bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

[1] Earlham Institute, Norwich Research Park, Norwich, NR4 7UZ, UK
[2] Quadram Institute, Norwich Research Park, Norwich, NR4 7UA, UK
[3] Norwich Medical School, University of East Anglia, Norwich, NR4 7TJ, UK
[4] School of Biological Sciences, University of East Anglia, Norwich, 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

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Background

Normal function of specific intestinal epithelial cell types, like the mucin producing goblet cells or the anti-microbial peptide producing Paneth cells, is key to the homeostasis of the gut. Dysfunction of these cells is often associated with severe gut pathologies, such as inflammatory bowel disease (IBD), including Crohn's disease and ulcerative colitis. Although transcriptional signatures of intestinal cells have been identified, an integrated and cell type specific network analysis that identifies the key regulators with their disease relevance has been lacking.

Method

In this study, 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 cell type specific regulatory networks by integrating expression data into multi-layered networks of regulatory interactions.

Results

By mapping cell type specific marker genes to the network, we were able to identify regulators potentially contributing to cell type specific functions. Among the seven putative master regulators (those targeting many of the markers), we identified four nuclear hormone receptors with links to IBD, immunity and autophagy: Vdr, Rxra, Nr1d1 and Nr3c1. We also found common regulators relevant in both Paneth and goblet cells that regulate different sets of cell type specific markers as a result of regulatory rewiring after differentiation.

Conclusions

We describe an integrative organoid study that combines -omics data with multi-layered networks to study the regulatory landscapes of Paneth cells and goblet cells. Using the developed computational workflow, we generated cell type specific regulatory networks encompassing transcription factor, microRNA and long non-coding RNA regulation at the transcriptional and post-transcriptional level. Analysis of the networks uncovered potential cell type specific master regulators. Specific investigation of four of these regulators identified links to IBD and to cellular phenotypes associated with IBD pathology. Therefore, we demonstrate that application of our workflow in a cell type specific context can be used to disentangle multifactorial mechanisms of IBD and improve our understanding of disease pathomechanisms.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

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

The diverse functions of the mammalian intestinal epithelium are attributed to the distinct cell types lining the epithelium (Okumura and Takeda 2017). To date, various intestinal epithelial cell types have been identified based on specific functional and gene expression signatures. Among the documented cell types, Paneth cells (which reside in the small intestinal crypts of Lieberkuhn) 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 villi, goblet cells secrete mucin, which aggregates to form the mucus layer. The mucus layer acts as a physical barrier between the intestinal lumen and the epithelial lining (Kim and Ho 2010). Both of these intestinal cell types 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 (Gersemann et al. 2009; Kim and Ho 2010).

To understand the function of intestinal epithelial cell types in a disease context, it is important to first identify their molecular signatures and regulatory pathways in a healthy context. At least three separate studies have captured RNA profiles of healthy and/or diseased intestinal cell types; capturing messenger RNA (mRNA), microRNA (miRNA) and long non coding RNA (lncRNA) signatures (Mirza et al. 2015; Haber et al. 2017; Peck et al. 2017). MiRNAs and lncRNAs, together with transcription factors (TFs) (proteins that regulate mRNA transcription), perform critical regulatory functions in maintaining intestinal homeostasis. Dysregulation of these functions has been associated with various gut pathologies (Chapman and Pekow 2015; Mirza et al. 2015). However, a comprehensive analysis of miRNAs, TFs and lncRNAs as well as their regulated genes has not yet been performed on a systems-level in a standardized manner. To identify connections between a regulator and its downstream pathological effect, a systems biology approach is often

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

required. 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.

Many recent studies, particularly those focusing on diseases, have used biopsy samples to produce -omics read-outs (Mirza et al. 2015; Balfe et al. 2018). When the expression profile of diseased tissue is greatly different from a healthy control, these experiments can yield interesting results. However, due to the cellular heterogeneity of the biopsies, the -omics readouts usually represent a combination of different cell types (including semi-differentiated cells). 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 issues often encountered with intestinal biopsies by employing the small intestinal organoid system, 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 organoids; normally differentiated (control) and those enriched in Paneth cells or goblet cells. In two recently published reports, we show that these organoid models are useful for the investigation of normal and disease processes in these intestinal cell types (Luu et al. 2018; Jones et al. 2019). We performed RNAseq transcriptomics analysis on the organoids and compared the readouts from enriched organoids to the control organoids. To the best of our knowledge, this is the first study profiling all the three RNA regulator types from across different intestinal epithelial cell types. To contextualise the sequencing data, we integrated the expression data with directed regulatory networks from transcriptional and post-transcriptional regulatory resources. Our integrative strategy to infer regulatory network landscapes (Figure 1) helped to identify putative master regulators of Paneth cells and goblet cells, and highlighted the contribution of regulatory rewiring to the diverse phenotypes of these cells.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Materials and Methods

Animal handling

C57/Bl6 mice of both sexes were used for organoid generation. All animals were maintained in accordance with the Animals (Scientific Procedures) Act 1986 (ASPA).

Small intestinal organoid culture

Murine small intestinal organoids were generated as described previously (Sato et al. 2009; Sato and Clevers 2013). 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 onto coverslips and incubated at 37°C for 20 minutes to allow Matrigel to polymerise. Organoid media containing EGF, Noggin and R-spondin (ENR; (Sato et al. 2009)) was then overlaid. Organoids were generated from three separate animals for each condition, generating three biological replicates.

Chemically inducing differentiation in organoid cultures

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).

RNA extraction

On day eight, organoids were extracted from Matrigel using Cell Recovery Solution (BD Bioscience), rinsed in PBS and RNA was extracted using Exiqon tissue kit according to the manufacturer's protocol.

Stranded RNA library preparation

The organoid transcriptomics libraries were constructed using the NEXTflex™ Rapid Directional RNA-Seq Kit (PN: 5138-07) using the polyA pull down beads from Illumina TruSeq RNA v2 library construction kit (PN: RS-122-2001) with the NEXTflex™ DNA Barcodes – 48 (PN: 514104) diluted to 6uM. 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).

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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 the 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 (PN: 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 PCR'd 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

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

processing enzymes, whilst most mature microRNAs have a 5' phosphate group. Reverse transcription is used to create cDNA, and 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 library sequencing on HiSeq 100PE

The final pool was quantified using a KAPA Library Quant Kit and found to be 2nM. 9.5µl of the pool was combined with 0.5µl 2N NaOH to make a 1.9nM dilution. This was incubated for five minutes at room temperature to denature the libraries before 940µl of HT1 was added to make a 20pM dilution. This dilution was stored at -20°C for ~2 weeks before being run. 60µl of the 20pM dilution was combined with 60µl of HT1 plus a 1% PhiX spike in three 200µl tubes of a strip of eight to make 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 library sequencing on HiSeq Rapid 50SE

The final pool was quantified using a KAPA Library Quant Kit and found to be 11.2nM. 1.8µl of the pool was combined with 0.5µl of 2N NaOH and 7.7µl EB to make a 2nM dilution. This was incubated for five minutes at room temperature to denature the libraries. 10µl of this was added to 990µl of HT1 to give 1ml at 20pM. 54.7µl of the 20pM dilution was combined in a single 200µl tube of a strip of eight with 7.3µl of PhiX at 20pM and 73µl of HT1. 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,

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

allowing for a one base-pair mismatch per library, and converted to FASTQ format by bcl2fastq.

