Insight Box

We address the technological advances and challenges for investigating complex pathological phenomena occurring within the gut. The intestine can be studied as a complex system with three major domains: the intestinal epithelium, the immune system, and the gut flora. Complexity emerges from numerous interactions that occur within and between each of these domains. The murine intestine is the closest and most applicable mammalian experimental system to study human conditions. Innovative, "big data" experimental technologies to interrogate this complex system can be viewed as a continuum, from those with genome-wide coverage to those with cellular and spatio-temporal resolution. We further discuss data integration techniques for deriving insight from large-scale data, and considerations for using mouse models as a platform for systems-level studies of human diseases.

Title: Multiscale Analysis of the Murine Intestine for Modeling Human Diseases

Authors: Jesse Lyons [1,2], Charles A. Herring [3,4], Amrita Banerjee [4,5], Alan J. Simmons [4,5], and Ken S. Lau [3,4,5]

[1] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge MA 02139, USA

[2] Cancer Research Institute, Beth Israel Deaconess Medical Center, Boston MA 02115, USA

[3] Department of Chemical and Physical Biology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA

[4] Epithelial Biology Center, Vanderbilt University Medical Center, Nashville, TN, 37232, USA

[5] Department of Cell and Developmental Biology, Vanderbilt University Medical Center, Nashville, TN, 37232, USA

Corresponding author: Ken S. Lau

                              Epithelial Biology Center

                              Vanderbilt University Medical Center

                              2213 Garland Ave

                              10475 MRB IV

                              Nashville, TN 37232-0441

                              Email: ken.s.lau@vanderbilt.edu

1

Abstract

When functioning properly, the intestine is one of the key interfaces between the human body and its environment. It is responsible for extracting nutrients from our food and excreting our waste products. It provides an environment for a host of healthful microbes and serves as a first defense against pathogenic ones. These processes require tight homeostatic controls, which are provided by the interactions of a complex mix of epithelial, stromal, neural and immune cells, as well as the resident microflora. This homeostasis can be disrupted by invasive microbes, genetic lesions, and carcinogens, resulting in diseases such *Clostridium difficile* infection, inflammatory bowel disease (IBD) and cancer. Enormous strides have been made in understanding how this important organ functions in health and disease using everything from cell culture systems to animal models to human tissue samples. This has resulted in better therapies for all of these diseases, but there is still significant room for improvement. In the United States alone, 14,000 people per year die of *C. difficile*, up to 1.6 million people suffer from IBD, and more than 50,000 people die every year from colon cancer. Because these and other intestinal diseases arise from complex interactions between the different components of the gut ecosystem, we propose that systems approaches that address this complexity in an integrative manner may eventually lead to improved therapeutics that deliver lasting cures. This review will discuss the use of systems biology for studying intestinal diseases *in vivo* with particular emphasis on mouse models. Additionally, it will focus on established experimental techniques that have been used to drive this systems-level analysis, and emerging techniques that will push this field forward in the future.

1. Introduction

1.1 Why use systems biology to study intestinal health and disease?

Evolution, due to its non-directed, non-design-driven nature, has generated substantial complexity in biological systems. Even the fundamental unit of biology, that of a cell, consists of many components that interact with each other to form dynamic complex networks[1]. Additional types of complexity emerge at the organ level *in vivo.* In the intestine, the *in vivo* environment is characterized by interactions between epithelial, immune, muscle, neural and stromal cells, as well as interactions between these cell types and extracellular matrix components, secreted factors, and micro-organisms (Figure 1). The reductionist view of biology maintains that the functional output of a single component or pathway dictates phenotypic behaviors. This view is drawn from *in vitro* cell culture experiments where simplicity is imposed by controlling individual components one at a time. However, because of the interconnectivity between components of *in vivo* biological systems, the effects of single perturbations are propagated throughout interaction networks[2]. Hence, macroscopic phenotypic outcomes, like those involved in diseases, can be viewed as network state effects versus single pathway effects.

Complexity of the intestinal system is best demonstrated by the interaction between the multitudes of bacterial species in the gut. Using a conservative estimate of ~500-1000 species in the gut[3] with ~5000 genes per species, the combinatorial potential of interaction is enormous. Combinatorial complexity is a feature of evolution for maintaining the stability of biological processes in the face of constant perturbations – a property known as robustness[4, 5]. Maintaining a homeostatic network of interactions within bacterial communities prevents colonization of the gut by harmful foreign pathogens via mechanisms of competitive exclusion, stimulation of host immunity, and direct antagonism[6]. These mechanisms are difficult to appreciate without

3

considering the underlying combinatorial complexity at the systems-level, and thus, remain far from full characterization.

Systems biology is an approach to studying biological and biomedical problems from an integrative perspective. There are three main ways in which a systems approach can be useful. First, large scale data collection and network level analysis are ideally suited for screening for molecules and pathways that contribute to a given phenotype or can be targeted to produce a desired outcome. Second, these networks can be viewed as phenotypes in and of themselves and can be used prognostically or diagnostically to determine the effects of an intervention on a pathway of interest. Third, network-level analyses facilitate the identification of unexpected effects resulting from a perturbation. This may be extremely useful in identifying pathways to drug resistance, and in identifying secondary targets that will subvert the resistance mechanisms.

An ideal outcome of a systems-level investigation would consist of a model that represents all of the species in a system and their interactions. Furthermore, it would describe how particular network states relate to given outcomes. By knowing how the pieces of the network relate to one another and how those relationships relate to particular outcomes, the network could be engineered in order to produce a desired outcome. This could be used for deriving combinations of therapies that target not just proteins or pathways, but entire network states, and by extension, phenotypic outcomes. While this represents the ideal outcome of systems-level analysis, in actual practice, researchers are still building the experimental and analytical tools that would enable the production of complete models.

The barriers to a true systems-level understanding are many. The first barrier is technical: even with advances in data collection, it is impossible to measure all of the components of a system that may contribute to a given outcome. RNA-based techniques bring us closest, with

typical RNAseq platforms capable of measuring hundreds of thousands of transcripts down to the level of splice variants[7]. Mass spectrometry has vastly accelerated our ability to interrogate protein abundance[8, 9], however there are many important low-abundance or poor-flying proteins that evade detection by the current state-of-the-art techniques. Fluorescence imaging facilitates temporal interrogation of signaling, but only a very limited number of analytes can be assessed at a time. The second barrier to a full systems understanding is analytical: as we increase the dimensionality of our measurement space, we must adapt our analytical frameworks to deal with this increased complexity. To be truly useful in generating actionable hypotheses, these analytical frameworks must not only be comprehensive, but predictive as well. Unfortunately, as more species are included in a network, it becomes increasingly difficult to predictively model their mechanistic interactions. From current sampling of mathematical models constructed from biological data (Figure 2), a general tradeoff can be observed between the mechanistic resolution of a model and the number of experimental conditions required to constraint each model parameter. To reduce the degrees of freedom and draw mechanistic conclusions from high dimensional data, the number of experimental conditions must increase in order to differentiate between correlative and causative interactions.

This is one of the main reasons that systems biology has been carried out primarily *in vitro*. Practically speaking, it is cheaper, faster and easier to get consistent results with large numbers of experimental perturbations using cell culture models than it is by using *in vivo* models or by testing human specimens. For these reasons, *in vitro* systems biology is extremely powerful for developing the experimental and analytical tools necessary for systems analysis. Furthermore, the large-scale connectivity networks that have been developed form the backbone of mechanistic systems models used for *in vivo* studies. However, while *in vitro* systems biology

is extremely useful for understanding intra-cellular signaling networks, it is very difficult to mimic the full range of inter-cellular interactions that ultimately control outcome in living tissues. Even improved culture methods that involve multiple cell types and three-dimensional structure often fail to fully capture the behavior of a living tissue or organ. Lau *et al.* provided a powerful example of unexpected inter-cellular interactions driving tissue behavior when they found that a lack of adaptive immune cells increased apoptosis in response to TNF-α in the intestinal epithelium[10]. Ultimately, this was the result of a compensatory proinflammatory environment (as underscored by IFN-γ) induced by other epithelial and immune cell subtypes in the tissue. While IFN-γ and TNF-α have been shown to cooperate to induce apoptosis *in vitro*, the increased IFN-γ and the mechanisms that induce its production would not have been predicted by current state-of-the art cell culture models. While this paper speaks to the importance of using systems approaches to identify unexpected and novel tissue responses, part of the reason the modeling was able to produce actionable hypotheses was that the number of analytes (cytokines, immune cells, epithelial response) was relatively few for a systems study. Identifying the key components of a network and their mechanistic interactions becomes increasingly difficult as the number of analytes increases. Still, this paper provides a valuable example of the importance using *in vivo* models when investigating tissue-level behavior using a systems approach.

In the same way that we think about molecular context in a cellular system, we must think about the tissue context when we try to expand to an organ or organismal system. While the outcome of activating a given pathway may be modified by the activities of other pathways, cell contact, or nutrients in an *in vitro* setting, an *in vivo* system is affected by these plus many other factors. Spatial and biophysical constraints, the presence of other cell types, access to

6

oxygen, and the presence of microbes are just a few of the additional factors that may drive cellular, organ and organismal behavior. This complexity provides one of the strongest arguments for the need for systems approaches in studying intestinal diseases, however, it is also one of the chief obstacles to creating meaningful, actionable models. Even comparatively simple *in vitro* systems cannot be mechanistically modeled at the scale of many thousands of molecular species. This difficulty is further compounded by the complexity of the *in vivo* tissue environment. While the ultimate goal of producing computable systems-level models at the tissue and organismal scale may be out of reach with current measurement and analytic techniques, systems-level correlative studies have already been powerful tools for hypothesis generation and have enhanced our knowledge of human gastrointestinal function and disease[10, 11]. In this review we will discuss some of the ways that even incomplete models have provided key insights into intestinal disease, with particular emphasis on the measurement techniques that generate systems scale data, and the mouse as a model system, which have been an invaluable tool for generating hypotheses using systems-level measurement and analysis.

1.2 Mouse as a model system for *in vivo* systems analysis

Every system used to understand human health and disease has its advantages and disadvantages. Though human subjects certainly provide the best reflection of human physiology, patient tissue is not always readily available or sufficiently robust for experimental purposes. On the other end of the spectrum, cell lines are far more readily available and provide some of the same network architecture observed in human beings. However, since they lack many of the critical features of physiological space, they are not always an ideal platform for simulating human disease. Model organisms, such as rodents or zebrafish, have long been used

to model human disease[12]. These animal models have provided the means for numerous breakthroughs yet there can be significant drawbacks in their use that should be kept in mind.

First, we want to touch upon the use of humans as a model system for studying intestinal biology. A major challenge for performing human studies is the deficiency of well-defined experimental controls. The natural genetic variation present in human populations adds tremendous confounding variation to biological phenotypes. Specifically in regards to the gut, environmental factors such as dietary influences on the microbiome can further confound biological variation[13]. Human social behavior, such as mental state and smoking habits, provides additional sources of variation and can also lead to non-compliance during intervention studies[14, 15]. Of course, the amount of variation that can be tolerated within human studies depends upon the phenotype being examined. A well-controlled study by Wu *et al.* examined the change in microbiome when human subjects were fed restricted diets while retained in a hospital setting for 10 days. Although changes in microbial species were examined, these changes were far outweighed by inter-individual variation, suggesting the genetics and environmental histories of the subjects play much more significant roles than the intervention[16]. Evidence from twin studies suggests that there is an important component of heritability in Crohn's disease[17]. Yet, similar concordance is not present for either severity or progression of disease[17, 18]. Furthermore, monozygotic twins display much less concordance in ulcerative colitis[19]. While genetic variation can often be a confounding factor in studying human disease directly, approaches such as Genome Wide Association Studies (GWAS) leverage this natural variation in order to identify rare and complexly interacting genes and genomic regions associated with a given phenotype. For example, IBD GWAS have been fruitful in identifying 163 disease-associated loci[20, 21]. However due to many non-genetic variations within human populations, larger and larger sample

sizes (some numbering in tens of thousands of patients) are required to find rare or interacting disease alleles[22]. Furthermore, although these analyses identify regions of the genome that are correlated with disease, follow up work, often using cell and animal models, is required in order to confirm a given gene as the driver of the genetic association.

At present, rodents and zebrafish are the prevailing model organisms utilized in studying gastrointestinal disorders. Various features of the zebrafish have made them a desirable candidate for modeling GI illnesses such as CD and UC[23, 24]. For instance, organism transparency during the larval stage lends itself to sophisticated imaging techniques[25]. Additionally, zebrafish are extremely fecund and long-term maintenance is relatively inexpensive. Their gastrointestinal tract is very similar to the human small and large intestine and the adult zebrafish possess an innate and adaptive immune system with many features resembling its human counterpart[26]. The high degree of gene and protein homology may lead to the use of zebrafish as an important tool for the study of IBD susceptibility genes. This platform could also prove beneficial for high-throughput screening of compounds suitable for IBD treatment[23]. An important caveat, however, is that, while the immune system is fully developed by adulthood, at this stage in their life cycle, the organism is no longer transparent. This greatly minimizes one of the aforementioned benefits of this model organism, which is the ease of real-time *in vivo* imaging. Moreover, the current protocols for modeling IBD in zebrafish are mostly limited to the use of chemicals known as "haptens," such as TNBS and oxazolone[27]. Though chemically inducing IBD is a laborious process, zebrafish larvae subjected to these treatments will develop many of the same histological characteristics of IBD observed in mice and humans[25]. These include neutrophil infiltration, altered intestinal lipid metabolism, and goblet cell hypertrophy. However, inflammatory response via haptens is little-understood since, at this

stage, the adaptive immune system is not fully functional[12]. More critically, these haptens do not cause epithelial cell damage, important hallmarks of human UC and CD[25, 27]. Additionally, the experimental tools such as surface protein antibodies and associated reagents are less developed for zebrafish than for mice or humans.

