNA expression datasets, in addition to mRNA data, as these molecules have been shown to perform critical regulatory and mediatory functions in maintaining intestinal homeostasis (44,66,82). However, only small proportions of the generated PCeE and the GCeE networks contained miRNA and lncRNA interactions (Figure 3), due to lack of published interaction information, particularly from murine studies. Both the application of human enteroid data and the future publication of high-throughput interaction studies involving miRNAs and lncRNAs will improve the ability to study such interactions. Nevertheless, these network approaches are beneficial for contextualising gene lists through annotating relevant signalling and regulatory pathways (31) and we can use them to represent and analyse current biological knowledge, to generate hypotheses and to guide further research.

In conclusion, we described an integrative systems biology workflow to compare regulatory landscapes between enteroids from different conditions, incorporating information on transcriptional and post-transcriptional regulation. We applied the workflow to compare Paneth cell and goblet cell enriched enteroids to conventionally differentiated enteroids and predicted Paneth cell and goblet cell specific regulators, which could provide potential targets for further study of IBD mechanisms. Application of this workflow to patient derived organoids, genetic knockout and/or microbially challenged enteroids, alongside appropriate validation and single cell sequencing if available, 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 Wileman, Haerty and Korcsmaros groups.

Competing interests
None

Author contributions

EJJ, ZJM and IH worked on the transcriptomics pipeline development; AT, PS, TWr, JB and MO designed and carried out specific network analysis tasks; SRC, UM, PPP and TWi established and supervised the organoid enrichment work; LJH and FDP supervised the experimental and transcriptomics analysis work; WH, TWi and TK designed the study, supervised the experiments and the analysis. 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). TWr 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.

Methods

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 (18,20,29). 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; (18)) 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] (25).

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.

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 (Corning, 356237) using Cell Recovery Solution (BD Bioscience, 354253), rinsed in PBS and RNA was extracted using miRCURY RNA Isolation Tissue Kit (Exiqon, 300115) according to the manufacturer's protocol.

Stranded RNA library preparation

The enteroid transcriptomics libraries were constructed using the NEXTflex™ Rapid Directional RNA-Seq Kit (Perkin Elmer, 5138-07) using the polyA pull down beads from Illumina TruSeq RNA v2 library construction kit (Illumina, RS-122-2001) with the NEXTflex™ DNA Barcodes – 48 (Perkin Elmer, 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 (CLS960010). 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 (Perkin Elmer, 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 (Illumina, 15004197), six-base indexes distinguish samples and allow multiplexed sequencing and analysis using 48 unique indexes ((Set A: indexes 1-12 (Illumina, RS-200-0012), Set B: indexes 13–24 (Illumina, RS-200-0024), Set C: indexes 25–36 (Illumina, RS-200-0036), Set D: indices 37–48 (Illumina, RS-200-0048)) (TruSeq Small RNA Library Prep Kit Reference Guide (Illumina, 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.

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,. 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 (Roche, 07960140001), 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 (Roche, 07960140001), 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) (83). 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) (84). 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 (85,86). Coding potential of each novel transcript was determined with CPC (version 0.9.2) and CPAT (version 1.2.2) (87,88). 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) (89). Sleuth (version 0.28.1) R library was used to perform differential gene expression (90). 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 (91) . 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 (92). 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 (93). miRNAs with an absolute log2 fold change of 1 and false discovery rate ≤ 0.05 were considered to be differentially expressed.

Differentially expressed genes were grouped by their presence in the PCeE dataset, the GCeE dataset or in both. Each group of differentially expressed genes was tested for functional enrichment (hypergeometric model, q value ≤ 0.1) based on Reactome and KEGG annotations using the ReactomePA R package (94–97) following conversion from mouse to human identifiers using Inparanoid (v8) (98,99).

Enrichment of marker genes

Cell type specific marker gene lists were obtained a mouse single cell sequencing survey (12). 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 (100).