Mapping and identification of 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; Pertea et al. 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) R library was used to perform differential gene expression (0.28.1) (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, which is dedicated for the mapping and identification of miRNAs from NGS data (Rueda et al. 2015) . The barcodes from the 5' end and adapter sequences (corresponding to Illumina RA3) from the 3' end were removed respectively. Zero mismatches were allowed in detecting the adapter sequences with a minimum adapter length set at 10. Post barcode and adapter trimming, only reads that have 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 normalized after multiple-mapping were calculated for all the libraries corresponding to the control and differentiated organoids. The multiple-mapping normalized read-counts from the corresponding differentiated organoids were compared against the control 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 in differentially expressed gene lists

Cell type specific marker gene lists were obtained from the mouse single cell sequencing survey published by Haber *et al.* (Haber et al. 2017). The cell type specific signature genes of the droplet based and the plate based method were obtained for Paneth, goblet and

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

enteroendocrine cell types. 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 background dataset and thus enable hypergeometric distribution testing, 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 the different regulatory layers were retrieved from multiple databases: miRNA-mRNA (ie., miRNAs regulating mRNAs) and lncRNA-miRNA (ie., miRNAs regulating lncRNAs) 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 specifically retrieved from TransmiR v1.2 (Wang et al. 2010), GTRD (Yevshin et al. 2017) and TRRUST v2 (Han et al. 2015; Han et al. 2018). Co-expression based inferences were ignored. Transcriptional regulatory interactions (ie., TFs regulating target genes) were inferred using data from ORegAnno 3.0 (Lesurf et al. 2016) and GTRD or 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). LncRNA-protein interactions (ie., lncRNAs regulating proteins) were retrieved from the NPINTER database v3.0 (Hao et al. 2016). Only those lncRNA-protein interactions were considered wherein the interactions were confirmed as physical interactions and did not involve non-specific RNA-binding proteins. Non-specific mice RNA-binding proteins were identified using the RBPDB and ATtRACT databases (Cook et al. 2011; Giudice et al. 2016). 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)

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

was used for the custom analyses involving looking 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 Supplementary table 1**.**

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 (Shannon et al. 2003; Goenawan et al. 2016). 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, p value <= 0.1) based on Reactome and KEGG annotations using the ReactomePA R package (Ogata et al. 1999; Yu and He 2016; Kanehisa et al. 2017; Fabregat et al. 2018) following conversion from mouse to human identifiers using Inparanoid (v8) (O'Brien et al. 2005; Sonnhammer and Östlund 2015).

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

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

We applied small intestines from mice to generate 3D self-organising, *in vitro* organoid cultures. By altering the stem cell culture medium, as described in (Yin et al. 2014) we were able to generate organoids with a standard distribution of cell types (normally differentiated) as well as organoids enriched with Paneth cells and organoids enriched with goblet cells. These organoids recapitulated features of the *in vivo* small intestinal epithelium, including formation of an internal lumen and distinct crypt- and villus-like domains (Figure 2).

Cell type specific marker genes are enriched in cell type enriched organoids compared to control organoids

Gene expression dynamics were studied by carrying out RNA extraction and sequencing on the normally differentiated, Paneth cell enriched and goblet cell enriched organoids. One technical repeat from the Paneth cell enriched organoids was removed due to poor read quality. Following mapping and identification of reads, differential expression was calculated. The cell type enriched organoid RNAseq data was compared to the normally differentiated data to obtain cell type specific gene expression signatures. In total 4,135 genes were differentially expressed in the Paneth cell enriched data at log2 fold change of 1 or greater (positive or negative) and 2,889 in the goblet cell enriched data (Figure 3). The larger number of differentially expressed genes (DEGs) in the Paneth cell data could be attributed to the highly specialised nature of the cell type (Stappenbeck and McGovern 2017; Clevers and Bevins 2013). 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 normally differentiated organoid data. In total, 1,363 genes were found upregulated in both the Paneth cell and goblet cell DEGs. 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 Paneth cells and goblet cells. The majority of the DEGs were annotated as protein coding. In addition, lncRNAs and miRNAs were identified and any gene not annotated as one of the above is classified as 'other', which included pseudogenes and antisense genes.

To validate the cell types present in the organoids, major cell type specific markers were observed across the organoids using transcript abundances and RNA differential expression results (Supplementary table 2). The control organoids and the cell type enriched organoids expressed markers for stem cells (*Lgr5*), enteroendocrine cells (*ChgA*), goblet cells (*Muc2*), Paneth cells (*Lyz1*) and enterocytes (*Vil1*). This confirms presence of all major small intestinal epithelial cell types in the organoids. In addition, we observed an upregulation of *Muc2, Lyz1* and *ChgA* in Paneth cell and goblet cell enriched organoids compared to the

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

control organoids. *Lgr5* is downregulated in the cell type enriched organoids compared to the control organoids, confirming the more pronounced differentiated status of the organoids. *Vil1* is not differentially expressed between the cell type enriched and the control organoids, indicating similar levels of enterocytes in organoids. Therefore, using primary cell type specific markers we show that the organoids 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 cell type specific marker genes in the upregulated DEG lists - given that DEGs represent the difference in cell type proportions between the cell type enriched and normally differentiated organoids. Marker gene lists were obtained from a high quality single cell sequencing survey carried out by Regev and colleagues (Haber et al. 2017). Hypergeometric distribution testing was used to generate a significance score for enrichment of Paneth cell and goblet cell type specific markers in each set of DEGs. While all tests were significant at corrected p 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 4). This confirms that differential expression tests successfully discerned cell type specific signatures from the RNAseq data, although the data also contains noise from other cell types. Carrying out the marker enrichment method using enteroendocrine markers, we identified that both of the upregulated DEG lists are additionally enriched with enteroendocrine markers (Supplementary table 3). The organoid differentiation protocols used have been shown previously, in addition to Paneth and goblet cell enrichment, to result in enteroendocrine enrichment at the mRNA level (Yin et al. 2014). In conclusion, we have used gene expression analysis to identify expression signatures of Paneth cells and goblet cells. Application of extensive cell type specific markers data shows that the Paneth and goblet expression datasets exhibit cell type specific signatures.

Reconstruction of cell type specific regulatory networks

A mouse molecular interaction network was reconstructed through collation of experimentally verified regulatory interactions from seven open source databases (see Methods, Supplementary table 1). This background network provided a framework on which we overlaid the differential expression data. All filtered interactions are directed and comprise the following types: transcription factor (TF) - target gene (TG), TF - lncRNA, miRNA - mRNA, TF - miRNA, lncRNA - miRNA and lncRNA - protein. All molecules are represented using their Ensembl gene ID. In total 1,383,896 unique interactions were combined to create the network (Supplementary table 1). Of these, the largest two contributors are TF-TG interactions at 77% and TF-lncRNA interactions at 11%. These interactions connect a total of 23,093 unique molecules.

To generate cell type specific regulatory networks, background interactions were filtered using the differential expression data. The assumption was made that if both nodes of a

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

particular interaction were expressed in the RNAseq data, the interaction was 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 interactions were included in the cell type specific networks. In total the Paneth network, generated using differential expression data from the Paneth cell enriched organoids compared to the normally differentiated organoids, contained 37,062 interactions, connecting 208 unique regulators with 3,023 unique targets (Figure 5). The goblet cell network, generated using differential expression data from the goblet cell enriched organoids compared to the normally differentiated organoids, contained 19,173 interactions connecting 125 unique regulators with 2,095 unique targets (Figure 5). 15.7% of all interactions (8,856 out of 56,235) were shared between the Paneth cell and goblet cell networks - although the direction of differential expression of the interacting nodes do not necessarily match between the networks. In each of these networks, some nodes act as both a regulator and a target and some were represented in more than one molecular form, for example, as a lncRNA as well as a target gene. Consequently, we have generated cell type specific regulatory interaction networks for Paneth cell and goblet cell types including transcriptional and post-transcriptional interactions.

Identification of potential cell type specific master regulators

Cell type specific markers represent genes and associated functions specific to a particular cell type. Therefore, we expect that the regulators of these marker genes have an important role in cell type specification and function. Consequently, to identify potential master regulators of the Paneth and the goblet cell types, the upstream regulators of cell type specific markers were investigated. To do this, all markers were mapped to the relevant networks and subnetworks were extracted consisting of markers and their regulators (Supplementary table 4). 33 Paneth cell markers were identified, all of which act as target genes in TF-TG interactions and 9 of which also act as mRNAs in miRNA-mRNA interactions (Figure 6). 62 regulators were identified for these markers in the Paneth cell network. In the goblet cell network 150 markers were identified with 63 regulators (Figure 6).

