CELL SPECIFIC RESPONSES TO MICROBIOTA PLAY GLOBAL ROLES IN

HOST DEVELOPMENT & HOMEOSTASIS

by

MICHELLE SCONCE MASSAQUOI

A DISSERTATION

Presented to the Department of Biology and the Division of Graduate Studies of the
University of Oregon in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

March 2022

DISSERTATION APPROVAL PAGE

Student: Michelle Sconce Massaquoi

Title: Cell Specific Responses to Microbiota Play Global Roles in Host Development & Homeostasis

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

Judith Eisen	Chairperson
Karen Guillemin	Advisor
Victoria Herman	Core Member
Kryn Stankunas	Core Member
Raghuveer Parthasarathy	Institutional Representative

and

Krista Chronister	Vice Provost for Graduate Studies

Original approval signatures are on file with the University of Oregon Division of Graduate Studies

Degree awarded March 2022

DISSERTATION ABSTRACT

Michelle Sconce Massaquoi

Doctor of Philosophy

Department of Biology

March 2022

Title: Cell Specific Responses to Microbiota Play Global Roles in Host Development & Homeostasis

Resident microbes are a fixture within all animal life and impact diverse aspects of host biology ranging from metabolism, training of the immune system to identify pathogens but tolerate commensals, and tissue development. An animal host's microbiota encompasses the consortium of bacteria, fungi, archaebacteria and viruses that live on and within them. Animal intestines harbor the highest density of microbes and across model organisms the microbiota has shown to play important roles in the development of organs both proximal and distant to the digestive system. Although pioneering work has significantly increased our understanding of the intricate dynamics within host-microbe interactions and has fundamentally altered how we define animal biology, the mechanisms behind these interactions and the extent that they influence host tissues globally is largely unknown. This dissertation describes the work that characterizes which cell types of the developing host are responsive to the microbiota in the model vertebrate, the larval zebrafish. This work also investigates the mechanism by which a bacterial-secreted protein induces the proliferation and development of the insulin-producing beta cells in the larval pancreas.

CURRICULUM VITAE

NAME OF AUTHOR: Michelle Sconce Massaquoi

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Pacific University of Oregon, Forest Grove

DEGREES AWARDED:

Doctor of Philosophy, Molecular Biology, 2022, University of Oregon
Bachelor of Science, Chemistry, 2012, Pacific University

AREAS OF SPECIAL INTEREST:

Cell & Developmental Biology
Microbiology

PROFESSIONAL EXPERIENCE:

Graduate Teaching Assistant, University of Oregon, September 2015 – June 2016

Research Assistant, Portland VA Medical Hospital, August 2012 – August 2015

Murdock Scholar, Oregon Health & Science University, May 2011 – August 2012

GRANTS, AWARDS, AND HONORS:

Pete von Hipple Graduate Scholar Award
        University of Oregon, Molecular Biology, 2020

Centurion Award
        University of Oregon, 2019

Keynote Speaker: Churchill WISE Symposium
        Churchill High School Women in Science & Engineering, 2019

Knight Campus Graduate Mentor
        University of Oregon, Phil & Penny Knight Campus for Accelerating
        Scientific Impact, 2018

UOWGS Travel Award
        University of Oregon Women of Graduate Sciences, 2018

Honorable Mention: NSF Fellowship
National Science Foundation, 2017

National Institute of Health (NIH) Developmental Biology Training Grant
Appointment, University of Oregon, 2016-2019

Borgen Adamson Award
University of Oregon, Molecular Biology, 2016

Outstanding Senior in Chemistry,
Pacific University of Oregon, Chemistry, 2012

PUBLICATIONS:

Doctoral work

Massaquoi, M.S., Kong, G., Chilin, D., Hamilton, M.K., Melancon, J., Eisen, J.S., & Guillemin, K.J. Global host responses to the microbiota at single cell resolution in whole organism gnotobiotic zebrafish. (*submitted*, Feb 2022)

Hill, J.H., **Massaquoi, M.S.,** Goers Sweeney, E., Wall, E.S., Jahl, J., Kallio, K., Derrick, D., Bell, R.L., Murtaugh, C., Parthasarathy, R., Remington, J.S., Round, J., & Guillemin, K.J. Insulin-producing beta cell mass increases in response to membrane-permeabilizing microbiome and host proteins. (*in review*, Jan 2022)

Postlethwait, J.H., **Massaquoi, M.S.,** Farnsworth, D., Yan, Y., Guillemin, K.J., Miller, A.D., (2021). The SARS-CoV-s receptor and other key components of the Renin-Angiotensin-Aldosterone System related to COVID-19 are expressed in enterocytes in larval zebrafish. *Biology Open*

Massaquoi, M.S., & Guillemin, K.J., Chapter 14 Right on Cue: Microbiota Promote Plasticity of Zebrafish Digestive Tract. *Cellular Dialogues in the Holobiont* (2020). *CRC Press, Monday August 24th*.

Predoctoral work

Massaquoi, M.S., Ligoure, W.A., Churchill, M.J., Moore, C., Melrose, H.D., & Meshul, C.K. (2020) Gait deficits and striatal tyrosine hydroxylase loss are restored following 7,8-dihydroxyflavone treatment in a progressive MPTP mouse model of Parkinson's disease. *Neuroscience*, *1;433*:53-71.

Tamburrino, A., Churchill, M.J., Wan, O.W., Colino-Sanguino, Y., Ippolito, R., Bergstrand, S., Wolf, D.A., Herz, N.J., **Sconce, M.D**., Björklund, A. & Meshul, C.K., (2015). Cyclosporin promotes neurorestoration and cell replacement therapy in pre-clinical models of Parkinson's disease. *Acta Neuropathologica Communications*, *3*(1), p.1.

Sconce, M. D., Churchill, M. J., Moore, C., & Meshul, C. K. (2015). Intervention with 7, 8-dihydroxyflavone blocks further striatal terminal loss and restores motor deficits in a progressive mouse model of Parkinson's disease. *Neuroscience*, *290*, 454-471.

*Pflibsen, L., *Stang, K. A., ***Sconce, M. D***., Wilson, V. B., Hood, R. L., Meshul, C. K., & Mitchell, S. H. (2015). Executive function deficits and glutamatergic protein alterations in a progressive 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine mouse model of Parkinson's disease. *Journal Of Neuroscience Research*. ***-these authors contributed equally to the work**

Sconce, M. D., Churchill, M. J**.,** Greene, R. E., & Meshul, C. K. (2015). Intervention with exercise restores motor deficits but not nigrostriatal loss in a progressive MPTP mouse model of Parkinson's disease. *Neuroscience*, *299*, 156-174.

*Bentea, E., ***Sconce, M. D***., *Churchill, M. J., Van Liefferinge, J., Sato, H., Meshul, C. K., & Massie, A. (2015). MPTP-induced parkinsonism in mice alters striatal and nigral xCT expression but is unaffected by the genetic loss of xCT. *Neuroscience Letters*, *593*, 1-6. ***-these authors contributed equally to the work**

Commentaries

Massaquoi, M.S., Is bar soap as gross as millennials say? Not really, and we're all covered with microbes anyway (2020). *The Conversation*, *Tuesday July 14th*.

Massaquoi, M.S., Why hand-washing really is as important as doctors say (2020). *The Conversation*, *Tuesday March 3rd* (updated article).

Massaquoi, M.S., & Guillemin, K.J., Packing in the Proteins (2019). *Developmental Cell*, *October 7th*.

Massaquoi, M.S., & Guillemin, K.J., Evolving in a microbial soup: you are what they eat (2018). *Developmental Cell*, *December 17th*.

Massaquoi, M.S., Why washing your hands well is so important to protect your family from the flu (2018). *The Conversation, Tuesday October 23rd* (updated article).

Massaquoi, M. S., You (and most of the millions of holiday travelers you encounter) are washing your hands wrong (2017). *The Conversation, Wednesday December 13th*.

Sconce, M. D., Banuelos, M., & Guillemin, K. (2017). GUEST OPINION: A roadmap for greater diversity in science at UO. *The Oregonian, Sunday June 11th, OP-ED*.

***Note that I have published under the surnames Massaquoi and Sconce

ACKNOWLEDGMENTS

This work is dedicated my loving family as well as my friends, colleagues and mentors
that include the additional family I got to choose

TABLE OF CONTENTS

LIST OF FIGURES

PREFACE

EVOLVING IN A MICROBIAL SOUP: YOU ARE WHAT THEY EAT

Introduction

Developmental biology aims to understand how a single cell can give rise to a diversity of differentiated cell types whose orchestrated movements and interactions produce the functionally distinct tissues and organs of a mature organism. This process is usually considered to be under the autonomous genetic control of the organism's genome. However, the organism's environmental conditions -- temperature, oxygen tension, nutrition -- all shape the expression of developmental programs. Not only abiotic factors, but also interactions with other organisms can shape developmental trajectories. A flamboyant example is the defensive neckteeth structures that the water flea *Daphnia* develops in response to the presence of its predator, the phantom midge *Chaoborus*. To hamper its ingestion, *Daphnia* has evolved an elaborate inducible developmental program that is expressed in response to chemical cues produced by actively predating *Chaoborus* (Weiss et al., 2018). Could associations with microscopic organisms play similar roles in the evolution and expression of developmental programs?

In this Spotlight piece, we describe recent work from the lab of Janelle Ayres that demonstrates how even among genetically homogeneous hosts, heterogeneities in their metabolic interactions with microbes can have profound impacts on host fitness (Sanchez et al., 2018). We discuss how such individual variability in host-microbe interactions is likely to have driven

1

animal and plant developmental programs by constraining options to those that promote tolerance of resident microbes (Ayres, 2016).

Those Left Standing, Iron-Clad and Sweet

Infectious disease investigators commonly use the metric of lethal dose 50 (LD50) to describe the virulence of a particular pathogen. Considered at a population level, this is simply the dose of pathogen that will kill half the inoculated hosts. Using an infection model with the mouse enteric pathogen *Citrobacter rodentium* (CR) in inbred C3H/HeJ mice, Ayres and colleagues went beyond this population level characterization to ask what distinguishes the individual mice that survive from those that perish. Strikingly, even though they observed a constant pathogen load across the mice, they saw a bimodal distribution of disease symptoms. Half of the mice had somehow managed to tolerate rather than succumb to the pathogen. A survey of these lucky mice's liver transcriptomes revealed a significant increase of genes involved in iron metabolism.

Iron is an important commodity that host and bacterial cells often try to horde, engaging in arms races of escalating iron-sequestration strategies (Lopez and Skaar, 2018). In this study, however, Ayres and colleagues found that while iron-depleted mice succumbed to the standard lethal dose of CR, sibling mice supplemented with dietary iron survived infections with 1000 times this CR dose. CR infected mice given dietary iron also displayed protection from intestinal damage and had attenuated host pro-inflammatory responses even with high pathogen load. Delving further into the molecular basis for this host tolerance of the pathogen, the investigators found that iron supplementation suppressed the expression of CR virulence factors in the mouse intestine. Inspired by clinical work linking iron metabolism and glucose homeostasis (Carlos et al., 2018), the Ayres group discovered that dietary iron induced insulin resistance in their mice.

The animals' decreased tissue glucose uptake resulted in increased glucose availability within the intestine, which in turn resulted in decreased CR virulence factor expression. Acute insulin resistance, such as Ayres et al. observed in their surviving mice, turns out to be a common response during infectious diseases (Carlos et al., 2018), suggesting that it may be an evolved strategy to placate pathogens with glucose availability.

Not only did this sweet-talking strategy pay off for the immediately infected mice, but it provided collateral benefits for their cage-mates and progeny by selecting for attenuated CR isolates that were less virulent even when transmitted to iron depleted, insulin sensitive hosts. This evolutionary trajectory played out differently from the typical arms race or "Red Queen" co-evolution dynamics that maintain a high level of disease pathology (Ayres, 2016). Instead, Ayres' insulin resistant CR infected mice seem to be on a trajectory to a stable new normal of co-existence.

The Roads Most Taken, Skirting Microbial Perils

All plants and animals are recent evolutionary newcomers to a planet dominated by microbes. We speculate that the evolutionary trajectories of their developmental programs are filled with accommodations to microbial co-existence like the one that Ayres' research captured. Learning to live with microbes has likely radically constrained the possible body plans and physiologies of multicellular organisms because the cost of choosing an option that promotes microbial virulence is astronomical. The evidence for such accommodations would be different from the hallmarks of evolutionary conflict (McLaughlin and Malik, 2017).

To find traces of microbial-imposed constraints on animal and plant development, we could start by looking for developmental programs that are responsive to the presence of microbes, just as *Daphnia*'s neckteeth sprout up at the whiff of an active predator. Such evidence

would come from comparing the developmental trajectories of animals in the presences versus the absence of microbes. To this end, many studies with germ-free mice, zebrafish, and fruit flies have identified traits that appear to be accommodations for co-existence with microbes. A major theme is the establishment of a comfortable balance of immune system activity with appropriate differentiation of T lymphocytes, census of intestinal neutrophils, and expression of intestinal antimicrobial peptides (Chu and Mazmanian, 2013). Another common theme is microbial stimulation of intestinal epithelial renewal (Jone and Guillemin, 2018) and, in vertebrates, increases in mucus secreting cells (Troll et al., 2018), accommodations that likely provides increased protection against epithelial damage while feeding luminal bacteria. Germ-free zebrafish also exhibit a paucity of insulin-producing pancreatic beta cells (Hill et al., 2016), suggesting that accommodating resident microbes involves finding an appropriate balance of circulating glucose, consistent with Ayres' work.

Further evidence for microbial tolerance adaptations driving animal and plant development come from knowledge of the microbial molecules that elicit these developmental programs. In contrast to the orchestrated signaling between cells of a single organism, such as the growth factors and cognate receptors familiar to developmental biologists, we would expect to find developmental programs elicited by inadvertently produced molecular cues of foreign organisms. A beautiful illustration are the molecules elucidated by Linda Weiss and colleagues (Weiss et al., 2018) that trigger *Daphnia*'s defensive developmental program, which turn out to be evolutionarily conserved byproducts of insect digestion, molecules that a water flea-munching midge can't help but spew. This surveillance of evolutionarily conserved signatures of a potentially predator resembles the strategy of the innate immune system. Innate immune pattern recognition receptors (PRRs) detect the presence of microbial associated molecular patterns

(MAMPs) such as cell wall components that a bacterial cell under turgor pressure cannot afford to do without (Chu and Mazmanian, 2013). In addition, a growing number of G-protein coupled receptors and hormone receptors are turning out to be detectors of bacterial metabolic byproducts, likely involved in microbial surveillance. Increasingly, these surveillance pathways and the microbial detritus they sense are being found to regulate host developmental programs, such as the innate immune suppression of Notch signaling to increase intestinal goblet cells (Troll et al., 2018). We anticipate that developmental biologists, working with microbiologists, will continue to find evidence of microbial constraints on the evolution of animal and plant development.

CHAPTER I

RIGHT ON CUE: MICROBIOTA PROMOTE PLASTICITY OF

ZEBRAFISH DIGESTIVE TRACT

Introduction

All animals on earth have evolved within a world teeming with microscopic life. The genesis of bacteria dates back 2.6 billion years, whereas primitive humans evolving only 315 thousand years ago from ancestors who had always co-existed with microbes. We refer to the community of bacteria, viruses, fungi and archaea inhabiting a multicellular host as its microbiota. Only recently have biologists considered an organism as an ecosystem of many, rather than an isolated individual. Although microbes were initially vilified as pathogens, just a small fraction of the thousands of microbial species cause disease in animals. Microbial life is now being appreciated for its multitude of roles in host homeostasis. As our knowledge of the intricate interactions between host organisms and resident microbiota increases, it is fundamentally changing how we view many aspects of animal biology, including animal development.

Viewed from an evolutionary perspective, microbes have shaped animal history by influencing their fitness throughout their lifespans. As described by the Modern Synthesis, natural selection dictates how organisms that are better adapted to their environment will pass down their genes to the next generation. Resident microbes can shift that fitness landscape for mature organisms, for example by supplying degradative enzymatic capacities to allow hosts to

access new sources of nutrition. Additionally, microbes can influence which specific genotypes survive and reproduce by shaping the developmental programs that determine how an organism forms from a single-cell embryo to a mature multi-cellular adult. Polyphenism is a biological phenomenon in which distinct phenotypes can arise from a single clonal genotype, demonstrating the plasticity of developmental biology. Predation, temperature, and nutrient availability are all direct environmental factors that induce polyphenism. Resident microbes can also be added to this list of factors that influence a host's developmental trajectories. For example, the presence or absence of the bacterium *Wolbachia* significantly impacts ovary and oocyte development in the parasitic wasp, *Aosbara tabida Nees* (Dedeine *et al*., 2001).

Vertebrates harbor dense and complex microbiota, especially in their digestive tracts (Ley *et al*., 2008). The impact of the microbiota can be studied through the use of microbiologically sterile, or "germ-free" animals and with gnotobiology experiments, using biological systems in which all members of a community are known. In germ-free mice, not only does the absence of the microbiota impair maturation of the gut and associated mucosal immune system (Belkaid and Harrison, 2017), but many distal organ systems are also impaired (Schroeder and Bäckhed, 2016). For example, germ-free mice have stunted development of their intestinal villi capillary network (Stappenbeck, Hooper and Gordon, 2002) as well as incomplete bone formation (Sjö *et al*., 2012). Although the majority of germ-free and gnotobiotic studies of vertebrates have been conducted in laboratory mice, comparisons across other vertebrate models are invaluable for understanding which host responses to microbiota are conserved across multiple host lineages.

In this chapter, we will discuss insights about the developmental impacts of the microbiota in the model vertebrate *Danio rerio*, the zebrafish. George Streisinger at the

University of Oregon pioneered the use of zebrafish as a model system for developmental and genetic research (Grunwald and Eisen, 2002). There are many advantages to working with zebrafish as a model vertebrate. In addition to genome conservation with mammals, zebrafish are genetically tractable, with many mutant and transgenic lines readily available. They also have high fecundity and are optically transparent during embryonic to larval stages, making this model organism ideal for studying developmental processes in real time. Additionally, because zebrafish develop *ex utero*, they can easily be derived germ-free by surface sterilization of the outer chorion for experiments aiming to understand host-microbe interactions (Melancon *et al*., 2017). Culture collections of zebrafish-associated bacterial isolates, with draft genome sequences, are available, enabling gnotobiotic experimentation (Stephens *et al*., 2016). Together, the transparent properties of larval zebrafish with the use of transgenic lines that allow tracking of specific cell types and gnotobiological experiments enable a high-resolution perspective of how the microbiota influence host development. Below we discuss the ways in which the gut microbiota impact different aspects of zebrafish larval development, following the animals' first exposure to environmental microbes upon hatching out of their chorions. We use insights gleaned from gnotobiotic zebrafish studies to speculate how host-microbe interactions evolved to modulate developmental program plasticity to optimize the organisms' fitness for different environments.

Development Under Immune Surveillance

The vertebrate intestine serves multiple roles, both as an organ for food digestion and nutrients absorption and as an immunological organ for harboring the body's most abundant microbial population. The intestinal epithelium can be thought of as the "inner skin" of the body because like the epidermis, it interacts with the outside environment, not only interfacing directly

with microbial cells inhabiting the intestinal lumen but also the rich source of bio-active molecules that they secrete (Fischbach and Segre, 2016). With the multitudes of microbial life that populate any given vertebrate, a healthy immune system continuously monitors the intestinal lumen and senses threats posed by resident or invading microbes. Appropriately balanced responses to the microbiota are critical for symbiosis because on the one hand, lack of defense can lead to microbial growth and on the other hand, excessive inflammatory responses can be detrimental to both host and microbe cell populations. As the host develops, the resident microbes help train the immune system to achieve this appropriately balanced response (Belkaid and Harrison, 2017).

The gnotobiotic zebrafish model has allowed a detailed characterization of the different immune responses elicited by individual members of the microbiota (Murdoch and Rawls, 2019). For example, zebrafish mono-associated with different zebrafish-derived bacterial isolates will exhibit different levels of immune gene expression (Rawls *et al*., 2006) and accumulate different numbers of neutrophils, a type of white blood cell that lead the immune system's inflammatory response (Rolig *et al*., 2015). These types of data inspire the question of how different bacterial residents elicit different immunological responses in the host. Two possible and not mutually exclusive mechanisms are that host immune sensors are differentially stimulated by different bacteria and that different bacteria produce immunomodulatory factors that alter host immune responses.

During larval development, zebrafish rely on their innate immune system for microbial sensing prior to the maturation of their adaptive immune system in juvenile stages. The best characterized of their innate immune sensors are the Toll-like family receptors (TLRs), which are part of an ancient pattern recognition family of receptors (Jault, Pichon and Chluba, 2004;

Deguine and Barton, 2014). TLR activation is mediated by the sensing of generic microbial products termed microbial-associated molecular patterns (MAMPs) such as cell wall components and flagellin, which subsequently regulates the appropriate immune response (Deguine and Barton, 2014). The specificity and downstream response of TLR signaling is partly dictated by the differential recruitment of intracellular adaptor molecules. Myeloid differentiation primary response 88 (MyD88) (Hall *et al*., 2009) is a common adaptor of TLRs and Interleukin-1 receptor that regulates the expression of pro- or anti-inflammatory cytokines and mitogen-activated protein kinase (MAPK) signaling for cell survival or proliferation (Akira, 2003; Larsson *et al*., 2012). The zebrafish genome has duplicated *tlr* genes with a single copy of *myd88*, that has conserved function in modulating innate immune responses (Jault, Pichon and Chluba, 2004; Meijer *et al*., 2004; van der Sar *et al*., 2006; Jennifer M Bates *et al*., 2007; Hall *et al*., 2009; Burns *et al*., 2017). Of note, Myd88-deficient zebrafish have a completely attenuated intestinal neutrophil influx, indistinguishable from germ-free animals (Bates *et al*., 2007; Burns *et al*., 2017). This indicates that much of the immunological responses to the intestinal microbiota are mediated through Myd88. One trait that varies dramatically across zebrafish bacterial isolates is their capacity for motility within the intestine (Schlomann *et al*., 2018; Wiles *et al*., 2018). Variation in bacterial *in vivo* production of motility machinery, such as flagellin subunits, could account for different capacities of different bacterial isolates to activate Myd88-dependent immune responses.

Resident bacteria also modulate host immune responses through specific secreted factors. For example, Rolig and colleagues demonstrated that a *Shewanella* isolate was an especially potent suppressor of neutrophil intestinal influx, a response that could be recapitulated with *Shewanella* secreted factors (Rolig *et al*., 2015). More recently, Rolig and colleagues showed

that several *Aeromonas* strains secrete a protein, *Aeromonas* immune modulator A (AimA), that dampens neutrophil influx and pro-inflammatory cytokine expression and also confers a colonization advantage to the bacteria (Rolig *et al*., 2018). The crystal structure of AimA revealed that it has two distinct domains with related folds, both individually retaining the capacity to regulate neutrophils. To investigate whether AimA confers a colonization advantage to *Aeromonas* through attenuating inflammation, the authors measured the abundance of *Aeromonas* strains with and without AimA in *myd88* mutant zebrafish. They found that in these immunocompromised hosts with a limited immune response, *Aeromonas* no longer required AimA for maximal intestinal colonization, suggesting that in the face of a normal host immune response, the bacteria benefit from AimA's ability to dampen inflammation (Rolig *et al*., 2018). This study illustrates that Myd88-mediated host responses to the microbiota modulate features of the host environment that impact the fitness of resident bacteria. This finding is corroborated by the fact that isolates of *Aeromonas* experimental evolved to colonize wild type zebrafish intestines are less fit when introduced into *myd88* deficient hosts (Robinson *et al*., 2018). Thus, the innate immune system acts as a conduit to intercept and respond to microbial cues and may underly microbiota-mediated developmental plasticity.

Developmental Plasticity at the Luminal Interface

In addition to the immune cells of the intestinal mucosa, the cells of the intestinal epithelium are situated to interact with the most abundant microbial population in the vertebrate body. The embryonic stages of zebrafish development happen within the sterile environment of the chorion, but once they hatch from their chorions into the larval stage (~2-3 days post fertilization (dpf)), they are inoculated with their first resident microbes from the surrounding environment. The load of microbes dramatically expands within the zebrafish intestine from 4-6

dpf (Bates *et al.*, 2006; Jemielita *et al.*, 2014; Stephens *et al.*, 2016), which coincides with

important developmental timepoints within the digestive system's intestine, pancreas, and liver.

By 6 dpf the intestine exhibits clear compartmentalization into morphologically and

transcriptionally distinct regions: the intestinal bulb (duodenum and jejunum-like), mid-intestine

(ileum-like), and posterior intestine (colon-like) (Ng *et al.*, 2005; Lickwar *et al.*, 2017).
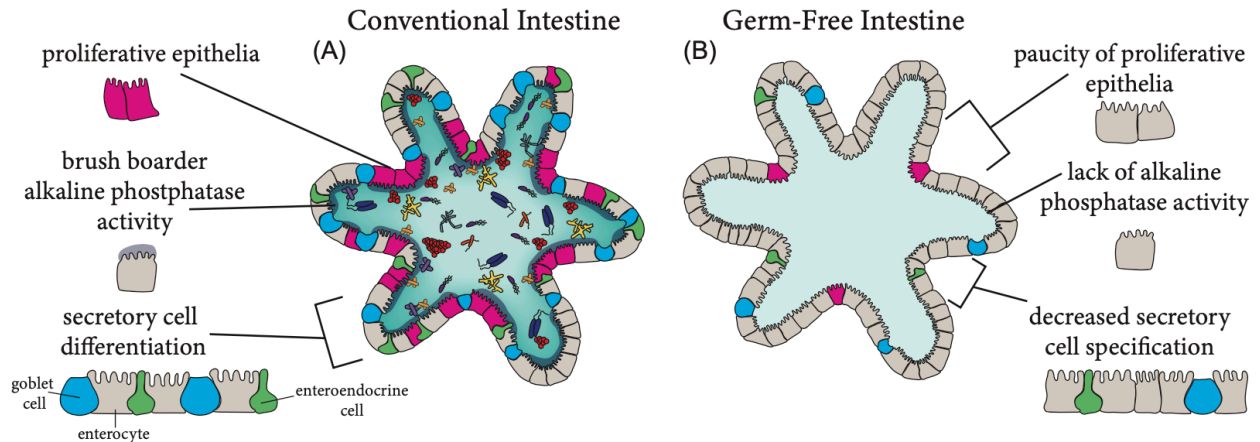
Figure 1.1 Intestinal Microbiota Induce Development of Host Intestinal Epithelia. In a cross-sectional view of a conventional intestine (A) the microbiota stimulates proliferation of the intestinal epithelia, expression of brush boarder intestinal alkaline phosphatase activity (AP, a marker of enterocyte maturation), and specification of secretory cell fates. In a germ-free intestine (B), epithelia exhibit less proliferation, decreased expression of markers of enterocyte maturity and a paucity of secretory cells.

Raised under germ-free conditions, larval (6dpf) zebrafish resemble their conventionally

reared counterparts at the level of gross morphology of their digestive organs, but the epithelium

is characterized by a paucity of proliferation (Rawls, Samuel and Gordon, 2004) and of secretory

cells, including mucus-secreting goblet cells and hormone secreting enteroendocrine cells (Bates

et al., 2006) (Fig1.1). These traits cannot be restored by exposure of germ-free animals to a

generic activator of TLR signaling, the Gram-negative bacterial cell wall component

lipopolysaccharide (LPS), but can be restored by mono-association with certain bacterial isolates

(Cheesman *et al.*, 2011; Bates *et al.*, 2006).

A prominent and highly conserved molecular pathway that modulates cell proliferation of

many tissues including the intestinal epithelium is Wnt signaling. Within the canonical pathway,

Wnt binds to the cell receptor Frizzled and activates the transcriptional regulator Beta-catenin by

preventing its proteasome-mediated degradation and allowing it to translocate to the nucleus to

turn on Wnt target genes. In the larval zebrafish intestine, both Wnt signaling and resident

microbiota promote epithelial cell proliferation and accumulation of Beta-catenin (Cheesman *et*

al., 2011). Mono-associated zebrafish with an *Aeromonas* strain or exposure to this strain's

secreted products was sufficient to rescue intestinal epithelial proliferation. This study also

discovered that Myd88 was required for intestinal epithelial cell proliferation, with *myd88*

deficient animals resembling the germ-free state. Together, the data indicate that host sensing of

microbial products stimulates intestinal epithelial proliferation through canonical Wnt signaling,

directly linking the innate immune sensing of the microbiota with a canonical animal

developmental program (Fig1.2). The continually renewing intestinal epithelial cells differentiate

into distinct functional cell types to carry out diverse functions such as absorbing nutrients,

secreting a protective mucus barrier, and relaying cell-to-cell communication. One marker of the

mature absorptive enterocyte lineage is the enzymatic activity of alkaline phosphatase (AP), the

expression of which requires Myd88 and is stimulated by LPS (Bates *et al*., 2007). This enzyme

in turn dephosphorylates LPS, rendering it less pro-inflammatory and promoting host tolerance

of the microbiota (Bates *et al*., 2007). Another indication of intestinal epithelial maturation is the

expression of glycoconjugates in different patterns along the length of the digestive tract (Falk,

Roth and Gordon, 1994; Wu *et al*., 2009). In the absence of microbiota, larval zebrafish show a

striking decrease in AP activity (Fig1.2) and levels of certain glucoconjugates, including N-

acetylgalactosamine and galactosea1,3galactosyl moieties (Bates *et al*., 2006). As the intestinal

epithelium matures, its distinct functional regions become more prominent. For example, in

larval zebrafish an ileum-like region of the mid-gut becomes apparent, where cells containing

large lysosomes engage in luminal protein absorption and degradation (Lickwar *et al*., 2017; Park *et al*., 2019). The microbiota are necessary for maturation of this region of the gut (Rawls *et al*., 2004; Bates *et al*., 2006). In addition to absorbing molecules from the lumen, intestinal cells also secrete molecules into the lumen and bloodstream. Secretory cells within the intestinal epithelial play many critical roles for intestinal and vertebrate homeostasis. Mucus-secreting goblet cells maintain a protective mucosal barrier between the luminal microbiota and epithelium. Enteroendocrine cells sample the luminal contents and communicate with the extra-intestinal cells by hormone secretion. In the absence of the microbiota, the specification and differentiation of goblet and serotonin-secreting enteroendocrine cells is reduced, with older germ free larvae having secretory cell numbers similar to younger conventionally reared animals (Bates *et al*., 2006).