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 S5): miRNA-mRNA (ie., miRNAs regulating mRNAs) and lncRNA-miRNA (ie., lncRNAs regulating miRNAs) interactions were downloaded from TarBase v7.0 (101) and LncBase v2.0 (102), respectively. Only miRNA-mRNA and lncRNA-miRNA interactions determined using HITS-CLIP (103) experiments were considered. Regulatory interactions between transcription factors (TFs) and miRNAs (ie. TFs regulating miRNAs) were retrieved from TransmiR v1.2 (104), GTRD (105) and TRRUST v2 (106,107). 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 (59), 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 (108) 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 PCeE and GCeE 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.Furthermore, to filter for the interactions of prime interest, only nodes which were differentially expressed and their associated interactors were included in the regulatory 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 (12)) were investigated. To do this, all markers were mapped to the relevant networks then subnetworks were extracted consisting of markers and their regulators (Table S5).

Regulatory rewiring analysis

To calculate rewiring scores for regulators, sub-networks were extracted (from the PCeE and GCeE regulatory networks) containing just the regulator of interest and its downstream targets. For each regulator of interest, the subnetworks from the PCeE and GCeE networks were compared using the Cytoscape app DyNet (109,110). The degree corrected D_n score was extracted for each regulator and used to quantify rewiring of the regulator's downstream targets between the PCeE and GCeE 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 PCeE network, only the GCeE 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 (94–97) following conversion from mouse to human identifiers using Inparanoid (v8) (98,99).

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 S3). Where stated, the hypergeometric distribution model was used to calculate significance using R.

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. Scripts to analyse the differentially expressed genes are available on GitHub: https://github.com/korcsmarosgroup/organoid_regulatory_networks.

Supplementary 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.

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: Differentially expressed antimicrobial peptide (AMP) and mucin related genes in Paneth cell enriched enteroids and goblet cell enriched enteroids (compared to conventionally differentiated enteroids). Only genes which are differentially expressed (log2fc ≥ 1 and false discovery rate ≤ 0.05) in at least one of the datasets was included. Lfc = log2 fold change; fdr = false discovery rate; DEG = differentially expressed gene; Paneth = Paneth enriched enteroid, goblet = goblet enriched enteroid.

Table S4: Hypergeometric distribution testing of cell type specific marker enrichment in upregulated differentially expressed gene lists.

Table S5: A summary of the molecular interactions compiled to generate the universal network.

Table S6: Interactions between markers and their regulators in the Paneth cell and the goblet cell subnetworks, including regulator specificity.

Table S7: Rewiring analysis results for the marker regulators present in the Paneth cell and the goblet cell subnetworks. D_n score generated using Cytoscape app DyNet.

Table S8: Functional enrichment analysis of the top five most rewired (shared) marker regulators (p ≤ 0.1).

Table S9: Crohn's disease associated genes in the enriched enteroid regulatory networks.

Figure S1: 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 enteroids, conventionally differentiated (left) and enriched for either Paneth cells or goblet cells (right). Lysozyme granules characteristic of Paneth cells are indicated with a green arrow. Goblet cells were identified using a specific anti-Muc2 mucin antibody (pink).

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|).

Figure S3: Principal component analysis of Paneth cell enriched enteroid transcriptomics data from each biological replicate. Related to mapping and identification of differentially expressed transcripts.

Item S1: Cytoscape file for the Paneth cell enriched enteroid and goblet cell enriched enteroid regulatory networks.