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) far fewer regulators are identified in the goblet cell subnetwork per marker than for Paneth cell. 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).

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Further investigation of the distinct regulator-marker interactions highlights a gradient of regulator specificity (Figure 7). 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. Klf15 is present in both subnetworks, regulating only two markers in each; Fgfrl1 and Pla2g2f in Paneth cells, Ggcx and Tor3a in goblet cells. These regulators likely have more functionally specific roles. Together, these results highlight potential cell type specific regulators which likely play key roles in specification and maintenance of Paneth and goblet cells and their functions.

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

Of all the regulators of cell type specific markers, we identified in the Paneth and the goblet regulatory networks approximately 1/3rd (33/92) as present in both networks (Figure 6). 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 between regulators in different environments - for example, in this case, between the Paneth cells and goblet cells.

The Cytoscape application, DyNet, was used to quantify the levels of regulatory rewiring for each of the 33 shared marker regulators in the Paneth cell and goblet cell regulatory networks (Supplementary table 5) (Goenawan et al. 2016). Using this analysis we identified the five most rewired regulators as Etv4, mmu-let-7e-5p, mmu-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, p value <= 0.1) (Supplementary table 6). Across all five regulators the general trend indicated that regulator targets specific to the Paneth cell regulatory network are associated with metabolism and 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 specific targets of mmu-miR-151-3p 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). Therefore we hypothesise that

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

mmu-miR-151-3p plays a role in the secretory pathway of goblet cells. In addition, different functional profiles were also observed for the targets of Rora in the Paneth and goblet regulatory networks. Targets present in both networks are significantly associated with mitosis, whereas those specific to the Paneth 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 cell cycle, signalling by Rho GTPases (associated with cell migration, adhesion and membrane trafficking), Notch signalling and interferon signalling. Altogether these observations suggest that regulatory rewiring has occurred between the Paneth and goblet cell types, resulting in functional differences in the actions of the same regulators. In particular, we see that metabolic functions are associated with Paneth cell regulators and the cell cycle is associated with goblet cell regulators.

Integrating our data with disease-related literature associations

To investigate the relevance of the identified regulators in human gastro-intestinal diseases, such as IBD, we carried out literature searches of the three groups of predicted master regulators: those specific to the Paneth marker subnetwork (Cebpa, Jun, Nr1d1 and Rxra), those specific to the goblet marker 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). Interestingly, five of these genes (*Ets1, Nr1d1, Rxra, Nr3c1 and Vdr*) encode transcription factors with connections to inflammation, autophagy and inflammatory bowel disease (IBD), as shown in Table 1. All five are associated with the Paneth cell subnetwork and apart from ETS1, all are nuclear hormone receptors. Paneth cell and goblet cell secretory functions, which play a major role in gut homeostasis, are highly dependent on autophagy (Cadwell et al. 2008; Stappenbeck 2010; Patel et al. 2013; Jones et al. 2018). Given that dysfunction of secretion and autophagy in Paneth and goblet cells has been associated IBD (Cadwell et al. 2009; Gersemann et al. 2009; Kim and Ho 2010; Ke et al. 2016; Stappenbeck and McGovern 2017; Jones et al. 2018), these putative master regulators highlight a link between Paneth and goblet cells, inflammation and IBD, providing an interesting avenue for further research.

Additionally, given the possible connections between identified master regulators and IBD, we tested the potential of our cell type specific regulatory networks for the study of Crohn's disease (CD). To do this, we observed the presence of known CD susceptibility genes in the networks, using lists obtained from two autoimmune disease studies (Jostins et al. 2012; Farh et al. 2015). Of the 81 CD susceptibility genes, 22 were identified in the Paneth cell network and 12 in the goblet cell network (Supplementary table 7). Of these, one gene acts as a regulator in each network. In the Paneth network the CD associated gene *Dbp,* coding the D site binding protein, regulates *Bik* which codes the BCL2 interacting killer, a pro-apoptotic, death promoting protein. In the goblet cell network the CD associated gene

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Notch2 regulates *Notch3* and *Hes1.* Interestingly, Real *et al*. demonstrated that through Notch signalling, HES1 can inhibit the expression of *NR3C1*, blocking glucocorticoid resistance in T-cell acute lymphoblastic leukaemia (Real et al. 2009). Additionally, the repression of *HATH1* by HES1 via Notch signalling has been previously associated with goblet cell depletion in humans (Zheng et al. 2011). Ultimately, the observation of IBD susceptibility genes in the regulatory networks of these organoids highlights possible application of this model system to the study disease regulation at the cellular level.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

Through the integration of organoid RNAseq datasets, we have generated Paneth cell type and goblet cell type specific regulatory networks. With network analysis, we predicted regulators of cell type specific marker genes and highlighted the role of regulatory rewiring in related secretory cells, such as Paneth cells and goblet cells. Furthermore, we showed that these regulatory networks have potential applications to study cellular functions in disease contexts, such as IBD.

The study of intestinal epithelia, in particular Paneth cells, has been hampered by difficulty culturing these non-proliferating cell types. In the last decade, development of organoid culturing techniques has enabled great progress in the understanding of intestinal epithelial physiology and function. These *ex vivo* intestinal models, grown from *Lgr5*+ stem cells, maintain basic small intestinal crypt physiology, recapitulating epithelial homeostasis (Zachos et al. 2016; Sato et al. 2009). Here, we utilised small intestinal organoid cultures to measure transcriptional signatures of Paneth cells and goblet cells. To obtain cell type specific data, organoid differentiation was directed to enrich for these specific cell types (Yin et al. 2014) and differential expression carried out by comparison to normally differentiated organoids. We showed that relevant cell type specific markers (Haber et al. 2017) are enriched in the differential expression datasets. However, we also found that enteroendocrine markers were significantly enriched in both datasets, which could suggest that the organoid differentiation protocols additionally affected this cell type; a result also seen in the original protocol paper and by others (Yin et al. 2014; Mead et al. 2018). Given that enteroendocrine cell lineage specification is closely related in differentiation pathway to both Paneth and goblet cell types, this is not surprising. In addition, the presence of noise in the differential expression data, as shown by the marker enrichment results, is likely to be a side effect of close functional relationships between the Paneth cells and goblet cells. Furthermore, genes in the applied marker lists may also be expressed (at lower levels) in other cell types. The gene *Itln1* provides a good example of this: it is given as a Paneth cell marker, but interrogation of the Haber *et al*. data shows the gene also has significant expression levels in goblet cells (Haber et al. 2017). Likely the gene, alongside many others, is important in both the Paneth cells and the goblet cells. However, despite the possible noise indicated by the cell-type specific marker analysis, our data provides far greater cell type specific resolution than using whole tissue samples. In future, we hope to significantly reduce this noise by applying fluorescence-activated cell sorting on the organoids to generate pools of cell types for subsequent sequencing - but this approach comes with its own challenges (Jung and Jung 2016). Transcriptomics of a cellular population is preferable to single cell sequencing as it mitigates effects of intracellular variation, for example due to spatio temporal asynchrony.

Through interrogation of transcriptional signatures of stranded and small RNAs, we were able to observe transcriptional (transcription factor - target gene) and post-transcriptional

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(e.g. miRNA-mRNA) regulation. miRNAs and lncRNAs have been shown to perform critical regulatory and mediatory functions in maintaining intestinal homeostasis (Chapman and Pekow 2015; Mirza et al. 2015). For example, a study by Farh *et al.* into the genetics of autoimmune diseases found that ~90% of causal variants map to non-coding regions of the genome (Farh et al. 2015), thus highlighting the importance of integrating miRNA and lncRNAs into regulatory networks. However, we believe our current study is the first to generate comprehensive transcriptional and post-transcriptional regulatory networks of Paneth cells and goblet cells. Interestingly, whilst absolute mRNA and protein copy numbers have previously been reported to correlate poorly within fibroblast samples (Schwanhäusser et al. 2011), recent publications found that within cell type enriched organoids, transcriptomic changes are well correlated to the proteome (Lindeboom et al. 2018) and with *in vivo* gene expression (Mead et al. 2018). 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.