A prominent, conserved developmental pathway known to regulate cell fate decisions in many tissues including the intestinal epithelium is Notch-Delta signaling. Within intestinal secretory cell development, appropriate regulation of Notch-Delta signaling is essential for the decision between absorptive versus secretory cell types. Notch receptors are transmembrane proteins that form a juxtracrine signaling response when bound by the membrane protein Delta on a neighboring cell. Notch-Delta binding on the outside of the two neighboring cells induces cleavage of the Notch receptor, releasing the intracellular domain to translocate to the nucleus and regulate gene expression. Inhibiting Notch signaling within zebrafish causes a conversion of the intestinal epithelium to a secretory cell fate (Crosnier *et al*., 2005). In the zebrafish intestinal epithelium, absence of the microbiota phenocopies activation of Notch-Delta signaling with an expansion of absorptive relative to secretory cells (Bates *et al*., 2006; Troll *et al*., 2018). Inhibiting Notch-Delta signaling specifically within the intestinal epithelium significantly

increases the number of secretory cells in both conventional and germ free reared zebrafish, indicating that the microbiota acts upstream of Notch-Delta signaling to specify intestinal secretory cell fate (Troll *et al.*, 2018). Echoing the microbiota's impact on host intestinal epithelial proliferation, Myd88-dependent host sensing of the microbiota is necessary for specification of appropriate proportions of secretory cells (Troll *et al.*, 2018). Double mutant animals lacking Myd88 and Notch signaling have high numbers of secretory cells, placing Myd88 upstream of Notch signaling (Troll *et al.*, 2018). In this model, the microbiota are perceived by the innate immune system, which inhibits Notch-Delta signaling to promote secretory cell differentiation (Fig1.2).
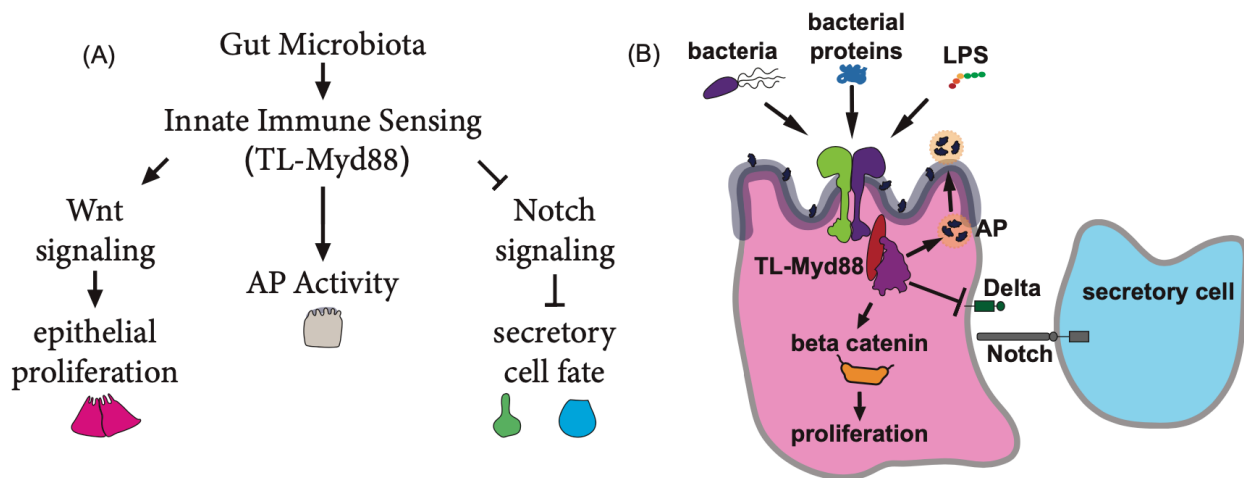
Figure 1.2: Innate Immune Sensing by Myd88 Drives Intestinal Epithelial Development in Response to Microbial Cues. Epistasis diagram (A) showing pathways downstream of Myd88 that impact epithelial proliferation, alkaline phosphatase (AP) activity and secretory cell fate. Microbial cues from bacteria or their secreted products differentially impact the Myd88-dependent downstream responses (B). Lipopolysaccharide (LPS) is only sufficient to stimulate AP activity, but does not rescue epithelia proliferation and secretory cell fates.

Developmental biologists typically study processes of organ growth, differentiation and functionalization under highly controlled conditions, allowing them to focus on the intrinsic programs of gene regulation through canonical pathways, such as the Wnt and Notch-Delta signaling that balance intestinal epithelial cell proliferation and differentiation (Sancho, Cremona and Behrens, 2015). Studies in the gnotobiotic larval zebrafish demonstrate that the presence and

composition of the microbiota modulate these canonical signaling pathways and modify intestinal developmental programs. In the case of both of Wnt and Notch signaling, their modulation by microbiota is mediated through Myd88-dependent processes. Since LPS is insufficient to activate Wnt and Notch signaling-mediated changes in the germ-free intestinal epithelium, although it is sufficient to up-regulate intestinal AP expression, this suggest that the microbiota produce multiple factors that non-redundantly stimulate innate immune sensing pathways and impact diverse aspects of host development.

Beyond The Lumen: A Secreted Bacterial Protein Impacts Pancreas Development

Branching from the direct host-microbe interface of the intestinal epithelium is the hepato-pancreatic ductal system. This ductal system does not harbor appreciable numbers of microbes, but it supplies the intestinal lumen with digestive enzymes from the pancreas and bile from the liver and gall bladder for digestion. The pancreas is a dual gland composed of exocrine tissue that secretes enzymes and endocrine tissue that secretes hormones and regulates blood sugar levels. The beta cell population is an essential host cell type as the sole provider of insulin for the body. After sensing blood glucose, beta cells release insulin into the blood stream allowing all host cell types to uptake blood-glucose for cellular respiration. As with the intestine, important developmental timepoints of the pancreas coincides with the assembly of the microbiota; beta cell mass dramatically expands during 4-6 dpf simultaneously with the increase of intestinal bacterial abundance (Kimmel and Meyer, 2010). In the absence the microbiota, the insulin-producing beta cells of the zebrafish pancreas fail to undergo their normal developmental expansion (Hill *et al*., 2016).

By mono-associating with zebrafish-associated bacterial isolates, Hill et al., 2016 found that only a few members of the zebrafish microbiota (several *Aeromonas* and *Shewanella spp*.)

were sufficient to rescue the normal expansion of beta cell mass in germ free zebrafish, suggesting involvement of a specific cue rather than a generic product of all microbiota members. From the known genome sequences of zebrafish bacterial isolates, Hill et al., 2016 compared the genomes of the mono-associated bacteria that rescued beta cell development to those that did not and discovered an un-characterized bacterial secreted protein, Beta Cell Expansion Factor A (BefA), that was necessary within *Aeromonas* and sufficient alone to rescue beta cell expansion within germ free derived zebrafish. Further, BefA facilitates expansion of beta cells by inducing their proliferation and homologs of BefA produced by members of the human gut microbiota have conserved function in facilitating zebrafish beta cell development. Luminal proteins have been shown to traffic to distant organs (Park *et al.*, 2019), suggesting a possible mechanism by which BefA could influence the developmental of an extra-intestinal tissue. The long-term consequences on host metabolism and fitness of this reduced beta cell expansion in larval life have not yet been characterized, but mice with reduced perinatal beta cell proliferation have reduced beta cell mass and impaired glucose homeostasis as adults (Berger *et al.*, 2015). More broadly, BefA provides an example of how the presence or absence of specific microbial members during developmental stages can impact organ development, further supporting the idea of microbiota composition modulating developmental plasticity.

Conclusions

The gnotobiotic zebrafish model reveals how the presence of microbiota modulate developmental programs of the digestive tract and how individual microbial isolates and community combinations can elicit unique host responses at the tissue and cellular levels (Rolig *et al.*, 2015, 2017; Burns and Guillemin, 2017; Burns *et al.*, 2017). We postulate that host developmental programs use microbial cues to interpret their environments and adapt tissue

development accordingly (Massaquoi and Guillemin, 2018). Viewed through this lens, microbiota-induced changes in the intestinal epithelium can be understood as accommodations for co-existing with an abundant microbial community. Up-regulation of the LPS-detoxifying enzyme AP prevents excessive intestinal inflammation in response to resident bacteria. Increasing epithelial cell turnover is a protective adaptation for removing damaged cells while also providing resident microbes with increased nutrients on shed cellular material. Increasing the proportion of secretory cells would also promote co-existence with microbes. Through the secretion of over thirty hormones, enteroendocrine cells dictate many aspects of the luminal environment from glucose metabolism, gut motility, digestion and absorption (Sikander, Rana and Prasad, 2009; Vincent, Sharp and Raybould, 2011; Tolhurst, Reimann and Gribble, 2012; Pais, Gribble and Reimann, 2016). The mucus secreted from goblet cells not only maintains a protective barrier between luminal contents and epithelia, it also provides nutrients and habitats for the microbiota (Kashyap *et al.*, 2013; Desai *et al.*, 2016; Johansson and Hansson, 2016). Host sensing of specific microbiota membership may also be a mechanism for interpreting host ecology and specifically nutritional landscapes, which will foster different environmental microbes that serve as source pools for host colonization. Modulation of beta cell mass in response to resident microbiota may represent an adaptation in anticipation of metabolic demands based on nutrient scarcity or abundance.

An important tenant in evolution is that natural selection acts on the genetic variation within a species population. We would argue that resident microbiota, which are highly variable between individuals, add another layer of phenotypic variation, beyond that encoded in genomes, that has shaped animal evolution. In particular, if certain developmental programs are especially sensitized to microbial cues, then the microbiota of particular developmental stages may dictate

functional adaptations that impact host fitness in given environments. We can think of this as a form of microbiota-induced polyphenism. Results from the zebrafish model provide molecular examples that could underly such polyphenism. Innate immune sensing of resident microbial products modulates canonical host developmental programs such as Wnt and Notch signaling to optimize host tissues for living with specific microbial consortia. A thorough understanding of how the microbiota impact host development will require discovering which microbial products are sensed by host cells, the identity and sensing mechanisms of host cells that perceive these microbial products, and the cellular responses and downstream consequences of microbial product perception.

Bridge

Chapter I reviews the different ways the microbiota influences the development of the digestive system in the host larval zebrafish. These pioneering works have established that the microbiota is important for the proliferation and specification of diverse digestive system tissue types both proximal and distal to the host-microbe interface within the intestinal lumen. However, little is known about how the microbiota influences gene expression networks within these cell types, let alone across all host cell types in the organism. This next chapter will present our work that catalogues which host cell types respond to the presence of the microbiota and describes the ways in which host cells respond globally across the entire larval zebrafish body at single cell resolution by transcriptomic sequencing.

CHAPTER II

GLOBAL HOST RESPONSES TO THE MICROBIOTA AT SINGLE CELL RESOLUTION

IN WHOLE ORGANISM GNOTOBIOTIC ZEBRAFISH

This chapter contains pre-published, co-authored material. I was the primary contributor for this work including, designing/conducting experiments, analyzing resulting data and writing the manuscript. Principle Investigator, and my mentor, Dr. Karen Guillemin contributed to experimental design, data interpretation and writing. Authors Garth Kong and Daisy Chilin curated the initial code for data analysis. Authors Ellie Melancon, Dr. Mary Hamilton and Dr. Judith Eisen provided unique expertise in data interpretation.

Introduction

Bacteria were the first organisms to evolve on earth, preceding the emergence of animals by roughly 3 billion years. All animals evolved amidst the pressures exerted by the abundant bacteria, fungi and viruses in their world, which required evolving mechanisms to shape and co-exist with resident microbial communities, or microbiota, that live on and within animal bodies. Emerging insights into the many roles that microbiotas play in host development and homeostasis are redefining how we view animal biology (Mcfall-Ngai 2014). The impacts of resident microbes on cellular transcription have not yet been describes globally across an entire vertebrate animal body. The goal of this research is to create a resource that catalogues which tissues of the larval zebrafish host are responsive to the microbiota and document cell type specific transcriptional responses. Zebrafish are an excellent model in which to study impacts of the microbiota on vertebrate biology because they are readily amendable to gnotobiotic manipulations, their small size allows single cell transcriptional profiling of the entire body, and

they share common developmental and physiological programs with larger vertebrates such as humans. In this study, we used the 10x Genomics pipeline to interrogate the transcriptomes of single cells dissociated from cohorts of larval zebrafish reared in the presence or absence of their normal microbiota. We report our analysis of the global, tissue, and cell type responses to the microbiota that emerge from this dataset, which will serve as a valuable hypothesis generating resource for the microbiome sciences field.

Results & Discussion

Responsiveness to microbes is widespread across the cells of the vertebrate body

The small size and tractable gnotobiology of zebrafish larvae afford an unprecedented opportunity to survey the responsiveness of every vertebrate host cell type to the presence of their microbiota. We derived sterile zebrafish embryos and reared them germ-free (GF) or conventionalized them (CVZ) them with parental tank water. At 6 days post fertilization (dpf) whole larvae were dissociated into single cells and processed for single cell RNA sequencing (Fig2.1A). This analysis yielded $33,992$ cells that averaged 23,485 mean reads/cell, with 879 median genes/cell, 2,396 unique transcript molecules (UMI)/cell and 22,953 and 22,898 total genes detected from CVZ and GF groups, respectively.

Based on transcriptional profiles, our dataset is composed of diverse host cell types with largely equal representation from CVZ and GF groups (Fig2.1B). We used the Seurat package in R to integrate cells from both treatments (Stuart et al. 2018), identifying common sources of

Figure 2.1 Single cell transcriptional analysis of whole gnotobiotic larval zebrafish. A) Whole conventional (CVZ) and germ free (GF) 6dpf zebrafish larvae were dissociated into individual cells prior to single cell RNA sequencing. B) uMAP plots display 78 clusters of cell types conserved between experimental groups. C) Cell types were identified by their transcriptomic profile and include enterocytes (*fabp2*), liver (*fetub*), acinar of the exocrine pancreas (*cpa5*), immune cells (*coro1a*) and neural cells (*elavl4*). D) Host cell types differentially respond to the presence of the microbiota throughout the body illustrated by the total number of genes significantly enriched within each experimental group (p<0.05). Black bars represent the sum of differentially expressed genes in CVZ and GF cells respective of each other for each cluster and the color bars represent the number of genes significantly expressed within CVZ cells over GF.
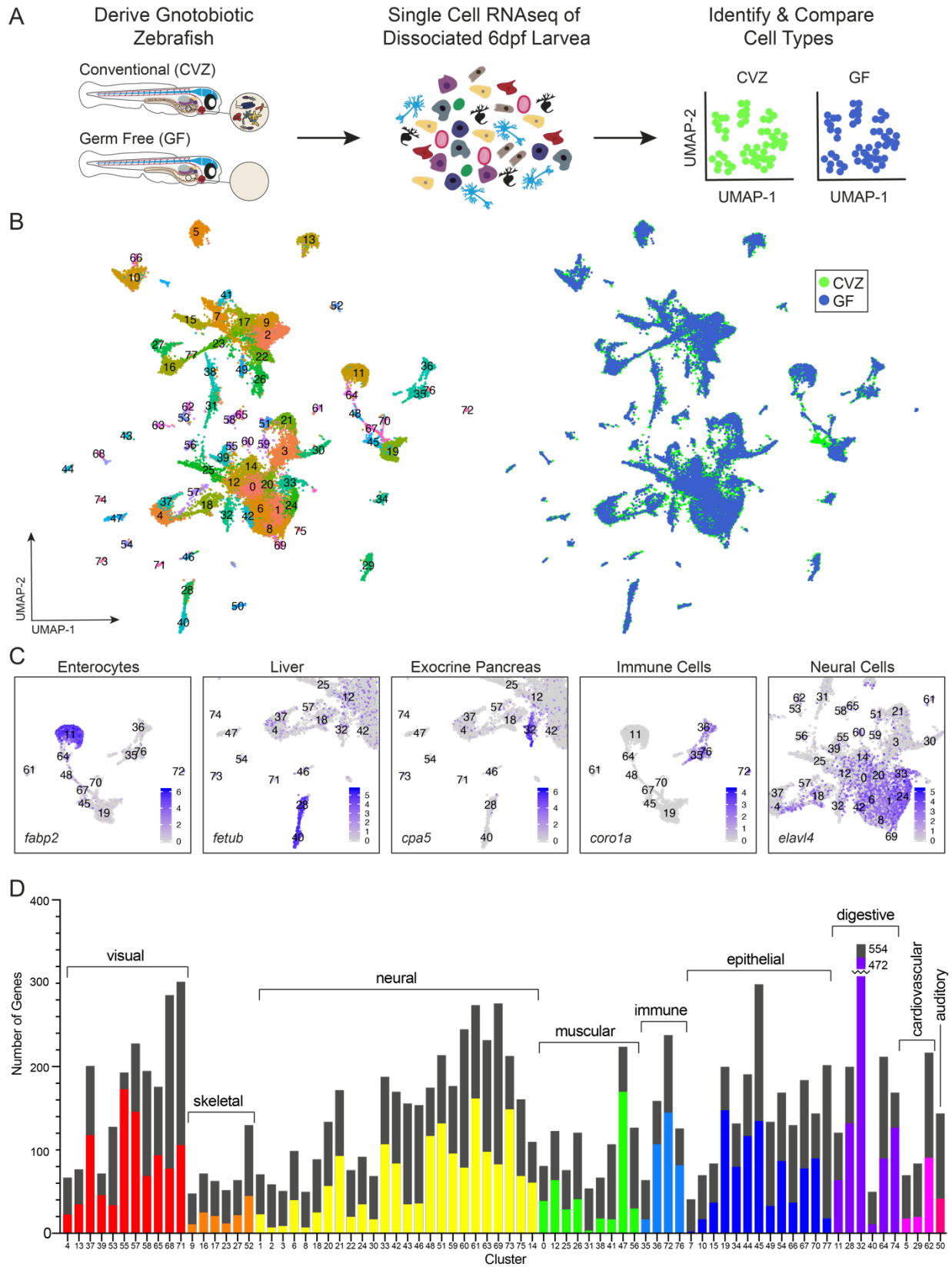
variation between CVZ and GF cells and performing a linear reduction of the data by principal

component analysis (PCA). The JackStrawPlot, ElbowPlot and HeatMap functions included in

Seurat deciphered which principal components (PC) represent a robust compression of the data

(FigS2.1&2). To confirm clustering of transcriptionally similar cells across experimental groups

by the Seurat integration strategy, we aligned 6 dpf CVZ cells with different developmental

timepoints (1, 2, 5 dpf) of whole organism dissociations from the Zebrafish Single Cell Atlas

(Farnsworth, Saunders, and Miller 2019b) (FigS2.3A&B). As anticipated, cells from our 6 dpf

CVZ zebrafish aligned more consistently with 5 dpf cells of the atlas compared to 1 and 2 dpf

cells.

By integrating CVZ and GF cells together, we generated lists of significantly enriched

genes shared by CVZ and GF cells for each cluster, allowing identification of different cell types

present in both treatment groups. Cell type identities were assigned based on tissue location

annotations from the Zebrafish Information Network (ZFIN) (Howe et al. 2013) of the top

enriched genes for each cluster and annotations from the Zebrafish Single Cell Atlas (S.Table1)

(Fig2.1C). Annotating clusters revealed that the dataset represents diverse cell types that can be

broadly categorized into tissue systems, enabling us to definitively compare transcriptional

differences between CVZ and GF treatments for cells across the vertebrate body.

Comparing differentially represented transcripts between CVZ and GF cells for each

cluster revealed that host cell responses to the microbiota occur across many cell types and are

not limited to cells that directly contact microbes (Fig2.1D). Additionally, the high variability in

the number of enriched genes across cell types demonstrates that different cells respond

differently to the microbiota even within a tissue type. One cluster, number 45, lacked GF cell

representation (Fig2.1B) in a pattern that was robust to different permutations of PC number

23

(FigS2.4A). This cluster is enriched for genes typically found within epithelia, such as *epithelial cell adhesion molecule* (*epcam*), and gene ontology (GO) analysis showed a signature of tissue regeneration and immune activation. In the GO analyses, input genes are assigned a GO term and are bioinformatically associated with GO annotations, which provides a statement about the function or biological context of genes based on current biological knowledge http://geneontology.org/. Although some of the cells express markers of skin epithelia, such as different types of *keratins*, the cluster lacks a strong signature of tissue identity (FigS2.4B-D). We postulate that this cluster consists of progenitor cells from multiple epithelial tissues that co-cluster based on their common cellular responses to microbiota that stimulate tissue growth and regeneration programs. The absence of this cell population in the GF dataset is consistent with the microbiota's role in stimulating epithelial cell proliferation in tissues including intestine (Cheesman 2011, Rawls 2006) and skin (Meisel et al., 2018). Together, these analyses illustrate that cells across the vertebrate body are responsive to the presence of the microbiota and that different cell types respond in distinct ways, as we describe below.

The microbiota stimulates intestinal epithelial cell differentiation and function

Transcriptional responses to the microbiota have been best characterized within the intestine (Heppert et al. 2021). To supplement our whole larval single cell analysis and compare it to previous analyses, we conducted an additional experiment in which we dissected larval digestive systems, comprised of intestines and associated liver and pancreas tissue, dissociated the cells and performed single cell sequencing (FigS2.5A&B). Data from dissected digestive tracts validated our annotation of intestinal epithelium, liver, and pancreas cells in the whole larva dataset (Fig2.2A&B). Comparing our lists of enriched transcripts in digestive tract clusters

with previous microarray analysis of dissected CVZ and GF larval zebrafish digestive tracts (Rawls et al., 2004) revealed good concordance between these datasets (STable. 2&3; FigS2.6).

Progenitor cells in the CVZ dataset include digestive system cells

Upon integrating the dissected digestive system cells with the whole organism dataset, we observed that CVZ epithelial cells from original cluster 45 maintained their segregation, to cluster 61, but contained fewer cells (Fig2.2A&B). We hypothesized that cells from the original cluster 45 may cluster with digestive cell types enriched in the dissected digestive system dataset. To investigate this possibility, we computationally isolated cluster 45 cells and used Seurat to integrate them with the dissected digestive system data (Fig2.2C). A majority of CVZ cells segregated to subcluster 7, indicating that most cells from original cluster 45 are extra-intestinal, likely including skin cells. However, 12% of the CVZ cells spread to diverse digestive system subclusters, suggesting that original cluster 45 includes progenitors of different digestive system tissues.

Intestinal enterocytes segregate by region and function

We identified two adjacent clusters of intestinal enterocytes based on regional specific genes (defined by Lickwar et al., 2017) including the proximal intestine genes *retinol binding protein 2a* (*rbp2a*) and *chitinase 2* (*chia.2*) (cluster 11, Fig2.3A) and the ileum-specific *fatty acid binding protein 6* (*fabp6*) (cluster 64, FigS2.7A). To further delineate cell types within clusters 11 and 64, we isolated and re-clustered these data (FigS2.7B&C), which resolved the proximal, mid, and distal intestine described in (Lickwar et al. 2017) and (Wen et al. 2021), and was consistent with further analysis of the Zebrafish Atlas intestinal epithelia clusters (Postlethwait et al., 2020). GO analysis of significantly enriched genes within cluster 11 CVZ versus GF enterocytes revealed that the microbiota stimulates specific cellular responses. We observed
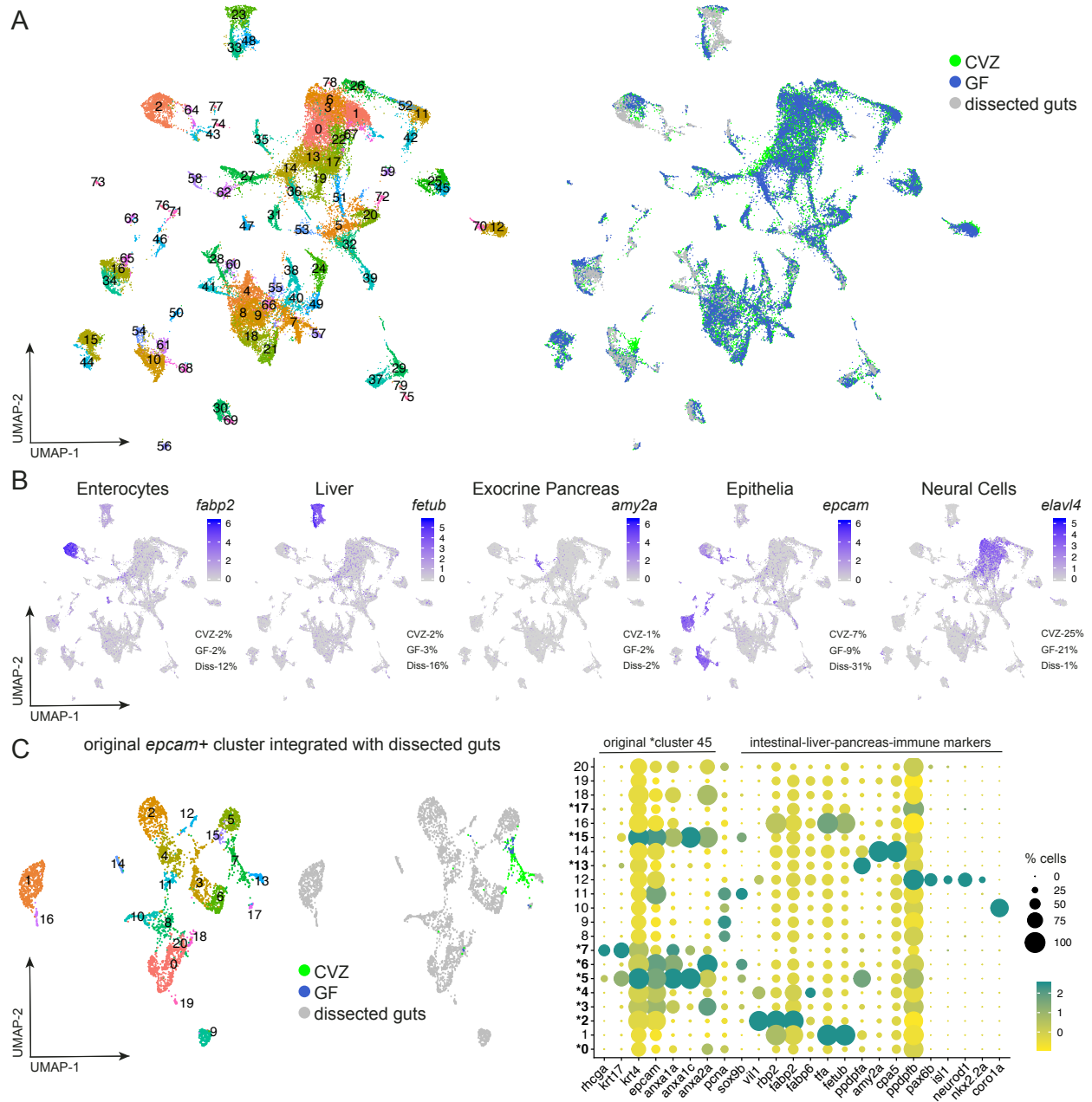
Figure 2.2 Integration of whole gnotobiotic larval zebrafish cells with dissected larval digestive systems. A) uMAP plots display the integration of single cells dissociated from whole larvae and dissected larval guts. B) Clusters populated by cells from each experimental group confirm and illustrate digestive cell types present from the dissociation of whole larvae including enterocytes (*fabp2*), liver (*fetub*), acinar of the exocrine pancreas (*amy2a*), and epithelia (*epcam*). Clusters primarily populated with cells from whole larvae dissociations and exclude cells from dissected guts, such as neural cells (elavl4) confirm the integration strategy as dissected digestive systems exclude neurons of the central nervous system. Numbers represent the percentage of cells within the designated clusters, illuminated by different biomarkers, with respect to the total cells for each experimental group. C) uMAP plots illustrate that cells from original epithelia cluster 45 include a minority of cells likely to be of the digestive system. The dotplot, and all following dotplots included in the manuscript, displays the percentage of cells within a cluster or subcluster that have a transcript present and its expression level. The expression of genes within this dotplot for a subcluster is relative to all the other subclusters. Bolded subcluster numbers designated with an * represent subclusters that include cell(s) from original cluster 45.

26

enrichment of genes involved in macromolecule catabolism, metallopeptidase activity, and lipid-protein assembly, including *apoa4b.1* and *apoa4b.2*, consistent with previous work illustrating that the microbiota regulate fatty acid metabolism and intestinal absorption (Semova et al. 2012). We also saw induction of responses to microbes such as chitinase genes *chia.1* and *chia.2*, and other cellular responses consistent with microbial challenge including responses to inorganic compounds, temperature, and ER stress (Fig2.3A). Consistent with our annotation of cluster 64 as mid intestine enterocytes, which includes lysosome rich enterocytes with high capacity for absorption and processing of luminal proteins (Park et al. 2019), we observed significant enrichment within CVZ cells of genes involved in coated vesicle and proteosome complex formation.

Intestinal goblet cells segregate with other mucus secreting cells

Goblet cells within the intestinal epithelium secrete mucins into the lumen, providing a nutrient source for microbes as well as a mucus barrier between the luminal microbiota and the epithelium. Agr2 is an endoplasmic reticulum protein disulfide-isomerase that is highly expressed in mucus secretory cells including goblet cells of the zebrafish intestine, pharynx and epidermis (Shih et al. 2007). Clusters 50 and 48 showed the highest enrichment for *agr2* (~60% and ~30%, *agr2*+ cells, respectively). To identify intestinal goblet cells within the *agr2*+ cell

Figure 2.3 Intestinal enterocytes and secretory cells characteristically respond to the microbiota at the transcriptional level. A) Cluster 11 is composed of enterocytes from the proximal intestine showing high expression of *rbp2a* and *chia.2*. Gene ontology analysis plot and genes included in the dotplot are based on genes that are significantly enriched within CVZ versus GF cells of cluster 11. Genes included in the dotplot designated with an * correspond to genes included in the gene ontology categories. B) Mucin-secreting cells are localized to cluster 48 indicated by the expression of *agr2* and *muc5.3*. The gene ontology analysis plot is based on genes that are significantly enriched within CVZ versus GF cells of cluster 48. C) uMAP plots display the integration of all *agr2*+ cells from whole larvae with the *agr2*+ goblet cells from dissected guts. By uMAP and feature plots, subcluster1 displays the most overlap of cells from each experimental group and the highest expression of *agr2*. D) Subcluster 1 demonstrates consistent pairing of cells from each experimental group from integration (ie similar numbers of cells). E) The dotplot illustrates the enriched expression of several genes within CVZ cells of subcluster 1 versus GF. The gene ontology plots in this figure and all following figures use bars to show the total number of genes included in a gene ontology category. The number in parentheses displayed after a gene ontology category title on the y-axis signify the number of gene ontology terms binned into the category. Gene ontology terms are considered significant at p<0.05 after an fdr p-corrected adjustment.