References

1. Zhang K, Hornef MW, Dupont A. The intestinal epithelium as guardian of gut barrier integrity. Cell Microbiol. 2015 Nov;17(11):1561–9.
2. Okumura R, Takeda K. Roles of intestinal epithelial cells in the maintenance of gut homeostasis. Exp Mol Med. 2017 May 26;49(5):e338.
3. Gerbe F, Jay P. Intestinal tuft cells: epithelial sentinels linking luminal cues to the immune system. Mucosal Immunol. 2016 Aug 24;9(6):1353–9.
4. Duerkop BA, Vaishnava S, Hooper LV. Immune responses to the microbiota at the intestinal mucosal surface. Immunity. 2009 Sep 18;31(3):368–76.
5. Mokry M, Middendorp S, Wiegerinck CL, Witte M, Teunissen H, Meddens CA, et al. Many inflammatory bowel disease risk loci include regions that regulate gene expression in immune cells and the intestinal epithelium. Gastroenterology. 2014 Apr;146(4):1040–7.
6. Bevins CL, Salzman NH. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. Nat Rev Microbiol. 2011 May;9(5):356–68.
7. Kim YS, Ho SB. Intestinal goblet cells and mucins in health and disease: recent insights and progress. Curr Gastroenterol Rep. 2010 Oct;12(5):319–30.
8. Gersemann M, Stange EF, Wehkamp J. From intestinal stem cells to inflammatory bowel diseases. World J Gastroenterol. 2011 Jul 21;17(27):3198–203.
9. Okamoto R, Watanabe M. Role of epithelial cells in the pathogenesis and treatment of inflammatory bowel disease. J Gastroenterol. 2016 Jan;51(1):11–21.
10. Liu T-C, Gurram B, Baldridge MT, Head R, Lam V, Luo C, et al. Paneth cell defects in Crohn's disease patients promote dysbiosis. JCI Insight. 2016 Jun 2;1(8):e86907.
11. Gersemann M, Becker S, Kübler I, Koslowski M, Wang G, Herrlinger KR, et al. Differences in goblet cell differentiation between Crohn's disease and ulcerative colitis. Differentiation. 2009 Jan;77(1):84–94.
12. Haber AL, Biton M, Rogel N, Herbst RH, Shekhar K, Smillie C, et al. A single-cell survey of the small intestinal epithelium. Nature. 2017 Nov 16;551(7680):333–9.
13. Smillie CS, Biton M, Ordovas-Montanes J, Sullivan KM, Burgin G, Graham DB, et al. Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. Cell. 2019 Jul 25;178(3):714-730.e22.
14. Parikh K, Antanaviciute A, Fawkner-Corbett D, Jagielowicz M, Aulicino A, Lagerholm C, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature. 2019 Feb 27;567(7746):49–55.
15. Lindeboom RG, van Voorthuijsen L, Oost KC, Rodríguez-Colman MJ, Luna-Velez MV, Furlan C, et al. Integrative multi-omics analysis of intestinal organoid differentiation. Mol Syst Biol. 2018 Jun 26;14(6):e8227.
16. Noben M, Vanhove W, Arnauts K, Santo Ramalho A, Van Assche G, Vermeire S, et al. Human intestinal epithelium in a dish: Current models for research into gastrointestinal pathophysiology. United European Gastroenterol J. 2017 Dec;5(8):1073–81.
17. Aberle MR, Burkhart RA, Tiriac H, Olde Damink SWM, Dejong CHC, Tuveson DA, et al. Patient-derived organoid models help define personalized management of gastrointestinal cancer. Br J Surg. 2018;105(2):e48–60.
18. Sato T, Vries RG, Snippert HJ, van de Wetering M, Barker N, Stange DE, et al. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature. 2009 May 14;459(7244):262–5.

19. Sato T, van Es JH, Snippert HJ, Stange DE, Vries RG, van den Born M, et al. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. Nature. 2011 Jan 20;469(7330):415–8.

20. Sato T, Clevers H. Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. Science. 2013 Jun 7;340(6137):1190–4.

21. Zachos NC, Kovbasnjuk O, Foulke-Abel J, In J, Blutt SE, de Jonge HR, et al. Human enteroids/colonoids and intestinal organoids functionally recapitulate normal intestinal physiology and pathophysiology. J Biol Chem. 2016 Feb 19;291(8):3759–66.

22. Schulte L, Hohwieler M, Müller M, Klaus J. Intestinal organoids as a novel complementary model to dissect inflammatory bowel disease. Stem Cells Int. 2019 Mar 19;2019:8010645.