Non-human primates, such as the rhesus macaque and the common marmoset, are immunologically and biologically closer to humans, certainly more so than rodents or zebrafish. The rhesus macaque and the cotton-top tamarin, a marmoset native to Colombia, are both known to develop spontaneous colitis when housed in captivity and display an increased incidence of chronic diarrhea[12, 28]. Primate models have been especially useful in the evaluation of gene therapy techniques and other intervention methods for IBD. In addition, the use of primate models has provided insight on the effect of sexual dimorphism in the development of IBD[29]. However, these organisms have many obvious drawbacks to their use, chiefly the increased expense and effort related to their care. Similarly most work in primate models has been primarily observational and limited to studying the pathology of spontaneous incidence of chronic inflammation. The dearth of genetic tools is a critical drawback of this model organism.

Ultimately, given trade-offs in cost versus homology, the mouse still represents the best model system for addressing questions of intestinal disorder via systems approaches. Mouse models are relatively inexpensive, have a strong experimental tool kit, and reflect many aspects of human immune and gastrointestinal physiology. Additionally, numerous genetic and chemical models have been developed to study various aspects of gut health and disease, including stem cell function[30], inflammation[31], and cancer[32] (Table 1). In the IBD field, for instance, there are five main classes of models available for the study of GI-related inflammation[12]. These include chemical induction models, cell transfer models, spontaneous models, congenital models, and

10

genetically engineered models. Certainly no other organism has an equally well-developed assortment of IBD and GI tumor models.

However, it is important to bear in mind that researchers must always exercise caution when extrapolating from the mouse system to that of the human. There are numerous examples of mouse models failing to accurately reflect human biochemistry and disease[33, 34]. For instance, the development of intervention methods for IBD have often yielded conflicting results in humans as opposed to mice. Experimental therapies that have proven efficacious in rodents have not always proven equally effective in clinical trials using human patients[12]. Additionally, significant evidence has implicated the importance of strain differences and diet modulation in murine research[35-39]. Modeling chronic human diseases may also be limited by the relatively short lifespan of mice. Specifically, late stage metastasis in colon cancer has been difficult to model with the mouse, perhaps due to insufficient time to accumulate mutations given similar mutation and cell division rates in the mouse as human[40]. With mice or any other model system, it is imperative to perform complementary studies in humans to confirm critical findings. We believe that systems approaches will be invaluable in this regard since they provide a more global outlook on a given health or disease state, enabling a more thorough comparison of how a mouse model may succeed or fail to reflect the human condition.

Below we will describe some of the powerful ways in which systems biology has been applied to mouse models of human intestinal biology. In particular, we will focus on the experimental techniques that have fueled this growing field and will propel it forward in the future (Table 2). We believe that this growing field, especially when combined with *in vitro* and human data, has tremendous promise for improving our understanding of intestinal diseases ranging from infection to inflammation to cancer. This enhanced understanding should provide

insight into novel methods of treating diseases at the whole organ level, hopefully leading to better therapeutic outcomes for human patients.

2. Tools for generating network-scale datasets from mouse tissues

2.1 RNA Tools

Quantitative, network-scale studies of mammalian tissues have been ongoing since the early 2000s via the use of gene expression arrays [41]. Of all the measurement techniques for interrogating biological systems, nucleic acid-based platforms including gene expression array and RNAseq provide the most comprehensive sampling. However, because gene expression is one level removed from the cellular machinery - the proteins and their modifications - that control function, signatures derived from these studies are correlative but do not necessarily reflect the causative factors that affect phenotypes. As such, these techniques have been most useful for classifying samples and for candidate gene identification. In the intestine for example, genes involved in neoplasia have been identified by comparing expressed genes between ApcMin tumors and adjacent normal tissues[42]. However, to best leverage the large scale of RNA-based measurements, computational tools for inferring signaling activity based on these profiles must be used. Using computational algorithms such as gene ontology enrichment[43, 44] and Gene Set Enrichment (GSEA)[45], it is possible to identify not only the individual genes that are shared/differ, but also to infer biological processes and pathways that contribute to different phenotypic outcomes between samples[46]. These techniques can be used for generating biological insights from multiple expression profile comparisons, for example, for distinguishing shared or divergent gene modules between mouse models and the human diseases they represent. This is important both as a basic scientific question and in choosing the best mouse to represent different aspects of human physiology or disease.

Gene expression analysis, when applied to small amounts of tissue, has the potential to expand current understanding of important biological phenomena. For instance, this technique could be of particular value in identifying and characterizing rare but significant cell populations, such as stem cells or circulating tumor cells[47, 48]. However, validation of single cell mRNA sequencing has not been without its challenges. Currently, many mRNA-sequencing protocols require anywhere from 1ng to 10ug of sample RNA[47]. Acquiring RNA quantities on the nano- or microgram level is often unfeasible for single-cell transcriptomic analysis[49]. As sufficient RNA cannot be directly obtained, a variety of protocols have been developed to sufficiently amplify the RNA pool in order to carry out next generation sequencing. mRNA transcripts undergo reverse transcription, after which double-stranded cDNA is amplified and fragmented for sequence library generation[50]. RNA amplification methods, however, vary in their workflow and can introduce variation in the downstream sequencing analysis[47]. In order to identify gene expression at single cell resolution, amplification protocols and sequencing approaches must be capable of accurately recapitulating the transcriptomic profile of individual cells. However, various factors should be taken into consideration when interpreting data from single cell sequence libraries. For instance, stochastic loss of low abundant transcripts is common when transcribing or amplifying cDNA[51]. Insufficient amplification and subsequent transcript fragmentation can impede read coverage of rare transcripts. This ultimately results in enrichment of high abundant transcripts and skewing of the generated sequence library. Moreover, amplification also tends to induce a bias by preferentially enriching towards the 3'end of transcripts. Generated fragments do not uniformly sample across the whole length of the transcript[50]. As a result, having adequate read coverage for long transcripts becomes increasingly

13

challenging as the levels of input RNA decrease. The 3'end bias also impedes the detection of alternatively spliced isoforms of RNA transcripts.

While the development of RNA amplification methods has brought the field closer to accurate representation of transcriptomics of single cells, there is still significant room for improvement. Smart-seq, for instance, has demonstrated capacity to overcome some common limitations of RNA amplification. Smart-Seq works by converting poly(A)+ RNA to full-length cDNA with the use of oligo(dT) primers and SMART template switching technology[49]. PCR amplification generates sufficient cDNA to then construct Illumina-based sequence libraries. Application of this technique to early mouse embryos has revealed the presence of random monoallelic gene expression in embryonic and mature mammalian cells[51]. Moreover, in comparison to various other mRNA amplification protocols adapted for single cells, Smart-Seq technology provides improved read coverage and enables quantitative assessment of single cell transcriptomes[49]. This improved read coverage augments identification and assessment of alternatively spliced exons and mRNA isoforms. However, there is still a sensitivity threshold where Smart-Seq fails. For instance, this platform possesses limited ability to accurately and efficiently read transcripts longer than 4 Kb, which can minimize its utility in certain contexts[50]. Furthermore, Smart-seq still demonstrates significant PCR biases during its amplification steps. Highly abundant transcripts are preferentially enriched while rarer transcripts are lost[50, 51]. Optimization of this and other amplification techniques remains vital for better characterization of single cell transcriptomes.

2.2 Protein Tools

While RNA-based measurement techniques are extremely powerful for generating global patterns of expression, most of the cellular work is carried out by proteins. Fortunately, the

number of tools becoming available to assess protein levels and modification at a large scale in tissues are increasing rapidly[52]. Candidate approaches such as reverse phase protein lysate arrays[53] and microwestern arrays[54] have all been applied to study cell culture phenomena at various multivariate scales. Furthermore, approaches such as multiplexed bead-based ELISA assays have been used to study acute and chronic inflammation in the mouse as well as culture models[11, 55, 56]. These technologies are candidate-based such that antibody probes can be selected to generate high confidence data, even for low abundance or post-translationally modified proteins. However, because of their candidate-based nature, the investigator must have some prior knowledge on the important analytes to examine for each study.

Moving towards higher coverage, advances in mass spectrometry have enabled researchers to produce a more quantitative global understanding of protein abundance. The two most accurate techniques for measuring relative protein abundances between samples are SILAC and isobaric tagging (iTraq, TMT). Both techniques work by differentially labeling proteins between samples with heavy metal isotopes and then analyzing all samples in one run. The main differences between the two techniques are the number of samples and the labeling procedure. For SILAC, proteins are tagged by providing isotopically labeled supplements live, such that the label is incorporated into proteins during metabolism. Because of the limited number of isotopes available for this type of labeling, only a few samples can be directly compared. Isobaric labeling is performed after removal of tissue, and because there are more isobaric tags available, can be performed with up to 8 different samples in one run[57]. SILAC has been applied to mouse models by feeding mice with a diet modified by isotope-labeled lysines. Using mass spectrometry, proteins that are differentially phosphorylated can be identified and quantified between different mice. This approach been used for studying the signaling network alterations that occur during

15

the progression of squamous cell carcinoma, using the 7,12-Dimethylbenz(a)anthracene (DMBA)/ 12-O-tetradecanoylphorbol-13-acetate (TPA) mouse model[58]. Isobaric tagging methods have also been applied to mouse models, and have been used to identify novel biomarkers of colorectal cancer from ApcMin mice[59].

Although mass spectrometry techniques allow for unbiased sampling of the proteome, this technique still carries certain limitations. For both SILAC and TMT labeling, only a few samples can be analyzed at a time. Without a sufficient amount of data points to constrain the large number parameters measured, it is not possible to build sophisticated models that depict biological complexity. This situation is similar to early microarray analyses performed with few samples, but many measured genes. This problem cannot be solved by simply performing multiple runs because the coverage of proteins detected by mass spectrometry can vary from run to run. At last, mass spectrometry tends to favor highly abundant proteins, while low abundant, but important proteins, are not detected[60].

3. Moving from bulk measurements to single cells

3.1 Flow cytometry based-approaches

Flow cytometry has been a powerful tool for moving from bulk measurements down to subpopulations or even single cells. This technique relies on fluidic systems to pass cells one at a time past lasers which can detect cell size and granularity as well as fluorescence derived from dyes or fluorophore conjugated antibodies. These properties allow flow cytometry to be used in one of 3 ways: (1) cell type identification and counting, (2) sorting cell mixtures into homogeneous populations or single cells for collection and downstream analyses, and (3) direct quantification of proteins in individual cells. The field of immunology has been built upon using flow cytometry to identify and enumerate specific immune cell subsets using cell type specific

16

markers (1). In the intestine, flow cytometry has been heavily used for profiling immune cells in the lamina propria. These include CD4+ helper, CD8+ cytotoxic T cells in response to food antigens in the gut[61], Foxp3+ T regulatory cells induced by microbiota[62, 63], dendritic cell interaction with the intestinal epithelium[64], and various innate lymphoid cell subsets interactions with the microbiot [65-67], macrophages and Tregs[68]. For intestinal epithelial cells, flow cytometry has mostly been utilized for sorting and collecting epithelial stem cell populations using surface markers such as CD24 and CD44 (2)[69]. Sorted intestinal cell populations can then be further analyzed, for example, by qPCR[70]. Alternatively, using multiplexed flow cytometry, protein abundances can be determined within those populations directly (3). For example, tumor initiating potential of colorectal cancer cell subpopulations can be determined by the levels of a panel of cancer stem cell surface markers[71]. Although the number of analytes has been limited by the number of unique fluorochromes and their spectral overlaps, the recent development of mass cytometry or Cytometry Time-of-Flight (CyTOF) has enabled the number of markers analyzed to increase several fold. This growing technology uses heavy metal-labelled reagents in lieu of fluorescently labelled antibodies. Because heavy metals have minimal mass overlap, up to 100 protein analytes can be identified and quantified in single cells[72]. Another method that overcomes fluorescence spectral overlap is DNA conjugation coupled to Nanostring barcode detection[73], which has been used to detect over 90 protein analytes in fine needle aspirates[74]. These highly multiplexable technologies enable high resolution phenotypic profiling of individual cells, which can lead to significant biological insight into how multiple cell types contribute to a tissue phenotype at a systems-level. Advantages of flow cytometry-based approaches include whole cell quantification (without sectioning cells into partial cell fragments as in *in situ* approaches), and the thorough sampling of cells from tissues. While these techniques

17

can provide high-dimensional information at the single-cell level, it should be noted that, once

the tissue is in single cell suspension, all spatial information is lost.

3.2 *In Situ* Microscopy Approaches

To maintain spatial information for single cell studies, microscopy-based approaches to

quantify protein and nucleic acid analytes at single cell resolution are constantly being improved.

Conventional and spectral deconvolution fluorescent microscopy, like flow cytometry, is limited

by spectral overlap for accurate quantification. One strategy to enable high multiplexity with

immunofluorescent microscopy is dye/reagent cycling. One example of this approach is the GE

MultiOmyx[TM] technology, which utilizes chemical-based photo-deactivation after every round of

imaging to allow multiple antibodies conjugated to the same set of fluorochromes to be used in

iterative cycles of staining. Using algorithmic software processing routines that register cells

from different staining rounds and segment individual cells by subcellular markers, MultiOmyx

has been used to quantify in excess of 60 different protein analytes from single cells in formalin-

fixed paraffin embedded tissue sections[75, 76]. Analogous to the development of CyTOF, the use

of heavy metal-labelled reagents in conjunction with mass spectrometry-based imaging can also

enable high multiplexity[77, 78]. A recent emerging imaging technology is to leverage tissue

clearing strategies and biophysical matrices to preserve the 3-dimensional architecture of a

cleared, transparent tissue for 3-D imaging (CLARITY)[79, 80]. Future developments of these

promising technologies for robust multiplex applications will further our understanding of spatial

relationships and communication mechanisms between individual cells in tissue contexts.