A

Cluster 11 Proximal Enterocytes

CVZ
GF

rbp2a   *chia.2*

UMAP-2
UMAP-1

Enriched Ontology
(adjusted p<0.05)

translation preinitiation (2)
macromolecule cat. (5)
response to inorg. sub. (2)
sensory organ morph. (2)
chitin metabolism (7)
response to temp. (3)
response to microbiota (5)
metallopeptidase activity (1)
ER stress (6)
lipid-protein assembly (77)

number of genes

Differential Expression
(p<0.05)

% cells
20
40
60
80

apoa4b.2
tmem176l.2
slc5a1
pvalb8
*apoa4b.1**
anxa4a
*chia.1**
ifi46
*chia.2**
muc13b

CVZ
(382 cells)
GF
(380 cells)

0.4
0.0
-0.4

B

proximal
enterocytes
11
mid/distal
64
secretory
48
ionocytes
70
67
45
gill/
epidermis
19

UMAP-2
UMAP-1

agr2
11
64
48
70
67
45
19
5
4
3
2
1
0

muc5.3
11
64
48
70
67
45
19
4
3
2
1
0

Enriched Ontology
(adjusted p<0.05)

snRNP complex (6)
protein catabolism (8)
H. peroxide/
antibiotic metab.(11)
ATP Metabolism (71)

number of genes

C

aligning all *agr2*+ cells with dissected intestinal *agr2*+ goblet cells

2
0
1
4
3

UMAP-2
UMAP-1

CVZ
GF
dissected guts

2
0
1
4
3

agr2
4
2
0

D

sub cluster	Number of cells		
	CVZ	GF	Dis.
0	224	227	184
*1	72	70	72
2	46	62	0
3	16	9	56
4	1	9	21

E

krt96 *fscn1a* *dynll2b* *ccl20a.3* *sesn1* *elf3* *marcksl1b* *b2m* *cfd* *psme2* *nucks1a* *anxa4* *ptgdsb.1* *avil* *mdkb* *gyg1b* *hmgb2a* *mdka* *grtp1a* *agr1* *muc5.1*

GF
(70 cells)

CVZ
(72 cells)

*p<0.05   p<0.09   ns

% cells
20
40
60
80

0.4
0.0
-0.4

28
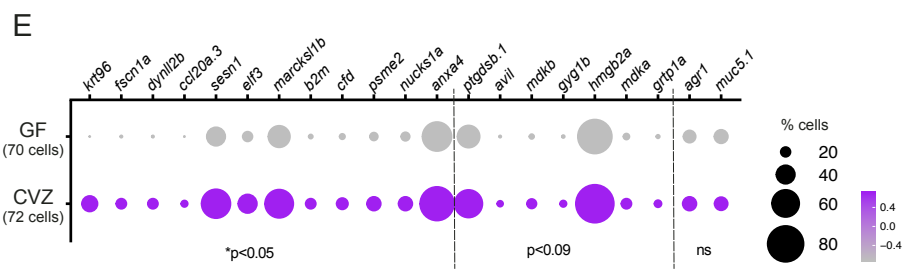
populations, we looked for expression of zebrafish orthologs of *Muc2* and *Muc5*, mucins specific

to mouse intestinal goblet cells (Larsson et al. 2011). We did not observe *muc2.2* expression in

our data set of whole or dissected digestive systems or in the Zebrafish Atlas, but saw enriched

muc5.1 and *muc5.3* expression within cluster 48, which lies between the intestinal and epidermal

epithelium clusters (S.Table1) (Fig2.3B). GO analysis showed significant enrichment of genes

involved in hydrogen peroxide and antibiotic metabolism within CVZ cells of cluster 48 relative

to GF, consistent with protective responses elicited in cells in close proximity to microbes.

Because expression of *muc5.3* in zebrafish has been found within the esophagus and intestine

(Udayangani et al. 2017; Jevtov et al. 2014), we speculate that cluster 48 includes extra-intestinal

goblet cells.

To refine our identification of intestinal goblet cells, we analyzed *agr2* expression in the

dissected digestive system dataset (FigS2.5B). In addition to digestive tracts, the dissections

included epidermal epithelia with two main populations of *agr2*+ cells. One of the *agr2*+

populations is localized within cluster 4 which has an enrichment for intestinal epithelial markers

(FigS2.5B). We predict this population of *agr2*+ are secretory cells likely to be intestinal. Using

the Seurat integration method, we combined all CVZ and GF *agr2*+ cells from whole larvae with

the dissected intestinal *agr2*+ population from cluster 4 (Fig2.3C; FigS2.5B). In the integration

strategy, matching cells from each dataset are anchored together based on the similarity of their

transcriptional profile. Within integrated datasets, clusters that include similar numbers of cells

from each treatment group represent the pairing of cells in a similar biological state. For

example, the CVZ data of our 6 dpf cells segregate consistently with the 5dpf cells of the

Zebrafish Atlas where there are similar cell types present in each dataset (FigS2.3). By

integrating the *agr2*+ cells from cluster 4 of the dissected digestive systems with the *agr2*+ cells

from whole gnotobiotic zebrafish, subcluster 1 exhibited the most similar number of cells between the datasets suggesting these cells have similar transcript expression patterns, including enriched expression of *agr2* relative to the other subclusters (Fig2.3C&D). The CVZ and GF cells within subcluster 1 may represent intestinal mucus secretory cells because they paired consistently with the intestinal *agr2+* cells of the dissected digestive system. Compared to GF cells within subcluster 1, CVZ cells had a significant enrichment of genes involved in cytoskeleton architecture (*krt96, fscn1a, dynll2b*), immune responses (*ccl20a.3, sesn1, b2m*), and transcription factor activity and development, including Notch signaling (*elf3, marcksl1b, cfd, psme2, nucks1a*) (Fig2.3E). Although not statistically significant, several genes involved in developmental pathway signaling and growth *(ptgdsb.1, avil, mdkb, mdka)*, glycogen synthesis (*gyg1b*), innate immune response (*hmgb2a*), vesicle trafficking (*grtp1a*) and markers of goblet cell function (*agr1, muc5.1*) were increased. The increased expression of genes involved in the Notch pathway is consistent with our previous finding that the microbiota promotes goblet cell fates through regulation of Notch signaling (Bates et al. 2006; Troll et al. 2018).

The microbiota stimulates functional maturation and activation of the immune and neural cells

Immune Cells

Immune cells are dedicated to surveilling and responding to microbial cues (Saliba et al. 2016; Stubbington et al. 2021; Carmona et al. 2017) and previous studies in gnotobiotic zebrafish have illustrated the responses of these cells to microbiota (Bates et al. 2007; Rolig et al. 2015; Wiles and Guillemin 2020; Murdoch and Rawls 2019; Kanther et al. 2014). Neutrophils, the major immune cell populations in the larval zebrafish, marked by *myeloid-specific peroxidase* (*mpo*) and macrophages, marked by *macrophage expressed 1, tandem duplicate 1 (mpeg1.1)*, segregated to clusters 72 and 36 respectively (Fig2.4A&B). We additionally annotated cluster 35
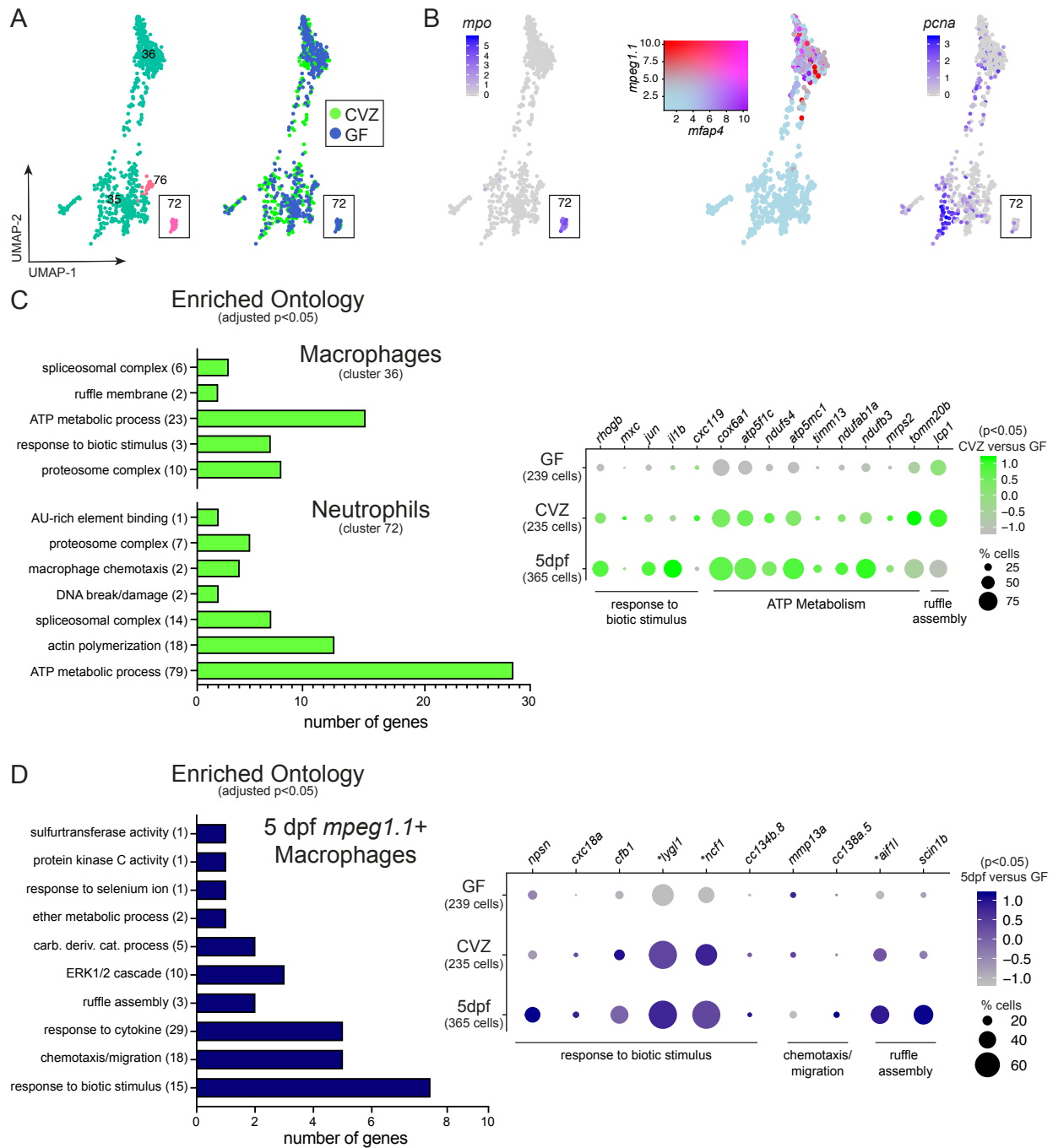
2.4 Immune cells characteristically respond to the microbiota at the transcriptional level. A) Clusters 35, 36, 72 and 76 include different immune cell types that differentially express B) neutrophil biomarker *mpo* and varying degrees of co-expression of macrophage biomarkers *mpeg1.1* and *mfap4*. The expression of pcna within cluster 35 display a population of immune progenitors. C) The gene ontology analysis plot and dotplot represent genes significantly enriched genes within CVZ macrophages and neutrophils versus GF, showing similar trends in gene expression when compared to conventional 5dpf larvae of the Zebrafish Atlas. D) The gene ontology analysis plot and dotplot represent genes significantly enriched genes within the 5dpf Zebrafish Atlas macrophages and neutrophils versus GF, showing similar trends in gene expression when compared to CVZ cells. Genes designated with a * were also statistically enriched within CVZ cells versus GF.

as immune cell progenitors as it showed an enrichment of several genes involved in larval

hematopoietic development including IKAROS *family zinc finger* 1(*ikzf1*) and genes involved in

cell cycle regulation and proliferation, such as the proliferation marker *proliferating cell nuclear*

antigen (*pcna*). Cluster 35 also contains markers of lymphocytes, which are just starting to

develop during larval life, including *recombination activating1* (*rag1*) and *zeta chain of T cell*

receptor associated protein kinase 70 (zap70) (Iwanami et al. 2020). In the neutrophil cluster 72,

genes enriched in the presence of microbiota included a striking number involved in

characteristic immune cell functions including threat sensing and processes involved in

chemotaxis such as cell shape changes and cellular processing (Fig2.4C). These findings were

similar for CVZ compared to GF *mpeg1.1+* macrophages and were cross validated with the

Zebrafish Atlas immune cells (Fig2.4C&D). Among the CVZ enriched genes within the

progenitor immune cell population of cluster 35 were genes involved in DNA recombination and

break repair consistent with the microbiota stimulating lymphocyte maturation.

Enteric Neurons

Mounting evidence indicates that both the peripheral (PNS) and central (CNS) nervous

system branches sense and respond to the microbiota (Hsiao et al. 2013; Arentsen et al. 2016;

Sampson and Mazmanian 2015; Jameson et al. 2020). Our dataset contained a large number of

neurons, marked by expression of the panneuronal, postmitotic marker *ELAV like neuron-*

specific RNA binding protein 4 (*elavl4*). Here we focus on intestine-resident neurons of the

enteric nervous system (ENS), a major component of the PNS that regulates diverse functions

within the intestine including secretion, motility and homeostasis (Spencer and Hu 2020).

Based on expression of *paired like homeobox 2* (*phox2bb* and *phox2a)*, enteric neurons

segregated to cluster 33 (Fig2.5A). Cluster 33 is composed of multiple types of PNS neurons as

co-clustering of enteric neurons and cranial ganglia is consistent with both cell types being

derived from neural crest. Re-analyzing cluster 33 alone resulted in 7 subclusters, with high co-

expression of *phox2bb* and *phox2a* in subclusters 4, 5, and 6 (Fig2.5B). These subclusters

showed expression of transcription factor *hoxb5a*, providing additional evidence they are

composed of enteric neurons (Shepherd and Eisen, 2011) (Fig2.5C). Transcriptional differences

between these three enteric neuron clusters reflected temporal progression of their maturation

(Taylor et al. 2016). Subcluster 6 is characterized by expression of early ENS fate determinants

(*hoxb5a*, *hand2*, and *sox10*) and limited neural receptor and enzyme expression, suggesting this

population contains enteric neuron progenitors. In subcluster 4, the combinatorial expression of

elavl4 and *ret*, with little expression of *sox10*, suggests a differentiating population (Taylor et al.

2016). Subcluster 4 shares some transcription factor expression with subcluster 6, but increased

expression of neural receptors (*ngfra*, *chrna5*, *chrna3*) and enzymes (*nos1*) is consistent with

increasing functionality. Subcluster 5 resembles fully mature enteric neurons lacking expression

of *ret* and *sox10* (Taylor et al. 2016) and enriched in expression of neural receptors (*htr3a*, *ngfra*,

chrna5) and neural markers (*calb2a*, *elavl4*).

     To further explore identities of these enteric neuron subpopulations, we performed GO

analysis on genes significantly expressed within each subcluster followed by comparing

transcriptional differences between CVZ and GF cells for each population (Fig2.5D). Within the

progenitor-like population of subcluster 6, several genes involved in ribosome biogenesis,

regulation of cell cycle and neurogenesis were enriched versus the other subclusters.

Figure 2.5 Enteric neurons are heterogenous transcriptionally and in response to the microbiota. A) Cluster 33 includes enteric neurons and cranial ganglia showing expression of *phox2bb* and *elavl4*. B) uMAP plots show the transcriptional heterogeneity of cluster 33 with a range in the co-expression of *phox2bb* and *phox2a*. C) Violin plots compare the expression of enteric neuron transcription factors, neural receptors, and neural markers within subclusters 4, 5, and 6. D) The gene ontology analysis plots represent genes enriched within subcluster 4, 5 or 6 relative to the other subclusters. The corresponding dotplots represents differential gene expression in CVZ versus GF cells for each subcluster.

A Cluster 33 Enteric Neurons & Cranial Ganglia

CVZ
GF

phox2bb   *elavl4*

UMAP-2
UMAP-1

B Sub Clusters of Cluster 33

phox2bb
phox2a

C

transcription factors

hox5b   *isl1*   *hand2*   *sox10*

Expression Level

receptor/enzymes

ret   *nos1*

Expression Level

neural receptors

htr3a   *ngfra*   *chrna5*   *chrna3*

Expression Level

neural markers

calb2a   *elavl4*

Expression Level

4   5   6

D Enriched Ontology
(adjusted p<0.05)

Subcluster 6
(progenitor-like)

retinoic acid biosyn. (1)
S-adenosylhomocys. (1)
reg.of neurogenesis (3)
actomyosin org. (14)
cartilage development (24)
regulation of cell cycle (1)
erythrocyte differentiation (5)
ribosome biogenesis (59)

Subcluster 4
(differentiating)

oxygen binding (3)
enhancer binding (4)
chromatin DNA binding (9)
proton trans./ATPase (12)
synapse/axon (7)

Subcluster 5
(functional)

autophagosome (3)
Ras prot. sig. (1)
growth factor sig. (3)
maint. of cell polarity (3)
transcription initiation (4)
response to metal ion (6)
cell resp. to NO levels (2)
organelle localization (4)
nuc./purinergic sig. (91)
actin cytoskeleton org. (41)
neurotransmitter trans. (105)   332
neuron projection (18)

0   50   100   150
number of genes

% cells (p<0.05)

34

Within this population, several genes involved in defense response to bacteria and viruses, metal ion binding activity and cell-cell adhesion were enriched within CVZ cells versus GF. Within the differentiating population of subcluster 4, GO analysis revealed an enrichment of genes associated with chromatin DNA binding and the synapse or axon. CVZ cells of subcluster 4 also showed enrichment of genes involved in defense response to microbes, as well as genes involved in cell maintenance and differentiation. GO analysis of genes enriched within subcluster 5 is consistent with mature neurons including: maintenance of cell polarity, purinergic signaling, neurotransmitter transport and axon projection. The presence of the microbiota induced the enrichment of many genes in CVZ cells including those involved in voltage-gated ion channel activity and neurotransmission. Our analysis of enteric neurons corroborates their transcriptional heterogeneity (Taylor et al., 2016) and illustrates how cells of different maturation states respond differently to the microbiota.

Central Neurons

Our analyses of CNS neurons also showed transcriptional responses to the microbiota that differed between cells in different maturity states. We identified a neuronal population likely to be composed of progenitors within cluster 21 by expression of the nervous system development gene *pleiotrophin* (*ptn*) and co-expression of *fatty acid binding protein 7 (brain a)* (*fapb7a*) and *hairy-related 4* (*her4.2*) (FigS2.8A&B) (Farnsworth, Saunders, and Miller 2019a). Similar to progenitor-like enteric neurons, several genes involved in regulation of neurogenesis were enriched within cluster 21 CVZ neurons, as were genes involved in immune function and development (FigS2.8C&D). These findings are consistent with mounting evidence implicating connections between immune response and neurodevelopment (Deverman and Patterson 2009; Straley et al. 2014; Seleme et al. 2017). We additionally observed that mature serotonergic

and/or dopaminergic neurons in cluster 69 responded to the microbiota by upregulating genes involved in synaptic machinery for dopamine transmission (FigS2.8E&F). CVZ neurons of cluster 69 also exhibited significant enrichment of genes involved in mitogen activated protein (MAP) kinase activity, negative regulation of proliferation, and calcium binding activity, suggesting that the microbiota plays a role in promoting characteristic function of mature neurons (FigS2.8G). Together these data suggest the microbiota plays specific and distinct roles within neuronal cell types, depending on their developmental state, and promotes neural functions at all developmental stages, including neurogenesis, neurodifferentiation, and neurotransmission.

Host tissues exhibit global patterns of microbiota responsiveness

Our GO analyses across each of the 78 clusters revealed two striking patterns. First, ATP metabolism genes such as *ATP synthase (atpf51b)*, *cytochrome c oxidase (COX5B)*, *voltage-dependent anion channel 2 (vdac2)*, which were widely expressed across all cells, were consistently expressed at higher levels in cells from CVZ animals. Second, whereas lens-associated crystallin genes were almost entirely restricted to the lens cell clusters 39 and 55 in CVZ animals, there was widespread expression of multiple crystallin genes across virtually all GF clusters, predominantly of the beta and gamma type (FigS2.6A-C, FigS2.9). We explored each of these transcriptional signatures further.

The microbiota elicit cell type specific ATP metabolism gene expression patterns

ATP is an ancient molecule whose production spurred one of the earliest examples of host-microbe interactions with the endosymbiosis event that produced mitochondria. ATP production has been shown to be modulated by the microbiota in a number of model organisms. In GF fruit flies, whole body ATP levels and mitochondrial respiration are significantly lower

than in conventional counterparts (Gnainsky et al. 2021). In mice, ATP levels and mitochondrial respiration are reduced in colonic tissue (Donohoe et al. 2011) and lower in feces from GF animals (Atarashi et al. 2008). We measured the concentration of ATP in whole zebrafish larvae and observed extensive inter-individual variation in levels of this rapidly metabolized molecule with no significant difference between GF and CVZ groups (Fig2.6B). We explored ATP turnover in dissected larval intestines and found that ADP:ATP ratios varied extensively across individuals, with a trend toward lower ratios in CVZ larvae. We tested the rate of ATP metabolism in the presence and absence of microbiota by microgavaging exogenous hydrolysable ATP into the proximal intestinal lumen, dissecting out the intestine, and measuring ATP levels 20-40 minutes later. The average amount of remaining ATP was significantly less in CVZ samples, indicating that ATP was consumed faster within CVZ intestines. These measurements are consistent with the transcriptional signature of upregulation of enterocyte cellular activities and ATP production genes.

ATP production for energy currency or cell-signaling can vary extensively across diverse cell types. For example, neurons use massive amounts of ATP for both neural activity (MacVicar, Wicki-Stordeur, and Bernier 2017) and as a signal co-released with various

Figure 2.6 The microbiota induce global patterns of host gene expression. A) The dotplot illustrates global trends of gene expression across all CVZ and GF cells from whole larvae. B) The boxplots show the concentration of ATP from whole larvae, the ratio of ADP:ATP of dissected larval intestines, and the concentration of ATP in dissected larval intestines challenged with exogenous ATP between CVZ and GF zebrafish. Each dot in the measurement of ADP:ATP, represents the average to two intestines together. C) The boxplots show the average relative optical density of beta and gamma crystallin protein expression in CVZ and GF larvae. Each symbol represents an individual flask of larvae processed together and matching symbols denote that flasks were derived from the same derivation. The colored dots represent the average expression between flasks in the same derivation with a line connecting the corresponding derivations between experimental groups. D) The uMAP plots represent the CVZ cells clustering based on the expression of genes involved in ATP metabolism. The feature plots illustrate cell types maintaining their original cell type identity. E) The hierarchal tree demonstrates the proximity of the clusters from the uMAP plot of cells clustered by the expression of ATP metabolism genes and the top gene ontology terms based on the enriched genes within a given cluster. F) The uMAP plots represent the corresponding GF cells from the original clusters used in 'D', clustering based on the expression of crystallin genes. The feature plots illustrate cell types lose their original cell type identity. G) The hierarchal tree demonstrates the proximity of the clusters from the uMAP plot of cells clustered by the expression of crystallin genes and the top gene ontology terms based on the enriched genes within a given cluster. Data from *in vivo* experiments, and for the remaining figure, are shown as boxplots with the middle line representing the median, the top and bottom of the box representing the upper and lower quartile, the whiskers representing the min and max values, and the '+' symbol representing the mean. Each dot symbol represents data from a single 6dpf zebrafish, unless specified otherwise. p-values displayed on the boxplots are the result of Students' Ttest, where $p<0.10$ is considered trending and $p<0.05$ is considered significant.

A

ak1 atp5f1b atp5j atp5mc3b atp5pb atp5po cox4il cox5aa COX5B cox6a1 cox8a gapdhs mdh1aa ndufa4l nme2b.2 vdac2 cryba4 cryba2b cryba1b crybn2 crygmx crygm2d1 crygm2d2 crygm2d7 crygm2d8 crygm2d10 crygm2d11 crygm2d17 crygm2d21

CVZ
(17,037 cells)

GF
(16,955 cells)

% cells
20
40
60
80

0.4
0.0
-0.4

B

[ATP] whole larvae

(nmol ATP/fish)

p=0.34

CVZ   GF

ADP:ATP intestine

Ratio

p=0.062

CVZ   GF

[ATP] intestines

nmol ATP / fish gut

p=0.0070

CVZ   GF   CVZ   GF
gavage
ATP

C

cryba4
(19kDa)

actin
(42kDa)

CVZ   GF

crygs
(24kDa)

actin
(42kDa)

CVZ   GF

Optical Density
(a.u.)

CVZ   GF

CVZ   GF

D

UMAP-2
UMAP-1

fabp7a

elavl4

coro1a

cpa5

E

Top Gene Ontology Terms
(adjusted p<0.05)

12 immune response; mig./chemotaxis
3 fatty acid/prostaglandin biosythesis
6 O₂/gas transport; nucleus/chromatin
4 resp. to EtOH/abx; chrom. packing
8 O₂/gas transport; iron binding
7 O₂/gas transport; lipid/sterol metabolism
16 lipid/sterol metabolism (1 gene, apoa1b)
14 O₂/gas transport; heme binding
13 endopeptidase activity
11 lateral line devo; immune activation
1 ribosome assembly; regeneration
2 purine nuc. metab; muscle/ECM
9 ribosome assembly; purine nuc. metab.
10 mRNA processing; splice/chrom. mod.
0 mRNA processing; neurotransmission
5 regulation of neurogenesis
15 resp. nutrients; membrane org.

F

UMAP-2
UMAP-1

fabp7a

elavl4

coro1a

cpa5

G

Top Gene Ontology Terms
(adjusted p<0.05)

32 lens development
26 collagen fibril org
19 gas/drug transport
13 intermediate filament
27 O₂ binding
8 mRNA met. process
29 lens development
14 visual perception
12 lens development
25 ribosome biogenesis
22 ribosome assembly
28 lens development
11 lens development
6 lens development
24 lens development
16 lens development
33 lens development
1 ATP metabolism
15 lens development
21 gas/drug transport
4 neurotransmitter trans.
2 pigment gran. org.
9 axon development
36 myeloid/leukocyte mig.
0 ribosome assembly
20 neg. reg. protein metab.
10 reg. mRNA splicing
3 lens development
5 mRNA processing
7 RNA splicing
17 mRNA processing
35 ret. gang. cell axon guid.
18 drug metab. proc.
23 RNA splicing
30 RNA splicing
34 neurotransmitter trans.
31 mRNA processing

neurotransmitters (Huang, Otrokocsi, and Sperlágh 2019). Neutrophil production of ATP

sustains their immune cell activation and also initiates purinergic signaling for chemotaxis (Bao

et al. 2014). CVZ cells of one third of the clusters in our dataset, representing diverse cell types,

exhibited enriched gene ontologies pertaining to nucleotide and/or ATP metabolism (FigS2.10).

To better understand how ATP metabolism genes are deployed in response to microbiota, we

bioinformatically isolated CVZ cells from the clusters with enriched ATP metabolism genes and

re-clustered them using a panel of 228 ATP metabolism-associated genes (Fig2. 6D). This

analysis revealed 17 new subclusters that largely maintained their original cell-type identity,

including neural progenitors (84% of original cluster within subcluster 5, *fabp7a*), neurons (73%

of original cluster within subcluster 0, *elavl4*) and immune cells (88% of original cluster

neutrophils, within subcluster 12, *coro1a*) (S.Table4). Some original populations were split into

different subclusters, such as acinar cells of original cluster 32 exocrine pancreas with 35%

within subcluster 15, 19% within subcluster 13 and 14% within subcluster 0, suggesting

heterogeneity of exocrine pancreas ATP metabolism gene utilization. GO analysis illustrated a

diversity of biological processes among the enriched genes within each subcluster, consistent

with cell type specific functions (e.g. upregulation of chemotaxis in neutrophils). This analysis

demonstrates that the microbiota stimulates cell type specific ATP metabolism gene expression.

The microbiota suppress cry gene expression across tissues

       Crystallins are a heterogenous group of extremely stable proteins that make up the

vertebrate eye lens. In most vertebrates, including zebrafish, rodents, and primates, the lens is

comprised of alpha crystallin small heat shock proteins and the beta and gamma crystallin

superfamily members. The zebrafish genome contains 3 alpha *cry* genes, 13 beta *cry* genes and

40 gamma *cry* genes (Farnsworth et al., 2021). Related *cry* genes were present in early

vertebrates prior to lens evolution, implying that they have additional functions (Zigler and Sinha 2015). In both rodents and zebrafish, alpha and beta crystallin are expressed and function in other tissues beside the lens, including the retina, brain, heart, and testes (Piri, Kwong, and Caprioli 2013; Sprague-Piercy et al. 2021; Zigler and Sinha 2015). At the protein level, we confirmed expression of Cryba4 is not restricted to the head or lens of GF larvae but is also detected within the body (FigS2.11A). Across several GF derivations, the levels of Cryba4 and Crygs were consistently increased in samples of pooled GF larvae.

Alpha crystallins function as chaperones that prevent aggregation of client proteins damaged by heat or oxidative stress (Sprague-Piercy et al. 2021). We explored whether *cry* genes were co-expressed with other chaperone genes in GF cells. Although we saw strong co-expression of *cry* genes with each other, they were not co-expressed with heat shock genes *hsp70.2* and *hsp70.3* (FigS2.10B). We also did not observed any relationship between the expression of crystallin genes and ATP metabolism genes (FigS2.10B). We next explored how *cry* genes are regulated in the absence of microbiota by re-clustering GF cells based solely on the 87 *cry* genes found within the sequencing data (Fig2.6F). In contrast to clustering with ATP metabolism genes, this analysis revealed clusters that largely lost their original cell-type identity; for example, neural progenitors (*fabp7a*), postmitotic neurons (*elavl4*), and immune cells (*coro1a*) no longer co-segregated within a specific sub-cluster (S.Table5), showing that *cry* genes are expressed across many cell types in the absence of the microbiota.