23. Dotti I, Salas A. Potential Use of Human Stem Cell-Derived Intestinal Organoids to Study Inflammatory Bowel Diseases. Inflamm Bowel Dis. 2018 Nov 29;24(12):2501–9.

24. Farin HF, Van Es JH, Clevers H. Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. Gastroenterology. 2012 Dec;143(6):1518-1529.e7.

25. Yin X, Farin HF, van Es JH, Clevers H, Langer R, Karp JM. Niche-independent high-purity cultures of Lgr5+ intestinal stem cells and their progeny. Nat Methods. 2014 Jan;11(1):106–12.

26. Mead BE, Karp JM. All models are wrong, but some organoids may be useful. Genome Biol. 2019 Mar 27;20(1):66.

27. Mead BE, Ordovas-Montanes J, Braun AP, Levy LE, Bhargava P, Szucs MJ, et al. Harnessing single-cell genomics to improve the physiological fidelity of organoid-derived cell types. BMC Biol. 2018 Jun 5;16(1):62.

28. Luu L, Matthews ZJ, Armstrong SD, Powell PP, Wileman T, Wastling JM, et al. Proteomic Profiling of Enteroid Cultures Skewed toward Development of Specific Epithelial Lineages. Proteomics. 2018 Aug 8;18(16):e1800132.

29. Jones EJ, Matthews ZJ, Gul L, Sudhakar P, Treveil A, Divekar D, et al. Integrative analysis of Paneth cell proteomic and transcriptomic data from intestinal organoids reveals functional processes dependent on autophagy. Dis Model Mech. 2019 Mar 18;12(3).

30. Kwoh CK, Ng PY. Network analysis approach for biology. Cell Mol Life Sci. 2007 Jul;64(14):1739–51.

31. Ideker T, Nussinov R. Network approaches and applications in biology. PLoS Comput Biol. 2017 Oct 12;13(10):e1005771.

32. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature. 2004 Sep 16;431(7006):308–12.

33. van der Wijst MGP, de Vries DH, Brugge H, Westra H-J, Franke L. An integrative approach for building personalized gene regulatory networks for precision medicine. Genome Med. 2018 Dec 19;10(1):96.

34. Módos D, Bulusu KC, Fazekas D, Kubisch J, Brooks J, Marczell I, et al. Neighbours of cancer-related proteins have key influence on pathogenesis and could increase the drug target space for anticancer therapies. npj Syst Biol Appl. 2017 Jan 24;3:2.

35. Stappenbeck TS, McGovern DPB. Paneth Cell Alterations in the Development and Phenotype of Crohn's Disease. Gastroenterology. 2017;152(2):322–6.

36. Clevers HC, Bevins CL. Paneth cells: maestros of the small intestinal crypts. Annu Rev Physiol. 2013;75:289–311.

37. Han P, Gopalakrishnan C, Yu H, Wang E. Gene Regulatory Network Rewiring in the Immune

Cells Associated with Cancer. Genes (Basel). 2017 Nov 7;8(11).

38. Park J, Wang HH. Systematic and synthetic approaches to rewire regulatory networks. Current Opinion in Systems Biology. 2018 Apr;8:90–6.

39. Leach MR, Williams DB. Calnexin and Calreticulin, Molecular Chaperones of the Endoplasmic Reticulum - Madame Curie Bioscience Database - NCBI Bookshelf. 2013;

40. Kaser A, Lee A-H, Franke A, Glickman JN, Zeissig S, Tilg H, et al. XBP1 links ER stress to intestinal inflammation and confers genetic risk for human inflammatory bowel disease. Cell. 2008 Sep 5;134(5):743–56.

41. Kaser A, Blumberg RS. Endoplasmic reticulum stress in the intestinal epithelium and inflammatory bowel disease. Semin Immunol. 2009 Jun;21(3):156–63.

42. Kaser A, Flak MB, Tomczak MF, Blumberg RS. The unfolded protein response and its role in intestinal homeostasis and inflammation. Exp Cell Res. 2011 Nov 15;317(19):2772–9.