In total we identified 4,135 differentially expressed genes (DEGs) in the Paneth data and cell 2,889 in the goblet data (Figure 3). The difference in quantity of DEGs between the datasets could be explained by the diverse selection of environmental signals integrated by Paneth cells. Additionally, as goblet cells are present in fewer numbers in the small intestine than the colon, we hypothesise that they could play a lesser role in this environment. (Stappenbeck and McGovern 2017; Clevers and Bevins 2013). These DEGs were subsequently mapped to a network of previously identified molecular interactions. This background network was collated by integrating experimentally verified mouse interactions from seven databases. To generate the cell type specific regulatory networks we extracted, from the background network, only those interactions where both nodes are differentially expressed in the relevant dataset. We cannot expect to reconstruct the exact structure of the underlying biological networks using this method, but we generate a collection of interactions likely present and important in the cell type of interest. For example, whilst regulators do not necessarily show strong co-expression with their targets, where co-expression exists there is greater chance 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.

Identification of potential cell type specific regulators was carried out by observing upstream regulators of marker genes in the relevant networks. Using this data, we highlighted regulatory rewiring between the cell type networks. Downstream rewiring scores were calculated (using the main regulatory networks) for the 33 marker regulators observed in both the Paneth and goblet marker analyses. This investigation yielded Etv4, mmu-let-7e-5p, mmu-miR-151-3p, Myb and Rora as the most rewired regulators. 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

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(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). Additionally, functional analysis revealed an overrepresentation of cell cycle associated targets in the goblet network. As terminally differentiated cells do not usually undergo mitosis, the cell cycle connection is likely explained by the presence of many goblet-like transit amplifying 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. Additionally we show the potential importance of regulatory rewiring in the evolution of cell type specific pathways and functions.

Furthermore, we used the marker regulators to predict three groups of master regulators: those specific to the Paneth marker subnetwork (Cebpa, Jun, Nr1d1 and Rxra), those specific to the goblet marker 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). Literature investigation of these regulators highlighted that many of these regulators, particularly those associated with Paneth cells, have connections to autophagy, inflammation and IBD. This highlights the importance of autophagy and inflammation in Paneth cell function and suggests that dysregulation of key cell master regulators could lead to IBD. In addition, we identified 22 (out of 81) CD susceptibility genes in the Paneth cell network and 12 in the goblet cell network (Supplementary table 7). This finding highlights the possible applications of these cell type specific regulatory networks for the study of IBD at a cellular level.

Conclusion

Growing availability of RNA sequencing platforms and datasets has facilitated the identification of transcriptional signatures of tissues, disease states and, more recently, single cells. However, functional association and contextualisation of these signatures has provided challenges for researchers and clinicians. We describe a combinatorial systems biology approach to integrate organoid transcriptomics data into cell type specific interaction networks, incorporating information on transcriptional and post transcriptional regulation. We applied this approach, using Paneth cell and goblet cell enriched organoids, to reconstruct regulatory landscapes and predict master regulators of these cell types. Adaptation of this workflow to patient derived, genetic knockout and/or microbially challenged organoids, alongside appropriate wet-lab validation, will aid discovery of key regulators and signalling pathways of healthy and disease associated intestinal cell types.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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, 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). 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.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figures

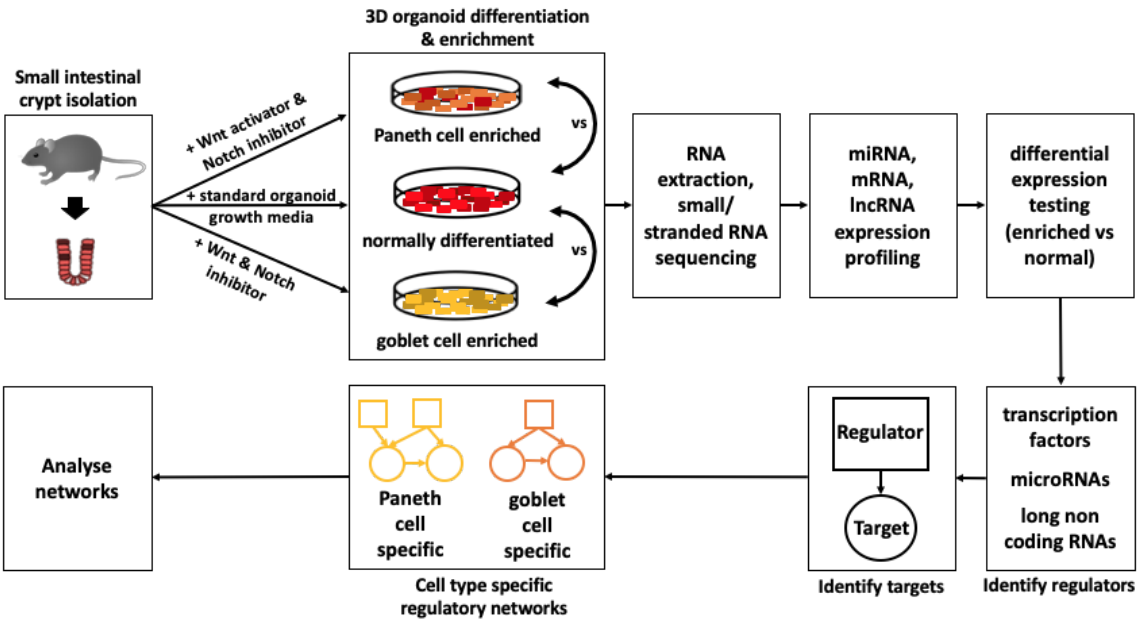

Figure 1: Schematic representation of the workflow to infer cell type specific regulatory networks from 3D organoids differentiated into specific intestinal cell types. TF - transcription factor; lncRNA - long non coding RNA; miRNA - microRNA; mRNA - messenger RNA.

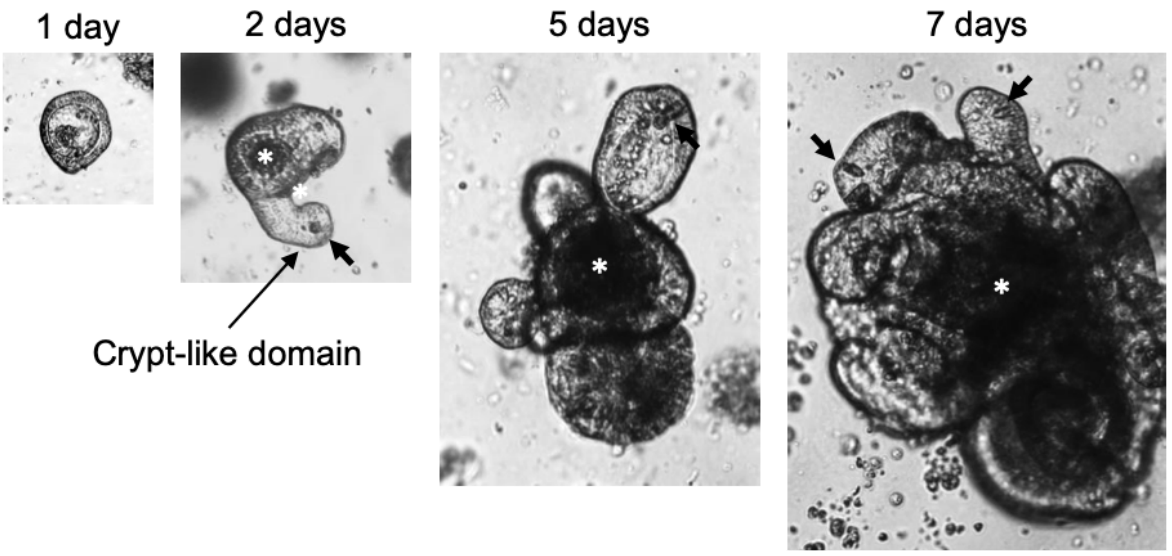

Figure 2: Small intestinal 3D organoid culture. Culture of isolated mouse small intestinal epithelial crypts in Matrigel matrix and A) ENR media or B) ENR media supplemented with CHIR99021 and DAPT (Paneth cell enrichment) or IWP-2 and DAPT (goblet and enteroendocrine cell enrichment) for 7 days. Isolated crypts form 3D cysts which bud after 2

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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.