Imaging gene expression has been made possible by RNA fluorescence *in situ*

hybridization (FISH). RNA-FISH has been used for visualizing the expression of stem cell

genes within in the intestinal epithelium in native tissue context at single-cell resolution[81].

Similar to protein immunofluorescence, RNA-FISH is also limited by the spectral properties of fluorochomes. To enable high multiplexity, George Church's group has developed a rolling circle-based RNAseq method (fluorescent *in situ* RNA sequencing FISSEQ), where amplification of the cDNA sequence as tandem repeats allows for multiple rounds of stripping and re-hybridization of probes with the same set of fluorochromes[82]. Likewise, Long Cai's group has developed RNA *in situ* hybridization techniques that either use combinatorial fluorescence[83] or sequential hybridization[84] to generate barcodes and enable multiplex quantification of gene expression via imaging.

While *in situ* imaging approaches maintain spatial resolution, sophisticated and powerful imaging processing algorithms are required for segmenting individual cells of analysis. Because each type of tissue has different cell morphologies, these algorithms have to be tailored made. Furthermore, the accuracy of quantification is reduced since sectioned cell fragments are analyzed; it is very difficult to control how much of each cell is sectioned during the tissue preparation process. This problem is further compounded by the sampling of a localized population of cells via tissue sectioning, thus, important details of heterogeneity may be missing unless serial sectioning of the whole tissue is performed. Lastly, both flow cytometry and *in situ* techniques are endpoint assays, meaning that one cannot follow individual cells over time. This limitation hampers dynamic analysis of single cells because one cannot directly link early events to later phenomena.

3.3 Intravital Imaging

At present, the only way to incorporate both spatial and temporal dynamics at the single cell level is via intravital imaging. Advances in imaging modalities such as two photon microscopy allow researchers to track populations of cells *in vivo* at depths of up to 1.6mm[85].

19

This can be used to track cell migration *in vivo* in real time and can be used to construct detailed three dimensional images of tissues with higher clarity than techniques such as confocal microscopy. Another approach for intravital imaging is confocal endoscopy. For this procedure, a confocal microscope is mounted directly to an endoscope, which can be used to fluorescently image the colon[86, 87]. Coupled with emerging varieties of live-cell biosensors that measure, for example, Erk activation[88] or cell proliferation[89, 90], intravital imaging may be used for identifying and quantifying cell types in the mucosa of living, sedated animals. Intravital imaging has been used in the intestine to follow single stem cells through an abdominal window in combination with a confetti reporter driven by the stem cell marker Lgr5[91]. This work provides strong evidence for a model called neutral competition. In this model, stem cells divide symmetrically but stochastically compete for space within the stem cell niche, such that differentiation and renewal occurs at equal probability to maintain homeostasis. This result confirms previous mathematical modeling of stem cell neutral competition[92], where individual outcomes of cell fate follow a stochastic process but stem cell population dynamics provide deterministic outcomes. While studies such as these show that intravital imaging can be extremely powerful in revealing spatiotemporal dynamics within living tissues, these approaches are not yet amenable to network-level studies because they cannot detect more than a handful of colors/analytes at a given time.

4. Microbiome and Metagenomics

In recent years, the improved cost and efficiency of next-generation sequencing platforms has greatly facilitated the study of the intestinal microbiome. This has vastly improved our ability to not only screen for and identify novel pathogens involved in gut diseases[93], but to study microbial communities at the population level as well. Due to the relative ease and lack of

invasiveness of sample collection, compared to other systems approaches, microbiome studies are perhaps most amenable to human studies. Changes in intestinal microbiome have been associated with everything from obesity[94, 95], to gestational diabetes[96], to schizophrenia[97] in human subjects. Generally, these effects are correlated with population-level changes in microbial population. The use of mice may help in identifying which segments of these pathogenic communities are directly causative of disease and the mechanisms by which they wield their effects. One of the key methods by which this can be accomplished is through the use of gnotobiotic mice with humanized microbiomes[98, 99]. Humanized mice are produced by maintaining recipient animals in germ-free conditions followed by transplant with human fecal innoculates. Through this method, population level structure of the microbiome is maintained, although there may be changes at the species or operational taxonomic unit (OTU) level[98]. These mice have been used to assess the microbiome's effects on traits such as obesity and susceptibility to infection by pathogenic bacteria such as *Salmonella*[100]. It should also be noted that this same study indicates that the mouse immune system does not mature as fully when it develops in the presence of a humanized microbiome. Furthermore, the methods to correctly identify bacteria at the species lever are still in development and shotgun approaches can both over-estimate species diversity[101, 102] and miss rare species that are present within samples[103]. While these mice can be very useful as hypothesis generators, as with all mouse models of human disease, results must be confirmed in human subjects before drawing any firm conclusions.

In addition to understanding the role of intestinal microbes in disease, there has been a great deal of interest in how the microbiome can be used to improve health. Intestinal microbiota are known to have important nutritional consequences, including vitamin synthesis and

21

fermentation of complex dietary carbohydrates. In the colonic lumen, clostridia and bifidobacteria ferment dietary fiber into short chain fatty acids (SCFAs) such as butyrate, propionate, and acetate[104]. Butyrate is a crucial component in the large intestine and contributes to colonocyte proliferation, differentiation, and maturation[105]. Additionally, propionate is known to induce Foxp3[+] IL-10 producing T-regulatory cells via the Ffar2 receptor[105]. Luminal levels of SCFAs play an important role in immune modulation, enteric neuron function[106], and epithelial cell maintenance. On a systems scale, metabolomics techniques such as correlative NMR have been used to profile the metabolic capability of the microbiome and its associations with pathologies such as infection [107]. The metabolic potential of the microbiome can also be inferred simply from metabolic genes expressed, using meta-genomic analyses such as PICRUSt[107-109]. However, clinical attempts to replace certain gut metabolites exogenously have largely been inconclusive, mainly to the vast combinatorial complexity in the microbial metabolome. As with other forms of high-dimensional measurement such as RNAseq and microarray, the techniques used to identify microbial molecular species have outpaced the computational and statistical tools needed to fully model and understand the interactions present in the system. Improved computational tools will be necessary if the field is to move from higher-level understanding of populations of bacteria to mechanistic modeling of the gene and metabolic interactions that define the microbial compartment.

Another approach that has not required gene-level understanding is replacement of the whole microbiome of a diseased individual with a healthy microbiome through fecal transplant. This approach has been extremely effective, particularly for treating infections like *Clostridium difficile[110-112]*. Patients who have shown poor response to conventional antibiotics and exhibit relapsing infection, have shown profound responses to fecal transplant from healthy donors. This

technique has already been implemented very successfully with up to 90% cure rates[112-114], and studies in humanized mice may enable researchers to gain insight into the populations of bacteria that produce the best patient outcomes. Additionally, a recent paper from Eric Pamer's group has shown how mouse models and meta-genomics can be used to identify single species, and particular metabolic functions of that species, that can enhance resistance to *C. difficile* infection[115]. In this paper, the authors used a series of antibiotic perturbations to alter the microbiome of mice prior to challenge with *C. difficile*. By correlating the resulting microbiomes with susceptibility to infection and comparing that with patient data, the authors identified a different species of *Clostridium* with an important role in bile acid metabolism that provided resistance to infection. Approaches such as this may eventually enable researchers and clinicians to identify the key components of the microbiome that provide resistance to infection, and to begin standardizing this very powerful therapeutic tool.

5. Systems Biology in Mouse Models

5.1 Advantages of Genetic Perturbations

One of the reasons that the mouse is such a powerful experimental tool is its genetic tractability and the large numbers of genetic models of human diseases that are already available to researchers (Table 1). Often based on human genetic data, mouse models have been generated that mimic features of inflammation and intestinal cancers. Intestinal-specific expression or knockout can be controlled by techniques such as the Cre-Lox recombination system[116]. Additionally, genes can be targeted by techniques such as RNA interference[117, 118] and CRISPR-CAS[119]. Using the experimental methods described above, these models can be used to generate systems-level understanding of the mechanisms of action of genetic lesions associated with human pathologies, or to identify unknown genetic events that may drive these diseases.

23

Alongside traditional transgenic approaches, an emerging discipline known as systems genetics may provide additional clues as to the genetic origins of intestinal function. This approach generates variation in mouse models for systems-level studies by harnessing the natural variation between different strains of mice[120]. Instead of strictly looking at genes and traits, for example, by using haplotype based computational genetic mapping[121], one can also leverage the natural variation in signaling sensitivity between different mouse strains to study network behaviors. Genetic variation between mouse strains can be used in lieu of targeted perturbations for systems-level studies. These variations can then be mapped back to pathways if genomic information is available. The Collaborative Cross, an effort to generate additional genomically characterized inbred lines from eight commonly used but diverse inbred mouse strains, will be a valuable resource in the future for studying mice with a spectrum of variation that can be leveraged for systems-level studies[122].

5.2 Computational modeling of *in vivo* datasets

Many theoretical approaches have been developed over the years to model biological behaviors at multiple scales. Much effort has been dedicated to deriving insights from systems-level data generated from cell culture. However, *in vivo* systems are poor candidates for traditional approaches that assume changes in the network state occur through predictable continuous pathways in predetermined network topologies due to their inherent heterogeneity. Probabilistic approaches based upon statistics are more appropriate for analyzing data that are generated from *in vivo* organisms by experimental systems biology approaches. Here, we give a brief introduction to some of the modeling approaches used for *in vivo* systems data sets. For a more thorough review, see *Wood et al.* in this issue.

Dimension reduction statistical approaches – An elegant approach to interpret large *in vivo* datasets is to classify/cluster phenotypic outcomes by measured components (protein/gene expression states, etc.) in multivariate space. A commonly used technique is principal component analysis (PCA), which uses a mathematical algorithm to reduce the dimensionality of the data while maintaining most of the variation in the dataset[123] (Figure 3A). PCA introduces new variables called principal components that are linear combinations of the original variables. Principal components can be assessed by the amount of information (variance) captured, and thus, plotting a few principal components with the maximal information content allows for the direct visualization of similarities and differences between samples in multi-dimensional space[123]. A supervised version of PCA, which allows for correlation between independent and dependent variables, called partial least squares regression (PLSR), has been used successfully to infer how changes in signaling network are correlated with cell death and proliferative outcomes[11, 124, 125]. Since PCA-based approaches use linear combinations of variables to form components, nonlinear relationships between the variables are not always apparent. t-SNE is a space-maintaining dimension reduction approach that preserves nonlinearity of the data[126]. For visualization, SPADE[127] combines spanning tree analysis with clustering to delineate relationships between samples and/or cells.

Bayesian networks – Bayesian-based methods are stochastic methods used to determine model parameters from a posterior probability distribution derived from experimental data[128, 129]. These methods utilize a likelihood function that increases the predictability with the number of datasets used to train the system. Graph theory and probability meet in Bayesian methods where variables are represented as nodes and their relationships through acyclic edges. The Bayesian parameters can be modeled as discrete, continuous, or both, and temporally, the system can be

25

modeled as static or dynamic. The result is an ensemble of feasible solutions that is more predictable and robust than a single best fit solution. Two popular and freely available software packages that incorporate Bayesian methods are SloppyCell and BioBayes[130, 131].

Logic modeling – An important goal of systems-level analysis is the derivation of a biological network's topology. Traditionally, a pathway's or network's topology is determined painstakingly from decades of experiments, which then are cataloged in the literature. Because the topology of a network can change depending on biological context (e.g., cell type), it is difficult to derive the correct network strictly from mining the literature. Logic-based modeling uses a technique borrowed from engineering that derives relationships from systems-level data with guidance from a prior knowledge network (PKN)[132] (Figure 3B). A PKN is first initiated by enumerating all components and connections that are included directly or indirectly in the data from the literature or databases. The PKN is then collapsed to exclude nodes and edges that have not been explicitly measured while maintaining logical consistency. Next, the modified PKN is fit to the experimental data using a logic-based objective function to determine an optimal set of networks, with each network in this set considered equally viable. Finally, the quality of each edge is statistically assessed by its representation in an ensemble of models. CellNetOptimizer is an example of a freely available logic-based modeling software [133].

5.3 Considerations for systems biology studies in mice

Studying biological phenomena in *in vivo* settings provides the advantage of physiological significance and relevance. However, studying organisms *in vivo* presents its own set of challenges for the design of controlled experiments. While one can largely control for extrinsic factors like diet and drug treatments when doing mouse experiments, even inbred mice have additional sources of variation intrinsic to each animal such as sex and age that can affect

the outcome of an experiment. These variations can sometimes override the effects of intended experimental perturbations. However, if one is aware of these variations from the outset of experimental design, one can adjust for these confounding effects with statistical analyses[134].

Sex – The most obvious controllable variable is the sex of the animal. Aside from organ-specific differences, there are many physiological differences prescribed by the presence of a Y-chromosome. Examples include differences in hormone levels[135], neuronal development and responses affecting behavior[136, 137], metabolism, and fat storage[138]. A particular hormonal influence that dramatically affects the interpretation of mouse studies is the use of tamoxifen to activate estrogen receptor-based inducible Cre recombinase, which can subsequently affect normal estrogen receptor function and hormonal balance. Tamoxifen use has also been associated with atrophy and metaplasia of the gastric epithelium[139], even in the absence of any floxed alleles.