43. Chandra V, Bhagyaraj E, Nanduri R, Ahuja N, Gupta P. NR1D1 ameliorates Mycobacterium tuberculosis clearance through regulation of autophagy. Autophagy. 2015 Nov 2;11(11):1987–97.

44. Mirza AH, Berthelsen CH, Seemann SE, Pan X, Frederiksen KS, Vilien M, et al. Transcriptomic landscape of lncRNAs in inflammatory bowel disease. Genome Med. 2015 May 13;7(1):39.

45. Oakley RH, Cidlowski JA. The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. J Allergy Clin Immunol. 2013 Nov;132(5):1033–44.

46. Rutgeerts P. The use of oral topically acting glucocorticosteroids in the treatment of inflammatory bowel disease. Mediators Inflamm. 1998;7(3):137–40.

47. Prantera C, Marconi S. Glucocorticosteroids in the treatment of inflammatory bowel disease and approaches to minimizing systemic activity. Therap Adv Gastroenterol. 2013 Mar;6(2):137–56.

48. De Iudicibus S, Franca R, Martelossi S, Ventura A, Decorti G. Molecular mechanism of glucocorticoid resistance in inflammatory bowel disease. World J Gastroenterol. 2011 Mar 7;17(9):1095–108.

49. Dubois-Camacho K, Ottum PA, Franco-Muñoz D, De la Fuente M, Torres-Riquelme A, Díaz-Jiménez D, et al. Glucocorticosteroid therapy in inflammatory bowel diseases: From clinical practice to molecular biology. World J Gastroenterol. 2017 Sep 28;23(36):6628–38.

50. Yemelyanov A, Czwornog J, Chebotaev D, Karseladze A, Kulevitch E, Yang X, et al. Tumor suppressor activity of glucocorticoid receptor in the prostate. Oncogene. 2007 Mar 22;26(13):1885–96.

51. Dinkel H, Van Roey K, Michael S, Kumar M, Uyar B, Altenberg B, et al. ELM 2016--data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Res. 2016 Jan 4;44(D1):D294-300.

52. Wu S, Zhang Y-G, Lu R, Xia Y, Zhou D, Petrof EO, et al. Intestinal epithelial vitamin D receptor deletion leads to defective autophagy in colitis. Gut. 2015 Jul;64(7):1082–94.

53. Bakke D, Lu R, Agrawal A, Zhang Y, Sun J. 18 myeloid vitamin d receptor signaling regulates paneth cell function and intestinal homeostasis. Gastroenterology. 2018 Jan;154(1):S41.

54. Wang T-T, Nestel FP, Bourdeau V, Nagai Y, Wang Q, Liao J, et al. Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression. J Immunol. 2004 Sep 1;173(5):2909–12.

55. Gombart AF, Borregaard N, Koeffler HP. 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. 2005 Jul;19(9):1067–77.

56. Pei FH, Wang YJ, Gao SL, Liu BR, Du YJ, Liu W, et al. Vitamin D receptor gene polymorphism and ulcerative colitis susceptibility in Han Chinese. J Dig Dis. 2011 Apr;12(2):90–8.

57. Abreu MT, Kantorovich V, Vasiliauskas EA, Gruntmanis U, Matuk R, Daigle K, et al. 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. 2004 Aug;53(8):1129–36.

58. Bovolenta LA, Acencio ML, Lemke N. HTRIdb: an open-access database for experimentally verified human transcriptional regulation interactions. BMC Genomics. 2012 Aug 17;13:405.

59. Lesurf R, Cotto KC, Wang G, Griffith M, Kasaian K, Jones SJM, et al. ORegAnno 3.0: a community-driven resource for curated regulatory annotation. Nucleic Acids Res. 2016 Jan 4;44(D1):D126-32.

60. Bettoun DJ, Burris TP, Houck KA, Buck DW, Stayrook KR, Khalifa B, et al. Retinoid X receptor is a nonsilent major contributor to vitamin D receptor-mediated transcriptional activation. Mol Endocrinol. 2003 Nov;17(11):2320–8.