Among the GO analysis of GF cells clustered by *crystallin* genes, reoccurring themes were related to cellular architecture (e.g. collagen fibril organization and intermediate filaments) and vesicle trafficking (e.g. neurotransmitter transmission and gas/drug transport) (Fig2.6G). Comparing differentially expressed genes across all CV and GF cells showed a significant co-

enrichment of several *keratin* genes and *crystallin* genes in GF cells (FigS2.12A-C). Keratins are intermediate filaments that confer structural integrity to postmitotic cells. Intermediate filaments are critical for neural development but their myriad functions have been under appreciated (Bott and Winckler 2020). Re-clustering *elavl4+* cells from the whole dataset that includes predominantly postmitotic neurons shows separation of GF and CV cells, with co-expression of certain *keratins* and *crystallins* in many *elavl4+* cells but especially within GF cells (FigS2.12D). This finding is consistent with an apparent decrease in arbor remodeling of specific CNS neurons in GF larvae (Bruckner et al, 2020) and the role intermediate filaments play in stabilizing microtubules and inhibiting synaptic vesicle trafficking during synaptic rewiring in postmitotic neurons (Kurup et al. 2018). Co-expression of *keratins* and *crystallins* extended across diverse cell types throughout the dataset (FigS2.12E).

We hypothesize that widespread expression of *cry* genes in GF cells counteracts cellular stresses induced by the artificial germ-free state. Beyond their roles in preventing lens protein aggregation and cataracts, Cry proteins chaperone cytoskeletal components and maintain cytoplasmic organization in diverse postmitotic cells (Slingsby and Wistow 2014). In a meta analysis of zebrafish proteins upregulated in response to various biological stressors, Cry proteins are among the top 25 protein families (Groh and J-F Suter 2015). Alpha crystallins have antiapoptotic and neuroprotective functions in the retina and other neuronal tissues (Phadte, Sluzala, and Fort 2021; Z. Zhu and Reiser 2018). In the retinal epithelium, both alpha and beta crystallins stabilize the vacuolar-ATPase, thereby protecting against lysosomal dysfunction (Cui et al. 2020; Valapala et al. 2016). In several mouse models of maternal immune activation, there is a transient upregulation of representative members of all *Cry* gene families within the embryonic brain, consistent with neuroprotection (Garbett et al. 2012). In zebrafish, alphaB-

crystallin confers protection against cardiac stress induced by crowding or cortisol (Mishra et al. 2018). More broadly, small heat shock proteins similar to alpha crystallins are abundant in cells experiencing low metabolic states, such as in nematodes dauers (Fu et al. 2021) and brine shrimp and insects in diapause (Khodajou-Masouleh et al. 2021) (King and Macrae 2015). Proteins sharing the tertiary structure of beta and gamma crystallin are also associated with stabilizing metabolically inert cells, for example Spherulin 3a, which is upregulated in cysts of the eukaryotic slime mold *Physarum polycephalum*, and Protein S, which is upregulated in *Myxococcus xanthus* spores (Clout et al. 2001; Wistow, Summers, and Blundell 1985). Our global transcriptional analysis of GF cells suggests they are less proliferative and less metabolically active than CV counterparts, potentially necessitating cellular mechanisms to maintain cytoplasmic organization, prevent protein aggregation, and stabilize stalled, energetically costly processes, such as vesicular trafficking.

Exocrine pancreas responses illustrate how the microbiota promotes tissue development and function

Of all the cell clusters, the exocrine pancreas acinar cell cluster 32, marked by high expression of digestive enzymes including *amylase alpha 2a* (*amy2a*) and *carboxypeptidase A5* (*cpa5*), had the largest number of microbiota-induced transcripts (Fig2.7A). The exocrine pancreas produces digestive enzymes that are packaged into vesicle-granules and delivered into the intestinal lumen by secretion. GO analysis indicated that the microbiota stimulate a diversity of biological processes in exocrine pancreas acinar cells including secretory vesicle formation, transporter activity and regulation of catabolism, consistent with several studies showing associations between exocrine pancreas function and microbiota composition (Pietzner et al. 2021; Adolph et al. 2019; Nishiyama et al. 2018; Frost et al. 2019). Pancreatic acinar cells are

not only responsible for organ function but also for organ development (Wan et al. 2006; Biemar et al. 2001; Field et al. 2003). Other genes upregulated within CVZ acini included genes
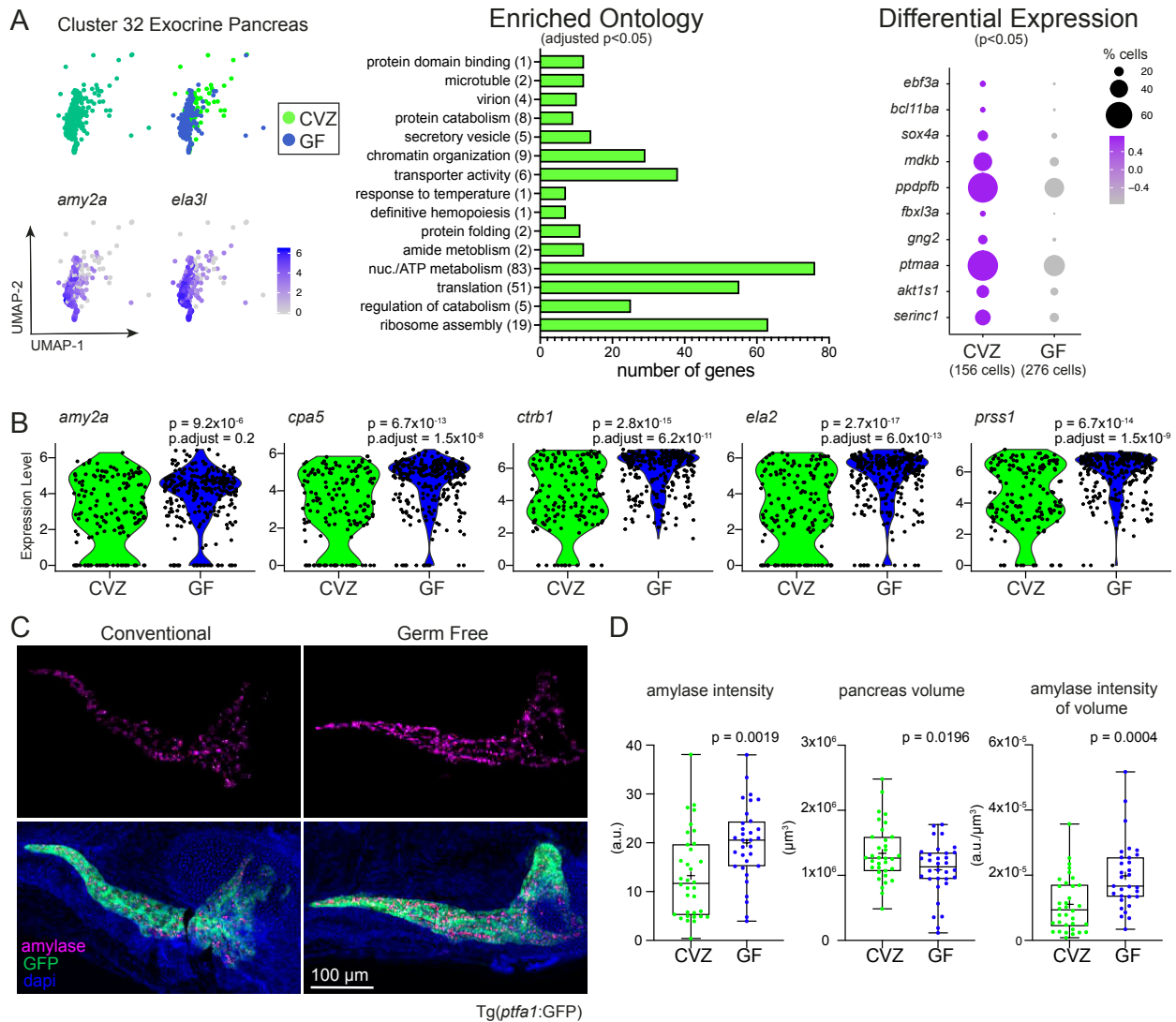
Figure 2.7 The microbiota promote tissue development and function within the exocrine pancreas. A) Cluster 32 is composed of acinar cells from the exocrine pancreas showing high expression of digestive enzymes *amy2a* and *ela3l*. Gene ontology analysis plot and genes included in the dotplot are based on genes that are significantly enriched within CVZ versus GF cells of cluster 32. B) Violin plots illustrate the difference in distribution of pancreatic digestive enzyme gene expression between CVZ and GF cells from cluster 32. C) Images display amylase protein expression within the exocrine pancreas of CVZ and GF larvae. Images are z-projections taken by tile-scan with a confocal microscope. D) The boxplots illustrate the average optical intensity of amylase expression, the volume of space taken by the pancreas and the intensity of amylase standardized to the volume of space taken by the pancreas. All images were taken with the same optical settings and larvae from each experimental group were imaged on the same day in the same session.

involved in DNA-binding transcription factor activity (*ebf3a*, *bcl11ba*, *sox4a*) as well as growth

factor activity and organ development (*mdkb*, *ppdpfb*, *fbxl3a*, *gng2*, *ptmaa*, *akt1s1*, *serinc1*).

Specifically, enriched expression of *pancreatic progenitor cell differentiation and proliferation factor b* (*ppdpfb*) and growth factor *midkine b* (*mdkb*) within CVZ cells suggests that the microbiota promote pancreas development.

In contrast to the increased expression of genes involved in acinar cell differentiation and function in CVZ cells, we noted a subtle but consistently elevated level of digestive enzyme transcripts in GF acinar cells (Fig2.7B). Expression of these digestive enzyme genes shows a bi-modal distribution within CVZ acinar cells, whereas GF acini showed consistently high expression. To validate transcriptional increases of digestive enzymes within GF exocrine pancreas, we derived Tg(*ptfa1*:GFP) zebrafish CVZ and GF to mark the exocrine pancreas and stained larvae with an antibody against Amylase to characterize expression of this digestive enzyme (Fig2.6C&D). GF larvae had higher intensity of Amylase labeling within the pancreas (Students' Ttest, p=0.0019) but the average pancreas volume was larger within CVZ larvae (Students' Ttest, p=0.0196). These analyses reveal how microbiota stimulate both tissue expansion and function, promoting developmental plasticity in response to the microbial environment. In the artificial GF state, the exocrine pancreas appears to be stunted and composed of fully differentiated acinar cells that, based on their transcriptional patterns, are poised in a functionally inactive state, a situation that may require upregulation of cytoprotective Cry proteins to maintain.

Conclusion

Our analyses reveal widespread and diverse host responses to the microbiota that are not limited to host cells in direct contact with resident microbes. Our study illustrates the utility of single cell transcriptomics to explore intricate interactions between vertebrate hosts and their

resident microbes. This dataset will be a valuable resource for generating new hypotheses about the impacts of the microbiota on animal development and physiology.

Materials & Methods

Gnotobiotic zebrafish

All protocols used for zebrafish experiments were approved by the University of Oregon Institutional Care and Use Committee (protocol number 14-14RR). Adult zebrafish were maintained using standard husbandry procedures (Westerfield., 2007) by the University of Oregon Zebrafish Facility. Zebrafish embryos were derived germ free (GF) as previously described (Melancon et al. 2017) and Conventionalized (CVZ) zebrafish were inoculated with parental tank water post derivation. Animals used for the single cell analysis of whole larvae were *Tg*(*nkx2.2a*:eGFP) (Pauls et al. 2007) and multiple clutches were collected from natural mating of adults from the same tank. The *Tg*(*ins*:eGFP) line of zebrafish were used for dissections of the digestive system prior to single cell dissociation. Animals used for western immunoblots were wild-type (ABCxTu strain). The Tg(*ptfa1*:GFP) line of zebrafish were used for all analyses with the exocrine pancreas. Unless specified, all zebrafish used were 6 days post fertilization (dpf).

Whole larval zebrafish dissociation

Larval zebrafish were anesthetized with tricaine (Western Chemical, Inc., Ferndale WA). 15 healthy zebrafish with inflated swim bladders were selected for dissociation and put into a sterile 1.5 ml Eppendorf tube. Stock solutions of 10mg/mL proteinase K (Millipore Sigma, St. Louis, MO) and 10mg/mL collagenase P (Millipore Sigma, St. Louis, MO) were prepared by dissolving in 1X HBSS (Thermo Fisher Scientific, Waltham, MA). Embryo media was pipetted out from each tube and 37C-warmed 1.3 ml of dissociation solution (0.12 mg/mL proteinase K

and 1 mg/mL collagenase P in 1x TrypLE (Thermo Fisher Scientific, Waltham, MA) was added

to each tube. Each tube was quickly added to a heat block at 37C, with a timer starter. To

dissociate the whole 15 zebrafish per tube into individual cells, each tube was mixed by pipetting

25x every 2 minutes until ~12 minutes is on the timer. To stop the dissociation reaction, 200 ul

of 4C 'stop solution' (5% calf serum, 1 mM CaCl2, and PBS) (Farnsworth, Saunders, and Miller

2019) was added to each tube, mixed by pipetting 5x, and immediately put into a pre-chilled 4C

table-top centrifuge. Dissociated cells were gently spun down for 3 minutes at 350g and the

supernatant removed. Cell pellets were gently rinsed and resuspended with 1mL of 1% Bovine

Serum Albumin (BSA) in HBSS and pelleted again for 3 minutes at 350g. The supernatant was

removed and the pellet resuspended in 100 μl of 0.04% BSA/HBSS. Cells were then filtered

through a pre-chilled 40 μM strainer into a pre-chilled Eppendorf tube, using a pre-chilled 1mL

syringe plunger to gently pestle the cells through. An additional 100 μl of fresh 0.04%

BSA/HBSS was used to rinse remnants of the strainer into the tube. A Bio Rad TC20 cell

counter was used to measure cell concentration and cell viability with Trypan Blue (Bio Rad,

Hercules, CA). Additionally, pilot experiments confirmed single cell dissociation with this

protocol by visual inspection using the DIC feature of a LEICA DM6 confocal microscope

(Leica Microsystems Inc., Buffalo Grove, IL). All groups of dissociated cells had over 80%

viability and were diluted to a final concentration of 3,500 cells/ul in 0.04% BSA/HBSS for

cDNA library preparation. The timing from euthanasia to handing off dissociated cells for library

preparation was under 30 minutes.

Larval zebrafish dissections & dissociation

For larval zebrafish dissections, the entire digestive system (intestine, pancreas, liver)

was dissected as previously described (Rolig et al. 2015). Briefly, zebrafish were derived germ

free and at 5 dpf were anesthetized in tricaine (Western Chemical, Inc., Ferndale WA), mounted on a slide and their digestive systems sterilely dissected. Dissected tissue was isolated and put into L-15 culture medium (Thermo Fisher Scientific, Waltham, MA) supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Waltham, MA), penicillin-streptomycin (MilliPore Sigma, St Louis, MO) and gentamycin (VWR, Radnor, PA). Dissected tissue was incubated overnight in culture media at room temperature prior to single cell dissociations.

Single-cell cDNA preparation

The University of Oregon Genomics and Cell Characterization Core Facility (https://gc3f.uoregon.edu/) performed the sample preparation by running the samples on a 10X Chromium Single Cell 3' platform using v2 chemistry. The goal was to target 10,000 cells per group and the resulting cDNA libraries were amplified with 10 cycles of PCR. The cDNA libraries were first sequenced with experimental groups combined onto a single Hi-seq 4000 lane. This resulted in a low coverage depth. We next had each experimental group sequenced onto its own Hi-seq 4000 lane. The output from all lanes was combined, which reached an optimal number of reads for each sample. All samples were prepared and sequenced on the same days.

Computational analysis

The sequencing data was aligned to the zebrafish genome (GFCz11_91) using the 10X Cell-ranger pipeline (3.0.2). The Seurat software package for R (3.1.4) was used to subject the data to standard pre-processing workflow. The data was filtered such that any cells expressing more than 3,000 genes, including more than 50,000 read counts, or more than 20% mitochondrial genes were not included in the final analysis. The expression level of each gene was normalized by total expression via log-transformation with a 10,000 scale factor. We performed a linear

dimensional reduction of the data by principle coordinate analysis (PCA) and calculated 150

principle components (PC). Based on Jack Straw methods to determine significance of PCs, the

first 122 PCs significantly explain the variance of the data (p<0.0001) but by Elbow Plot most of

the variation can be explained between 30 to 60 PCs (FigS2.1-2). To better grasp how adding

additional PCs impact our data, we performed multiple clustering analyses comparing the

inclusion of 30, 60, 90, and 122 PCs using all genes in the dataset. The data we are reporting in

subsequent figures includes a conservative 60 PCs with a resolution of 3.0, which produces 78

clusters where we can conservatively identify distinct cell types and avoid possible technical

noise.

Differential gene expression between the clusters was performed using the

FindAllMarkers function in Seurat, which used Wilcoxon rank sum tests. We then isolated the

raw data from specific cell types by sub-setting a cluster(s) or subset cells globally that were

positive for a biomarker(s) of interest. Re-analyzing specific cell type populations allowed us to

understand the extent of gene expression heterogeneity within a cell population and the

FindAllMarkers function was used to find differentially expressed genes between CV and GF

cells in the sub populations. Gene ontology (GO) analysis on the lists of differentially expressed

genes (p<0.05) was done using the ClusterProfiler software package for R (3.14.3) and the

genome wide annotation for zebrafish org.Dr.eg.db (3.10.0). Resulting GO terms were subject to

FDR p-value correction to limit potential false-positives. To distill the repetitive GO annotations

and terms, we manually triaged through each analysis and binned similar annotations that were

described by the same sets of genes into broader categories. GO analyses are shown as bar

graphs displaying the number of cumulative genes that contribute to the gene GO category. All

GO terms, genes that contribute to the terms and which terms were binned into our broad

categories is made available.

Western immunoblot

Larval zebrafish were euthanized by tricaine and ~30-40 zebrafish were added to an

Eppendorf tube. Leftover embryo media was removed from each tube and replaced with 400-600

μl of 4C from a 10mL stock of cold lysis buffer (RIPA buffer (Boston BioProducts, Ashland,

MA) and ½ protease cocktail inhibitor (Thomas Scientific, Swedesboro, NJ)). Samples were

sonicated with a microtip at 20% amplitude, 1 second pulse on, 0.3 second off for 30 seconds.

This step was repeated until fish were completely dissociated, taking care that a sample would

not become warm. Tubes were then put in the freezer for 15 minutes, followed by thawing on ice

prior centrifugation at 14,000 rpm for 20 minutes at 4C. The supernatant from each tube was

isolated and protein concentration for each sample was quantified using the Pierce BCA Protein

Assay Kit (Thermo Scientific, Rockford, IL). 20 μg of each sample was loaded onto a 4-20% Gel

(Bio Rad, Hercules, CA) for electrophoresis and transferred to a PVDF membrane (GE

Healthcare Amersham, Chicago, IL). For Cryba4 (Thermo Fisher Scientific, Waltham, MA)

detection, membranes were blocked in 5% milk in standard tris base, saline tween (TBST) and

then probed with a rabbit polyclonal antibody (Thermo Fisher Scientific, Waltham, MA) at 1:100

in blocking buffer overnight at 4C. For Crygs (Thermo Fisher Scientific, Waltham, MA)

detection, separate membranes were blocked in 5% BSA in TBST and then probed with a rabbit

polyclonal antibody (Thermo Fisher Scientific, Waltham, MA) at 1:100 in blocking buffer

overnight at 4C. The resulting bands were visualized with a secondary antibody Anti-rabbit IgG

HRP-linked antibody (Thermo Fisher Scientific, Waltham, MA) at 1:1000 for 1 hour at room

temperature. Membranes where then stripped and re-probed for the loading control Actin or

Tubulin. An Anti-Actin polyclonal antibody (Millipore Sigma, St. Louis, MO) made in mouse was diluted at 1:1000 in TBST and incubated with the membrane overnight at 4C. The resulting bands were visualized with a secondary antibody Anti-mouse IgG HRP-linked antibody (Cell Signaling Technology, Dancers, MA) at 1:1000 for 1 hour at room temperature. Protein densities for Cryba and Cryg2 bands were measured and normalized to the protein densities of the corresponding actin bands. Each sample was at least triplicated and averaged according to each gnotobiotic derivation.

ATP assays

For measuring total ATP concentration of whole larvae, 6/7 dpf zebrafish were euthanized by tricaine. Individual zebrafish were put into an Eppendorf tube, pipetting out as much excess liquid as possible and 100 μl of 95C distilled water was added to each fish. The tubes were immediately but on the 95C heat-block. All tubes of zebrafish were incubated at 95C for 20-30 minutes, mixing by pipette every 5 minutes. All sample carcass remnants were pellet by centrifugation at 14,000 rpm for 30 minutes at 4C. The ATP concentration of the supernatant was calculated by the ATP Determination Kit (Thermo Scientific, Rockford, IL) relative to a standard. The same protocol was used to measure the ATP concentration of dissected digestive systems of larvae but dissected tissue was boiled in 30 μl of 95C water and centrifugation was omitted. The ADP/ATP Ratio Assay Kit (Bioluminescent) (Abcam, Cambridge, MA) was used to measure the average ratio of ADP to ATP of larvae within dissected digestive systems from 2 zebrafish per sample using the same tissue preparation with 95C water as described above.

Gavage of ATP

Larval zebrafish were challenged by direct administration of 1 or 10mM of ATP (Millipore Sigma, St. Louis, MO), dissolved in embryo media, directly into the proximal

intestine by gavage as previously described (Cocchiaro and Rawls 2013). Approximately 20-40 minutes post gavage, the digestive systems of zebrafish were dissected and processed for measuring ATP concentration as described above.

Statistical analysis

Data from *in vivo* experiments are displayed as boxplots with the data median as the line within the box, the top and bottom of the box representing the upper and lower quartile, the whiskers representing the min and max values, and the '+' symbol representing the mean. A Students' Ttest was used to gauge statistical differences of means between CVZ and GF groups in non-sequencing experiments. For all statistical tests, $p<0.05$ is considered significant

Supplemental Figures
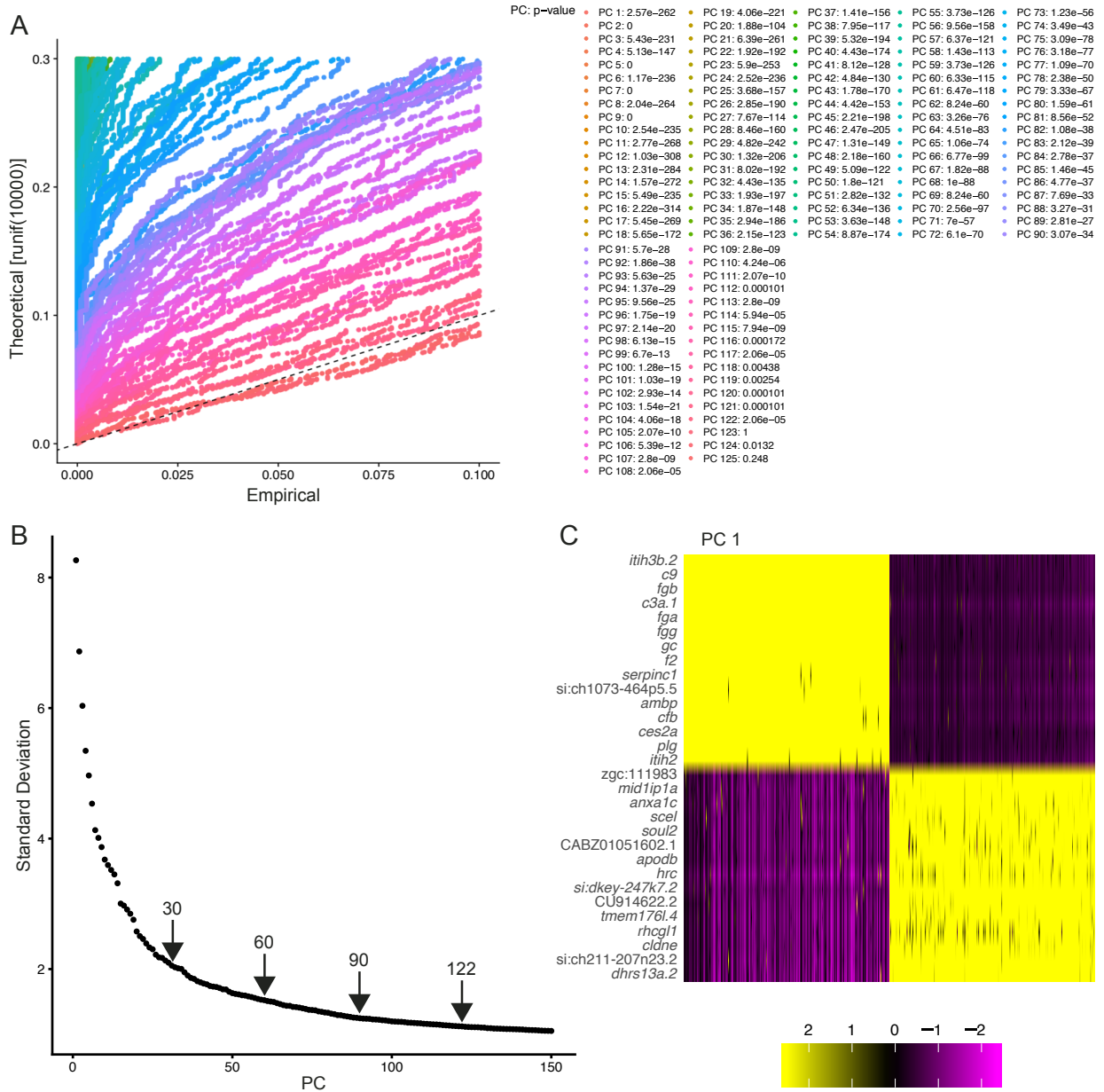
Figure S2.1 Determination of principle components to include for downstream analyses. A) The Jackstraw plot illustrates the p-value for each principle component (PC) and the B) Elbow plot displays how much variation is attributed to each subsequent PC. C) The heatmap of the genes significantly contributing to the variation within PC1 are those expressed within the liver and integument.
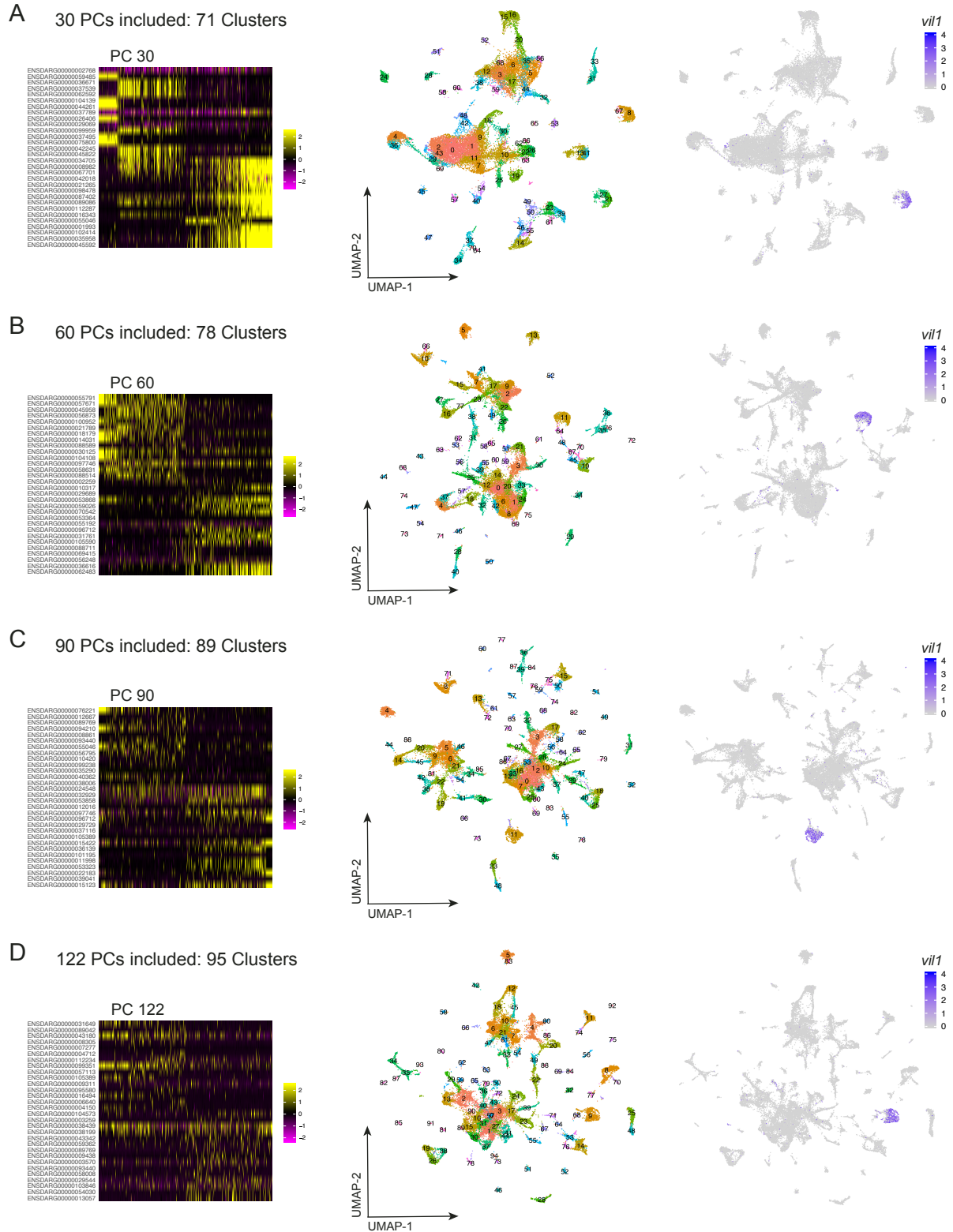
Figure S2.2 The inclusion of additional PCs increases the number of clusters. The Heatmap, resulting uMAP and expression of intestinal biomarker *vil1* are displayed after the inclusion of the first A) 30, B) 60, C) 90, and D) 122 PCs.
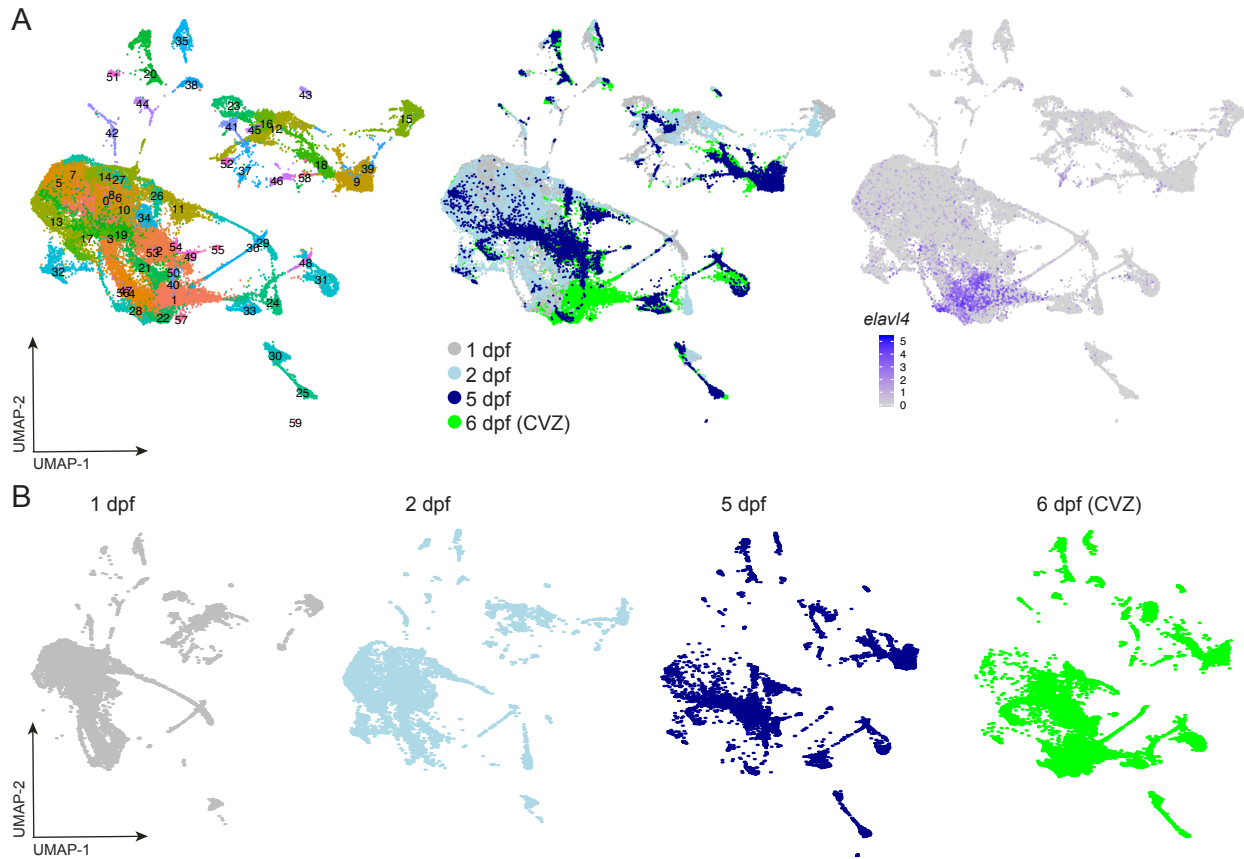
Figure S2.3 The transcriptomic profile of the 6 dpf CVZ cells are similar to the 5dpf larvae of the Zebrafish Atlas. A) Integration of the Zebrafish Atlas 1, 2, 5 dpf whole organism single cell RNAseq data with the CV 6dpf data illustrates that the 6dpf cells align more closely with the 5dpf cells than the earlier timepoints with the exception of the neural cells (*elavl4+*) which are B) largely absent within the 5dpf data sets.
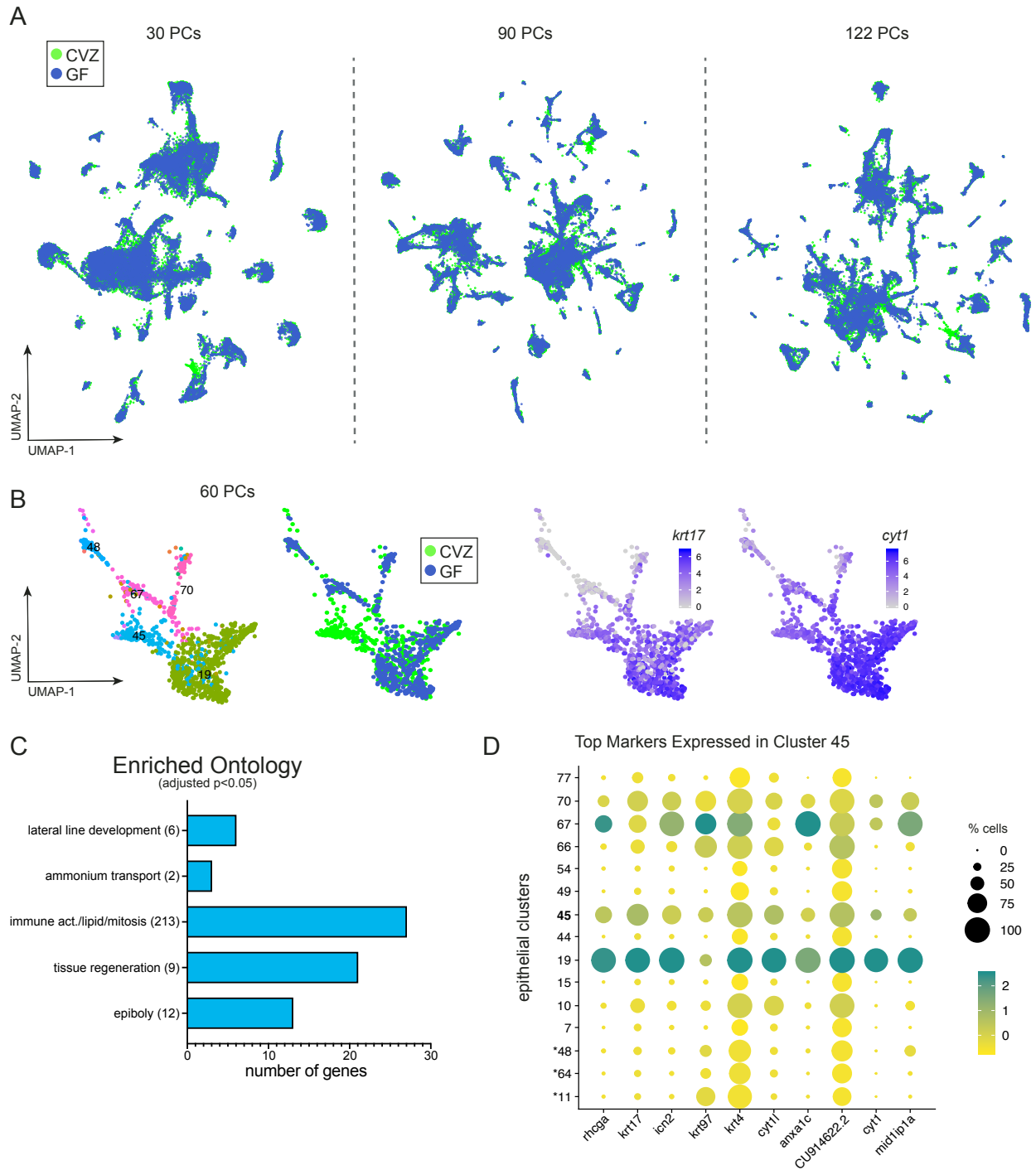
Figure S2.4 GF zebrafish lack representation of an epithelia cell type. A) The uMAP plots display a transcriptionally distinct type of epithelia is overrepresented within CVZ cells despite different permutations of PC inclusion. B) Cluster 45 has high expression of *krt17 cyt1* similar to cluster 19 epithelia. C) The gene ontology analysis plot is based on genes enriched within cluster 45 versus all other clusters and D) the dotplot illustrates that the top markers expressed within cluster 45 are broadly expressed across the different clusters that are categorized as epithelia. The * designates clusters that contain cells from the intestinal epithelia.