Age – Aging is associated with cellular and physiological changes. As such, aging itself is a systems-level phenomenon that results from the integration of multiple inputs and outputs occurring at different levels. Changes in oxidative stress, endocrine functions, and metabolism occurring at the organismal level can have effects down to the cellular level. Stem cell function is drastically reduced in aged mice, resulting in defects in bone marrow[140] and intestinal regeneration[141].   The latter effect contributes to malabsorption of nutrients and subsequent changes in metabolism[142].   Thymus involution occurs in early adulthood resulting in the exhaustion of T cells later in life, a major contributor to immunosenescence[143]. Imbalance of the immune system with age can lead to chronic, low grade inflammation called "inflammaging"[144]. Changes in the immune system, diet, and gastrointestinal environment with age lead to alterations in the microbiome[145, 146]. These are a few specific examples, but it should be born in

27

mind that the multitude of changes that occur with age may have profound effects on many biological processes.

Strain differences – Mice from different strains and sub-strains are produced via multiple generations of inbreeding. Their overall genetic compositions can vary significantly, with some even harboring homozygous recessive alleles and mutations. As such, phenotypes observed in one mouse strain either at steady state or under perturbation are very likely to be influenced by its genetic background. Thus, comparing the effect of a specific perturbation should be performed under one background or with added caution if performed across multiple backgrounds. An example of a strain-specific effect is the susceptibility to dextran sodium sulfate-induced colitis due to variation in immunity genes[147, 148]. As described above, systems genetics and strain-specific differences can be leveraged to link suites of genes to changes in phenotype.

Microbiome – Recent research has shed light on the drastic influences of the microbiome on vertebrate organism physiology, including metabolism and immunity. The establishment of the microbiome is a function of not only inherent features such as genetics, but the environment where the organism is housed and raised as well. Mice from the same strain exhibit different biological behaviors if they are acquired through different sources. For example, C57BL6 mice exhibit divergent Th17 differentiation in the gut depending on whether they originate from Taconic Farms, Jackson Laboratories or Charles River[149]. Human studies suggest that the adult microbiome is relatively stable[16], and is established through a chaotic process during the first year of life where dietary richness and environmental exposure increase. Importantly, dizygotic twins show significant similarity in early temporal profile gut microbiome development, demonstrating the importance of early fostering[150]. There has been evidence suggesting that the

28

adult mouse microbiome is not as stable and can be changed within days[151]. Cage specific effects are strong and can account for up to 30% of variation. These effects can be reduced by mixing bedding and/or medium term co-housing (weeks), although these approaches may not result in perfect normalization[152]. A better strategy may be to begin with littermates or germ free mice followed by co-housing, although only a limited number of mice can be studied by this strategy.

Circadian Rhythm – The day-night cycle also greatly affects vertebrate organism physiology. For example, light-dark cycles neurologically affect feeding behaviors[153], and circadian genes like Bmal1 cyclically regulate cytokine secretion and immune function[154, 155]. Cyclic regulation of toll-like receptors can also affect bacterial pathogenesis and microbiome balance [156]. To control for these variations, *in vivo* experiments should be performed and repeated at similar stages of the circadian cycle.

Discussion

Technological advances during the latter half of the 20th century enabled scientists to interrogate biological systems down to the molecular level and ushered in a golden age of reductionist biomedicine. By understanding how genes and proteins are structured and function at the molecular level, we have been able to design targeted therapies for many clinically important diseases including cancer[157], inflammation[158], and infection[159]. While these reductionist approaches have been extremely successful in increasing life span and quality for many patients, there are many others for whom these treatments eventually fail. In colon cancer or IBD, treatments based on initiator genes and proteins often show periods of efficacy followed by relapse[160]. In some cases this relapse is triggered by compensatory mutations within the target gene or pathway, while in other cases, mutations activating independent pathways or network adaptation without mutation can drive resistance. Understanding and targeting these

compensatory mechanisms may prove effective in driving disease remission, with the caveat that these successes might be met with secondary resistance resulting in cyclical rounds of treatment and resistance.

Drugs that have even a partial effect or a short-term cure provide immense benefits to patients. However, better treatments and complete cures remain the goal for scientists, physicians and the patients they serve. For systems biologists, the best route to improving therapies is to view these diseases not from the perspective of the single genes and proteins that drive them, but from the integrative perspective of the whole disease network. This has been born out at the intracellular level using *in vitro* cell culture systems to map interaction networks and to identify network architectures that drive disease-related phenotypes. As our understanding of these architectures and the available tools for measuring nodes (whether they be at the level of RNA, protein or cell type) *in vivo* have improved, we are entering an era where we will possess an integrated understanding of disease and health at the organ and organismal level.

The field of systems biology, particularly *in vivo*, is still in its infancy and exactly how this type of understanding will translate to better therapies remains an open question. In the ideal scenario, computable tissue-level networks would be leveraged to design combination therapies capable of shifting a tissue from a disease state to a healthy one. Several systems biology driven companies have been founded based on the principles and techniques described above, and have achieved success in bringing drugs to the clinic based on *in vitro* systems modeling[161]. However, it is important to recognize that the field of systems biology is still growing and there are many limitations to reaching a fully realized systems understanding of any disease. At present, it is impossible to measure and meaningfully model all of the components of any given system. However, by incorporating as many of those system-level interactions into our understanding of

a given organ, we will be able to better design therapies that attack the disease network, and we will be able to better understand how the network responds to drive resistance.

Here, we have discussed some of the ways in which systems biology has been used to study intestinal disease in the mouse and some of the techniques that will drive this growing field forward in the future. This approach will be extremely useful in identifying paradigms of tissue-level systems behavior and to identify candidate networks that may be operative in human disease. Furthermore, systems-level understanding will enable us to calibrate our mouse models to determine which ones best recapitulate human tissue behavior and which parts of human networks are well-represented by a given model. This will be of benefit directly to systems biologists and to researchers utilizing mouse models for reductionist approaches. This knowledge can be leveraged to identify new therapeutic approaches for diseases affecting not only the intestine directly, but many diseases of distant organ systems that are caused or modified by intestinal health as well. While the field of *in vivo* systems biology is still in its infancy, there is every reason to hope that it will provide massive value not only to researchers, but also to patients suffering from diseases such as cancer, *C. difficile* infection and IBD.

Author Information

The authors declare no competing financial interests. Correspondence should be addressed to K.S.L. (ken.s.lau@vanderbilt.edu).

Acknowledgement

31

Abbreviations

IFN	Interferon
TNF	Tumor necrosis factor
ApcMin	Adenomatous polyposis coli multiple intestinal neoplasia
iTraq	Isobaric tag for relative and absolute quantitation
TMT	Tandem mass tag
ELISA	Enzyme-linked immunosorbent assay
SILAC	Stable isotope labeling by/with amino acids in cell culture
qPCR	Quantitative polymerase chain reaction
RNAseq	Ribonucleic acid sequencing
PICRUSt	Phylogenetic investigation of communities by reconstruction of unobserved states
CRISPR	Clustered regularly interspaced short palindromic repeats
Cas	CRISPR associated
t-SNE	t-distributed stochastic neighbor embedding
SPADE	Sequential pattern discovery using equivalence classes

References

1. A. L. Barabasi and Z. N. Oltvai, *Nat Rev Genet*, 2004, 5, 101-113.
2. J. R. Pritchard, D. A. Lauffenburger and M. T. Hemann, *Drug Resist Updat*, 2012, 15, 249-257.
3. J. Xu and J. I. Gordon, *Proc Natl Acad Sci U S A*, 2003, 100, 10452-10459.

4.   N. Ay, J. Flack and D. C. Krakauer, *Philos Trans R Soc Lond B Biol Sci*, 2007, 362, 441-
        447.
5.   J. Kim, D. Vandamme, J. R. Kim, A. G. Munoz, W. Kolch and K. H. Cho, *PLoS Comput
        Biol*, 2014, 10, e1003763.
6.   C. G. Buffie and E. G. Pamer, *Nat Rev Immunol*, 2013, 13, 790-801.
7.   Z. Wang, M. Gerstein and M. Snyder, *Nat Rev Genet*, 2009, 10, 57-63.
8.   N. Dephoure and S. P. Gygi, *Sci Signal*, 2012, 5, rs2.
9.   R. Wu, W. Haas, N. Dephoure, E. L. Huttlin, B. Zhai, M. E. Sowa and S. P. Gygi, *Nat
        Methods*, 2011, 8, 677-683.
10.   K. S. Lau, V. Cortez-Retamozo, S. R. Philips, M. J. Pittet, D. A. Lauffenburger and K. M.
        Haigis, *PLoS Biol*, 2012, 10, e1001393.
11.   K. S. Lau, A. M. Juchheim, K. R. Cavaliere, S. R. Philips, D. A. Lauffenburger and K. M.
        Haigis, *Sci Signal*, 2011, 4, ra16.
12.   P. M. Conn, *Animal models for the study of human disease*, Elsevier, London ;
        Waltham, MA, 2013.
13.   L. G. Albenberg and G. D. Wu, *Gastroenterology*, 2014, 146, 1564-1572.
14.   S. Nahon, P. Lahmek, C. Saas, C. Durance, A. Olympie, B. Lesgourgues and J. P. Gendre,
        Inflamm Bowel Dis, 2011, 17, 1270-1276.
15.   V. Pittet, G. Rogler, C. Mottet, F. Froehlich, P. Michetti, P. de Saussure, B. Burnand and
        J. P. Vader, *Scand J Gastroenterol*, 2014, 49, 662-673.
16.   G. D. Wu, J. Chen, C. Hoffmann, K. Bittinger, Y. Y. Chen, S. A. Keilbaugh, M. Bewtra, D.
        Knights, W. A. Walters, R. Knight, R. Sinha, E. Gilroy, K. Gupta, R. Baldassano, L.
        Nessel, H. Li, F. D. Bushman and J. D. Lewis, *Science*, 2011, 334, 105-108.
17.   M. Orholm, V. Binder, T. I. Sorensen, L. P. Rasmussen and K. O. Kyvik, *Scand J
        Gastroenterol*, 2000, 35, 1075-1081.
18.   L. Halme, P. Paavola-Sakki, U. Turunen, M. Lappalainen, M. Farkkila and K. Kontula,
        World J Gastroenterol, 2006, 12, 3668-3672.
19.   M. Horiya, S. Kakizaki, K. Teshigawara, Y. Kikuchi, T. Hashida, Y. Tomizawa, T. Iida, T.
        Masuo, H. Takagi and M. Mori, *World J Gastroenterol*, 2005, 11, 7547-7549.
20.   B. Khor, A. Gardet and R. J. Xavier, *Nature*, 2011, 474, 307-317.
21.   L. Jostins, S. Ripke, R. K. Weersma, R. H. Duerr, D. P. McGovern, K. Y. Hui, J. C. Lee, L.
        P. Schumm, Y. Sharma, C. A. Anderson, J. Essers, M. Mitrovic, K. Ning, I. Cleynen, E.
        Theatre, S. L. Spain, S. Raychaudhuri, P. Goyette, Z. Wei, C. Abraham, J. P. Achkar, T.
        Ahmad, L. Amininejad, A. N. Ananthakrishnan, V. Andersen, J. M. Andrews, L. Baidoo,
        T. Balschun, P. A. Bampton, A. Bitton, G. Boucher, S. Brand, C. Buning, A. Cohain, S.
        Cichon, M. D'Amato, D. De Jong, K. L. Devaney, M. Dubinsky, C. Edwards, D.
        Ellinghaus, L. R. Ferguson, D. Franchimont, K. Fransen, R. Gearry, M. Georges, C.
        Gieger, J. Glas, T. Haritunians, A. Hart, C. Hawkey, M. Hedl, X. Hu, T. H. Karlsen, L.
        Kupcinskas, S. Kugathasan, A. Latiano, D. Laukens, I. C. Lawrance, C. W. Lees, E.
        Louis, G. Mahy, J. Mansfield, A. R. Morgan, C. Mowat, W. Newman, O. Palmieri, C. Y.
        Ponsioen, U. Potocnik, N. J. Prescott, M. Regueiro, J. I. Rotter, R. K. Russell, J. D.
        Sanderson, M. Sans, J. Satsangi, S. Schreiber, L. A. Simms, J. Sventoraityte, S. R.
        Targan, K. D. Taylor, M. Tremelling, H. W. Verspaget, M. De Vos, C. Wijmenga, D. C.
        Wilson, J. Winkelmann, R. J. Xavier, S. Zeissig, B. Zhang, C. K. Zhang, H. Zhao, M. S.
        Silverberg, V. Annese, H. Hakonarson, S. R. Brant, G. Radford-Smith, C. G. Mathew, J.

D. Rioux, E. E. Schadt, M. J. Daly, A. Franke, M. Parkes, S. Vermeire, J. C. Barrett and J. H. Cho, *Nature*, 2012, 491, 119-124.

22.   W. A. Walters, Z. Xu and R. Knight, *FEBS Lett*, 2014, 588, 4223-4233.

23.   S. H. Oehlers, M. V. Flores, C. J. Hall, S. Swift, K. E. Crosier and P. S. Crosier, *Dis Model Mech*, 2011, 4, 832-841.

24.   I. C. Allen, *Dis Model Mech*, 2011, 4, 711-712.

25.   A. Fleming, J. Jankowski and P. Goldsmith, *Inflamm Bowel Dis*, 2010, 16, 1162-1172.

26.   Y. Yang, S. Tomkovich and C. Jobin, *Inflamm Bowel Dis*, 2014, 20, 956-966.

27.   J. R. Goldsmith and C. Jobin, *J Biomed Biotechnol*, 2012, 2012, 817341.

28.   R. S. Blackwood, R. P. Tarara, K. L. Christe, A. Spinner and N. W. Lerche, *Comp Med*, 2008, 58, 81-87.

29.   B. A. t'Hart, M. Vervoordeldonk, J. L. Heeney and P. P. Tak, *Gene Ther*, 2003, 10, 890-901.

30.   N. Barker, J. H. van Es, J. Kuipers, P. Kujala, M. van den Born, M. Cozijnsen, A. Haegebarth, J. Korving, H. Begthel, P. J. Peters and H. Clevers, *Nature*, 2007, 449, 1003-1007.