61. Grenningloh R, Kang BY, Ho I-C. Ets-1, a functional cofactor of T-bet, is essential for Th1 inflammatory responses. J Exp Med. 2005 Feb 21;201(4):615–26.

62. Mouly E, Chemin K, Nguyen HV, Chopin M, Mesnard L, Leite-de-Moraes M, et al. The Ets-1 transcription factor controls the development and function of natural regulatory T cells. J Exp Med. 2010 Sep 27;207(10):2113–25.

63. Konno S, Iizuka M, Yukawa M, Sasaki K, Sato A, Horie Y, et al. Altered expression of angiogenic factors in the VEGF-Ets-1 cascades in inflammatory bowel disease. J Gastroenterol. 2004 Oct;39(10):931–9.

64. Li D, Haritunians T, Potdar A, McGovern DPB. 16 Genetic analysis identified novel loci associated with IBD. Inflamm Bowel Dis. 2018 Jan 18;24(suppl_1):S14–S14.

65. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012 Nov 1;491(7422):119–24.

66. Farh KK-H, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015 Feb 19;518(7539):337–43.

67. Thorne CA, Chen IW, Sanman LE, Cobb MH, Wu LF, Altschuler SJ. Enteroid monolayers reveal an autonomous WNT and BMP circuit controlling intestinal epithelial growth and organization. Dev Cell. 2018 Mar 12;44(5):624-633.e4.

68. Davis TL, Rebay I. Master regulators in development: Views from the Drosophila retinal determination and mammalian pluripotency gene networks. Dev Biol. 2017 Jan 15;421(2):93–107.

69. Mendoza-Parra M-A, Malysheva V, Mohamed Saleem MA, Lieb M, Godel A, Gronemeyer H. Reconstructed cell fate-regulatory programs in stem cells reveal hierarchies and key factors of neurogenesis. Genome Res. 2016 Sep 20;26(11):1505–19.

70. Cadwell K, Liu JY, Brown SL, Miyoshi H, Loh J, Lennerz JK, et al. A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. Nature. 2008 Nov 13;456(7219):259–63.

71. Rodríguez-Colman MJ, Schewe M, Meerlo M, Stigter E, Gerrits J, Pras-Raves M, et al. Interplay between metabolic identities in the intestinal crypt supports stem cell function.

Nature. 2017 Mar 16;543(7645):424–7.

72. Paulus U, Loeffler M, Zeidler J, Owen G, Potten CS. The differentiation and lineage development of goblet cells in the murine small intestinal crypt: experimental and modelling studies. J Cell Sci. 1993 Oct;106 (Pt 2):473–83.

73. Jung J, Jung H. Methods to analyze cell type-specific gene expression profiles from heterogeneous cell populations. Animal Cells Syst (Seoul). 2016 May 3;20(3):113–7.

74. Brazovskaja A, Treutlein B, Camp JG. High-throughput single-cell transcriptomics on organoids. Curr Opin Biotechnol. 2019;55:167–71.

75. Chen G, Ning B, Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front Genet. 2019 Apr 5;10:317.

76. Stappenbeck TS. The role of autophagy in Paneth cell differentiation and secretion. Mucosal Immunol. 2010 Jan;3(1):8–10.

77. Patel KK, Miyoshi H, Beatty WL, Head RD, Malvin NP, Cadwell K, et al. Autophagy proteins control goblet cell function by potentiating reactive oxygen species production. EMBO J. 2013 Dec 11;32(24):3130–44.

78. Nunes T, Bernardazzi C, de Souza HS. Cell death and inflammatory bowel diseases: apoptosis, necrosis, and autophagy in the intestinal epithelium. Biomed Res Int. 2014 Jul 14;2014:218493.

79. Aghdaei HA, Kadijani AA, Sorrentino D, Mirzaei A, Shahrokh S, Balaii H, et al. An increased Bax/Bcl-2 ratio in circulating inflammatory cells predicts primary response to infliximab in inflammatory bowel disease patients. United European Gastroenterol J. 2018 Aug;6(7):1074–81.