Figure S2.5 Dissected intestines of larval zebrafish prior to single cell sequencing enrich for a diversity of cell types within the digestive system. A) The digestive systems of larval zebrafish were dissected and isolated prior to single cell dissociation and RNA sequencing. B) The uMAPs shows the diversity and enrichment of intestinal epithelia and associated digestive system cells.
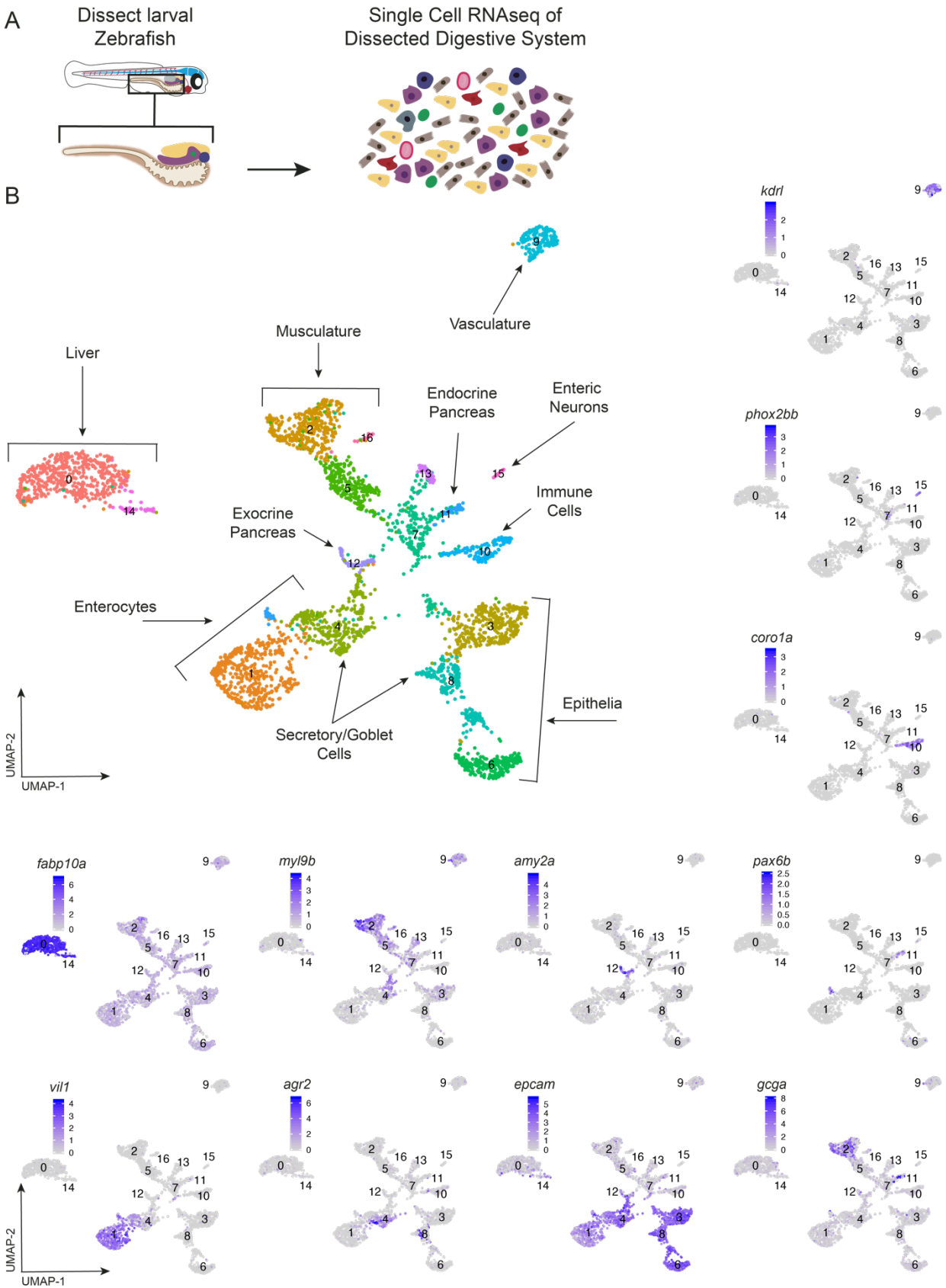
Figure S2.6 Comparison of differentially expressed genes in gnotobiotic microarray versus single cell RNA-sequencing of digestive systems. Common genes between gnotobiotic dissected intestines of 6dpf zebrafish larvae from Rawls et al., 2004 microarray analysis and the cell types from single cell dissociations likely included in dissected digestive systems. Congruent genes highlighted in green indicate a similar enrichment within the respective CVZ treatments while genes highlighted in red indicate the opposite trend (i.e. an increase of expression within GF single cells).

Comparing Enriched Gene Expression in CV Cells
(Rawls et al., 2004 Microarray on Dissected Guts versus Single Cell Clusters)

Similar Trend (green) **Opposite Trend** (red)

| Rawls et al 2004 DEGs	Cluster 11 & 64 Enterocytes	Cluster 28 & 40 Liver	Cluster 48 & 61 Secretory Cells	Cluster 33 Enteric Neurons	Cluster 32 Exocrine Pancreas	Cluster 72 Neutrophils	Cluster 36 & 76 Macrophages
ANXA2	anxa4	anxa4	anxa1a	anxa1a	anxa1a	anxa4	anxa1a, anxa11b
AP1S1					ap2m1a		ap1s2, ap2s1, ap3s2
Apob	apoa4b.1, apoa4b.2 (2x), apoc2, apoa1b, apoa2	apoc1, apom	apoa1b, apoa2	apoa1b, apoa2, apoe	apoc1, apoda.2	apoa1b, apoa2, apoc1, apom	
ARP2	arpc1a				arpc2		
C3 C4		c3a.3 c4b					
Calr[g]	calr		calr3a	calr	calr		
CBX1		cbx7a			cbx1a, cbx1b, cbx7a		
CORO1C						coro1a	
Dlc2		dynll2b					
DNAJB11		dnajc5aa	dnajc5aa	dnaja2a, dnajc7, dnajc8	dnaja2b, dnajc5aa		
			dnajb12a				
Gpx2	gpx1b	gpx1a	gpx1a, gpx1b / gpx4a				gpx1a
HMGA1	hmgb1b	hmgb1a, hmgb1b	hmga1a, hmgb1b		hmgb1a, hmgb1b		
HMGN2		hmgn7			hmgn2, hmgn6, hmgn7		
HSPD1[h]	hsp10.1, hsp70.2, hsp90b1, hspa5	hsp70.2, hsp90aa1.2	hsp90b1	hsp70l	hsp70l, hsp90ab1	hsp90ab1	hsp70.2, hsp70.3
HSPD1[h]		hspd1	hspd1				
IF2	eif2s1b, eif3ha / eif3c, eif3d, eif3ja, eif3m, eif4abp2	eif1b, eif4a1b, eif4bb					
Ifit1	ifi45, ifi46		ifi46				
KPNA2			kpna3				
LSM6			lsm5, lsm6				
MAPRE1[g]	map1ab				map1aa, map1lc3b, map4l	mapre1a	
Mcm5			mcm5				
MFAP4							MFAP4 (1 of many)
MSN	msna						
NUCKS		nucks1a			nucks1a		
PABPC1			pabpc1b		pabpc1a		
Pcna			pcna				
PFDN2		pfdn1			pfdn1	pfdn6	
PHB	phb						
PPP1R3B			ppp1r7		ppp2cb, ppp2r1ba		ppp2cb
PPP4R2	ppp3r1b						
Psma5	psma1, psma3, psma6a		psma4, psma8		psma3, psma4, psma5	psma3	psma4, psma5, psma6a, psma6l
Psmb3			psmb4	psmb1, psmb6		psmb1	psmb6
Psmd12	psmd2, psmd14	psmd6	psmc6		psmd1, psmd3	psmd7, psmd11a	
Psme3	psme2		psme1				psme1, psme2
PTGDS	ptgdsb.1						
SDF2L1				sdf2l1			
SF3B4		sf3b6	sf3a2				sf3a1, sfb1
SMARCA5			smarce1		smarce1		
SNRPD1			snrpd1		snrpb, snrpd2		snrpd1, snrpd2
			snrpd2				
SNRPE						snrpe	
SPC18	spcs3						
SRI						sri	
TOMM34			tomm20b	tomm5			tomm20b
Tpm3	tpm1				tpma		
TSPAN-1	tspan13a, tspan15	tspan3a			tspan2a, tspan3a, tspan7b		
UBE2N	ube2na	ube2e3	ube2ib	ube2ia	ube2ib, ube2v1	ube2ia	ube2v1
ZNF259		znf395b	ZNF276	znf536	znf395a, znf395b, znf593, znf609b		

Figure S2.7 Enterocytes transcriptionally segregate by proximal to distal localization within the intestine. A) Cluster 11 and 64 are composed of intestinal enterocytes with enriched expression of *vil1*, *muc13b*, *fabp6* and *anxa4*. B) The uMAP plots show the transcriptional heterogeneity of intestinal enterocytes from cluster 11 and 64 which C) segregate by the expression of proximal to distal intestinal epithelia biomarkers. D) The uMAP plots shows minimal co-expression of secretory cell biomarkers within cluster 11 and 64. E) subcluster 7 has the most expression of secretory cell biomarkers and the gene ontology analysis plot is based on the genes enriched within subcluster 7 versus the other subclusters by pink bars as well as genes enriched within CVZ cells of subcluster 7 versus GF by green bars.
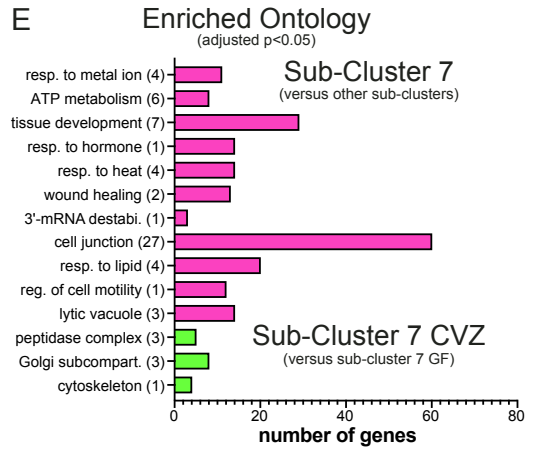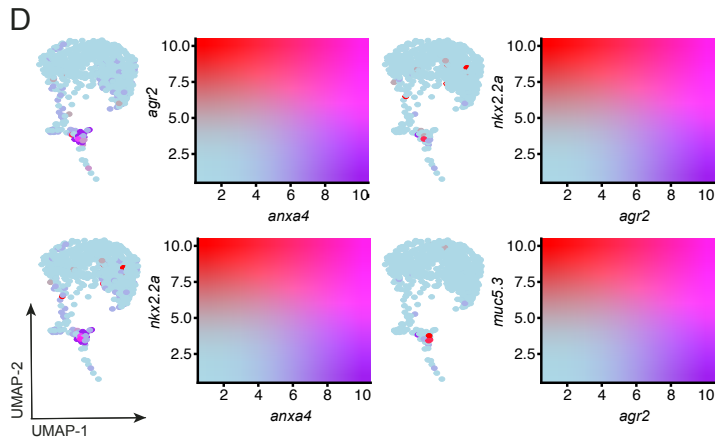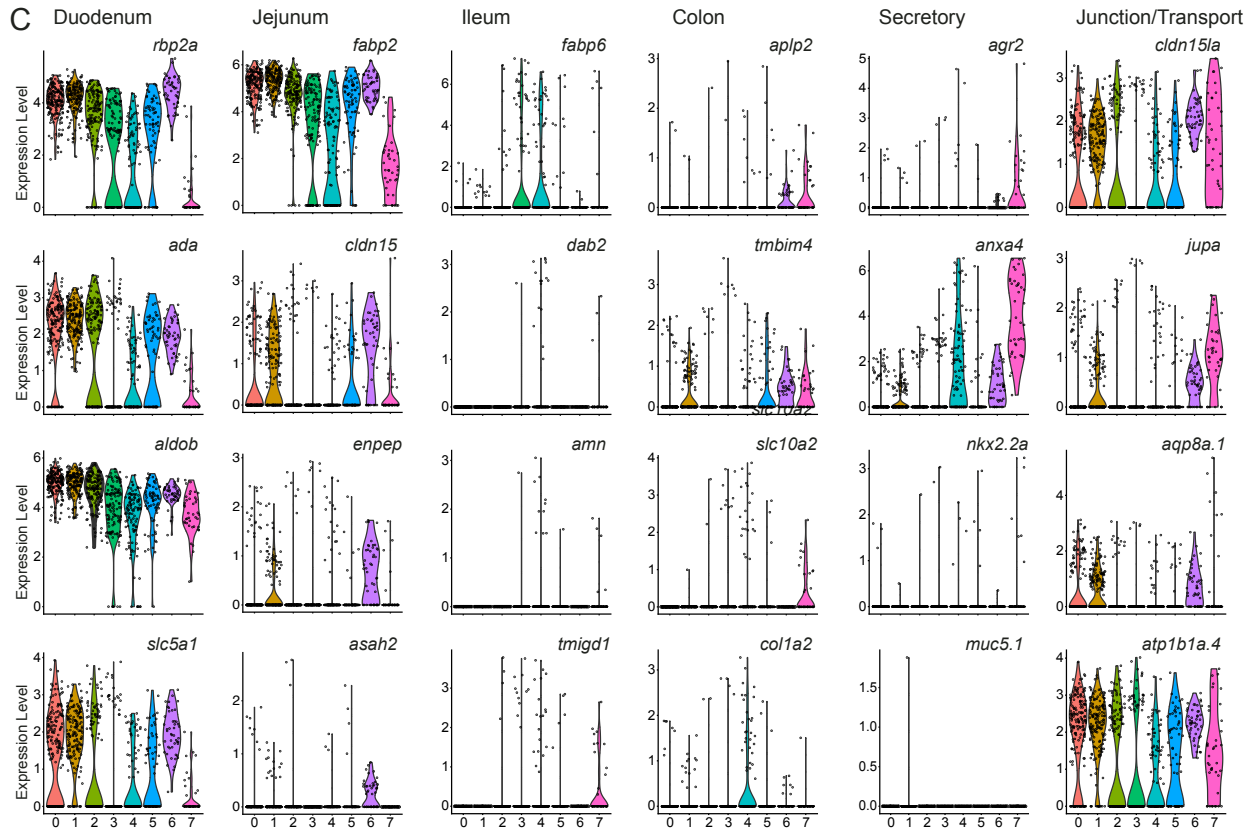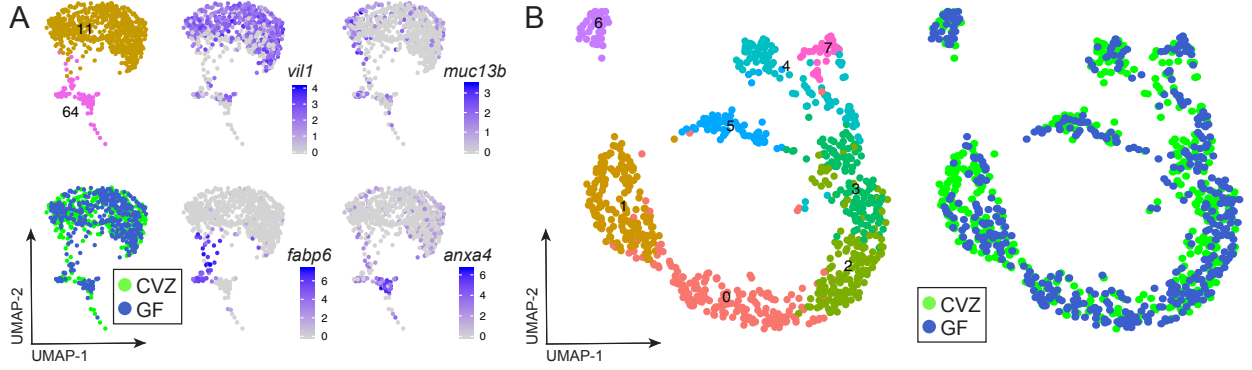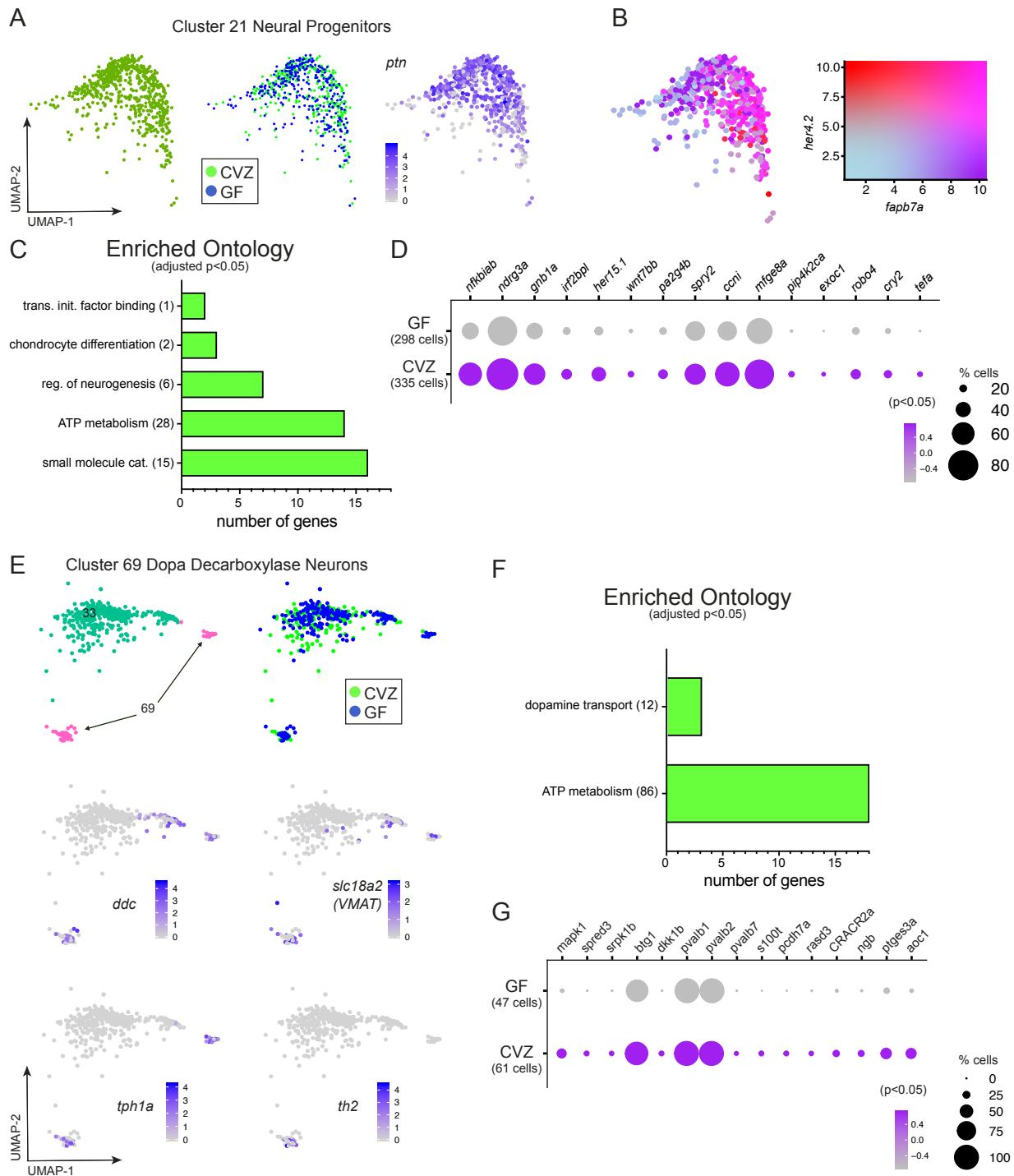
Figure S2.8 Neurons of the CNS respond to the microbiota. A) Cluster 21 is composed of neural progenitors showing enriched expression of *ptn* and B) the co-expression of *fapb7a* and *her4.2*. C) The gene ontology analysis plot and D) dotplot is based on the enrichment of genes within CVZ cells of cluster 21 versus GF. E) Cluster 69 is populated with serotonergic and dopaminergic neurons displaying expression of *ddc*, *slc18a2*, *tph1a*, and *th2*. F) The gene ontology analysis plot and G) dotplot is based on the enrichment of genes within CVZ cells of cluster 69 versus GF.