31.   A. Mizoguchi, *Prog Mol Biol Transl Sci*, 2012, 105, 263-320.

32.   M. K. Washington, A. E. Powell, R. Sullivan, J. P. Sundberg, N. Wright, R. J. Coffey and W. F. Dove, *Gastroenterology*, 2013, 144, 705-717.

33.   J. Seok, H. S. Warren, A. G. Cuenca, M. N. Mindrinos, H. V. Baker, W. Xu, D. R. Richards, G. P. McDonald-Smith, H. Gao, L. Hennessy, C. C. Finnerty, C. M. Lopez, S. Honari, E. E. Moore, J. P. Minei, J. Cuschieri, P. E. Bankey, J. L. Johnson, J. Sperry, A. B. Nathens, T. R. Billiar, M. A. West, M. G. Jeschke, M. B. Klein, R. L. Gamelli, N. S. Gibran, B. H. Brownstein, C. Miller-Graziano, S. E. Calvano, P. H. Mason, J. P. Cobb, L. G. Rahme, S. F. Lowry, R. V. Maier, L. L. Moldawer, D. N. Herndon, R. W. Davis, W. Xiao and R. G. Tompkins, *Proc Natl Acad Sci U S A*, 2013, 110, 3507-3512.

34.   A. Rangarajan, S. J. Hong, A. Gifford and R. A. Weinberg, *Cancer Cell*, 2004, 6, 171-183.

35.   E. S. Rinella, E. D. Bankaitis and D. W. Threadgill, *Cancer Biol Ther*, 2012, 13, 130-137.

36.   M. La Merrill, R. Harper, L. S. Birnbaum, R. D. Cardiff and D. W. Threadgill, *Environ Health Perspect*, 2010, 118, 596-601.

37.   R. R. Gordon, M. La Merrill, K. W. Hunter, P. Sorensen, D. W. Threadgill and D. Pomp, *Clin Exp Metastasis*, 2010, 27, 279-293.

38.   C. L. Powell, B. U. Bradford, C. P. Craig, M. Tsuchiya, T. Uehara, T. M. O'Connell, I. P. Pogribny, S. Melnyk, D. R. Koop, L. Bleyle, D. W. Threadgill and I. Rusyn, *Toxicol Sci*, 2010, 115, 131-139.

39.   M. La Merrill, B. S. Kuruvilla, D. Pomp, L. S. Birnbaum and D. W. Threadgill, *Environ Health Perspect*, 2009, 117, 1414-1419.

40.   I. Bozic, T. Antal, H. Ohtsuki, H. Carter, D. Kim, S. Chen, R. Karchin, K. W. Kinzler, B. Vogelstein and M. A. Nowak, *Proc Natl Acad Sci U S A*, 2010, 107, 18545-18550.

41.   W. Zhang, Q. D. Morris, R. Chang, O. Shai, M. A. Bakowski, N. Mitsakakis, N. Mohammad, M. D. Robinson, R. Zirngibl, E. Somogyi, N. Laurin, E. Eftekharpour, E. Sat, J. Grigull, Q. Pan, W. T. Peng, N. Krogan, J. Greenblatt, M. Fehlings, D. van der Kooy, J. Aubin, B. G. Bruneau, J. Rossant, B. J. Blencowe, B. J. Frey and T. R. Hughes, *J Biol*, 2004, 3, 21.

42. D. Leclerc, L. Deng, J. Trasler and R. Rozen, *J Cell Biochem*, 2004, 93, 1242-1254.
43. W. Huang da, B. T. Sherman and R. A. Lempicki, *Nat Protoc*, 2009, 4, 44-57.
44. W. Huang da, B. T. Sherman and R. A. Lempicki, *Nucleic Acids Res*, 2009, 37, 1-13.
45. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander and J. P. Mesirov, *Proc Natl Acad Sci U S A*, 2005, 102, 15545-15550.
46. A. J. Cornish and F. Markowetz, *PLoS Comput Biol*, 2014, 10, e1003808.
47. D. G. Rothwell, Y. Li, M. Ayub, C. Tate, G. Newton, Y. Hey, L. Carter, S. Faulkner, M. Moro, S. Pepper, C. Miller, F. Blackhall, G. Bertolini, L. Roz, C. Dive and G. Brady, *BMC Genomics*, 2014, 15, 1129.
48. K. A. Janes, C. C. Wang, K. J. Holmberg, K. Cabral and J. S. Brugge, *Nat Methods*, 2010, 7, 311-317.
49. D. Ramskold, S. Luo, Y. C. Wang, R. Li, Q. Deng, O. R. Faridani, G. A. Daniels, I. Khrebtukova, J. F. Loring, L. C. Laurent, G. P. Schroth and R. Sandberg, *Nat Biotechnol*, 2012, 30, 777-782.
50. V. Bhargava, S. R. Head, P. Ordoukhanian, M. Mercola and S. Subramaniam, *Sci Rep*, 2014, 4, 3678.
51. Q. Deng, D. Ramskold, B. Reinius and R. Sandberg, *Science*, 2014, 343, 193-196.
52. K. A. Janes, J. G. Albeck, L. X. Peng, P. K. Sorger, D. A. Lauffenburger and M. B. Yaffe, *Mol Cell Proteomics*, 2003, 2, 463-473.
53. M. Sevecka and G. MacBeath, *Nat Methods*, 2006, 3, 825-831.
54. M. F. Ciaccio, J. P. Wagner, C. P. Chuu, D. A. Lauffenburger and R. B. Jones, *Nat Methods*, 2010, 7, 148-155.
55. P. K. Kreeger, R. Mandhana, S. K. Alford, K. M. Haigis and D. A. Lauffenburger, *Cancer Res*, 2009, 69, 8191-8199.
56. M. E. McBee, Y. Zeng, N. Parry, C. R. Nagler, S. R. Tannenbaum and D. B. Schauer, *PLoS One*, 2010, 5, e13277.
57. E. R. Miraldi, H. Sharfi, R. H. Friedline, H. Johnson, T. Zhang, K. S. Lau, H. J. Ko, T. G. Curran, K. M. Haigis, M. B. Yaffe, R. Bonneau, D. A. Lauffenburger, B. B. Kahn, J. K. Kim, B. G. Neel, A. Saghatelian and F. M. White, *Integr Biol (Camb)*, 2013, 5, 940-963.
58. S. Zanivan, A. Meves, K. Behrendt, E. M. Schoof, L. J. Neilson, J. Cox, H. R. Tang, G. Kalna, J. H. van Ree, J. M. van Deursen, C. S. Trempus, L. M. Machesky, R. Linding, S. A. Wickstrom, R. Fassler and M. Mann, *Cell Rep*, 2013, 3, 552-566.
59. A. Hammoudi, F. Song, K. R. Reed, R. E. Jenkins, V. S. Meniel, A. J. Watson, D. M. Pritchard, A. R. Clarke and J. R. Jenkins, *Biochem Biophys Res Commun*, 2013, 440, 364-370.
60. K. Kito and T. Ito, *Curr Genomics*, 2008, 9, 263-274.
61. N. Korneychuk, E. Ramiro-Puig, J. Ettersperger, J. Schulthess, N. Montcuquet, H. Kiyono, B. Meresse and N. Cerf-Bensussan, *Gastroenterology*, 2014, 146, 1017-1027.
62. K. Atarashi, T. Tanoue, K. Oshima, W. Suda, Y. Nagano, H. Nishikawa, S. Fukuda, T. Saito, S. Narushima, K. Hase, S. Kim, J. V. Fritz, P. Wilmes, S. Ueha, K. Matsushima, H. Ohno, B. Olle, S. Sakaguchi, T. Taniguchi, H. Morita, M. Hattori and K. Honda, *Nature*, 2013, 500, 232-236.
63. A. Cebula, M. Seweryn, G. A. Rempala, S. S. Pabla, R. A. McIndoe, T. L. Denning, L. Bry, P. Kraj, P. Kisielow and L. Ignatowicz, *Nature*, 2013, 497, 258-262.

64.   M. Shan, M. Gentile, J. R. Yeiser, A. C. Walland, V. U. Bornstein, K. Chen, B. He, L. Cassis, A. Bigas, M. Cols, L. Comerma, B. Huang, J. M. Blander, H. Xiong, L. Mayer, C. Berin, L. H. Augenlicht, A. Velcich and A. Cerutti, *Science*, 2013, 342, 447-453.

65.   C. S. Klose, E. A. Kiss, V. Schwierzeck, K. Ebert, T. Hoyler, Y. d'Hargues, N. Goppert, A. L. Croxford, A. Waisman, Y. Tanriver and A. Diefenbach, *Nature*, 2013, 494, 261-265.

66.   C. S. Klose, M. Flach, L. Mohle, L. Rogell, T. Hoyler, K. Ebert, C. Fabiunke, D. Pfeifer, V. Sexl, D. Fonseca-Pereira, R. G. Domingues, H. Veiga-Fernandes, S. J. Arnold, M. Busslinger, I. R. Dunay, Y. Tanriver and A. Diefenbach, *Cell*, 2014, 157, 340-356.

67.   M. R. Hepworth, L. A. Monticelli, T. C. Fung, C. G. Ziegler, S. Grunberg, R. Sinha, A. R. Mantegazza, H. L. Ma, A. Crawford, J. M. Angelosanto, E. J. Wherry, P. A. Koni, F. D. Bushman, C. O. Elson, G. Eberl, D. Artis and G. F. Sonnenberg, *Nature*, 2013, 498, 113-117.

68.   A. Mortha, A. Chudnovskiy, D. Hashimoto, M. Bogunovic, S. P. Spencer, Y. Belkaid and M. Merad, *Science*, 2014, 343, 1249288.

69.   A. D. Gracz, M. K. Fuller, F. Wang, L. Li, M. Stelzner, J. C. Dunn, M. G. Martin and S. T. Magness, *Stem Cells*, 2013, 31, 2024-2030.

70.   S. T. Magness, B. J. Puthoff, M. A. Crissey, J. Dunn, S. J. Henning, C. Houchen, J. S. Kaddis, C. J. Kuo, L. Li, J. Lynch, M. G. Martin, R. May, J. C. Niland, B. Olack, D. Qian, M. Stelzner, J. R. Swain, F. Wang, J. Wang, X. Wang, K. Yan, J. Yu and M. H. Wong, *Am J Physiol Gastrointest Liver Physiol*, 2013, 305, G542-551.

71.   P. Dalerba, S. J. Dylla, I. K. Park, R. Liu, X. Wang, R. W. Cho, T. Hoey, A. Gurney, E. H. Huang, D. M. Simeone, A. A. Shelton, G. Parmiani, C. Castelli and M. F. Clarke, *Proc Natl Acad Sci U S A*, 2007, 104, 10158-10163.

72.   S. C. Bendall, E. F. Simonds, P. Qiu, A. D. Amir el, P. O. Krutzik, R. Finck, R. V. Bruggner, R. Melamed, A. Trejo, O. I. Ornatsky, R. S. Balderas, S. K. Plevritis, K. Sachs, D. Pe'er, S. D. Tanner and G. P. Nolan, *Science*, 2011, 332, 687-696.

73.   G. K. Geiss, R. E. Bumgarner, B. Birditt, T. Dahl, N. Dowidar, D. L. Dunaway, H. P. Fell, S. Ferree, R. D. George, T. Grogan, J. J. James, M. Maysuria, J. D. Mitton, P. Oliveri, J. L. Osborn, T. Peng, A. L. Ratcliffe, P. J. Webster, E. H. Davidson, L. Hood and K. Dimitrov, *Nat Biotechnol*, 2008, 26, 317-325.

74.   A. V. Ullal, V. Peterson, S. S. Agasti, S. Tuang, D. Juric, C. M. Castro and R. Weissleder, *Sci Transl Med*, 2014, 6, 219ra219.

75.   D. A. Nelson, C. Manhardt, V. Kamath, Y. Sui, A. Santamaria-Pang, A. Can, M. Bello, A. Corwin, S. R. Dinn, M. Lazare, E. M. Gervais, S. J. Sequeira, S. B. Peters, F. Ginty, M. J. Gerdes and M. Larsen, *Biol Open*, 2013, 2, 439-447.

76.   M. J. Gerdes, C. J. Sevinsky, A. Sood, S. Adak, M. O. Bello, A. Bordwell, A. Can, A. Corwin, S. Dinn, R. J. Filkins, D. Hollman, V. Kamath, S. Kaanumalle, K. Kenny, M. Larsen, M. Lazare, Q. Li, C. Lowes, C. C. McCulloch, E. McDonough, M. C. Montalto, Z. Pang, J. Rittscher, A. Santamaria-Pang, B. D. Sarachan, M. L. Seel, A. Seppo, K. Shaikh, Y. Sui, J. Zhang and F. Ginty, *Proc Natl Acad Sci U S A*, 2013, 110, 11982-11987.

77.   C. Giesen, H. A. Wang, D. Schapiro, N. Zivanovic, A. Jacobs, B. Hattendorf, P. J. Schuffler, D. Grolimund, J. M. Buhmann, S. Brandt, Z. Varga, P. J. Wild, D. Gunther and B. Bodenmiller, *Nat Methods*, 2014, 11, 417-422.

78.   M. Angelo, S. C. Bendall, R. Finck, M. B. Hale, C. Hitzman, A. D. Borowsky, R. M. Levenson, J. B. Lowe, S. D. Liu, S. Zhao, Y. Natkunam and G. P. Nolan, *Nat Med*, 2014, 20, 436-442.