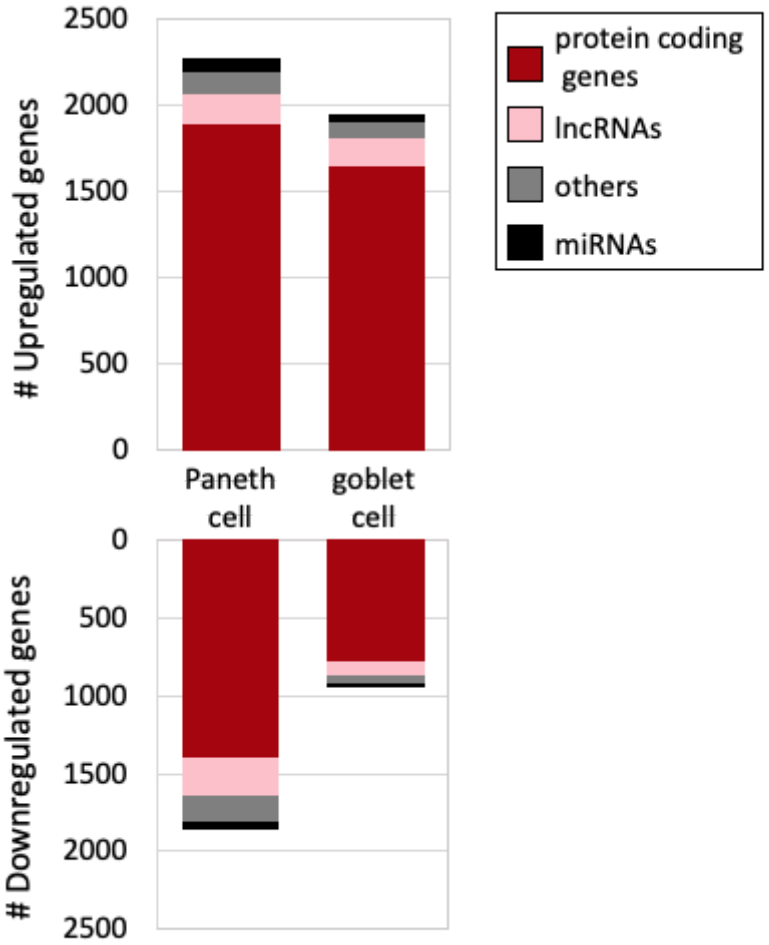

Figure 3: Number of differentially expressed genes in Paneth cells and goblet cells. Differential expression was defined as log2 fold change > |1| and false discovery rate <= 0.05. miRNA = microRNA; lncRNA = long non-coding RNA; Genes annotated as 'other' include pseudogenes and antisense genes. goblet = goblet enriched organoids compared to normally differentiated; Paneth = Paneth enriched organoids compared to normally differentiated; downregulated = log2 fold change <= -1; upregulated = log2 fold change >= 1.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

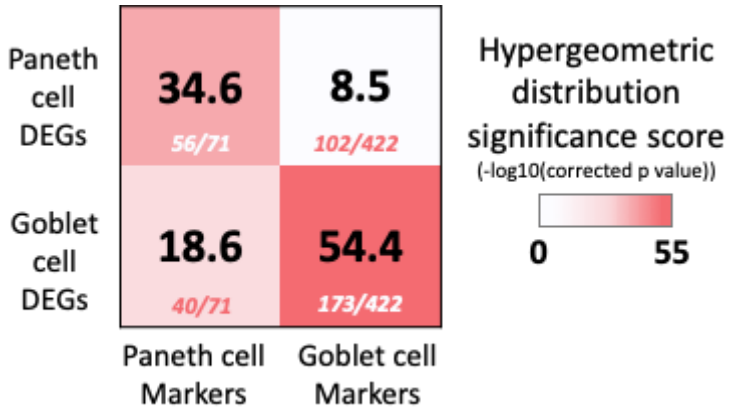

Figure 4: Enrichment of cell type specific marker genes in the differentially expressed gene (DEG) lists. Higher significance scores indicate greater enrichment. Number of markers in DEG list out of total number of markers shown below significance score.

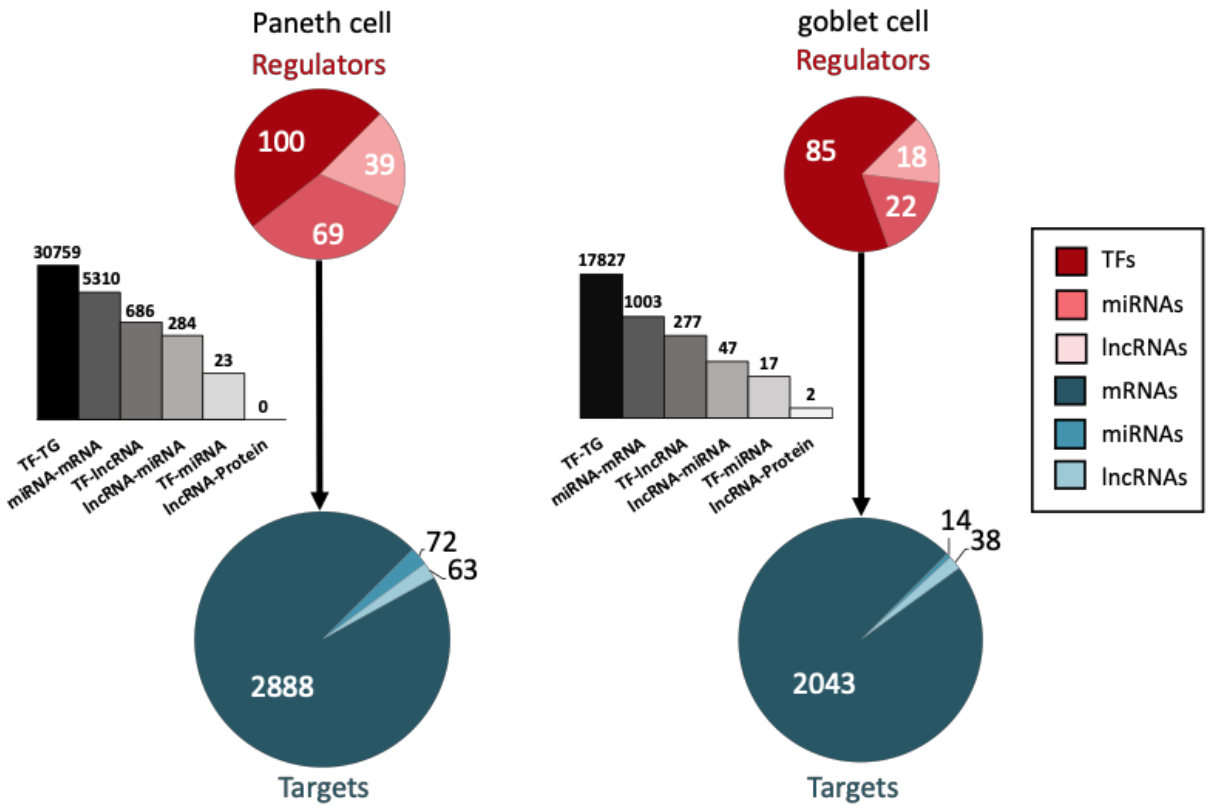

Figure 5: Cell type specific regulatory network summary for Paneth (left) and goblet (right) datasets. TF = transcription factor; miRNA = microRNA; lncRNA = long non-coding RNA. Total number of each regulator type shown in red, number of each target type shown in blue. In the targets pie-chart, mRNAs 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).