Figure S2.9 Crystallin genes are enriched within GF cells. A) The violin plot shows expression of *crygn2* within lens fiber clusters 39 and 55 with enriched expression in GF cells across all clusters. B) The uMAP plots compare the global expression of crystallin in CVZ and GF cells.
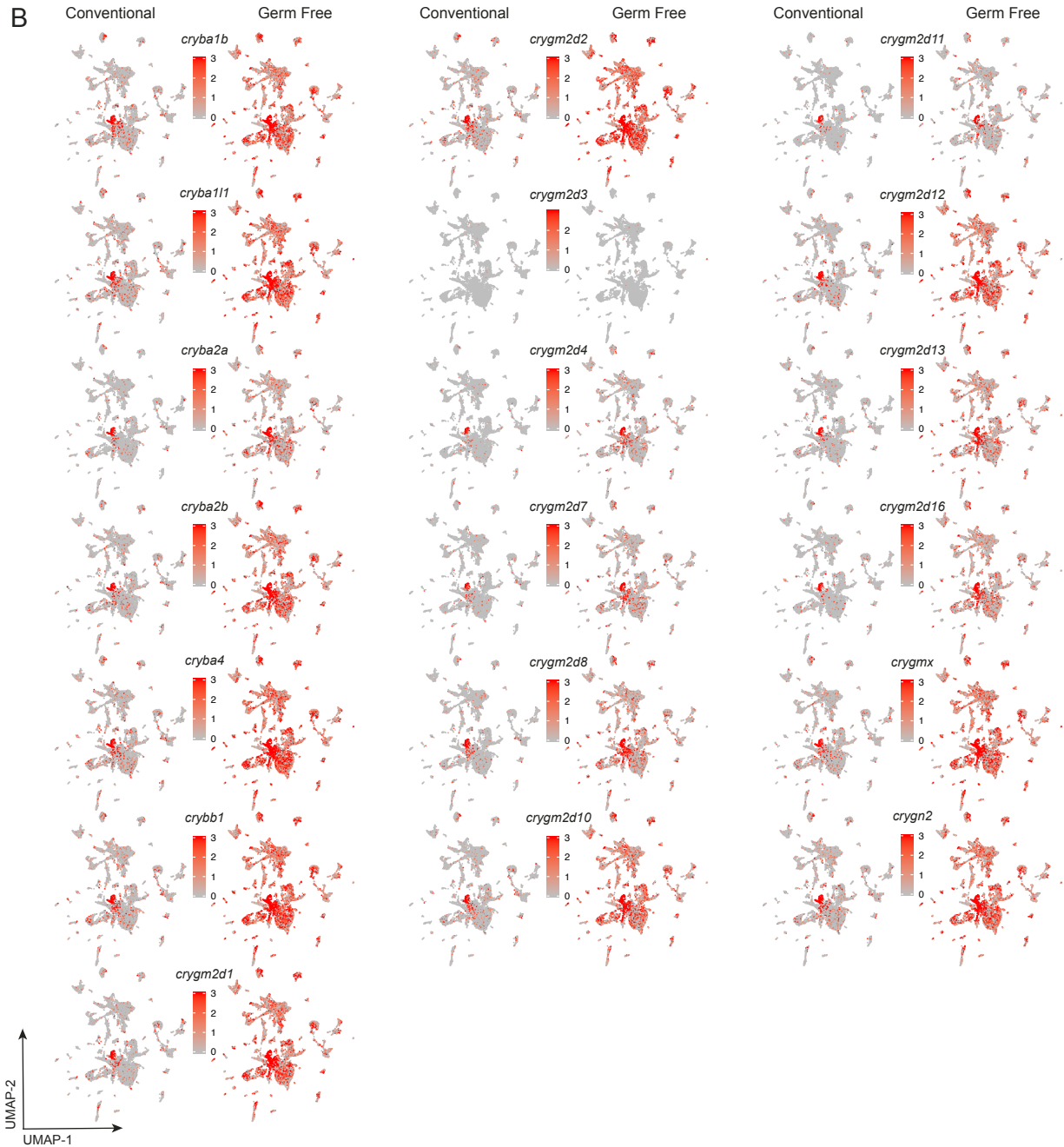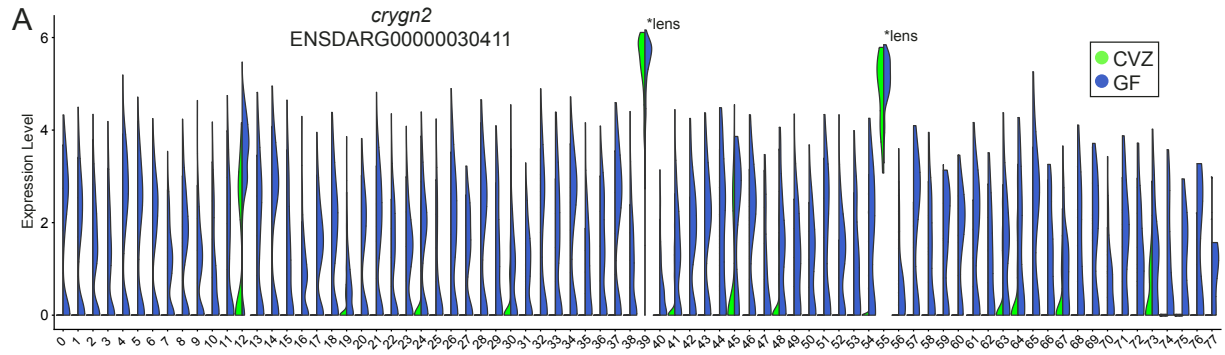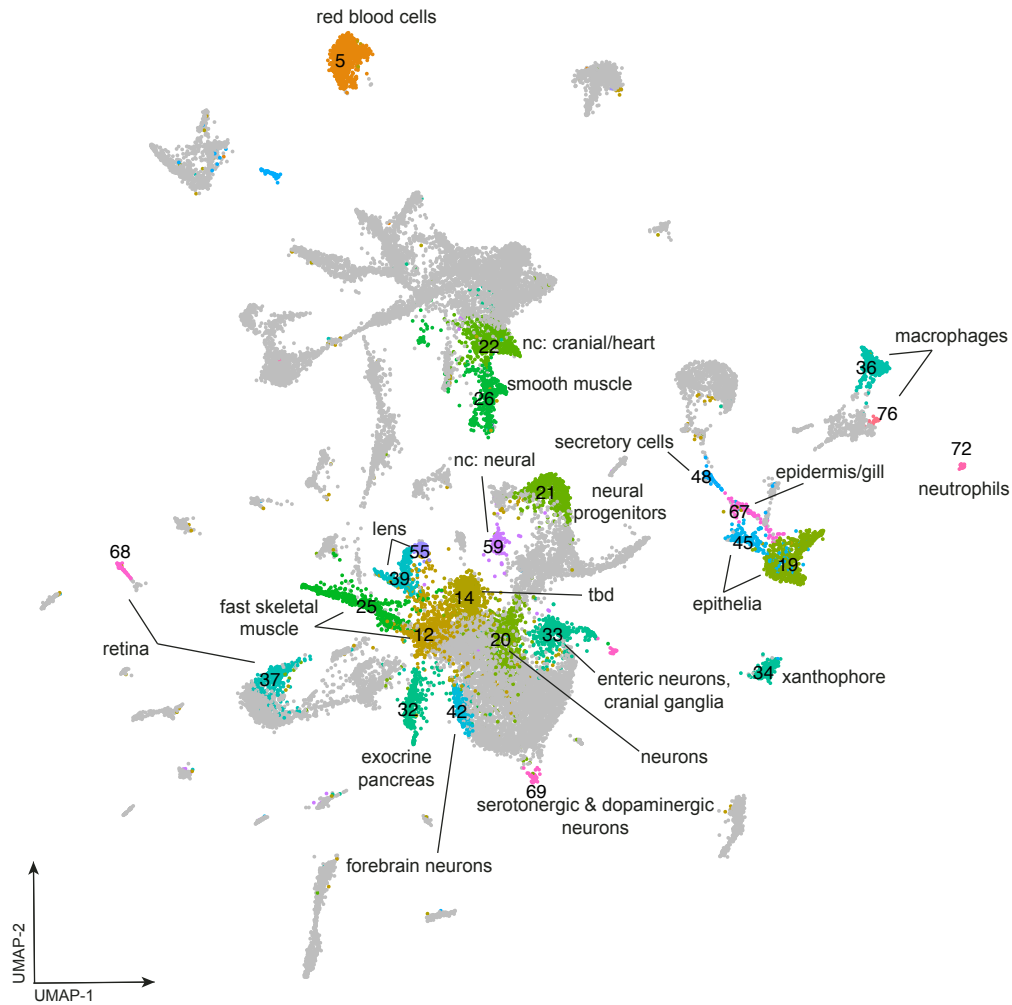
Figure S2.10 Diverse cell types respond to the microbiota by an enrichment of genes involved in ATP and nucleotide metabolism. The uMAP highlights the clusters composed of diverse cell types within the data that show an enrichment of gene expression within CVZ versus GF cells that are involved in ATP metabolism.
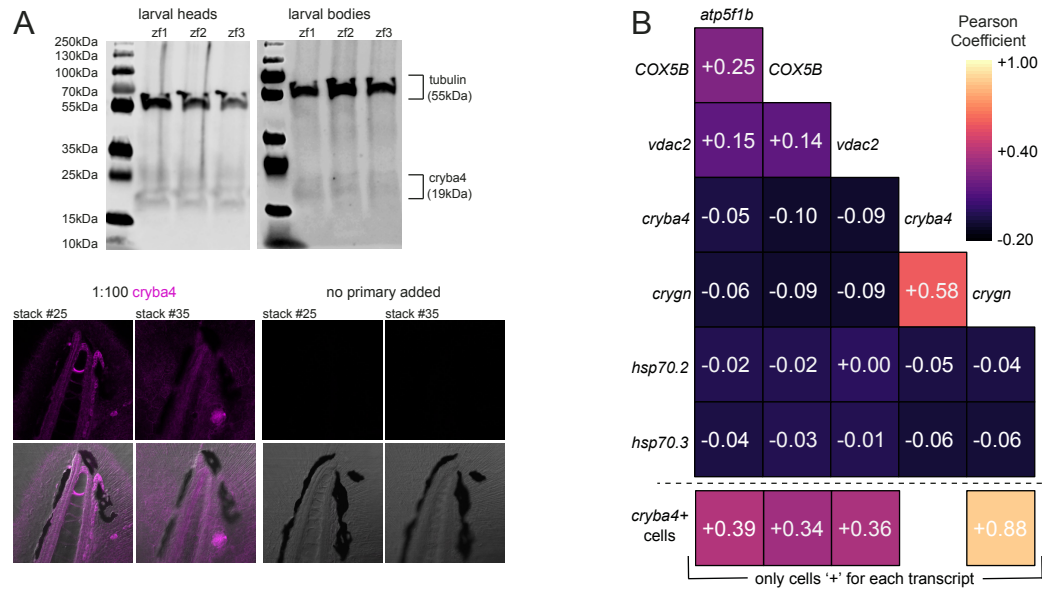
Figure S2.11 Crystallin expression occurs outside of the lens in GF larvae and lacks correlation to heat-shock and ATP metabolism genes. A) the western blot shows the protein expression of Cryba4 within the heads versus corresponding bodies of GF larvae. Images display the expression of Cryba4 in different tissue structures of the larval tail. B) The heatmap plot shows Pearson Correlational coefficients across all CVZ and GF cells in the experiment with respect to *cryba4* and *crygn* to heat-shock and ATP metabolism genes. The heatmap also displays Pearson Correlational coefficients in analyses that only include cells that are positive for both transcripts.

Figure S2.12 The co-expression of crystallin and select keratins are enriched within GF cells. A) The dotplot features the transcripts significantly enriched within GF cells versus CVZ accounting for all cells in the whole larvae. B) The uMAP plots show that specific keratins (*cyt* and *krt5*) are broadly expressed across GF cells versus epidermal specific keratin (krt17) that maintains its original cell type identity despite clustering GF cells by crystallin gene expression. C) The corresponding uMAP plots show co-expression of *cyt1* and *krt5* with *cryba4* across GF cells. D) The uMAP plots show CVZ and GF *elavl4+* cells separate based on their experimental group and that co-expression of *cyt1* and *krt5* with *cryba4* is enriched within GF *elavl4+* cells. E) The uMAP plots show that *cyt1* is co-expressed with *cryba4* across diverse cell types within whole larvae.

67
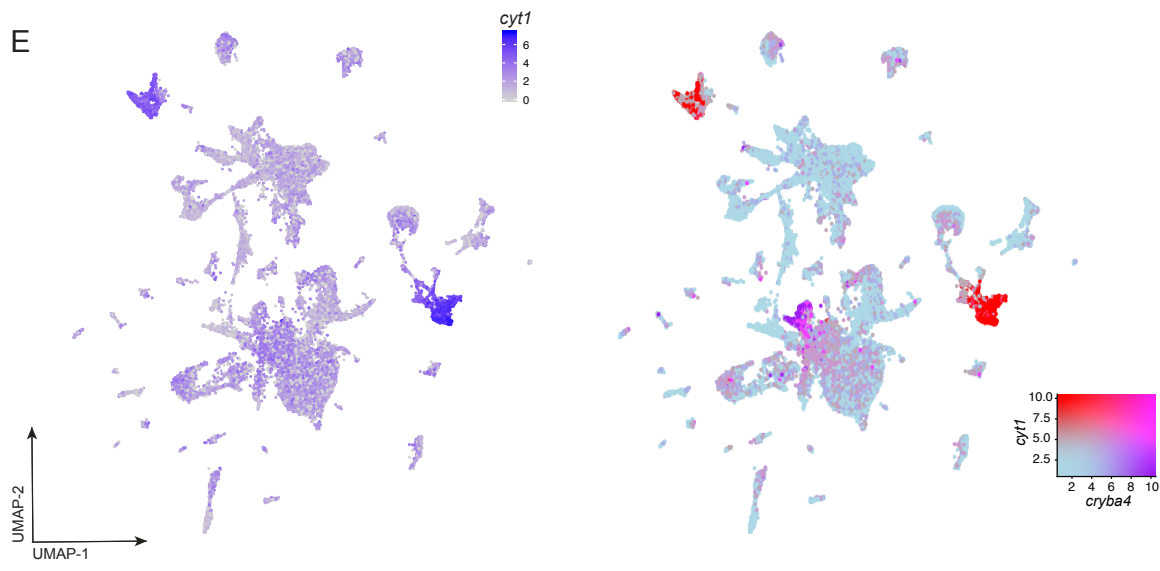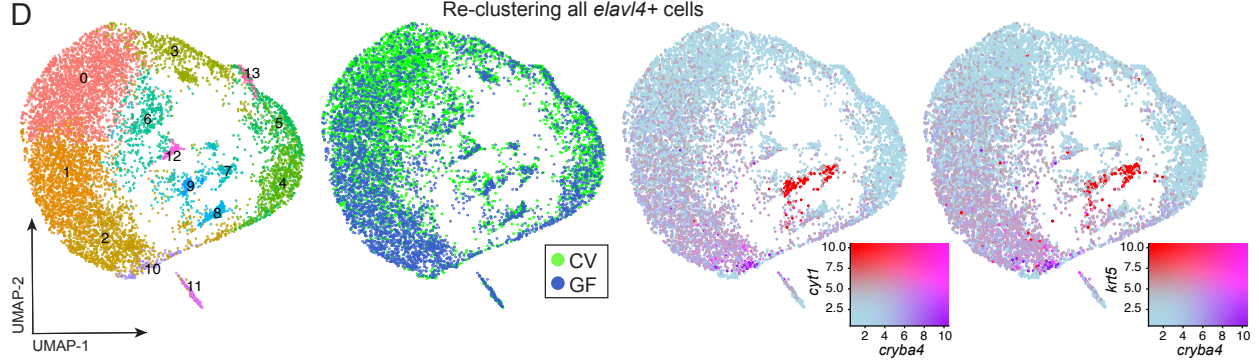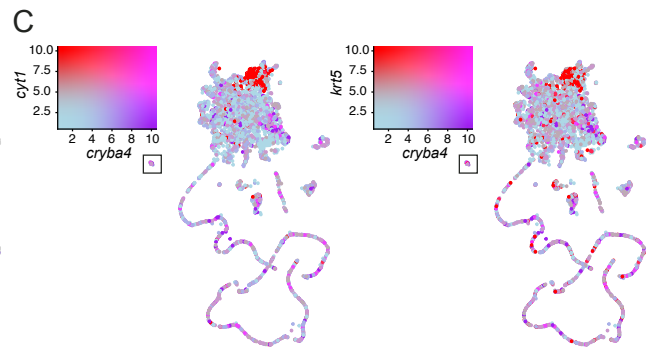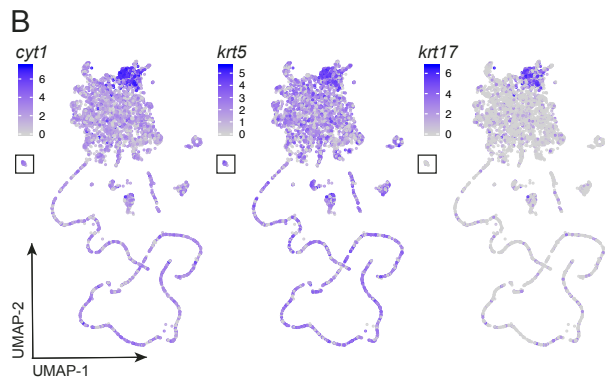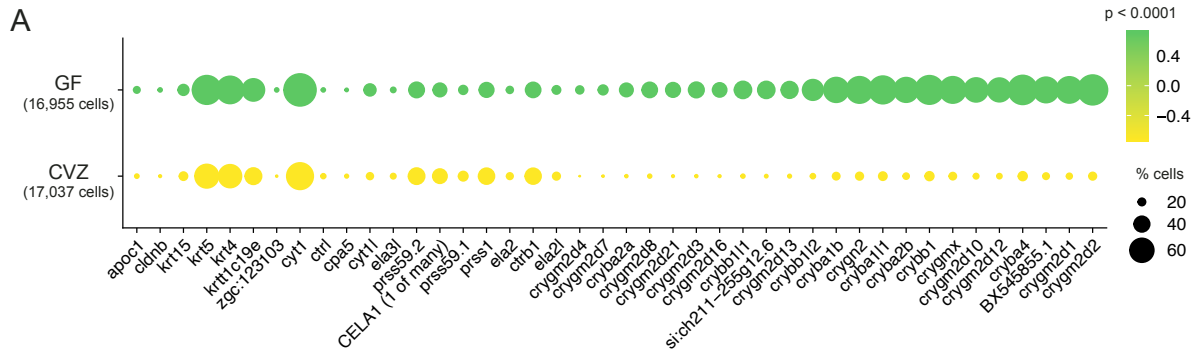
Bridge

Chapter II describes the competency of diverse host cell types throughout the body to the presence of the microbiota. Host cells differentially responded to the microbiota in tissue-type specific ways and responsive cells were not limited to the cell types at barrier sites that are in close proximity to the microbiota. As evidence is mounting that the microbiota plays critical roles in diverse aspects of host biology, the work of understanding the mechanisms that drive the communication between hosts and their resident microbes is only beginning. The next chapter will describe the structure and function of a bacterial-secreted protein that is sufficient to induce the proliferation and development of insulin-producing beta cells within larval zebrafish (Hill et al. 2016).

CHAPTER III

A MICROBIOTA MEMBRANE DISRUPTER DISSEMINATES TO THE PANCREAS

AND INCREASES β-CELL MASS

This chapter is a culmination of years of experiments and team-work from multiple scientists spanning across disciplines. Dr. Jennifer Hampton Hill (Guillemin Lab, University of Oregon, followed by Round & Murtaugh Labs, University of Utah) lead the project and directed the manuscript. After discovering the novel bacterial-secreted protein, BefA, Dr. Hill produced crystals of the purified protein and Dr. Emily Goers Sweeney (Guillemin Lab, University of Oregon) solved the BefA crystal structure in collaboration with Dr. Jim Remington (University of Oregon). I performed extensive investigations testing the capacity of BefA to induce zebrafish larvae beta cell expansion in several paradigms and Dr. Hill performed these congruent experiments within mice. Dr. Goers Sweeney performed extensive biochemical assays of BefA interactions with membranes *in vitro* and Elena Wall (Guillemin Lab, University of Oregon) undertook complementary microbiological experiments.

Introduction

β-cells are an essential pancreatic endocrine cell type necessary for producing insulin to modulate blood glucose levels. Patients with diabetes lack functional β-cells due to either autoimmune destruction or long-term insulin resistance. Strategies for replacement of lost β-cells are sought as a means of treatment, particularly for type 1 diabetes (T1D), to restore endogenous insulin production (Aguayo-Mazzucato et al., 2018). Extensive research has focused on understanding endogenous pathways of β-cell development and proliferation to harness these existing mechanisms for β-cell replacement therapies (Aguayo-Mazzucato et al., 2018, Murtaugh

et al., 2007). One example of an initially promising class of potential therapeutics was the **Reg**enerating islet-derived proteins, originally discovered in pancreatic tissue for their upregulation during the regenerative response of pancreatic β-cells (Terazono et al., 2988). These proteins however, were found to have pleiotropic effects, causing proliferation of several other tissues such as exocrine cells, intestinal epithelium, hepatocytes, and nerves [reviewed in (Chen et al., 2019)]. Subsequent to their characterization in the context of the pancreas, the Reg3 proteins were found to be potent pore-forming antimicrobial host defense proteins, essential for the maintenance of intestinal barrier function (Cash et al., 2006; Vaishnava et al., 2011; Mukherjee et al., 2014). More recently, disease-protective microbiomes in the NOD mouse model of T1D were found to induce Reg3 expression (Zhang eta l., 2021). However, no connections have been made between Reg3's pore forming activity and its impact on the pancreas.

We recently showed that the resident microbiota promotes β-cell development in zebrafish larvae and discovered a secreted bacterial protein that could rescue germ free (GF) larval β-cell development, which we named BefA, for Beta-cell Expansion Factor A (Hill et al., 2016). BefA is a 29 kDa protein produced by a limited number of intestinal bacteria from various hosts including zebrafish, mice and humans (Hill et al., 2016). A BefA homolog from a human-derived *Klebsiella*, with only 34% amino acid sequence similarity to our original *Aeromonas veronii*-derived BefA, was sufficient to rescue GF zebrafish β-cell development (Hill et al., 2016). This result suggested that the mechanism by which BefA stimulates expansion of insulin-producing cells is mediated through structural features of the protein shared across distant homologs. It also raised the possibility that these features of BefA homologs from diverse host-

associated bacteria may stimulate β-cell expansion across multiple host species, a question we explore in the current study.

Analysis of the amino acid sequence of the original BefA isolate and its functional homologs revealed little about the proteins' activities, but uncovered sequence similarity to a lipid-binding SYLF domain (named for the first four proteins found to contain it, **S**H3YL1, **Y**sc84p, **L**sb4p, and **F**YVE), which is prevalent across the kingdoms of life (Hasegawa et al., 2011). Four eukaryotic SYLF-containing proteins, found in yeast, plants, and mammals, have been studied to date, and all are associated with cellular or organellar membranes (Hasegawa et al., 2011; Robertson et al., 2009; Sutipatanasomboon et al., 2017; Tonikian et al., 2009; Urbanek et al., 2015; Dewar et al., 2002; Wywail et al., 2010). Here we describe the atomic structure and biochemical features of a bacterial SYLF-domain protein. We find that BefA has an intrinsic affinity for membranes, inducing vesiculation and leakiness of synthetic membranes as well as loss of integrity of bacterial cells. We further identify a point mutation within the SYLF domain of BefA that reduced its membrane permeabilization activity. We also show that BefA can disseminate from the intestinal lumen via multiple anatomic routes to act directly on pancreatic tissue. Although small molecular metabolites of the gut microbiota are known to exert systemic effects on the host, especially in shaping immunity (Levy et al., 2016), the dissemination and functional impact of microbiota-derived proteins like BefA has not been described. Finally, we show that BefA's β-cell expansion activity is conserved between zebrafish and mice and is dependent upon its membrane permeabilization activity, whereas the host-derived membrane permeabilizing Reg3 protein is sufficient to mimic BefA's impact on islets. By investigating the structural and biochemical properties of BefA, our work reveals membrane permeabilization as a

common mechanism of action by which both bacterial and host proteins stimulate pancreatic β-cell expansion.

Results

The structure of BefA reveals the SYLF domain that confers its function

To better understand how BefA functions, we determined its 3D atomic structure to 1.3 Å resolution, revealing a compact partial β-barrel carboxy-terminal domain with three flanking α-helices and five α-helices in the amino-terminal region (Fig3.1A&B). The location of Helix 1 (H1) is likely constrained due to crystal packing contacts, and has been grayed out to show that the location in solution is unknown (Fig3.1A&D). The high-resolution crystal structure of BefA confirms the backbone fold very recently reported for the solution NMR structure of a distantly related protein, BPSL1445 from *Burkholderia pseudomallei* (Quilici et al., 2021), which taken together are the first representatives of a novel protein fold.

To define the structural features corresponding to BefA's β-cell expanding activity, we performed a structural comparison between the original *Aeromonas veronii* BefA and a distant BefA homolog identified in *Klebsiella aerogenes* (originally annotated as *Enterobacter aerogenes* (Hill et al., 2016)). Although the two amino acid sequences have only 34% identity (FigS3.1A), the *K. aerogenes* homolog is sufficient to increase β-cell numbers in GF larval zebrafish. The structure of the *K. aerogenes* homolog was modeled using the Iterative Threading ASSEmbly Refinement (I-TASSER) server with BefA as a template (Krissinel et al., 2004; Krissinel et al., 2005; Roy et al., 2010; Zhang et al., 2018; Yang et al., 2015). Overlaying the structure of BefA and the model of the BefA homolog revealed that despite the low amino acid sequence conservation between these two proteins, they are predicted to share a high degree of structural similarity starting at residue 99 at the beginning of H4 (RMSD = 0.84 Å) (Fig3.1C).

The absence of the three N-terminal α-helices (H1-H3) from the shorter *K. aerogenes* homolog model suggests that these helices are dispensable for function. Alignment with the NMR structure for BPSL1445 (PBD ID: 7OFN), a protein of unknown function from *B. pseudomallei* that shares 22% sequence identity with BefA, showed a common backbone fold across the SYLF domain with an RMSD value of 1.8 Å (FigS3.1B). Additionally, we selected two well-studied eukaryotic SYLF-containing proteins from yeast (Ysc84) and humans (SH3YL1) and again used I-TASSER to predict the 3D structures of their SYLF domains. Both eukaryotic models are predicted to superimpose closely with our proposed BefA SYLF domain (RMSD values of 1.62 Å for Ysc84 and 1.92 Å for SH3YL1), beginning at residue 99 (FigS3.1C), suggesting a high degree of structural conservation in SYLF domains across kingdoms. The Conserved Domain Databank (Marchler-Bauer et al., 2017) defines the SYLF domain in BefA as starting at residue 114, midway through H4 (FigS3.1A, magenta arrowhead), but our structural analysis suggests that the conserved SYLF domain includes the entire helix, starting from residue 99 in BefA (Fig3.1C).

With these new insights into the structure of the SYLF domain and the availability of many new bacterial genome and metagenome sequences, we expanded on our homology searches for bacterial BefA homologs (Hill et al., 2016), focusing on the SYLF sequence. Genes encoding SYLF-containing proteins were found in both environmental and host-associated bacteria across a variety of aquatic and terrestrial environments (FigS3.1C). Our analysis identified many more examples of homologs like the *K. aerogenes* BefA and the *B. pseudomallei* BPSL1445, consisting almost entirely of the SYLF domain. Indeed, a survey of the average amino acid length of prokaryotic and eukaryotic SYLF-containing proteins shows the typical prokaryotic ones to be approximately the length of the 195 aa SYLF domain, whereas the

eukaryotic proteins are significantly longer (FigS3.1E), being comprised of one or more additional domains.

We predicted that the SYLF domain of BefA would be sufficient to mediate the pro-proliferative effect on larval zebrafish β-cells. To test this, we cloned two truncated BefA proteins. One truncation comprised amino acids 99 – 261, which encode the entire SYLF region by our definition (BefA[99]) (Fig3.1D). The second shorter truncation incorporated amino acids 133-261, which correspond to the C-terminus compact partial β-barrel with three flanking α-helices (BefA[133]) (Fig3.1D).
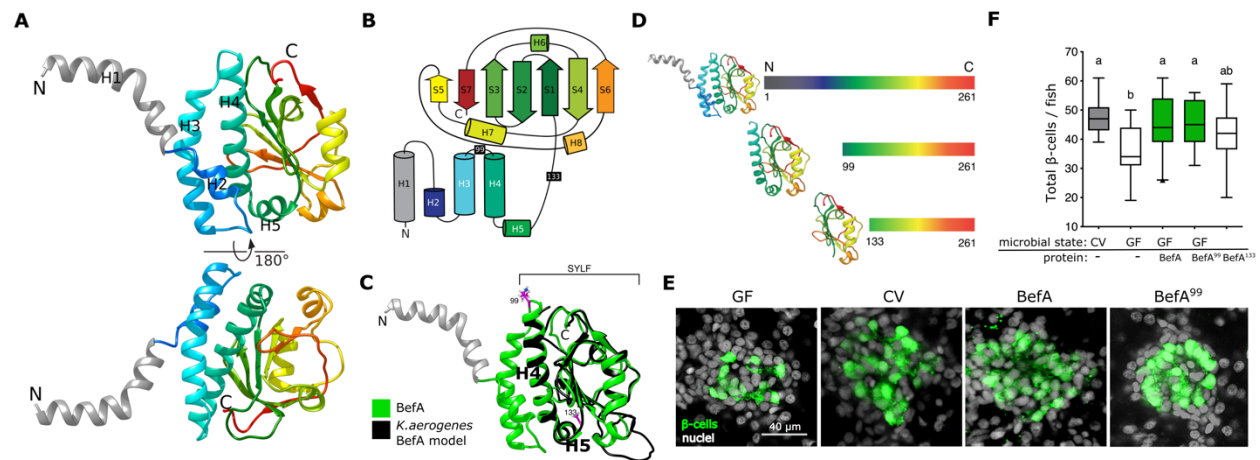
Figure 3.1 The structure of BefA reveals the SYLF domain that confers its function. (A) Ribbon diagram of the full length BefA protein in two orientations (rotated by 180º along the indicated axis) with the five amino terminal helices labelled H1-H5. H1 is colored gray to demonstrate that the position is constrained due to crystal packing. (B) 2D domain architecture map of BefA, to scale. Truncation locations from panel D are marked with corresponding amino-acid numbers (99 and 133) inside black rectangles. (C) The structure of BefA (green) overlaid with the predicted I-TASSER model of the structure of the K. aerogenes BefA homolog (black). Note that the amino terminus of the K. aerogenes model starts at H4 of the BefA structure. Residues 99 and 133 are shown as pink sticks in order to highlight that they denote the beginning of the BefA[99] and BefA[133] truncations (in panel D) respectively. (D) Scale schematics of the BefA truncation proteins, BefA[99] and BefA[133], accompanied by ribbon diagrams of their predicted structure. (E) Representative 2D slices from confocal scans through the primary islets of 6 dpf ins:gfp larvae which are either CV, GF, or GF treated with purified BefA or BefA[99]. Each slice is taken from the approximate center of the islet. Insulin promoter expressing β-cells are in green and nuclei are white. Scale bar = 40 μm. (F) Boxplots illustrating total β-cell quantifications from 6 dpf larvae in each treatment group, n≥15 larvae for each group from at least 2 experimental replicates. In this, and in all subsequent figures containing boxplots, CV data are colored grey, GF or control treatment groups are colored white, and BefA treated, or statistically similar groups, are colored green. In all relevant panels and remaining figures, box plot whiskers represent the 95% confidence interval of the data set from pooled replicate experiments. In this and in all subsequent figures, lowercase letters above groups indicate the results of post hoc means testing (Tukey) to denote statistically significant groupings.

We tested each purified truncated BefA protein for the ability to expand β-cell mass when incubated with GF *Tg(-1.0insulin:eGFP)* (*ins:gfp*) larvae from 4 to 6 days post fertilization (dpf).

74

Quantifications of the total β-cell numbers in these fish confirm that BefA[99] was equally effective at rescuing GF β-cell expansion as full length BefA (Fig3.1F), whereas the BefA[133] truncation only conferred a partial rescue of GF β-cell numbers (Fig3.1F). Collectively, these results illustrate that the functional region of the BefA protein for inducing β-cell expansion is contained within the C-terminal SYLF domain.

BefA induces membrane permeabilization

While SYLF domains are found across diverse forms of life, their mechanisms of action have only been investigated in a few proteins. Studies in yeast, plants, and mammals suggest that the SYLF domain commonly drives interactions at lipid membranes, helping to orchestrate complex biological processes such as motility and endocytosis (Urbanek et al., 2015). Previously published evidence suggests that SYLF domain containing proteins mediate these important events by binding to membrane lipids (Hasegawa et al., 2011; Sutipatanasomboon et al., 2017; Urbanek et al., 2015; Quilici et al., 2021) and/or actin filaments (Robertson et al., 2017; Urbanek et al., 2015). Therefore, we looked for biochemical interactions between BefA and lipids or F-actin.