79.   E. A. Susaki, K. Tainaka, D. Perrin, F. Kishino, T. Tawara, T. M. Watanabe, C. Yokoyama, H. Onoe, M. Eguchi, S. Yamaguchi, T. Abe, H. Kiyonari, Y. Shimizu, A. Miyawaki, H. Yokota and H. R. Ueda, *Cell*, 2014, 157, 726-739.

80.   K. Chung, J. Wallace, S. Y. Kim, S. Kalyanasundaram, A. S. Andalman, T. J. Davidson, J. J. Mirzabekov, K. A. Zalocusky, J. Mattis, A. K. Denisin, S. Pak, H. Bernstein, C. Ramakrishnan, L. Grosenick, V. Gradinaru and K. Deisseroth, *Nature*, 2013, 497, 332-337.

81.   S. Itzkovitz, A. Lyubimova, I. C. Blat, M. Maynard, J. van Es, J. Lees, T. Jacks, H. Clevers and A. van Oudenaarden, *Nat Cell Biol*, 2012, 14, 106-114.

82.   J. H. Lee, E. R. Daugharthy, J. Scheiman, R. Kalhor, J. L. Yang, T. C. Ferrante, R. Terry, S. S. Jeanty, C. Li, R. Amamoto, D. T. Peters, B. M. Turczyk, A. H. Marblestone, S. A. Inverso, A. Bernard, P. Mali, X. Rios, J. Aach and G. M. Church, *Science*, 2014, 343, 1360-1363.

83.   E. Lubeck and L. Cai, *Nat Methods*, 2012, 9, 743-748.

84.   E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad and L. Cai, *Nat Methods*, 2014, 11, 360-361.

85.   D. Kobat, N. G. Horton and C. Xu, *J Biomed Opt*, 2011, 16, 106014.

86.   J. W. Choi, J. K. Kim, M. Choi, Y. R. Kim and S. H. Yun, *Endoscopy*, 2014, 46, 1110-1116.

87.   P. Kim, E. Chung, H. Yamashita, K. E. Hung, A. Mizoguchi, R. Kucherlapati, D. Fukumura, R. K. Jain and S. H. Yun, *Nat Methods*, 2010, 7, 303-305.

88.   J. G. Albeck, G. B. Mills and J. S. Brugge, *Mol Cell*, 2013, 49, 249-261.

89.   O. Basak, M. van de Born, J. Korving, J. Beumer, S. van der Elst, J. H. van Es and H. Clevers, *Embo J*, 2014, 33, 2057-2068.

90.   A. Sakaue-Sawano and A. Miyawaki, *Cold Spring Harb Protoc*, 2014, 2014.

91.   L. Ritsma, S. I. Ellenbroek, A. Zomer, H. J. Snippert, F. J. de Sauvage, B. D. Simons, H. Clevers and J. van Rheenen, *Nature*, 2014, 507, 362-365.

92.   C. Lopez-Garcia, A. M. Klein, B. D. Simons and D. J. Winton, *Science*, 2010, 330, 822-825.

93.   A. S. Bhatt, S. S. Freeman, A. F. Herrera, C. S. Pedamallu, D. Gevers, F. Duke, J. Jung, M. Michaud, B. J. Walker, S. Young, A. M. Earl, A. D. Kostic, A. I. Ojesina, R. Hasserjian, K. K. Ballen, Y. B. Chen, G. Hobbs, J. H. Antin, R. J. Soiffer, L. R. Baden, W. S. Garrett, J. L. Hornick, F. M. Marty and M. Meyerson, *N Engl J Med*, 2013, 369, 517-528.

94.   F. H. Karlsson, V. Tremaroli, I. Nookaew, G. Bergstrom, C. J. Behre, B. Fagerberg, J. Nielsen and F. Backhed, *Nature*, 2013, 498, 99-103.

95.   M. D. Schulz, C. Atay, J. Heringer, F. K. Romrig, S. Schwitalla, B. Aydin, P. K. Ziegler, J. Varga, W. Reindl, C. Pommerenke, G. Salinas-Riester, A. Bock, C. Alpert, M. Blaut, S. C. Polson, L. Brandl, T. Kirchner, F. R. Greten, S. W. Polson and M. C. Arkan, *Nature*, 2014, 514, 508-512.

96.   O. Koren, J. K. Goodrich, T. C. Cullender, A. Spor, K. Laitinen, H. K. Backhed, A. Gonzalez, J. J. Werner, L. T. Angenent, R. Knight, F. Backhed, E. Isolauri, S. Salminen and R. E. Ley, *Cell*, 2012, 150, 470-480.

97.   B. Pyndt Jorgensen, L. Krych, T. B. Pedersen, N. Plath, J. P. Redrobe, A. K. Hansen, D. S. Nielsen, C. S. Pedersen, C. Larsen and D. B. Sorensen, *Physiol Behav*, 2014, 141C, 32-39.

98.   P. J. Turnbaugh, V. K. Ridaura, J. J. Faith, F. E. Rey, R. Knight and J. I. Gordon, *Sci Transl Med*, 2009, 1, 6ra14.

99.   F. P. Martin, Y. Wang, N. Sprenger, I. K. Yap, S. Rezzi, Z. Ramadan, E. Pere-Trepat, F. Rochat, C. Cherbut, P. van Bladeren, L. B. Fay, S. Kochhar, J. C. Lindon, E. Holmes and J. K. Nicholson, *Mol Syst Biol*, 2008, 4, 205.

100.   H. Chung, S. J. Pamp, J. A. Hill, N. K. Surana, S. M. Edelman, E. B. Troy, N. C. Reading, E. J. Villablanca, S. Wang, J. R. Mora, Y. Umesaki, D. Mathis, C. Benoist, D. A. Relman and D. L. Kasper, *Cell*, 2012, 149, 1578-1593.

101.   S. M. Huse, D. M. Welch, H. G. Morrison and M. L. Sogin, *Environ Microbiol*, 2010, 12, 1889-1898.

102.   S. P. Preheim, A. R. Perrotta, J. Friedman, C. Smilie, I. Brito, M. B. Smith and E. Alm, *Methods Enzymol*, 2013, 531, 353-370.

103.   L. J. Forney, X. Zhou and C. J. Brown, *Curr Opin Microbiol*, 2004, 7, 210-220.

104.   C. Huttenhower, A. D. Kostic and R. J. Xavier, *Immunity*, 2014, 40, 843-854.

105.   P. M. Smith, M. R. Howitt, N. Panikov, M. Michaud, C. A. Gallini, Y. M. Bohlooly, J. N. Glickman and W. S. Garrett, *Science*, 2013, 341, 569-573.

106.   E. Suply, P. de Vries, R. Soret, F. Cossais and M. Neunlist, *Am J Physiol Gastrointest Liver Physiol*, 2012, 302, G1373-1380.

107.   E. Le Chatelier, T. Nielsen, J. Qin, E. Prifti, F. Hildebrand, G. Falony, M. Almeida, M. Arumugam, J. M. Batto, S. Kennedy, P. Leonard, J. Li, K. Burgdorf, N. Grarup, T. Jorgensen, I. Brandslund, H. B. Nielsen, A. S. Juncker, M. Bertalan, F. Levenez, N. Pons, S. Rasmussen, S. Sunagawa, J. Tap, S. Tims, E. G. Zoetendal, S. Brunak, K. Clement, J. Dore, M. Kleerebezem, K. Kristiansen, P. Renault, T. Sicheritz-Ponten, W. M. de Vos, J. D. Zucker, J. Raes, T. Hansen, P. Bork, J. Wang, S. D. Ehrlich and O. Pedersen, *Nature*, 2013, 500, 541-546.

108.   M. G. Langille, J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. Vega Thurber, R. Knight, R. G. Beiko and C. Huttenhower, *Nat Biotechnol*, 2013, 31, 814-821.

109.   G. D. Wu, C. Compher, E. Z. Chen, S. A. Smith, R. D. Shah, K. Bittinger, C. Chehoud, L. G. Albenberg, L. Nessel, E. Gilroy, J. Star, A. M. Weljie, H. J. Flint, D. C. Metz, M. J. Bennett, H. Li, F. D. Bushman and J. D. Lewis, *Gut*, 2014.

110.   X. Peer and G. An, *J Pharmacokinet Pharmacodyn*, 2014, 41, 493-507.

111.   V. Shankar, M. J. Hamilton, A. Khoruts, A. Kilburn, T. Unno, O. Paliy and M. J. Sadowsky, *Microbiome*, 2014, 2, 13.

112.   E. van Nood, A. Vrieze, M. Nieuwdorp, S. Fuentes, E. G. Zoetendal, W. M. de Vos, C. E. Visser, E. J. Kuijper, J. F. Bartelsman, J. G. Tijssen, P. Speelman, M. G. Dijkgraaf and J. J. Keller, *N Engl J Med*, 2013, 368, 407-415.

113.   E. Gough, H. Shaikh and A. R. Manges, *Clin Infect Dis*, 2011, 53, 994-1002.

114.   I. Youngster, J. Sauk, C. Pindar, R. G. Wilson, J. L. Kaplan, M. B. Smith, E. J. Alm, D. Gevers, G. H. Russell and E. L. Hohmann, *Clin Infect Dis*, 2014, 58, 1515-1522.

115.   C. G. Buffie, V. Bucci, R. R. Stein, P. T. McKenney, L. Ling, A. Gobourne, D. No, H. Liu, M. Kinnebrew, A. Viale, E. Littmann, M. R. van den Brink, R. R. Jenq, Y. Taur, C. Sander, J. R. Cross, N. C. Toussaint, J. B. Xavier and E. G. Pamer, *Nature*, 2015, 517, 205-208.

116.   J. J. Gierut, T. E. Jacks and K. M. Haigis, *Cold Spring Harb Protoc*, 2014, 2014, 339-349.

117.   S. Beronja, P. Janki, E. Heller, W. H. Lien, B. E. Keyes, N. Oshimori and E. Fuchs, *Nature*, 2013, 501, 185-190.

118.   L. Zhang, T. Zhang, L. Wang, S. Shao, Z. Chen and Z. Zhang, *Gene Ther*, 2014, 21, 709-714.

119. H. Wang, H. Yang, C. S. Shivalila, M. M. Dawlaty, A. W. Cheng, F. Zhang and R. Jaenisch, *Cell*, 2013, 153, 910-918.
120. D. Quigley and A. Balmain, *Nat Rev Genet*, 2009, 10, 651-657.
121. G. Peltz, A. K. Zaas, M. Zheng, N. V. Solis, M. X. Zhang, H. H. Liu, Y. Hu, G. M. Boxx, Q. T. Phan, D. Dill and S. G. Filler, *Infect Immun*, 2011, 79, 4472-4479.
122. C. E. Welsh, D. R. Miller, K. F. Manly, J. Wang, L. McMillan, G. Morahan, R. Mott, F. A. Iraqi, D. W. Threadgill and F. P. de Villena, *Mamm Genome*, 2012, 23, 706-712.
123. I. T. Jolliffe, *Principal component analysis*, Springer, New York, 2nd edn., 2002.
124. K. A. Janes, J. G. Albeck, S. Gaudet, P. K. Sorger, D. A. Lauffenburger and M. B. Yaffe, *Science*, 2005, 310, 1646-1653.
125. K. A. Janes, S. Gaudet, J. G. Albeck, U. B. Nielsen, D. A. Lauffenburger and P. K. Sorger, *Cell*, 2006, 124, 1225-1239.
126. A. D. Amir el, K. L. Davis, M. D. Tadmor, E. F. Simonds, J. H. Levine, S. C. Bendall, D. K. Shenfeld, S. Krishnaswamy, G. P. Nolan and D. Pe'er, *Nat Biotechnol*, 2013, 31, 545-552.
127. P. Qiu, E. F. Simonds, S. C. Bendall, K. D. Gibbs, Jr., R. V. Bruggner, M. D. Linderman, K. Sachs, G. P. Nolan and S. K. Plevritis, *Nat Biotechnol*, 2011, 29, 886-891.
128. K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger and G. P. Nolan, *Science*, 2005, 308, 523-529.
129. J. P. Wagner, A. Wolf-Yadlin, M. Sevecka, J. K. Grenier, D. E. Root, D. A. Lauffenburger and G. MacBeath, *Sci Signal*, 2013, 6, ra58.
130. K. S. Brown and J. P. Sethna, *Phys Rev E Stat Nonlin Soft Matter Phys*, 2003, 68, 021904.
131. V. Vyshemirsky and M. Girolami, *Bioinformatics*, 2008, 24, 1933-1934.
132. M. K. Morris, J. Saez-Rodriguez, D. C. Clarke, P. K. Sorger and D. A. Lauffenburger, *PLoS Comput Biol*, 2011, 7, e1001099.
133. C. Terfve, T. Cokelaer, D. Henriques, A. MacNamara, E. Goncalves, M. K. Morris, M. van Iersel, D. A. Lauffenburger and J. Saez-Rodriguez, *BMC Syst Biol*, 2012, 6, 133.
134. A. Schneider, G. Hommel and M. Blettner, *Dtsch Arztebl Int*, 2010, 107, 776-782.
135. E. L. Ding, Y. Song, V. S. Malik and S. Liu, *Jama*, 2006, 295, 1288-1299.
136. B. Cooke, C. D. Hegstrom, L. S. Villeneuve and S. M. Breedlove, *Front Neuroendocrinol*, 1998, 19, 323-362.
137. L. M. Monteggia, B. Luikart, M. Barrot, D. Theobold, I. Malkovska, S. Nef, L. F. Parada and E. J. Nestler, *Biol Psychiatry*, 2007, 61, 187-197.
138. C. T. Montague, J. B. Prins, L. Sanders, J. E. Digby and S. O'Rahilly, *Diabetes*, 1997, 46, 342-347.
139. W. J. Huh, S. S. Khurana, J. H. Geahlen, K. Kohli, R. A. Waller and J. C. Mills, *Gastroenterology*, 2012, 142, 21-24 e27.
140. S. M. Chambers, C. A. Shaw, C. Gatza, C. J. Fisk, L. A. Donehower and M. A. Goodell, *PLoS Biol*, 2007, 5, e201.
141. F. Sipos, K. Leiszter and Z. Tulassay, *World J Gastroenterol*, 2011, 17, 2981-2986.
142. L. Drozdowski and A. B. Thomson, *World J Gastroenterol*, 2006, 12, 7578-7584.
143. D. P. Shanley, D. Aw, N. R. Manley and D. B. Palmer, *Trends Immunol*, 2009, 30, 374-381.