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

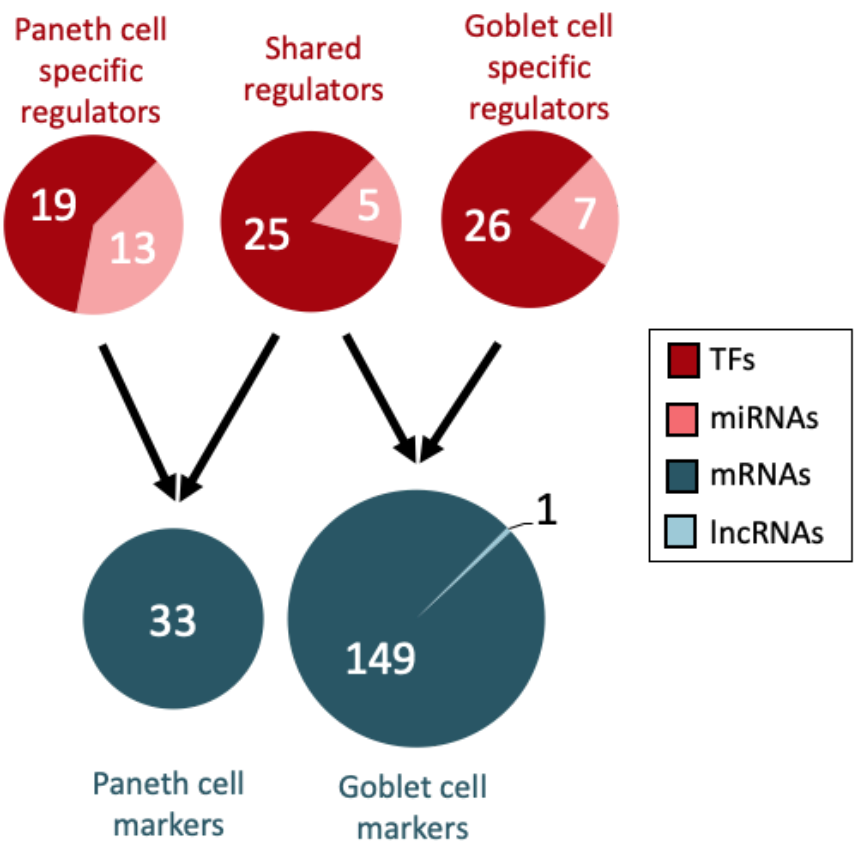

Figure 6: Cell type specific regulator-marker subnetwork summary for Paneth cell (left) and goblet cell (right) datasets. TF = transcription factor; miRNA = microRNA; lncRNA = long non-coding RNA. Total number of each regulator type shown in red, number of each target type shown in blue. Regulators have 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).

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

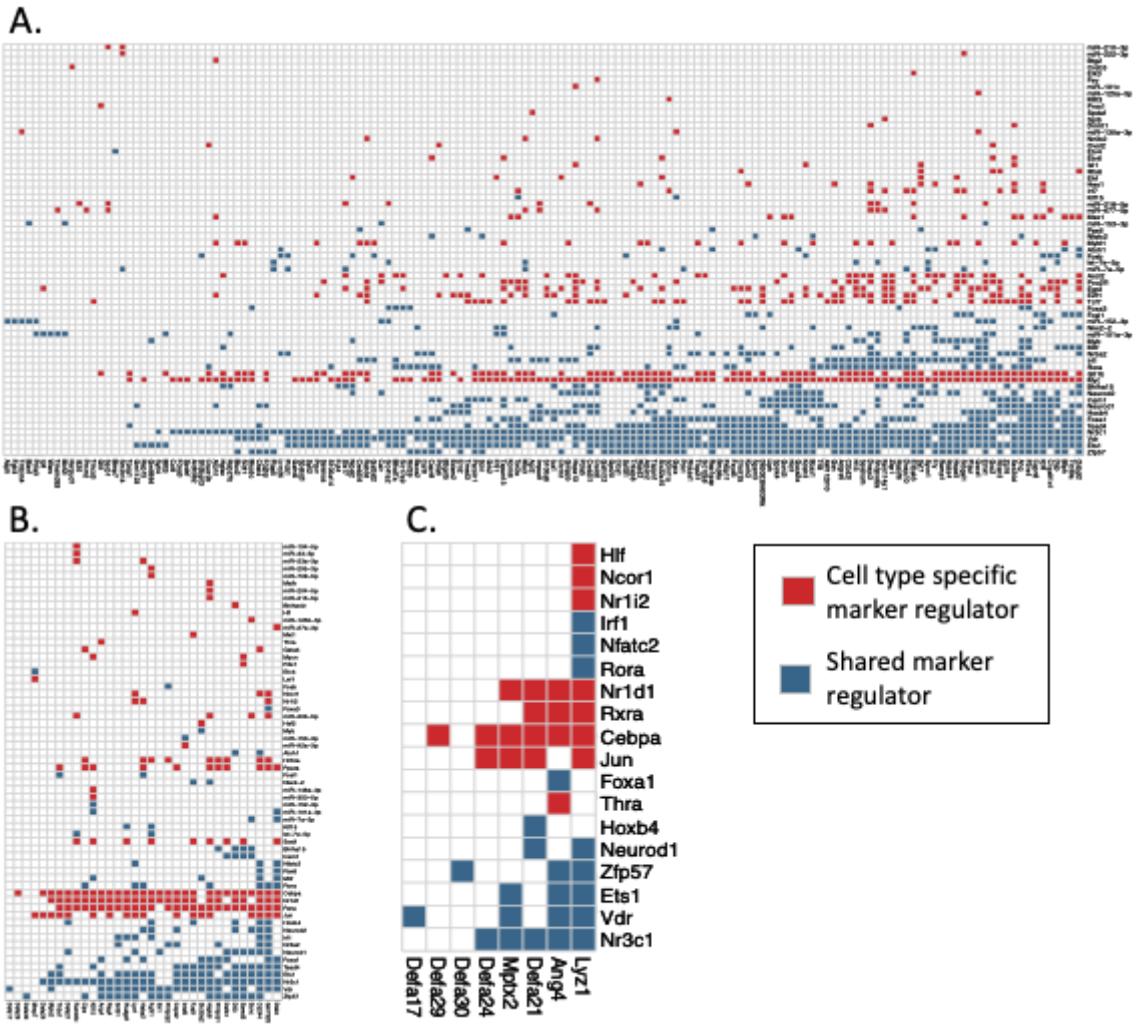

Figure 7: Matrices of regulator-marker interactions. Regulators on Y-axis, markers (regulator targets) on X-axis. Red boxes indicate regulator-marker interactions where the regulator is only found in one cell type marker subnetwork. Blue boxes signify that the regulator is found in both the Paneth and the goblet marker subnetworks. A) All goblet cell markers from Haber *et al.* (Haber et al. 2017) and their regulators. B) All Paneth cell markers from Haber *et al.* and their regulators. C) High quality Paneth cell markers from Haber *et al.* and their regulators. Table of interactions in Supplementary table 4.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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	(Rutgeerts 1998; Prantera and Marconi 2013))
	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 NFK β 1, a SNP affected gene in ulcerative colitis	(Yemelyanov et al. 2007; Dinkel et al. 2016)
VDR (Vitamin D Receptor)	Regulates autophagy in Paneth cells through ATG16L1 – dysfunction of autophagy in Paneth cells has been linked to Crohn's disease	(Sun 2016; Bakke et al. 2018)
	Induces antimicrobial gene expression in other cell lines	(Wang et al. 2004; Gombart et al. 2005)
	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)

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplementary

Supplementary table 1: A summary of the physical interactions compiled to generate the background network.

Supplementary table 2: Expression and differential expression values for primary cell type markers. Expression given as mean transcripts per million (TPM) for each organoid type. Log2 fold changes values given where differential expression criteria passed (p value <= 0.05 and log2 fold change >= |1|. Control = normally differentiated organoids; Paneth = Paneth cell enriched organoids; goblet = goblet cell enriched organoids.

Supplementary table 3: Hypergeometric significance testing of cell type specific marker enrichment in upregulated differentially expressed gene lists.

Supplementary table 4: Regulator-marker interactions in the Paneth and the goblet subnetworks, including regulator specificity.