The yeast protein Ysc84p contains actin binding function in its N-terminal SYLF domain (Hasegawa et al., 2011; Robertson et al., 2017) and some of these actin binding residues are conserved in BefA. Therefore, we performed an F-actin co-sedimentation assay. Unlike our positive control, cortactin, we could not detect any significant enrichment of either BefA or BefA[99] in the presence of F-actin (FigS3.2A), indicating that BefA does not share Ysc84p's ability to bind F-actin.

Since several other SYLF domain containing proteins can bind to phosphatidyl inositols (PIPs) (Hasegawa et al., 2011; Sutipatanasomboon et al., 2017; Urbanek et al., 2015; Quilici et

al., 2021), we next investigated the potential lipid and membrane binding activities of BefA. We used commercial lipid strip assays, which indicated no binding of BefA to PIPs, but consistent binding to phosphatidyl serine (PS) and cardiolipin, both negatively charged lipids (FigS3.2B). The addition of calcium, which commonly mediates binding to negatively charged lipids, did not result in additional binding activity or increased affinity (FigS3.2B). Several lipid binding residues previously identified in eukaryotic SYLF domains are conserved in BefA. To test whether any of these residues are necessary for BefA to bind to lipids, we created point mutations in BefA at several of these sites. We noted that several of these point mutants appeared to reduce BefA's affinity for PS (FigS3.2B), although conclusions from this assay are limited by the fact that these strips do not display lipids in the biologically relevant organization of membrane bilayers.

To test BefA's interactions with lipid bilayers, we incubated either BefA or BefA[99] with synthetic Texas Red-labelled giant lipid vesicles (Veatch et al., 2007) (Fig3.2A). In the presence of BefA or BefA[99] we observed a high frequency of large vesicles joined to smaller vesicles, a morphology we refer to as multi-vesicular vesicles (Fig3.2A-B). We observed similar vesiculation with both neutral and negatively charged vesicles containing PS. We also noted that BefA treatment resulted in an increase in the total concentration of single and multi-vesiculated vesicles (Fig3.2C; FigS3.2C), suggesting that multi-vesiculated vesicles are formed by the budding and fragmentation of larger vesicles, as schematized in Fig3.2D.

Induction of membrane curvature, which occurs in the formation of multi-vesicular vesicles, is a common characteristic of membrane permeabilizing antimicrobial proteins (Guha et al., 2019). To determine whether BefA could impair membrane integrity, we performed a dye release assay on vesicles encapsulating fluorescent dye. While little to no dye was released by

smaller vesicles ($\leq$100 nm in diameter) upon treatment with either BefA, BefA[99], or a water control (Fig3.2E), a significant amount of dye was released in a concentration dependent manner when BefA or BefA[99] was added to larger vesicles ($\leq$400 nm in diameter) (Fig3.2F), suggesting BefA may have greater affinity for lower membrane curvature. We next tested whether a BefA mutant with reduced PS binding in the lipid strip assay would have reduced vesicle permeabilizing activity. We found that BefA[R195A] released significantly less dye from both neutral and negatively charged PS-containing vesicles compared to wild type BefA (Fig3.2G&H). Arginine 195 is located on the same protein face immediately adjacent to a stretch of residues shown in BPSL1445 to interact in solution with phospholipids (Quilici et al., 2021).

Many antimicrobial proteins target bacterial cell membranes by inducing membrane curvature (Mukherjee et al., 2015). To assess whether BefA could compromise bacterial membrane integrity, we used the CyQUANT Direct Red cell permeability assay, whereby cell leakiness is measured as quenching of a membrane permeable fluorescent DNA dye by a cell-impermeant suppressor. We tested two novel Gram-positive isolates from zebrafish belonging to the genera *Bacillus* and *Staphylococcus*. These cells were readily permeabilized by the detergent SDS, as well as by the well-characterized pore-forming antimicrobial protein, Reg3α (Fig3.2I&J), which is secreted by host enterocytes to maintain barrier function against the gut microbiota (Vaishnava et al., 2014). BefA also induced bacterial membrane permeabilization in *Bacillus* and *Staphylococcus* with levels of permeabilization comparable to that of Reg3α in the case of *Staphylococcus* (Fig3.2I&J). We also tested whether BefA bound directly to bacterial cells, using an mCherry (mCh)-BefA fusion protein. We observed clear binding of mCh-BefA to the outside of the *Bacillus sp*. and very little to no binding of the unconjugated mCh protein (Fig3.2K&L). BefA binding appeared to be enriched at junctions between cells, where nascent

cell wall synthesis during septation may allow greater access to the cell membrane (Pasquina-Lemonche et al., 2011).

We speculated that BefA's capacity to permeabilize membranes was the basis for its β-cell expanding activity. If this were the case, then other unrelated membrane-permeabilizing proteins should also be able to increase β-cell mass. To test this, we exposed GF larvae to purified Reg3γ, the murine homolog of human Reg3α. The addition of Reg3γ was sufficient to rescue β-cell numbers to that of CV or BefA treated levels (Fig3.2M). Overall morphology of larval islets treated with Reg3γ was similar to that of CV, suggesting that addition of the AMP did not cause significant injury or damage to the developing cells (Fig3.2N).

Host AMPs are often produced downstream of Toll-like receptor (TLR) activation signaling cascades. Both pancreatic β-cells and intestinal cells express TLRs, which recognize microbial products (Garay-Malpartida et al., 2011; Vives-Pi et al., 2003), making TLR signaling a candidate mechanism to explain both BefA and Reg3γ activity on β-cells. To test whether TLR signaling was required for BefA sensing, we quantified β-cell numbers in zebrafish larvae carrying a mutation in the *myd88* gene (Burns et al., 2017). Myd88 is a universal adaptor protein downstream of most TLR and IL1R activation. We did not observe any difference in the number of β-cells between wild-type (WT) and *myd88*[-/-] larvae that were raised conventionally (Fig3.2O). Furthermore, treatment of GF *myd88*[-/-] larvae with BefA resulted in an increase of β-cells to numbers similar to CV fish (Fig3.2O). Together these results suggest that TLR signaling is not required for BefA sensing or possibly that BefA acts through the repression of TLR signaling, as has been shown to induce β-cell replication in response to obesity (Ji et al., 2019).
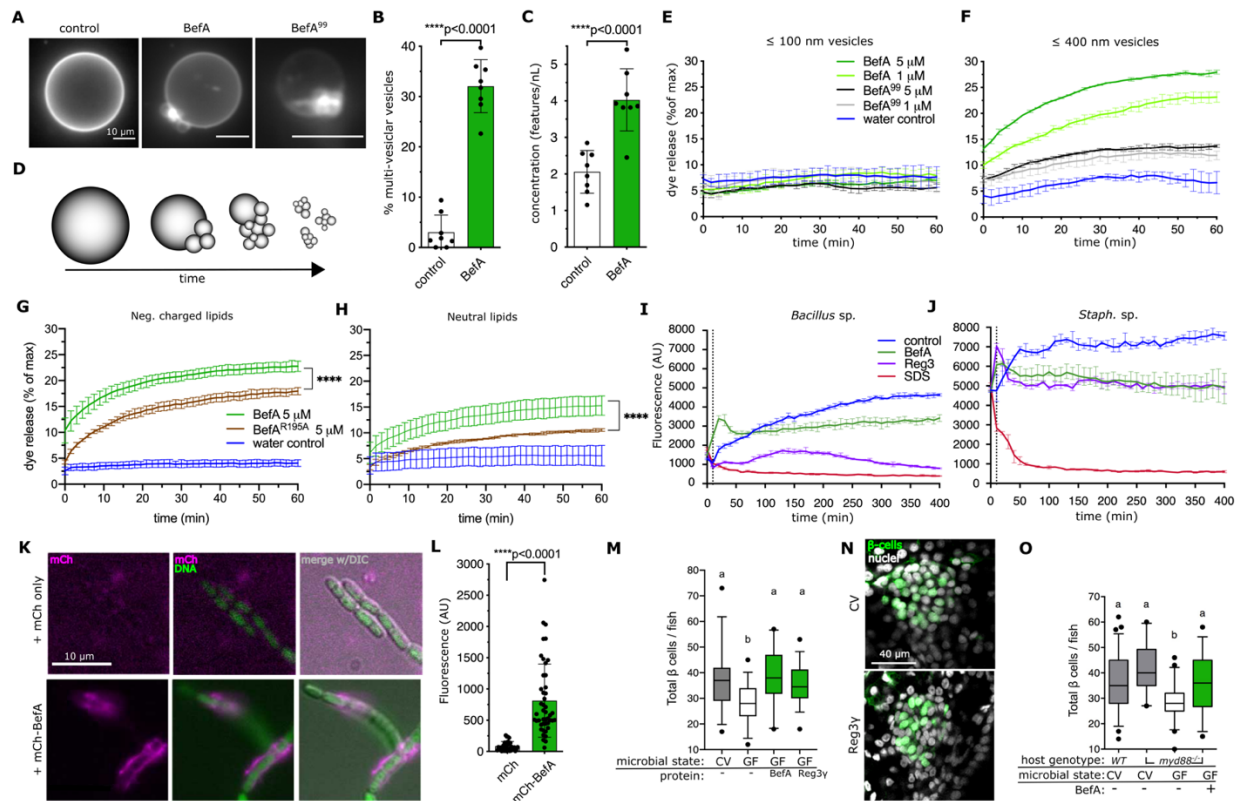
Figure 3.2 BefA induces membrane permeabilization. (A) Neutral (DOPC), fluorescently labeled vesicles in a range of sizes were treated with water (representative vesicle on left), 2 μM BefA (representative vesicle center), or 9 μM BefA[99] (representative vesicle right) and imaged using light sheet fluorescence microscopy. All scale bars = 10 μm. (B) Model showing predicted vesiculation and subsequent fragmentation over time upon BefA treatment. (C) Each dot represents an image stack at either 35 or 50 min after treatment with 2 μM BefA, 32-150 vesicles per stack. (n = 8 stacks and total vesicles quantified $\geq$ 457 per treatment) Error bars represent the standard deviation from the mean. (D) The concentration of treated or untreated vesicles was quantified by summing the number of single plus multi-vesiculated vesicles (features) per volume for each image stack. (n = 8 stacks). Error bars represent the standard deviation from the mean. (E&F) Neutral, extruded vesicles were loaded with quenched dye and treated with water (control, blue), BefA at 1 and 5 μM (green traces), or BefA[99] at 1 and 5 μM (gray traces). Leakiness of vesicles over time was monitored by release of dye as a percentage of maximal dye release upon addition of detergent. Vesicle sizes for (E) were less than or equal to 100 nm in diameter or less than or equal to 400 nm for (F). Lines follow the mean of duplicate experiments with error bars denoting the standard deviation of the mean between experiments for each time point. (F) Two-way ANOVA between treatment groups p<0.0001. (G&H) Negatively charged vesicles containing PS (G) or neutral vesicles (H) were loaded with quenched dye and treated with water (control, blue), BefA at 5 μM (green traces), or BefA[R195A] at 5 μM (brown traces). Leakiness of vesicles over time was monitored by release of dye as a percentage of maximal dye release upon addition of detergent. Vesicles were less than or equal to 400 nm in size. Lines follow the mean of duplicate or triplicate experiments with error bars denoting the standard deviation at each time point. ****p value<0.0001 based on Two-Way ANOVA multiple comparisons test. (I&J) Bacterial cell permeabilization was monitored for Bacillus sp. and Staphylococcus sp. using CyQUANT Direct Red plate assay for negative control (black), 7 μM BefA (green) and 3 μM Reg3α (purple) treated cells. 0.08% SDS detergent (red) was used as a positive control. In contrast to panels E&F, permeabilization is detected as dye released from cells becomes quenched, resulting in decreasing fluorescent signal. Lines follow the mean of 3 replicate experiments with error bars denoting the min and max spread of the data between experiments for each time point. (K) Representative images of Bacillus sp. incubated with mCh only (top) or mCh-BefA (bottom). Magenta = mCh protein, green = DNA stained with SYBRGreen, DIC = differential interference contrast microscopy. Scale bar = 10 μm. (L) Quantification of BefA binding to Bacillus sp. after incubation with mCh-BefA or control mCh alone. The mean fluorescence intensity of the mCh channel along the length of each cell was measured and graphed. (n$\geq$33 individual bacterial cells per treatment) Error bars represent the standard deviation from the mean. (M) Boxplots illustrating total β-cell quantifications from 6 dpf larvae in each treatment group, n$\geq$25 larvae for each group. (N) Representative 2D slices from confocal scans through the primary islets of 6 dpf ins:gfp larvae which are either CV, or Reg3γ treated. Each slice is taken from the approximate center of the islet. Insulin promoter expressing β-cells are in green and nuclei are white. Scale bar = 40 μm. (O) Boxplots illustrating total β-cell quantifications from 6 dpf larvae in each treatment group, n$\geq$19 larvae for each group.

79

BefA interacts directly with pancreatic β-cells

Our finding that BefA and another membrane-permeabilizing protein could expand pancreatic β-cells suggested that BefA's effects would be mediated through direct interaction with target host cells. To visualize interactions between BefA's SYLF domain and host cells, we utilized fluorescently tagged versions of BefA[99] (mCh- or mNeonGreen (mNG)-BefA[99]), which we showed induced a similar increase in β-cell numbers in GF larvae as the untagged full length BefA; unconjugated fluorescent proteins did not increase β-cell number (Fig3.3A). First, we added mNG-BefA[99] to dissociated zebrafish cells cultured from dissected intestines of larval zebrafish. These preparations contained insulin positive pancreatic β-cells and a variable mixture of cells from other digestive tract tissues. While we detected little to no mNG signal in cultures incubated with mNG alone, we saw distinct punctate labeling of mNG-BefA[99] co-localized with β-cells (Fig3.3B, yellow arrowheads) as well as other non-insulin expressing cells (Fig3.3B, white arrowheads). We found that an average of 21 (std. dev. = 14) percent of β-cells associated with mNG-BefA[99], whereas only 1 (std. dev. = 3) percent of β-cells associated with mNG alone, demonstrating that the SYLF domain of BefA is sufficient to facilitate a direct interaction with β-cells. To determine the temporal dynamics of this interaction, we performed a longitudinal imaging experiment. Dissociated larval islet cells were treated with mNG-BefA[99] and imaged using confocal microscopy for a period of 30 minutes. A subset of cells in the islet displayed an initial uniform surface fluorescence with mNG-BefA99, followed by the appearance of surface puncta as early as 10 minutes post-exposure (FigS3.3A). After 20 minutes, the puncta appear to become internalized into the cells (FigS3.3A), consistent with a process of protein endocytosis.

We next tested if direct interaction with BefA is sufficient to exert a proliferative effect on larval β-cells using primary zebrafish islet explant cultures. We micro-dissected the primary

islet from GF 4 dpf *ins:gfp* larvae (FigS3.3B) and maintained them *ex vivo* in sterile cell culture media for 48 hours untreated or in the presence of full length BefA or BefA[99]. *Ex vivo* larval β-cells appeared healthy after 48 hours as indicated by robust insulin promoter driven *gfp* transgene expression, absence of TO-PRO-3 Iodide (TOPRO) incorporation marking dead or dying cells, and an otherwise normal appearance (FigS3.3C). After the treatment period, we observed significantly more β-cells in explants treated with either BefA or BefA[99] than in those that received only rich cell culture medium (Fig3.3C&D), showing that BefA can elicit β-cell expansion via a direct interaction, and that the SYLF domain is sufficient to mediate this effect.

After finding BefA's direct effect on pancreatic cells, we sought to determine whether this protein, secreted by gut bacteria, could disseminate to the pancreas. Plausible routes for dissemination from the gut lumen would be via the vasculature or through reflux up the hepatopancreatic duct (HPD), which connects the liver, the gallbladder, and the pancreas to the gut for the passage of bile and digestive enzymes (Field et al., 2003). Additionally, because BefA could elicit β-cell expansion when added to the water column, we considered the possibility that it could be absorbed through the skin and gills into the bloodstream. To confirm that gut lumenal BefA could affect pancreatic β-cells, we used oral micro gavage to deliver different doses of BefA directly into the intestinal lumen of 4 dpf larval zebrafish, before quantifying cell numbers at 6 dpf. We found that administration of BefA directly into the gut lumen was sufficient to increase β-cell numbers in GF fish compared to the gavage vehicle control of PBS, even at the lowest dose tested of 0.2 pg (Fig3.3E).

We next sought to visualize systemic BefA dissemination following oral micro gavage of *Tg*(*Nkx2.2*:*gpf*) GF larvae, which express GFP in intestinal and pancreatic endocrine tissues, allowing us to visualize both intestinal and pancreatic anatomical landmarks. Using either mCh-

BefA or mNG-BefA and corresponding mCh or mNG control proteins, fluorescence was readily visualized within the intestinal lumen (Fig3.3F) and in lysosome rich enterocytes (LREs) (FigS3.3D) that have recently been described as a site of active, non-selective protein internalization (Park et al., 2019). Given the strong lumenal and LRE fluorescent signals and our finding that even very low doses of BefA will induce β-cell expansion, we concluded that visualizing functionally relevant, BefA-specific host cell interactions was beyond our *in vivo* imaging capabilities. We next explored BefA's routes of dissemination using zebrafish mutants.

To test the possibility of BefA reflux through the ductal network connecting the intestine and pancreas, we used *sox9b*[-/-] mutant zebrafish in which HPD development is compromised (Delous et al., 2012; Manfroid et al., 2012) such that bile secretion into the intestine is blocked in adulthood (Delous et al., 2012). We tested the patency of the HPD in *sox9b*[-/-] 4 dpf larval zebrafish using the short chain fluorescent lipid analog, BODIPY-FL C_2, which fills lumenal spaces of the GI tract because it is not appreciably metabolized (Carten et al., 2011). After feeding BODIPY-FL C_2 to WT larvae, it could be detected in intrapancreatic lumenal structures likely to be the intra-pancreatic ducts (IPD), which are contiguous with the intestinal lumen (Fig3.3G), whereas in *sox9b*[-/-] larvae, little to no fluorescent signal from BODIPY-FL C_2 could be found in the pancreas despite its obvious ingestion and presence in the intestinal lumen (Fig3.3G). We next tested whether this loss of patency through the HPD in *sox9b*[-/-] mutant larvae would affect the ability of either the microbiota or BefA to influence the growth of the larval β-cell population. CV *sox9b*[-/-] larval β-cell numbers were significantly reduced compared to CV *WT* animals, and were equivalent to GF *sox9b*[-/-] larvae (Fig3.3H), consistent with *sox9b*[-/-] larvae being unable to respond to BefA from their endogenous gut microbiota. However, GF *sox9b*[-/-] larvae that were treated with BefA in their water had significantly increased β-cell numbers

(Fig3.3H), indicating that *sox9b*[-/-] larvae were competent to respond to BefA and suggesting that BefA added to the water column may act through a different route than BefA produced by gut bacteria. To test this hypothesis, we delivered a high dose of BefA via oral microgavage to 4 dpf GF *sox9b*[-/-] larvae and quantified β-cell numbers at 6 dpf. When delivered directly to the intestinal lumen, BefA was not capable of increasing β-cell numbers in *sox9b*[-/-] larvae (Fig3.3H). Together, these results suggest that patency of the HPD is required for delivery of BefA from the gut lumen, but that BefA is also able to reach its target cell population through an alternative route.

To test whether the vasculature was an alternative route through which BefA could disseminate, we injected BefA[99] into the region of the developing circulation valley of GF larval zebrafish. Larvae injected with BefA[99] had significantly greater β-cell numbers compared to an injection control (Fig3.3I). We next injected BefA into *sox9b*[-/-] larvae and found a significant increase in β-cells (Fig3.3I). Together, these results suggest that BefA added to the water column is absorbed and disseminated through the blood stream to elicit β-cell growth in *sox9b*[-/-] larvae.

To further investigate the role of the vasculature in BefA dissemination, we utilized *cloche* mutant zebrafish which have grossly abnormal hematopoietic development and lack blood vessels, resulting in severe edema and early lethality (Reischauer et al., 2016; Stainier et al., 1995) (Fig3.3J). Because of these defects and low yields of offspring, *cloche* mutants were not amenable to our GF derivation protocols and we were limited to experiments in CV larvae. CV *cloche* mutant larvae had significantly reduced numbers of β-cells compared to their CV *WT* clutch-mates (Fig3.3K), and addition of a concentrated dose of BefA to the water column was not sufficient to fully rescue this effect. However, we were able to increase β-cell numbers more if we delivered a high dose of BefA to the *cloche* mutants via oral microgavage (Fig3.3K),

83

suggesting that these mutants are still competent to respond to BefA via the HPD, but fail to do so when BefA is produced in lesser amounts by endogenous gut microbiota or added in the water column. Collectively, these data support the idea that secreted BefA can reach distant β-cells in the pancreas both by traveling from the gut lumen through the HPD or by being absorbed into the vasculature. Furthermore, these results suggest that access to the pancreas is required for BefA to exert its function on β-cells, supporting the hypothesis that it acts in a direct manner *in vivo*.

The intestinal lumen, the HPD, and the blood stream represent drastically different chemical microenvironments that could influence protein folding. When purifying BefA protein, we noted that it was soluble in pure water without ions or buffer at concentrations exceeding 20 mg/mL, indicating high structural stability. Using a thermal stability assay, we determined BefA's melting temperature to be approximately 68° C, comparable to the highly stable protein lysozyme, which melts at 65° C (Ku et al., 2009) (Fig3.3L), suggesting that BefA's structure would be maintained while disseminating through potentially harsh host environments.

Figure 3.3 BefA interacts directly with host β-cells. (A) Boxplots illustrating total β-cell quantifications from 6 dpf larvae in each treatment group, n≥10 larvae for each group. (B) Confocal images of dissociated larval gastrointestinal cell cultures, treated with either mNG or mNG-BefA[99]. Panels from left to right: Insulin in magenta, mNG in green, nuclei in blue, and DIC overlay with fluorescent channels merged. White arrowheads indicate some of the mNG puncta associated with non-insulin expressing cells. Yellow arrowheads indicate some of the mNG puncta associated with β-cells. Scale bar = 50 μm. (C) Representative 2D slices from confocal scans through islet explants after treatment. Insulin promoter expressing β-cells are in green. Scale bar = 40 μm. (D) Boxplots illustrating total β-cell quantifications from primary islet explants in each treatment group, n≥20 explants for each group. (E) Boxplots illustrating total β-cell quantifications from 6 dpf larvae treated by oral gavage with varying concentrations of BefA (in picograms (pg)), n≥18 larvae for each group. (F&G) Cartoon larvae aged 4 dpf with gray dotted outline denoting the trunk region imaged. For orientation, the swim bladder (sb) is outlined in large white dashed lines, the intestinal lumen is outlined in small white dashed lines, and the pancreas (pan) is outlined in a solid white line, ib = intestinal bulb, pi = primary islet. (F) Top panel = merge of bottom two panels with DIC overlay. Lower left panel = Nkx2.2 promoter expressing endocrine tissue in green. Lower right panel = mCh-BefA in magenta. * = reflection of mCh signal off swim bladder. (G) Left panels represent wild-type (top) and sox9b[-/-] (bottom) larvae fed BODIPY FL-C2 (green), insulin-expressing cells = magenta, DIC overlay = grey. Scale bar = 200 μm. White box around the primary islet indicates the region of the zoom inset in right panels. Scale bar = 20 μm. (H&I) Boxplots illustrating total β-cell quantifications from 6 dpf larvae in each treatment group, (H) n≥17 or (I) n≥6 larvae for each group. (J) Representative confocal scans through primary islets of WT and cloche mutant larvae. Insulin promoter expressing β-cells are in magenta, vasculature (anti-kdrl) staining in green, nuclei in blue, scale bar = 100 μm. (K) Boxplots illustrating total β-cell quantifications from 6 dpf larvae in each treatment group, n≥15. (L) The thermal stability of BefA (green trace) and lysozyme (gray trace), measure by the Thermofluor assay. *T_m value of BefA as indicated by the inflection point of the curve = 68° C.

A — Total β-cells / fish — microbial state: CV, GF, GF, GF, GF, GF, GF, GF — protein: -, -, BefA, mCh, mCh-BefA, mCh-BefA99, mNG, mNG-BefA99

B — mNG, β-cells, nuclei, merge — + mNG only (50 μm), + mNG-BefA99

C — β-cells — untreated, BefA99 (40 μm)

D — Total β-cells / explant — untreated, BefA, BefA99

E — Total β-cells / fish — microbial state: CV, GF, GF, GF, GF, GF — oral gavage: -, -, PBS, 0.2, 2.0, 20 — pg of BefA

F — merge (sb, pi, pan, ib, 200 μm) — endocrine (pi), mCh-BefA (ib, *)

G — lumens — WT (sb, ib, 200 μm / 20 μm), sox9b−/−

H — Total β-cells / fish — host genotype: WT, sox9b−/− — microbial state: CV, CV, GF, GF, GF, GF — treatment: -, -, BefA, PBS, BefA (oral gavage)

I — Total β-cells / fish — host genotype: WT, sox9b−/− — microbial state: GF, GF, GF, GF, GF, GF — injection: -, PBS, BefA99, -, PBS, BefA99

J — vasculature, β-cells, nuclei — WT, cloche (100 μm)

K — Total β-cells / fish — p=0.0658 — host genotype: WT, cloche — microbial state: CV, CV, CV, CV, CV — treatment: -, -, BefA, PBS, BefA (oral gavage)

L — Fluorescence (AU) vs Temperature (°C) — BefA, lysozyme, *

BefA's direct activity on host β-cells is conserved in mice

Our analysis of SYLF domain encoding genes uncovered multiple homologs in

Enterobacteriaceae members of mammalian microbiotas (FigS3.1C), suggesting the possibility

that the mammalian pancreas is exposed to BefA homologs during neonatal β-cell development. Therefore, we sought to determine whether the effects of the microbiota and BefA on β-cell development were conserved in mammals.

Like zebrafish, mice are born with a population of fully differentiated, insulin-secreting β-cells (Larsen et al., 2017). During neonatal life, which is comparable to the post-hatching age of zebrafish, mammalian β-cells undergo high levels of proliferation to establish a healthy amount of insulin-producing tissue capable of maintaining glucose homeostasis throughout life (Miller et al., 2009). To test for the ability of BefA to increase murine β-cell mass, we engineered the probiotic *E. coli* Nissle 1917 (Nissle) strain to express *befA* from a chromosomally integrated transgene. Importantly, wild-type Nissle (Nissle^{WT}) has no endogenous copy of *befA* in its genome, it robustly colonizes both zebrafish and mice, and it is not pathogenic. Furthermore, since we found homologs of BefA within the genomes of other *E. coli* strains (FigS3.1C) (Hill et al., 2016) we predicted that Nissle would be able to functionally produce and secrete BefA. To confirm this, we performed Western blots on bacterial cell free supernatants (CFS) using a polyclonal anti-BefA antibody (FigS3.4A). CFS concentrated from wild-type *Aeromonas veronii* HM21 (*A. veronii*), the strain from which BefA was originally isolated, showed a strong signal at the expected size of BefA (29 kDa), while no signal was present in the lane containing CFS from our *befA* knockout *A. veronii* strain (FigS3.4A), validating the specificity of the anti-BefA antibody. Similarly, we saw a strong band in the lane containing CFS from Nissle^{befA} and no signal in the lane containing CFS from Nissle^{WT}, indicating that our new transgenic Nissle strain was capable of secreting BefA protein (FigS3.4A). To test the functionality of BefA protein secreted by Nissle^{befA}, we mono-associated GF zebrafish larvae with either Nissle^{WT} or Nissle^{befA}. Larvae colonized with Nissle^{befA} had significantly more β-cells than either GF larvae or those

colonized with Nissle^{WT} (FigS3.4B), confirming that the BefA secreted by our transgenic strain was functional, and that the wild type strain was not capable of inducing β-cell expansion via an alternative mechanism.

We next examined postnatal β-cell development in wild type C57/BL6 mice with different microbial associations at postnatal day 12 (P12), just after neonatal β-cell proliferation levels have peaked (Miller et al., 2009). Importantly, mice of this young age are highly variable in overall weight due to differences such as litter size, which can drive similar differences in pancreas mass across our experimental groups (FigS3.4C). Therefore, we quantified the average ratio of insulin-expressing area to total pancreatic cross-sectional area for each pup, rather than β-cell mass, which is the conventional unit in adult mice. We found that mice with a conventional or specific pathogen free (SPF) microbiota had significantly increased ratios of insulin-positive area compared to mice that were either GF or had been given antibiotics (ABX) from birth (Fig3.4A&B), indicating that resident bacteria play a stimulatory role in mammalian β-cell development. To test whether BefA is sufficient to rescue reduced β-cell levels in mice lacking resident bacteria, we mono-associated GF mice with our Nissle strains by inoculating breeder pairs, and allowing them to transmit Nissle vertically to their offspring. As before, we analyzed β-cell area in P12 aged mice from each mono-association and found that pups colonized with Nissle^{befA} had significantly higher insulin positive area than Nissle^{WT} (Fig3.4B&C). We also treated SPF pups concurrently with both antibiotics and purified BefA protein administered by oral gavage during postnatal life, and again analyzed insulin content of the pancreas. Purified BefA treatments were sufficient to rescue the total β-cell content of antibiotic-treated pups to levels similar to those of SPF (Fig3.4A&B). We also allowed the mice from some of our Nissle mono-associated litters to grow to adulthood before we analyzed total

87

β-cell mass. Like the pups, adult mice mono-associated with Nissle[befA] had significantly greater

β-cell mass compared to those mono-associated with Nissle[WT] (Fig3.4D). The overall pancreas

mass of adult mice with these mono-associations was not different (Fig3.4E), suggesting that

mouse pancreas size is not affected by BefA. We also examined alpha-cell mass in adult GF and

SPF mice, and found no significant differences (FigS3.4D&F), similar to our previous findings

in zebrafish (Hill et al., 2016). Collectively, these results demonstrate that BefA is capable of

increasing the β-cell mass of mice, and that these effects are specific and long lasting.

We next performed several experiments to determine if BefA acts directly on murine β-

cells. First, we exposed cultured βTC-6 cells, an immortalized murine β-cell line (Poitout et al.,

1995), to mCh alone or mCh-BefA[99]. Following an incubation of several hours, βTC-6 cells were

washed, and fractions of the cell membranes, cytosol, and rinse were analyzed by Western blot.