144. C. Franceschi, M. Capri, D. Monti, S. Giunta, F. Olivieri, F. Sevini, M. P. Panourgia, L. Invidia, L. Celani, M. Scurti, E. Cevenini, G. C. Castellani and S. Salvioli, *Mech Ageing Dev*, 2007, 128, 92-105.
145. E. B. Hollister, C. Gao and J. Versalovic, *Gastroenterology*, 2014, 146, 1449-1458.
146. T. Yatsunenko, F. E. Rey, M. J. Manary, I. Trehan, M. G. Dominguez-Bello, M. Contreras, M. Magris, G. Hidalgo, R. N. Baldassano, A. P. Anokhin, A. C. Heath, B. Warner, J. Reeder, J. Kuczynski, J. G. Caporaso, C. A. Lozupone, C. Lauber, J. C. Clemente, D. Knights, R. Knight and J. I. Gordon, *Nature*, 2012, 486, 222-227.
147. M. Mahler, I. J. Bristol, E. H. Leiter, A. E. Workman, E. H. Birkenmeier, C. O. Elson and J. P. Sundberg, *Am J Physiol*, 1998, 274, G544-551.
148. R. Suzuki, H. Kohno, S. Sugie, H. Nakagama and T. Tanaka, *Carcinogenesis*, 2006, 27, 162-169.
149. Ivanov, II, L. Frutos Rde, N. Manel, K. Yoshinaga, D. B. Rifkin, R. B. Sartor, B. B. Finlay and D. R. Littman, *Cell Host Microbe*, 2008, 4, 337-349.
150. C. Palmer, E. M. Bik, D. B. DiGiulio, D. A. Relman and P. O. Brown, *PLoS Biol*, 2007, 5, e177.
151. R. N. Carmody, G. K. Gerber, J. M. Luevano, Jr., D. M. Gatti, L. Somes, K. L. Svenson and P. J. Turnbaugh, *Cell Host Microbe*, 2015, 17, 72-84.
152. F. Hildebrand, T. L. Nguyen, B. Brinkman, R. G. Yunta, B. Cauwe, P. Vandenabeele, A. Liston and J. Raes, *Genome Biol*, 2013, 14, R4.
153. D. F. Patton and R. E. Mistlberger, *Front Neurosci*, 2013, 7, 185.
154. A. M. Curtis, M. M. Bellet, P. Sassone-Corsi and L. A. O'Neill, *Immunity*, 2014, 40, 178-186.
155. K. D. Nguyen, S. J. Fentress, Y. Qiu, K. Yun, J. S. Cox and A. Chawla, *Science*, 2013, 341, 1483-1488.
156. A. Mukherji, A. Kobiita, T. Ye and P. Chambon, *Cell*, 2013, 153, 812-827.
157. P. B. Chapman, A. Hauschild, C. Robert, J. B. Haanen, P. Ascierto, J. Larkin, R. Dummer, C. Garbe, A. Testori, M. Maio, D. Hogg, P. Lorigan, C. Lebbe, T. Jouary, D. Schadendorf, A. Ribas, S. J. O'Day, J. A. Sosman, J. M. Kirkwood, A. M. Eggermont, B. Dreno, K. Nolop, J. Li, B. Nelson, J. Hou, R. J. Lee, K. T. Flaherty and G. A. McArthur, *N Engl J Med*, 2011, 364, 2507-2516.
158. M. J. Elliott, R. N. Maini, M. Feldmann, J. R. Kalden, C. Antoni, J. S. Smolen, B. Leeb, F. C. Breedveld, J. D. Macfarlane, H. Bijl and et al., *Lancet*, 1994, 344, 1105-1110.
159. R. D. Moore and R. E. Chaisson, *Aids*, 1999, 13, 1933-1942.
160. R. Altwegg and T. Vincent, *Mediators Inflamm*, 2014, 2014, 172821.
161. S. Wang, J. Huang, H. Lyu, B. Cai, X. Yang, F. Li, J. Tan, S. M. Edgerton, A. D. Thor, C. K. Lee and B. Liu, *Breast Cancer Res*, 2013, 15, R101.
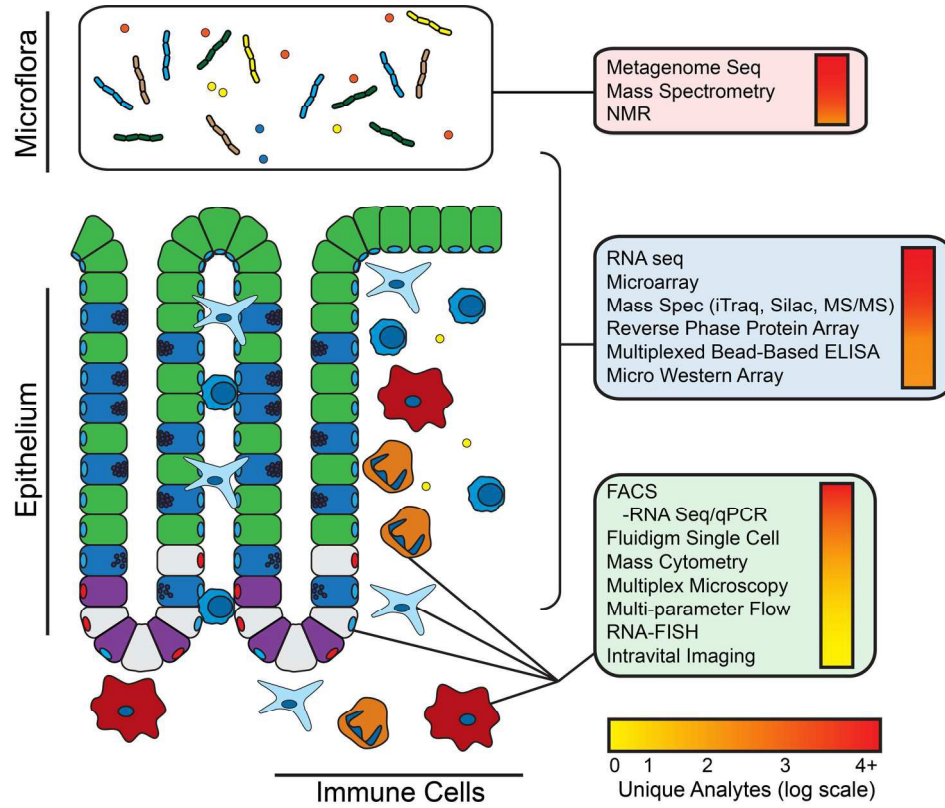
Figure Legends

Figure 1: Gut function and disease is governed by interactions between the epithelium, the immune system and the gut microflora. These interactions are mediated by cell-cell contact, cytokines, metabolites, and microbial products. We propose that an integrative systems approach that engages with this complexity is necessary to fully understand these interactions and their impact on gut homeostasis and disease. This figure lists some of the techniques that can be applied to systems-level studies in the gut. Techniques in the red box are amenable to studies of the microflora; techniques in the blue box are suitable for bulk measurements encompassing all components of the gut ecosystem; techniques in the green box are suitable for single cell or population-based analysis. Experimental approaches are listed in descending order of coverage with the number of unique analytes denoted by the color bar.

Figure 2: Increased mechanistic resolution of mathematical models require larger amounts of data to constrain each model parameter. The ratio of the number of parameters per model to number of experimental data points to build each model from a collection of mathematical models in the literature.
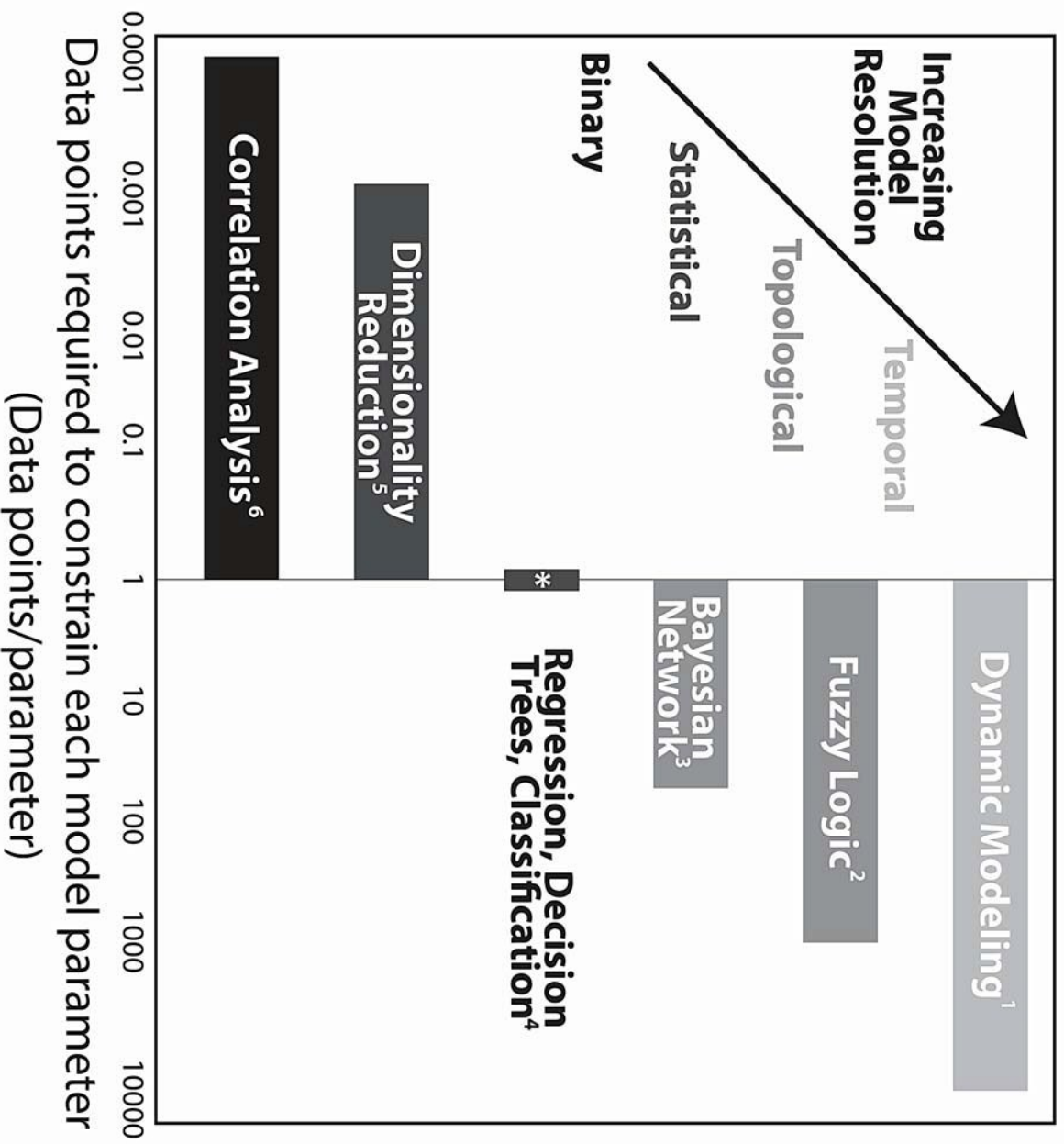
1.   G. Neuert, B. Munsky, R. Z. Tan, L. Teytelman, M. Khammash and A. van Oudenaarden, *Science*, 2013, **339**, 584–7.

2.   M. K. Morris, J. Saez-Rodriguez, D. C. Clarke, P. K. Sorger and D. A. Lauffenburger, *PLoS Comput. Biol.*, 2011, **7**, e1001099.

3.   K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger and G. P. Nolan, *Science*, 2005, **308**, 523–9.

4.   S. Hautaniemi, S. Kharait, A. Iwabu, A. Wells and D. a. Lauffenburger, *Bioinformatics*, 2005, **21**, 2027–2035.

5. K. S. Lau, S. B. Schrier, J. Gierut, J. Lyons, D. A. Lauffenburger and K. M. Haigis, *Integr. Biol. (Camb).*, 2013, **5**, 1355–65.