Supplementary table 5: Rewiring analysis results for the marker regulators present in the Paneth and the goblet marker subnetworks. D_n score generated using Cytoscape app DyNet.

Supplementary table 6: Functional enrichment analysis of the top 5 most rewired (shared) marker regulators (p <= 0.1).

Supplementary table 7: Crohn's disease associated genes in the cell type specific regulatory networks.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References

Abreu, M.T., Kantorovich, V., Vasiliauskas, E.A., Gruntmanis, U., Matuk, R., Daigle, K., Chen, S., Zehnder, D., Lin, Y.C., Yang, H., Hewison, M. and Adams, J.S. 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(8), pp. 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., Flak, M.B., Cusick, J.L., Kohno, K., Iwawaki, T., Billmann-Born, S., Raine, T., Bharti, R., Lucius, R., Kweon, M.-N., Marciniak, S.J. and Blumberg, R.S. 2013. Paneth cells as a site of origin for intestinal inflammation. *Nature* 503(7475), pp. 272–276.

Andrews, S. 2010. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [Online]. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed: 28 September 2018].

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(1), p. 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. *Clinical and experimental gastroenterology* 11, pp. 29–37.

Bevins, C.L. and Salzman, N.H. 2011. Paneth cells, antimicrobial peptides and maintenance of intestinal homeostasis. *Nature Reviews. Microbiology* 9(5), pp. 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, p. 405.

Bray, N.L., Pimentel, H., Melsted, P. and Pachter, L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* 34(5), pp. 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., Stone, C.D., Brunt, E.M., Xavier, R.J., Sleckman, B.P., Li, E., Mizushima, N., Stappenbeck, T.S. and Virgin, H.W. 2008. A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. *Nature* 456(7219), pp. 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(2), pp. 250–252.

Chandra, V., Bhagyaraj, E., Nanduri, R., Ahuja, N. and Gupta, P. 2015. NR1D1 ameliorates

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mycobacterium tuberculosis clearance through regulation of autophagy. *Autophagy* 11(11), pp. 1987–1997.

Chapman, C.G. and Pekow, J. 2015. The emerging role of miRNAs in inflammatory bowel disease: a review. *Therapeutic advances in gastroenterology* 8(1), pp. 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(7254), pp. 479–486.

Clevers, H.C. and Bevins, C.L. 2013. Paneth cells: maestros of the small intestinal crypts. *Annual Review of Physiology* 75, pp. 289–311.

Cook, K.B., Kazan, H., Zuberi, K., Morris, Q. and Hughes, T.R. 2011. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Research* 39(Database issue), pp. D301-8.

Crosnier, C., Stamataki, D. and Lewis, J. 2006. Organizing cell renewal in the intestine: stem cells, signals and combinatorial control. *Nature Reviews. Genetics* 7(5), pp. 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 Journal of Gastroenterology* 17(9), pp. 1095–1108.

Dinkel, H., Van Roey, K., Michael, S., Kumar, M., Uyar, B., Altenberg, B., Milchevskaya, V., Schneider, M., Kühn, H., Behrendt, A., Dahl, S.L., Damerell, V., Diebel, S., Kalman, S., Klein, S., Knudsen, A.C., Mäder, C., Merrill, S., Staudt, A., Thiel, V. and Gibson, T.J. 2016. ELM 2016--data update and new functionality of the eukaryotic linear motif resource. *Nucleic Acids Research* 44(D1), pp. 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 Journal of Gastroenterology* 23(36), pp. 6628–6638.

Duerkop, B.A., Vaishnava, S. and Hooper, L.V. 2009. Immune responses to the microbiota at the intestinal mucosal surface. *Immunity* 31(3), pp. 368–376.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C.D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G. and D'Eustachio, P. 2018. The reactome pathway knowledgebase. *Nucleic Acids Research* 46(D1), pp. 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., Hatan, M., Carrasco-Alfonso, M.J., Mayer, D., Luckey, C.J., Patsopoulos, N.A., De Jager, P.L., Kuchroo, V.K., Epstein, C.B., Daly, M.J., Hafler, D.A. and Bernstein, B.E. 2015. Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature* 518(7539), pp. 337–343.

Gerbe, F. and Jay, P. 2016. Intestinal tuft cells: epithelial sentinels linking luminal cues to the immune system. *Mucosal Immunology* 9(6), pp. 1353–1359.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gersemann, M., Becker, S., Kübler, I., Koslowski, M., Wang, G., Herrlinger, K.R., Griger, J., Fritz, P., Fellermann, K., Schwab, M., Wehkamp, J. and Stange, E.F. 2009. Differences in goblet cell differentiation between Crohn's disease and ulcerative colitis. *Differentiation; Research in Biological Diversity* 77(1), pp. 84–94.

Gersemann, M., Stange, E.F. and Wehkamp, J. 2011. From intestinal stem cells to inflammatory bowel diseases. *World Journal of Gastroenterology* 17(27), pp. 3198–3203.

Giudice, G., Sánchez-Cabo, F., Torroja, C. and Lara-Pezzi, E. 2016. ATtRACT-a database of RNA-binding proteins and associated motifs. *Database: the Journal of Biological Databases and Curation* 2016.

Goenawan, I.H., Bryan, K. and Lynn, D.J. 2016. DyNet: visualization and analysis of dynamic molecular interaction networks. *Bioinformatics* 32(17), pp. 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. *The FASEB Journal* 19(9), pp. 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. *The Journal of Experimental Medicine* 201(4), pp. 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., Tirosh, I., Beyaz, S., Dionne, D., Zhang, M., Raychowdhury, R., Garrett, W.S., Rozenblatt-Rosen, O., Shi, H.N., Yilmaz, O., Xavier, R.J. and Regev, A. 2017. A single-cell survey of the small intestinal epithelium. *Nature* 551(7680), pp. 333–339.

Han, H., Cho, J.-W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., Lee, S., Kang, B., Jeong, D., Kim, Y., Jeon, H.-N., Jung, H., Nam, S., Chung, M., Kim, J.-H. and Lee, I. 2018. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Research* 46(D1), pp. D380–D386.

Han, H., Shim, H., Shin, D., Shim, J.E., Ko, Y., Shin, J., Kim, H., Cho, A., Kim, E., Lee, T., Kim, H., Kim, K., Yang, S., Bae, D., Yun, A., Kim, S., Kim, C.Y., Cho, H.J., Kang, B., Shin, S. and Lee, I. 2015. TRRUST: a reference database of human transcriptional regulatory interactions. *Scientific reports* 5, p. 11432.

Hao, Y., Wu, W., Li, H., Yuan, J., Luo, J., Zhao, Y. and Chen, R. 2016. NPInter v3.0: an upgraded database of noncoding RNA-associated interactions. *Database: the Journal of Biological Databases and Curation* 2016.

Jones, E., Matthews, Z., Gul, L., Sudhakar, P., Divekar, D., Buck, J., Jefferson, M., Armstrong, S., Watson, A., Carding, S., Mayer, U., Powell, P., Hautefort, I., Wileman, T. and Korcsmaros, T. 2018. Integrative analysis of Paneth cell proteomic data from intestinal organoids reveals functional processes affected in Crohn's disease due to autophagy impairment. *BioRxiv*.

Jones, E.J., Matthews, Z.J., Gul, L., Sudhakar, P., Treveil, A., Divekar, D., Buck, J., Wrzesinski, T., Jefferson, M., Armstrong, S.D., Hall, L.J., Watson, A.J.M., Carding, S.R., Haerty, W., Di

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Palma, F., Mayer, U., Powell, P.P., Hautefort, I., Wileman, T. and Korcsmaros, T. 2019. Integrative analysis of Paneth cell proteomic and transcriptomic data from intestinal organoids reveals functional processes dependent on autophagy. *Disease Models & Mechanisms*, p. dmm.037069.

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., Essers, J., Mitrovic, M., Ning, K., Cleynen, I., Theatre, E., Spain, S.L., Raychaudhuri, S., Goyette, P., Wei, Z., Abraham, C. and Cho, J.H. 2012. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 491(7422), pp. 119–124.