80. Real PJ, Tosello V, Palomero T, Castillo M, Hernando E, de Stanchina E, et al. Gamma-secretase inhibitors reverse glucocorticoid resistance in T cell acute lymphoblastic leukemia. Nat Med. 2009 Jan;15(1):50–8.

81. Zheng X, Tsuchiya K, Okamoto R, Iwasaki M, Kano Y, Sakamoto N, et al. Suppression of hath1 gene expression directly regulated by hes1 via notch signaling is associated with goblet cell depletion in ulcerative colitis. Inflamm Bowel Dis. 2011 Nov;17(11):2251–60.

82. Chapman CG, Pekow J. The emerging role of miRNAs in inflammatory bowel disease: a review. Therap Adv Gastroenterol. 2015 Jan;8(1):4–22.

83. Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Internet]. 2010 [cited 2018 Sep 28]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

84. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015 Apr;12(4):357–60.

85. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015 Mar;33(3):290–5.

86. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016 Aug 11;11(9):1650–67.

87. Kong L, Zhang Y, Ye Z-Q, Liu X-Q, Zhao S-Q, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007 Jul;35(Web Server issue):W345-9.

88. Wang L, Park HJ, Dasari S, Wang S, Kocher J-P, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013 Apr

1;41(6):e74.

89. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016 Apr 4;34(5):525–7.

90. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. Nat Methods. 2017 Jul;14(7):687–90.

91. Rueda A, Barturen G, Lebrón R, Gómez-Martín C, Alganza Á, Oliver JL, et al. sRNAtoolbox: an integrated collection of small RNA research tools. Nucleic Acids Res. 2015 Jul 1;43(W1):W467-73.

92. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014 Jan;42(Database issue):D68-73.

93. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139–40.

94. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1999 Jan 1;27(1):29–34.

95. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. Mol Biosyst. 2016 Feb;12(2):477–9.

96. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. 2017 Jan 4;45(D1):D353–61.

97. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2018 Jan 4;46(D1):D649–55.

98. O'Brien KP, Remm M, Sonnhammer ELL. Inparanoid: a comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 2005 Jan 1;33(Database issue):D476-80.

99. Sonnhammer ELL, Östlund G. InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res. 2015 Jan;43(Database issue):D234-9.

100. Mudunuri U, Che A, Yi M, Stephens RM. bioDBnet: the biological database network. Bioinformatics. 2009 Feb 15;25(4):555–6.

101. Vlachos IS, Paraskevopoulou MD, Karagkouni D, Georgakilas G, Vergoulis T, Kanellos I, et al. DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. Nucleic Acids Res. 2015 Jan;43(Database issue):D153-9.

102. Paraskevopoulou MD, Vlachos IS, Karagkouni D, Georgakilas G, Kanellos I, Vergoulis T, et al. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. Nucleic Acids Res. 2016 Jan 4;44(D1):D231-8.

103. Chi SW, Zang JB, Mele A, Darnell RB. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature. 2009 Jul 23;460(7254):479–86.

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

105. Yevshin I, Sharipov R, Valeev T, Kel A, Kolpakov F. GTRD: a database of transcription factor binding sites identified by ChIP-seq experiments. Nucleic Acids Res. 2017 Jan 4;45(D1):D61–7.

106. Han H, Shim H, Shin D, Shim JE, Ko Y, Shin J, et al. TRRUST: a reference database of human transcriptional regulatory interactions. Sci Rep. 2015 Jun 12;5:11432.

107. Han H, Cho J-W, Lee S, Yun A, Kim H, Bae D, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res. 2018 Jan 4;46(D1):D380–6.

108. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.

Bioinformatics. 2010 Mar 15;26(6):841–2.

109. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003 Nov;13(11):2498–504.

110. Goenawan IH, Bryan K, Lynn DJ. DyNet: visualization and analysis of dynamic molecular interaction networks. Bioinformatics. 2016 Sep 1;32(17):2713–5.