6. W. Zhang, Q. D. Morris, R. Chang, O. Shai, M. a Bakowski, N. Mitsakakis, N. Mohammad, M. D. Robinson, R. Zirngibl, E. Somogyi, N. Laurin, E. Eftekharpour, E. Sat, J. Grigull, Q. Pan, W.-T. Peng, N. Krogan, J. Greenblatt, M. Fehlings, D. van der Kooy, J. Aubin, B. G. Bruneau, J. Rossant, B. J. Blencowe, B. J. Frey and T. R. Hughes, *J. Biol.*, 2004, **3**, 21.
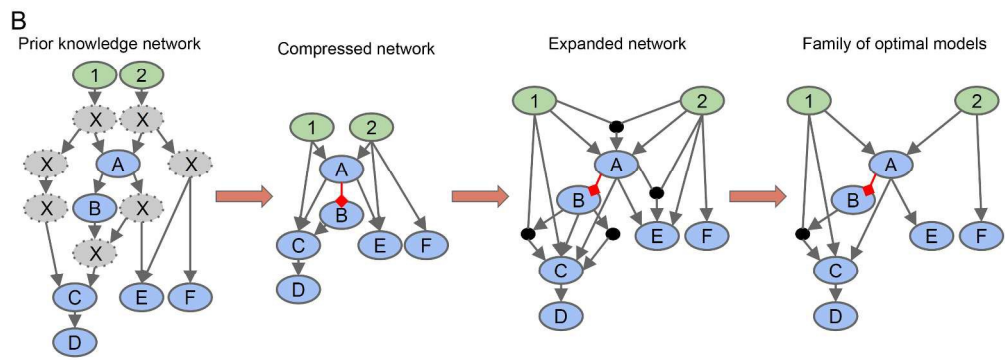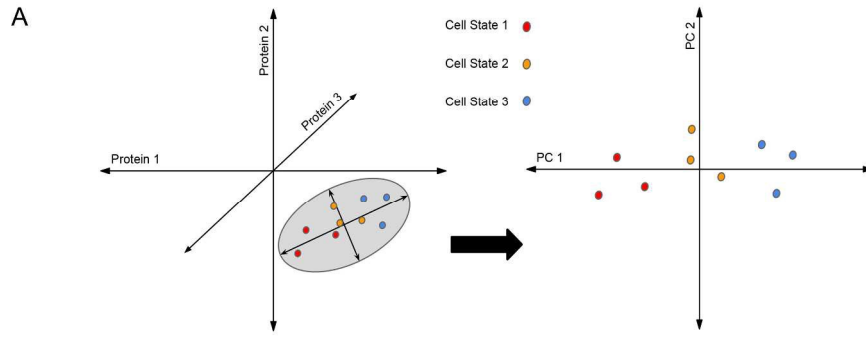
Figure 3: Quantitative modeling of multidimensional data. (A) Principal component analysis can be used for interpreting multivariate data by reducing dimensionality. In this simple example, cells are plotted (left plot) in 3-D according to expression levels of three proteins. The grey ellipse encircles the axes that contain the most variation in the data, the principal components. The data aligned to the principal components in 2-D (right plot) allows for easier visualization of the difference(s) in the data. In this case, the three cell states are more easily identified. (B) Logic-based modeling can be used for constructing a family of optimal models from a prior knowledge network (PKN) with a given set of inputs (green nodes) and observables (blue nodes). The PKN is compressed by removing all unobservables (grey nodes) while maintaining connectivity of nodes downstream. Grey arrows represent activation, while red arrows represent inhibition. To allow for all possible connections, the network is expanded. A family of optimal models is derived using experimental data.

2

173x140mm (300 x 300 DPI)

254x182mm (300 x 300 DPI)

Model	Disease Modeled	Location	Details
Inflammatory Bowel Disease			
DSS	Chemical damage	Colon	Epithelial cell toxicity followed by bacterial penetration and inflammatory reaction[1]
TNBS	Chemical damage	Colon	Mucosal barrier damage followed by increased antigenicity[2]
T-cell transfer	Ulcerative colitis or Crohn's disease	Colon and Ileum	1-3 month latency, high penetrance, severe mucosal inflammation. Crohn's-like cytokine profile; UC-like hyperproliferation, goblet cell loss[3]
IL-10$^{-/-}$	Crohn's disease (Colitis)	Colon and cecum	Variable mucosal inflammation dependent on microbiota, epithelial hyperplasia, inflammatory cell infiltration, loss of goblet cells[4]
TNF$^{\Delta ARE/+}$	Crohn's disease (Ileitis)	Terminal ileum (proximal colon)	Severe ileal transmural inflammation, mucosal/submucosal infiltration of inflammatory cells, increased goblet cells[5]
Colorectal Cancer			
AOM/DSS	Colonic adenocarcinoma	Colon	Tubular adenoma, dysplasia, and colitis with mucosal ulceration[6]
Lrig1^{CreERT2}/Apc$^{flox/+}$	Adenoma and adenocarcinoma	Distal colon (small intestine)	Mainly distal colonic adenomas with high-grade dysplasia[7]
Apc$^{Min/+}$	Adenoma	Small intestine and colon	Low-grade adenoma mainly in the small intestine[8]
Fabpl-Cre;Apc$^{flox/+}$;LSL-Kras$^{G12D/+}$	Adenoma and adenocarcinoma	Distal colon	High-grade dysplasia[9]
Apc$^{flox/+}$;LSL-K-ras$^{G12D/+}$; Adeno-cre after colonic abrasion	Adenoma and adenocarcinoma	Distal colon	Colonic tumors with high grade dysplasia and metastasis[10]
Infection Models			
Citrobacter rodentium	Enteropathogenic E. coli (EPEC) and Enterohemorrhagic E. coli (EHEC)	Colon and cecum	Crypt hyperplasia, goblet cell loss, mucosal infiltration, diarrhea[11]
Salmonella-induced colitis (with antibiotic treatment)	Early (4-72 hours) intestinal events in salmonella infection	Colon and cecum	Mainly in cecum, epithelial hyperplasia and erosion, acute inflammation, goblet cell lost[12]

Table 1: Common mouse models of intestinal pathologies.

1. I. Okayasu, S. Hatakeyama, M. Yamada, T. Ohkusa, Y. Inagaki and R. Nakaya, *Gastroenterology*, 1990, 98, 694-702.
2. M. F. Neurath, I. Fuss, B. L. Kelsall, E. Stuber and W. Strober, *J Exp Med*, 1995, 182, 1281-1290.
3. D. V. Ostanin, J. Bao, I. Koboziev, L. Gray, S. A. Robinson-Jackson, M. Kosloski-Davidson, V. H. Price and M. B. Grisham, *Am J Physiol Gastrointest Liver Physiol*, 2009, 296, G135-146.
4. R. Kuhn, J. Lohler, D. Rennick, K. Rajewsky and W. Muller, *Cell*, 1993, 75, 263-274.
5. D. Kontoyiannis, G. Boulougouris, M. Manoloukos, M. Armaka, M. Apostolaki, T. Pizarro, A. Kotlyarov, I. Forster, R. Flavell, M. Gaestel, P. Tsichlis, F. Cominelli and G. Kollias, *J Exp Med*, 2002, 196, 1563-1574.
6. T. Tanaka, H. Kohno, R. Suzuki, Y. Yamada, S. Sugie and H. Mori, *Cancer Sci*, 2003, 94, 965-973.
7. A. E. Powell, G. Vlacich, Z. Y. Zhao, E. T. McKinley, M. K. Washington, H. C. Manning and R. J. Coffey, *Am J Physiol Gastrointest Liver Physiol*, 2014, 307, G16-23.
8. L. K. Su, K. W. Kinzler, B. Vogelstein, A. C. Preisinger, A. R. Moser, C. Luongo, K. A. Gould and W. F. Dove, *Science*, 1992, 256, 668-670.
9. K. M. Haigis, K. R. Kendall, Y. Wang, A. Cheung, M. C. Haigis, J. N. Glickman, M. Niwa-Kawakita, A. Sweet-Cordero, J. Sebolt-Leopold, K. M. Shannon, J. Settleman, M. Giovannini and T. Jacks, *Nat Genet*, 2008, 40, 600-608.
10. K. E. Hung, M. A. Maricevich, L. G. Richard, W. Y. Chen, M. P. Richardson, A. Kunin, R. T. Bronson, U. Mahmood and R. Kucherlapati, *Proc Natl Acad Sci U S A*, 2010, 107, 1565-1570.
11. D. Borenshtein, M. E. McBee and D. B. Schauer, *Curr Opin Gastroenterol*, 2008, 24, 32-37.
12. J. U. Que and D. J. Hentges, *Infect Immun*, 1985, 48, 169-174.

Experimental Approach	Approximate Number of Analytes
Microbiome	
Metagenomic Sequencing[1]	4×10^6
Mass Spectrometry[2]	500-6000
NMR[3]	500
Bulk Cell Analysis	
RNASeq[4]	25000
Microarray[5]	40000
Mass Spec (iTraq, SILAC)[6]	500-6000
Reverse Phase Protein Array[7]	150
Multiplex Bead Based ELISA[8]	25-100
Microwestern Array[9]	100
Single Cell Analysis	
FACS RNASeq[10]	10000
Fluidigm Single Cell RNASeq[4]	6000
Fluidigm Biomark[11]	96
CyTOF[12]	50
Multiplex Imaging[13]	60
Multiparameter FACS[14]	18
Multiplex RNA FISH[15]	30
Intravital Imaging[16]	1-2

Table 2: A subset of multiplex and -omics experimental approaches for obtaining quantitative data for systems biology, with approximate numbers of analytes measured from cited studies.

1.   J. Qin, Y. Li, Z. Cai, S. Li, J. Zhu, F. Zhang, S. Liang, W. Zhang, Y. Guan, D. Shen, Y. Peng, D. Zhang, Z. Jie, W. Wu, Y. Qin, W. Xue, J. Li, L. Han, D. Lu, P. Wu, Y. Dai, X. Sun, Z. Li, A. Tang, S. Zhong, X. Li, W. Chen, R. Xu, M. Wang, Q. Feng, M. Gong, J. Yu, Y. Zhang, M. Zhang, T. Hansen, G. Sanchez, J. Raes, G. Falony, S. Okuda, M. Almeida, E. LeChatelier, P. Renault, N. Pons, J.-M. Batto, Z. Zhang, H. Chen, R. Yang, W. Zheng, S. Li, H. Yang, J. Wang, S. D. Ehrlich, R. Nielsen, O. Pedersen, K. Kristiansen and J. Wang, *Nature*, 2012, **490**, 55–60.
2.   C. L. Wilkins and J. O. Lay, *Identification of Microorganisms by Mass Spectrometry*, 2005.
3.   S. Fukuda, H. Toh, K. Hase, K. Oshima, Y. Nakanishi, K. Yoshimura, T. Tobe, J. M. Clarke, D. L. Topping, T. Suzuki, T. D. Taylor, K. Itoh, J. Kikuchi, H. Morita, M. Hattori and H. Ohno, *Nature*, 2011, **469**, 543–547.
4.   J. Franklin, Personal Communication.
5.   W. Zhang, Q. D. Morris, R. Chang, O. Shai, M. a Bakowski, N. Mitsakakis, N. Mohammad, M. D. Robinson, R. Zirngibl, E. Somogyi, N. Laurin, E. Eftekharpour, E. Sat, J. Grigull, Q. Pan, W.-T. Peng, N. Krogan, J. Greenblatt, M. Fehlings, D. van der Kooy, J. Aubin, B. G. Bruneau, J. Rossant, B. J. Blencowe, B. J. Frey and T. R. Hughes, *J. Biol.*, 2004, **3**, 21.
6.   C. Jørgensen, A. Sherman, G. I. Chen, A. Pasculescu, A. Poliakov, M. Hsiung, B. Larsen, D. G. Wilkinson, R. Linding and T. Pawson, *Science*, 2009, **326**, 1502–1509.
7.   R. Akbani, N. Carragher, T. Goldstein, L. de Koning, U. Korf, L. Liotta, G. B. Mills, S. S. Nishizuka, E. F. Petricoin III, H. B. Pollard, B. Serrels and J. Zhu, *Mol. Cell. Proteomics*, 2014, **13**, 1625–43.
8.   K. S. Lau, V. Cortez-Retamozo, S. R. Philips, M. J. Pittet, D. A. Lauffenburger and K. M. Haigis, *PLoS Biol.*, 2012, **10**, e1001393.
9.   M. F. Ciaccio, J. P. Wagner, C.-P. Chuu, D. a Lauffenburger and R. B. Jones, *Nat. Methods*, 2010, **7**, 148–155.

10. A. K. Shalek, R. Satija, X. Adiconis, R. S. Gertner, J. T. Gaublomme, R. Raychowdhury, S. Schwartz, N. Yosef, C. Malboeuf, D. Lu, J. J. Trombetta, D. Gennert, A. Gnirke, A. Goren, N. Hacohen, J. Z. Levin, H. Park and A. Regev, *Nature*, 2013, **498**, 236–40.

11. S. J. A. Buczacki, H. I. Zecchini, A. M. Nicholson, R. Russell, L. Vermeulen, R. Kemp and D. J. Winton, *Nature*, 2013, **495**, 65–9.

12. B. Becher, A. Schlitzer, J. Chen, F. Mair, H. R. Sumatoh, K. Wei, W. Teng, D. Low, C. Ruedl, P. Riccardi-castagnoli, M. Poidinger, M. Greter, F. Ginhoux and E. W. Newell, *Nat. Immunol.*, 2014, **15**, 1181–9.

13. M. J. Gerdes, C. J. Sevinsky, A. Sood, S. Adak, M. O. Bello, A. Bordwell, A. Can, A. Corwin, S. Dinn, R. Filkins, D. Hollman, V. Kamath, S. Kannumalle, K. Kenny, M. Larsen, M. Lazare, Q. Li, C. Lowes, C. C. McCulloch, E. McDonough, M. C. Montalto, Z. Pang, J. Rittscher, A. Santamaria-Pang, B. D. Sarachan, M. L. Seel, A. Seppo, K. Shaikh, Y. Sui, J. Zhang and F. Ginty, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 11982–7.

14. E. A. O'Donnell, D. N. Ernst and R. Hingorani, *Immune Netw.*, 2013, **13**, 43–54.

15. E. Lubeck and L. Cai, *Nat. Methods*, 2012, **9**, 743–8.

16. M. J. Pittet and R. Weissleder, *Cell*, 2011, **147**, 983–91.