Jung, J. and Jung, H. 2016. Methods to analyze cell type-specific gene expression profiles from heterogeneous cell populations. *Animal cells and systems* 20(3), pp. 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 Research* 45(D1), pp. D353–D361.

Ke, P., Shao, B.-Z., Xu, Z.-Q., Chen, X.-W. and Liu, C. 2016. Intestinal autophagy and its pharmacological control in inflammatory bowel disease. *Frontiers in immunology* 7, p. 695.

Kim, D., Langmead, B. and Salzberg, S.L. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 12(4), pp. 357–360.

Kim, Y.S. and Ho, S.B. 2010. Intestinal goblet cells and mucins in health and disease: recent insights and progress. *Current Gastroenterology Reports* 12(5), pp. 319–330.

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 Research* 35(Web Server issue), pp. 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. *Journal of Gastroenterology* 39(10), pp. 931–939.

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

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 Research* 44(D1), pp. D126-32.

Li, D., Haritunians, T., Potdar, A. and McGovern, D.P.B. 2018. 16 Genetic analysis identified novel loci associated with IBD. *Inflammatory Bowel Diseases* 24(suppl_1), pp. S14–S14.

Lindeboom, R.G., van Voorthuijsen, L., Oost, K.C., Rodríguez-Colman, M.J., Luna-Velez, M.V.,

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Furlan, C., Baraille, F., Jansen, P.W., Ribeiro, A., Burgering, B.M., Snippert, H.J. and Vermeulen, M. 2018. Integrative multi-omics analysis of intestinal organoid differentiation. *Molecular Systems Biology* 14(6), p. e8227.

Liu, T.-C., Gurram, B., Baldridge, M.T., Head, R., Lam, V., Luo, C., Cao, Y., Simpson, P., Hayward, M., Holtz, M.L., Bousounis, P., Noe, J., Lerner, D., Cabrera, J., Biank, V., Stephens, M., Huttenhower, C., McGovern, D.P.B., Xavier, R.J., Stappenbeck, T.S. and Salzman, N.H. 2016. Paneth cell defects in Crohn's disease patients promote dysbiosis. 1(8), p. e86907.

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(16), p. e1800132.

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., Wadsworth, M.H., Ahmad, R., Rakoff-Nahoum, S., Carr, S.A., Langer, R., Collins, J.J., Shalek, A.K. and Karp, J.M. 2018. Harnessing single-cell genomics to improve the physiological fidelity of organoid-derived cell types. *BMC Biology* 16(1), p. 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 Medicine* 7(1), p. 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(4), pp. 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. *The Journal of Experimental Medicine* 207(10), pp. 2113–2125.

Mudunuri, U., Che, A., Yi, M. and Stephens, R.M. 2009. bioDBnet: the biological database network. *Bioinformatics* 25(4), pp. 555–556.

Oakley, R.H. and Cidlowski, J.A. 2013. The biology of the glucocorticoid receptor: new signaling mechanisms in health and disease. *The Journal of Allergy and Clinical Immunology* 132(5), pp. 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 Research* 27(1), pp. 29–34.

Okamoto, R. and Watanabe, M. 2016. Role of epithelial cells in the pathogenesis and treatment of inflammatory bowel disease. *Journal of Gastroenterology* 51(1), pp. 11–21.

Okumura, R. and Takeda, K. 2017. Roles of intestinal epithelial cells in the maintenance of gut homeostasis. *Experimental & Molecular Medicine* 49(5), p. e338.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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

Paraskevopoulou, M.D., Vlachos, I.S., Karagkouni, D., Georgakilas, G., Kanellos, I., Vergoulis, T., Zagganas, K., Tsanakas, P., Floros, E., Dalamagas, T. and Hatzigeorgiou, A.G. 2016. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Research* 44(D1), pp. 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., Dinauer, M.C., Virgin, H.W. and Stappenbeck, T.S. 2013. Autophagy proteins control goblet cell function by potentiating reactive oxygen species production. *The EMBO Journal* 32(24), pp. 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. *Journal of Cell Science* 106 (Pt 2), pp. 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. *The Journal of Biological Chemistry* 292(7), pp. 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. *Journal of digestive diseases* 12(2), pp. 90–98.

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. *Nature Protocols* 11(9), pp. 1650–1667.

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. *Nature Biotechnology* 33(3), pp. 290–295.

Pimentel, H., Bray, N.L., Puente, S., Melsted, P. and Pachter, L. 2017. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* 14(7), pp. 687–690.

Prantera, C. and Marconi, S. 2013. Glucocorticosteroids in the treatment of inflammatory bowel disease and approaches to minimizing systemic activity. *Therapeutic advances in gastroenterology* 6(2), pp. 137–156.

Quinlan, A.R. and Hall, I.M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26(6), pp. 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., Meijerink, J., Aifantis, I., Basso, G., Cordon-Cardo, C., Ai, W. and Ferrando, A. 2009. Gamma-secretase inhibitors reverse glucocorticoid resistance in T cell acute lymphoblastic leukemia. *Nature Medicine* 15(1), pp. 50–58.

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

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(1), pp. 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., Verhoeven-Duif, N., Fodde, R. and Burgering, B.M.T. 2017. Interplay between metabolic identities in the intestinal crypt supports stem cell function. *Nature* 543(7645), pp. 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 Research* 43(W1), pp. W467-73.

Rutgeerts, P. 1998. The use of oral topically acting glucocorticosteroids in the treatment of inflammatory bowel disease. *Mediators of Inflammation* 7(3), pp. 137–140.

Sato, T. and Clevers, H. 2013. Growing self-organizing mini-guts from a single intestinal stem cell: mechanism and applications. *Science* 340(6137), pp. 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. and Clevers, H. 2009. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* 459(7244), pp. 262–265.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. and Selbach, M. 2011. Global quantification of mammalian gene expression control. *Nature* 473(7347), pp. 337–342.

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 Research* 13(11), pp. 2498–2504.

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

Stappenbeck, T.S. 2010. The role of autophagy in Paneth cell differentiation and secretion. *Mucosal Immunology* 3(1), pp. 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(2), pp. 322–326.

Sun, J. 2016. VDR/vitamin D receptor regulates autophagic activity through ATG16L1. *Autophagy* 12(6), pp. 1057–1058.

Vanuytsel, T., Senger, S., Fasano, A. and Shea-Donohue, T. 2013. Major signaling pathways in intestinal stem cells. *Biochimica et Biophysica Acta* 1830(2), pp. 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., Fevgas, A., Dalamagas, T. and Hatzigeorgiou, A.G. 2015. DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Research* 43(Database

bioRxiv preprint doi: https://doi.org/10.1101/575845; this version posted March 13, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

issue), pp. D153-9.

Wang, J., Lu, M., Qiu, C. and Cui, Q. 2010. TransmiR: a transcription factor-microRNA regulation database. *Nucleic Acids Research* 38(Database issue), pp. 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 Research* 41(6), p. 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. and White, J.H. 2004. Cutting edge: 1,25-dihydroxyvitamin D3 is a direct inducer of antimicrobial peptide gene expression. *Journal of Immunology* 173(5), pp. 2909–2912.

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(13), pp. 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 Research* 45(D1), pp. 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. *Nature Methods* 11(1), pp. 106–112.

Yu, G. and He, Q.-Y. 2016. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Molecular Biosystems* 12(2), pp. 477–479.

Zachos, N.C., Kovbasnjuk, O., Foulke-Abel, J., In, J., Blutt, S.E., de Jonge, H.R., Estes, M.K. and Donowitz, M. 2016. Human enteroids/colonoids and intestinal organoids functionally recapitulate normal intestinal physiology and pathophysiology. *The Journal of Biological Chemistry* 291(8), pp. 3759–3766.

Zhang, K., Hornef, M.W. and Dupont, A. 2015. The intestinal epithelium as guardian of gut barrier integrity. *Cellular Microbiology* 17(11), pp. 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. *Inflammatory Bowel Diseases* 17(11), pp. 2251–2260.