A band at the expected size of mCh-BefA[99] (45 kDa) localized to all the βTC-6 culture

components, including the cell membrane fraction (Fig3.4F), whereas bands of the expected size

for mCh alone (29 kDa) were not found in the cell membrane pellet (Fig3.4F). We next imaged

mouse βTC-6 cells incubated with our brighter mNG- BefA[99] protein. Similarly to our

observations with dissociated zebrafish cells (Fig3.3B, FigS3.3A), we detected little to no mNG

puncta associated with β-cells in cultures incubated with mNG alone (mean percentage of cells

co-localized with mNG = 0.2, std. dev = 0.6), but observed significant punctate labeling of

mNG-BefA[99] co-localized with β-cells (mean percentage of cells co-localized with mNG = 11,

std. dev = 5) (Fig3.4G), indicating that BefA is capable of directly interacting with cells of

murine origin, and that the SYLF domain is sufficient for this activity.

We next tested whether direct interaction of BefA's SYLF domain with murine pancreas

tissue would increase β-cell proliferation. Primary mouse islets were isolated from wild type SPF

Swiss Webster mice aged P12 and treated *in vitro* with either mNG-BefA[99], mNG alone, 10 μM glucose, or were left untreated for 48 hours. To measure islet cell proliferation, we treated with the nucleotide analog EdU during the final 4-6 hours of incubation. Islets treated with mNG-BefA[99] showed distinct BefA[99] puncta, similar to the cultured βTC-6 cells and absent from mNG treated islets (Fig3.4H). The addition of either mNG-BefA[99] or our positive control, glucose, resulted in significantly higher levels of EdU incorporation per islet than either untreated or mNG treated groups (Fig3.4I), indicating that the SYLF domain of BefA is sufficient to elicit proliferation of neonatal β-cells in mice.

To determine whether membrane permeabilization is sufficient to elicit mammalian islet cell turnover, we next treated primary neonatal mouse islets with the pore-forming protein, Reg3α. Reg3α treatment resulted in a significant dose-dependent increase in islet cell proliferation as measured by EdU incorporation (Fig3.4J). To test whether membrane permeabilization was necessary for BefA-mediated islet cell turnover, we treated primary neonatal mouse islets with the BefA[R195A] mutant that permeabilizes lipid vesicles less effectively (Fig3.2G&H). When compared to full-length wild type BefA, BefA[R195A] elicited significantly reduced levels of EdU incorporation (Fig3.4K). Taken together, our results indicate that BefA elicits β-cell proliferation via a mechanism of increasing membrane permeabilization. Lastly, to test whether BefA could disseminate through alternate anatomical routes to elicit a response from β-cells *in vivo*, as we had observed in our zebrafish model, we injected
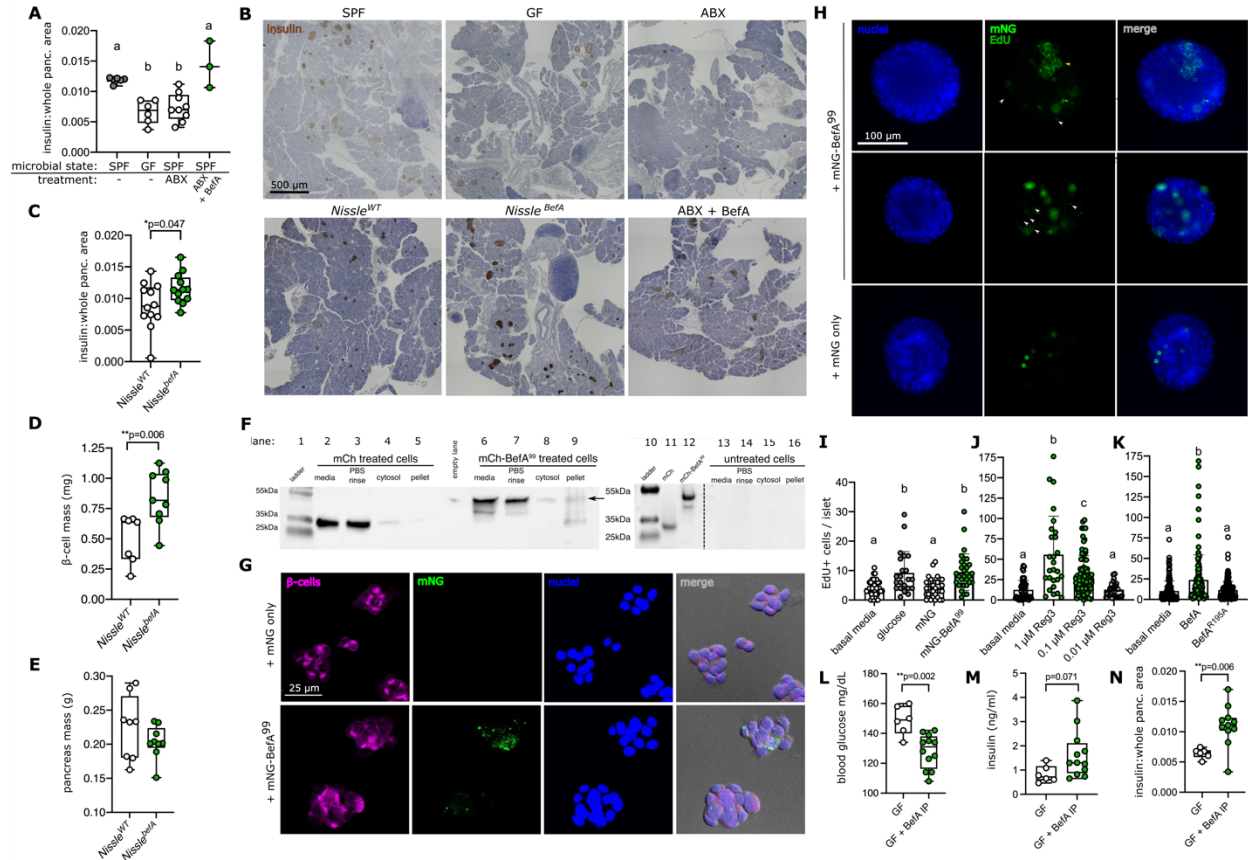
Figure 3.4 BefA's direct activity on host beta-cells is conserved in mice. (A) Boxplots of quantifications of average ratio of insulin area to whole pancreas area across at least three cross sections per mouse pup aged P12. n≥5 mice in each treatment group except ABX + BefA where n=3. (B) Representative cross sections through postnatal day 12 (P12) mouse pancreata, ABX = Antibiotic treated, Hematoxylin staining nuclei (blue) and Insulin (brown). Because insulin staining in GF mice is faint, dashed brown outlines are meant to help distinguish insulin+ areas. Scale bar = 500 μm. (C) Boxplots of quantifications of average ratio of insulin area to whole pancreas area across at least three cross sections per mouse pup mono-associated with Nissle strains, aged P12. n=12 mice in each treatment group. *p value for Student's T-test. (D) Boxplot illustrating total β-cell mass in milligrams (mg) of adult mice, mono-associated with Nissle strains, n≥8 for each group, *p value for Student's T-test. (E) Boxplot illustrating total pancreas mass in grams (g) in adult mice mono-associated with Nissle strains, n≥8 for each group. (F) Western blots for mCherry on βTC-6 cell cultures treated with either mCh (29 kDa) or mCh-BefA[99] (45 kDa) proteins. Lane 1 & 10: ladder. Lanes 2-5: mCherry treated culture components. Lanes 6-9: BefA[99FL] treated culture components. Lane 11: purified mCh protein. Lane 12: purified mCh-BefA protein. Lanes 13-16: untreated culture components. Arrow: mCh-BefA[99] band within cell pellet fraction. (G) Confocal images of βTC-6 treated with either mNG or mNG-BefA[99]. Panels from left to right: Insulin in magenta, mNG in green, DAPI in blue, and DIC overlay with fluorescent channels merged. Scale bar = 25 μm. (H) Fluorescent images of primary mouse islets from P12 SPF pups, after 48-hour treatment with either mNG-BefA[99] (top and middle rows), or mNG only/unconjugated (bottom row). Left column = DAPI stain for nuclei in blue, middle column = mNG (either punctate as pointed out by white arrowheads and outer membrane associated as denoted by yellow arrowhead) and EdU (nuclear) in green, right column = merge of blue and green channels. Scale bar = 100 μm. (I-K) Bar graphs illustrating mean EdU labelled cells in each primary islet per treatment group, each dot represents one islet, whiskers denote std. dev. n≥22 islets. (L) Blood glucose readings (mg/dL) from nonfasted mice of each treatment group. (M) Serum insulin measurements from nonfasted mice of each treatment group. (N) Boxplots of quantifications of average ratios of insulin area to whole pancreas area across at least two cross sections per mouse. (L-N) n≥6 mice in each treatment group, aged P21. *p value for Student's T-test.

90

purified BefA protein intraperitoneally (IP) into GF Swiss Webster neonatal mice and analyzed β-cell mass at weaning (P21). Blood glucose measurements taken from mice injected with BefA were significantly lower than untreated mice (Fig3.4L), which was consistent with increased serum insulin levels (Fig3.4M). Image analyses of the pancreata from these mice also revealed significantly higher levels of insulin expressing tissue in BefA injected samples (Fig3.4N), indicating that IP delivery of BefA is sufficient to exert the same effect on the pancreas as when secreted by gut bacteria. Collectively, these data show that BefA is able to induce β-cell expansion in mice via multiple routes of delivery, highlighting the conserved capacity of a gut microbiota derived protein to disseminate through the body and impact the development of distant tissues.

Discussion

The many bacterial genomes that make up the human microbiome encode a largely unexplored repertoire of diverse bioactive molecules that can impact animal biology. One such example is the secreted bacterial protein, BefA, which we showed can induce expansion of pancreatic β-cells (Hill et al., 2016). The novelty of such microbial products presents a unique challenge for uncovering their bioactive mechanisms. In the current study, we utilized a broad array of molecular, genetic and biochemical approaches to elucidate the mechanism of action of the BefA protein. Our findings reveal that BefA is a compact and stable membrane permeabilizing protein that directly stimulates β-cell expansion in vertebrates. We show BefA's membrane permeabilizing activity is necessary to stimulate β-cell proliferation and that a non-homologous membrane permeabilizing protein, Reg3, recapitulates BefA's impact on β-cells.

The BefA crystal structure reveals atomic features of the lipid-binding SYLF domain, which is widely distributed in predicted proteomes across the kingdom of life. While eukaryotic

SYLF domain-containing proteins are typically multidomain proteins, the SYLF region comprises the bulk of BefA and its prokaryotic homologs. This suggests that the SYLF's intrinsic membrane permeabilizing activity confers a selective advantage for bacteria, possibly as an antimicrobial weapon against competitors in multispecies communities (Granato et al., 2019) or as an agent of programmed cell death (Bayles et al., 2014). *befA* deficient *Aeromonas* mutants are at a slight competitive disadvantage in di-association with wild type *Aeromonas* (Hill et al., 2016) under conditions of normal host β-cell census, indicating that BefA benefits *Aeromonas* independently of bacterial competitors or insulin paucity. Although we do not yet fully understand BefA's role in *Aeromonas* fitness, our studies reveal how the protein's biochemical properties of stability and membrane permeabilizing activity mediate its impact on host biology.

BefA's compact and stable structure likely endows it with the capacity to persist in diverse extra-cellular environments. Our finding that BefA can directly induce β-cell proliferation in primary islet cultures suggests that it disseminates to the pancreas from the intestinal lumen and water column. This conclusion is supported by imaging of fluorescently labelled BefA and functional characterization of BefA through multiple routes of administration. Furthermore, the BefA recalcitrance of larval mutants lacking a patent HPD or vasculature indicates that both of these routes are involved in the pancreatic response to BefA. Reflux up the HPD has long been suspected as a potential route of bacterial transmission to both the pancreas and the liver, where bacteria have been found to drive carcinogenesis (Pushalkar et al., 2018; Castillo et al., 2019; Tilg et al., 2016; Riquelme et al., 2019). In addition, Zhang and colleagues recently highlighted the bloodstream as a mode of delivery for bacterial products directly to the pancreas. Their study showed that microbial ligands of the Nod1 receptor could access the blood stream in order to subsequently bind to β-cells and promote insulin trafficking (Zhang et al.,

2019). Together with our findings, these results further our mechanistic understanding of how microbial products from the gut impact distant host tissue biology.

Our studies in simplified synthetic lipid vesicle systems revealed that BefA and its SYLF domain have intrinsic membrane vesiculating and permeabilizing activity. Membrane permeabilizing proteins and peptides are produced across the kingdom of life, from viral encoded viroporins to animal antimicrobial proteins. Many of these proteins act by changing membrane curvature (Guha et al., 2019). A common mechanism of permeabilization involves oligomerization on membranes, which may be the phenomenon we observed with BefA puncta that form on eukaryotic cells. Alternatively, these BefA puncta may represent endocytic or exocytic responses of cells repairing BefA-induced membrane damage (Brito et al., 2019).

Lipid membrane repair is critical for the survival of all cells. Animal cells, which lack defensive cell walls, have evolved extensive mechanisms for monitoring and maintaining their membrane integrity, including responding to increases in intracellular calcium and extracellular ATP that result from plasma membrane ruptures (Horn et al., 2018). Pancreatic β-cells rely on these same cues to regulate their insulin secretion and thus may be especially sensitive to perturbations in their membrane integrity (Noguchi et al., 2019). Membrane attack can come from microbial products, such as pore forming toxins and antimicrobial peptides, or from host defense proteins like the pore-forming Reg family antimicrobial lectins. The functions and cellular targets of these host-derived proteins are not fully understood. For example, while Reg proteins are highly expressed and secreted at the apical border of intestinal epithelial cells, where their antimicrobial activity plays a key role in the maintenance of a healthy barrier against microbes (Vaishnava et al., 2014; Wang et al., 2016), they are also expressed in tissues lacking a resident microbial population such as the developing mammalian pancreas (Bartoli et al., 1998;

Mally et al., 1994; Perfetti et al., 1996; Unno et al., 2002). Reg proteins are also upregulated during β-cell regeneration (Terazono et al., 1988), and are capable of stimulating β-cell expansion (Unno et al., 2002). Type I, or RegI, proteins are associated with β-cell turnover and protection from diabetes in animal models (Okamoto et al., 1999). In the T1D NOD mouse model, antibiotic-induced disease acceleration was associated with decreased Reg3 expression in both the ileum and pancreas, and upregulated by early life exposure to a protective microbiota (Zhang et al., 2021). Furthermore, RegI and Reg3 proteins are upregulated in humans with diabetes (Astorri et al., 2010; Bacon et al., 2012; Yang et al., 2015) and are hypothesized to play a role in promoting regeneration and recovery of β-cells during type 2 diabetes remission (Sala et al., 2017). Similarly, another antimicrobial peptide, Cathelicidin, is expressed in rat and human β-cells and treatment with this pore-forming peptide stimulates short term insulin secretion and longer term β-cell expansion (Pound et al., 2015). Until now, neither the pore forming activities of Reg nor Cathelicidin have been linked to their roles in the pancreas. Our insights into the biochemical properties of the bacterial BefA protein demonstrate that diverse membrane-permeabilizing proteins can stimulate β-cell development, pointing to membrane manipulations as a new approach for stimulating β-cell renewal to treat diabetes and motivating future experiments to test whether BefA can prevent or ameliorate diabetes in animal disease models.

Membrane permeabilizing activity is a ubiquitous feature of multispecies microbial communities, representing what we call a Microbial Associated Competitive Activity (MACA) (Wiles et al., 2020). Membrane disrupting MACAs are detected by host immune sensors such as the NLRP3 inflammasome (Swanson et al., 2019), similar to host immune reception of Microbial Associated Molecular Patterns (MAMPs). We hypothesize that manipulation of membrane integrity is an underexplored mechanism in tissue development and plasticity. The earliest

eukaryotic cells, evolving among MACA-rich bacterial communities, would have required membrane defense and counterattack strategies. Extant eukaryotes have elaborate repertoires of MACA mimics, including many membrane permeabilizing antimicrobial proteins (AMPs). Strikingly, these host-derived AMPs, like Reg proteins and Cathelicidin, have a breadth of immunomodulatory activities at lower concentrations and tissue locations distant from where they function in microbial defense (Hancock et al., 2016). We hypothesize that microbiota induction of host defense proteins, along with microbiota-derived secreted proteins, contribute to programs of tissue development, such as early life β-cell expansion.

The developmental plasticity of the pancreas in response to microbial membrane permeabilizing activities may confer a selective advantage by allowing developing animals to match their metabolic capacities to their nutritional environment. Bacterial abundance is a signature of nutrient availability. Membrane permeabilizing activity specifically is a characteristic of bacteria that compete successfully in dense, multispecies communities found in animal digestive tracts. For example, of all the bacterial taxa encoding pore-forming MACPF superfamily members, the most abundant representatives are the Firmicutes, Bacteroidetes, Actinobacteria, and Gammaproteobacteria (Moreno-Hagelsieb et al., 2017), the most common residents of vertebrate digestive tracts. Almost all the BefA homologs we found in human-associated bacteria were produced by the Gammaproteobacteria, which are metabolically versatile, fast-growing, "weedy" species that are often abundant members of the neonatal gut microbiome, but become rarer in healthy adults (Yassour et al, 2018; Bokulich et al., 2016). The exposure of developing pancreatic cells to bacterial-derived membrane-perturbing MACAs and to self-produced antimicrobial defense MACA mimics (such as Regs) would provide information about the abundance of bacteria and by extension nutrients in the environment.

The recent surge of type 1 diabetes in high income countries is strongly associated with increased hygiene measures that reduce microbial exposure during early infant life. In a refinement of the hygiene hypothesis (Strachan 1989), we hypothesize that increased early life hygiene has resulted in mismatched microbial and nutritional developmental cues. Low exposure to membrane-perturbing MACAs would set up developmental programs for conditions of nutrient scarcity, restricting the β-cell population, and leaving individuals ill-prepared for future diets of caloric excess and with limited β-cell reserves to counteract future autoimmune attacks or hyperglycemia that would drive diabetes development. Understanding how different membrane permeabilizing activities stimulate β-cell expansion will spur new strategies to prevent or reverse the consequence of developing with mismatched microbial and nutritional cues.

Materials & Methods

BefA protein expression & purification

The *befA* gene was expressed and purified as previously described (Hill et al., 2016). Additionally, BefA protein containing selenium methionine was produced as described by Van Duyne et al. and purified using the same methods as native BefA (Hill et al., 2016; Van Duyne et al., 1993). The nucleotide sequences corresponding to the Befa[99] and BefA[133] truncations were amplified using the same reverse PCR primer previously published for amplifying BefA (Hill et al., 2016) which was paired with the following forward PCR primers for each truncation respectively: 5'-GGCCATATGATGaagacggcgaaagaggcgagg-3' and 5'-GGCCATATGATGggttatgcggtgttcgattcgcgc-3'. A SOE (splicing by overlap extension) PCR reaction (Horton et al., 1990) was used to fuse the *mCherry* or *mneongreen* genes to the 5' end of either the *befA* or the *befA*[99] gene with a six amino acid linker (AAAGGG) connecting them.

Each construct was then cloned into the pET-21b+ plasmid (Novagen, Darmstadt, Germany) and expressed and purified using a C-terminal HisTag® as previously described for the native BefA protein (Hill et al., 2016).

Human Reg3α protein expression & purification

Human Reg3α (hReg3α) lacking the N-terminal inhibitory pro-peptide was expressed and purified from a pET3a vector (Addgene plasmid ID 64937, deposited by the laboratory of Dr. Lora Hooper) in *E. coli* BL21 DE3. hReg3α was purified based on the denaturing protocol in Cash et al. 2006, with some differences provided by personal communication with the Hooper laboratory. Briefly, hReg3α was expressed in a 1 L culture of *E. coli* with 1mM IPTG for 3 hrs at 37°C shaking, cells were pelleted and stored at -20°C. The cell pellet was resuspended at 4°C in 50 mL of 20 mM Tris pH 7.5, 10 μM EDTA pH 8 and 1% Triton x-100. All subsequent steps were performed at 4°C unless otherwise noted. Cells were lysed using a sonicator, pelleted, supernatant removed and the pellet resuspended using a dounce homogenizer in 100 mL of 20 mM Tris pH 7.5, 10 μM EDTA pH 8, 500 mM NaCl and 1% Triton x-100. The pelleting step was repeated, the supernatant was removed and resuspend again in the previous solution. The solution was centrifuged and the pelleted inclusion bodies were resuspended gently in 20 mL of 7M guanidine-HCl, 10 μM reduced glutathione, 20 mM Tris pH 8.0 and 2 mM EDTA. The solution was rotated for 24 hrs at room temperature. The solubilized inclusion bodies were spun down and 4 mL of the supernatant was added, dropwise, to 200 mL of refolding buffer: 50 mM Tris pH 8.0, 10 mM KCl, 2 mM MgCl$_2$, 2 mM CaCl$_2$, 240 mM NaCl, 500 mM Guanidine-HCl, 400 mM sucrose, 500 mM arginine-HCl, 1 mM reduced glutathione, 0.75% Triton x-100 and 0.1 mM oxidized glutathione. Solution was left to stand, covered for 24 hrs. The solution was centrifuged and the supernatant collected and then a series of dialysis steps were performed.

First, the protein solution was dialyzed against 25 mM Tris pH 7, 2 mM $CaCl_2$ and 25 mM NaCl for 8-24 hrs, then repeated with a fresh batch of dialysis solution. Then the solution was dialyzed into 25 mM MES pH 6.0, 2 mM $CaCl_2$, and 25 mM NaCl for 8-24 hrs in preparation for cation exchange chromatography. After dialysis finished, the solution was centrifuged and the supernatant collected. The solution was passed over a 2 mL SP-Sepharose column (cation exchange) equilibrated with the last dialysis solution. The column was washed with 25 mL of 25 mM MES pH 6.0, 2 mM $CaCl_2$, 25 mM NaCl and 150 mM NaCl. The protein was eluted with 15 mL of 25 mM MES pH 6.0, 2 mM $CaCl_2$, 25 mM NaCl and 400 mM NaCl. The eluted protein was dialyzed into water or 25 mM MES pH 5.5 and 25 mM NaCl and concentrated using a Vivaspin 20 centrifugal concentrator with a 3 kDa molecular weight cutoff (Millipore Sigma, St. Louis, MO).

BefA crystallization

Purified BefA and selenium methionine substituted BefA (SeMetBefA) were crystalized from a starting concentration of 10-16 mg/mL in ddH_20 in a reservoir solution of: 24-25% PEG 3350 (Hampton Research, Aliso Viejo, CA), 0.1M citric acid pH 3.5, and for semetBefA, 1mM TCEP (tris(2-carboxyethyl)phosphine). Hanging drop vapor diffusion with ratios of protein to reservoir solution of either 1:1 or 2:1, resulted in crystals within 7-10 days. Crystals were cryoprotected in the reservoir solution plus 20% PEG 200 (Hampton Research, Aliso Viejo, CA) and were flash frozen in liquid nitrogen for data collection at the Advanced Light Source in Berkeley, CA beamline 5.0.2 using the Pilatus detector at a wavelength of 1 Å.

BefA structure determination

The space group is C2 with a=94.49, b=64.28, c=42.60 Å and β=113.43, with one molecule in the asymmetric unit. Various diffraction data sets were collected to ~1.3 Å

resolution at the Advanced Light Source in Berkeley, CA beamline 5.0.2 using the Pilatus detector at a wavelength of 0.9795 Å and were integrated and scaled using the program package HKL2000 (Otwinowski et al., 1997). The initial structure determination was based on a single selenomethionine data set using Autosol from the PHENIX program package (Liebschner et al., 2019) using the SIRAS approach. The Autosol/Autobuild procedures produced a model with 202 residues identified and placed in 5 fragments, including 203 water molecules. Rwork/Rfree for that model were 0.198/0.210 at 2.0 Å resolution. Several cycles of manual model corrections using the program Coot (Emsley et al., 2010) against the SeMet data set resulted in a nearly complete model (residues 36-258) with Rwork/Rfree 0.174/0.197 at 1.31 Å resolution (Table S1).

BefA structure refinement

Initial rigid body refinement using the semetBefA structure, merged data from two native data sets and Refmac 5 (Collaborative Computational Project., 1994) resulted in Rwork/Rfree values of 0.297/0.298. Manual model building was performed using Coot 0.8.1 (Emsley et al., 2010) and the bulk of the refinement was carried out using PHENIX 1.9-1692 (Adams et al., 2010). The high-resolution cutoff was determined by the method of Karplus and Diederichs 2012 using CC1/2 of > 0.3 and completeness of > 50% in the highest resolution shell. Two citrate molecules, found in the crystallization conditions at 100 mM, are bound near Ser218 (FigS3.5). They likely coordinate a metal that presents as a positive electron density peak if left out. Attempts to place a magnesium ion between the two citrates, as our best guess, resulted in higher Rwork/Rfree values, and therefore was left out. In addition, the distances between the metal and the citrate oxygens are between 1.4-1.7 Å, which is particularly short. In later stages of refinement, hydrogens, and individual anisotropic B-factors were included and all decreased the

Rwork/Rfree values. The final refinement round resulted in Rwork/Rfree of 0.132/0.154 and is included in Table S1. The structure is deposited in the PDB as 7RFQ.

Thermofluor melting temperature assay

The thermofluor assay was performed as previously described (Henderson et al., 2013) with the following alterations. A final concentration of 0.3-0.5 mg/mL of BefA or lysozyme (chicken egg white lysozyme, VWR, Radnor, PA) was incubated on ice with 16X SYPRO Orange protein dye (Thermo Fisher Scientific, Waltham, MA), 150 mM NaCl, and 25 mM HEPES or Tris (pH 7-7.5). 20 μL samples were pipetted into a 364 well PCR plate and covered with a clear, optically transparent seal for reading fluorescence. Melting temperature data for BefA and lysozyme, as a control, were collected on a StepOnePlus real time PCR machine (Thermo Fisher Scientific, Waltham, MA). Reactions were run using the following thermal denaturation protocol: 4°C for 1 min, slow ramp rate (~0.03 °C/sec) from 4°C to 80°C with fluorescence data collected ~8.5 sec. Fluorescence data was plotted against temperature and approximate melting temperature (T_m) determined using the following equation: max – ½*(max – min). Max refers to fluorescence intensity maximum value (peak) and min refers to minimum intensity value prior to intensity increase. The BefA thermofluor assay was run on two separate days with a total of 12 replicates. The lysozyme thermofluor assay was run in triplicate on two separate days.

Actin co-pelleting assay

The actin co-pelleting assay was performed as described (Hasegawa et al., 2011) with the following changes: BefA, BefA[99], and cortactin (gift from Dr. Brad Nolen), were pre-clarified to remove aggregate proteins by spinning at 65,000 rpm for 30 min at 4°C in a TLA100 rotor (Beckman Coulter, Brea, CA) before incubating with F-actin. F-actin filament formation was

encouraged by incubating purified rabbit actin (gift from Dr. Brad Nolen) in a solution of 10 mM imidazole (pH 7.0), 50 mM KCl, 1 mM EGTA, 1 mM $MgCl_2$ and 1 mM DTT at room temperature for 45 min. The supernatants of the clarified proteins were incubated at a final concentration of 2.5 μM, with or without F-actin at 5 μM, in the buffer described above for 1.5 h at room temperature. All reactions were centrifuged at 65,000 rpm for 30 min at 4°C in a TLA100 rotor. The supernatant (unbound) and pellet (bound) proteins were separated and analyzed by SDS-PAGE and stained with coomassie brilliant blue.

PIP strip assay

BefA, negative control AimA (Rolig et al., 2018) and positive control $PI(4,5)P_2$ Grip (Echelon catalog number G4501) all contain 6X-His tags and were incubated with individual membrane strips or arrays at a final concentration of 0.75 μg/mL for 1 hr at RT as directed by the manufacturer (Echelon Biosciences, Inc., Salt Lake City, UT). For visualization, an anti-His rabbit polyclonal antibody (ABM, Richland, BC, Canada) in combination with a polyclonal goat anti-rabbit HRP conjugated secondary antibody (Thermo Fisher Scientific, Waltham, MA) were used to detect the presence of His-tagged proteins after washing the membranes with 1X PBS-T (phosphate buffered saline with 0.1% Tween-20) as directed. Membrane and PIP lipid strips were performed with BefA 3-5 times, once with AimA and once with Grip for the PIP lipid strip. The membrane lipid arrays with or without calcium were performed once with BefA.

Vesicle preparation, treatment, and quantification

Vesicles were made as previously described using electroformation (Veatch 2007; Andelova et al., 1992), with the following lipid molar ratio to obtain red fluorescent vesicles: 99.5 % 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) (Avanti Polar Lipids Inc., Alabaster, Alabama) and 0.5 % Texas Red 1,2 Dihexadecanoyl-sn-Glycero-3-Phosphoethanolamin (Texas

Red DHPE) (Thermo Fisher Scientific, Waltham, MA). Fresh vesicles in 0.1 M sucrose were made the previous day and incubated with either the same volume of water (control) or BefA in water to a final concentration of 2 μM for 35 min at room temperature. Treated vesicles were loaded into a 350 μL glass cuvette (Starna Cells, part number 3-3.45-SOG-3, Atascadero, CA) and imaged on a home-built light sheet fluorescence microscope described in Jemielita et al. 2014. Four, three dimensional stacks of fluorescent images, spaced 1 μm apart, were collected across the middle of the cuvette 35 and 50 min after addition of BefA protein or water (control). Phenotypes, such as percent of multi-vesicular vesicles and vesicle concentration, were quantified using a custom-made Python script (available here: https://github.com/rplab/vesicle_detector) that identified potential vesicles and then sorted by hand, removing false positives and sorting features as single vesicles or multi-vesiculated vesicles. Percentages were calculated as the number of multi-vesiculated vesicles divided by the total number of features (sum of multi-vesicular vesicles and single vesicles). The total number of single and multi-vesicular vesicles (three or more vesicles attached) were quantified and graphed in GraphPad Prism.

Vesicle dye release assay

The vesicle dye release assay was done according to Jimah et al. 2017 with some slight changes. Briefly, a total of 1 mg of DOPC lipids (Avanti Polar Lipids Inc., Alabaster, Alabama) in chloroform was dried under vacuum for 3 hrs, resuspended in 1 mL of 20 mM carboxyfluorescein, 25 mM Tris pH 7.5 and 25 mM NaCl with a brief vortex, sealed with parafilm and left overnight at 4 °C. The next day, lipids were extruded 10 times through 100 nm or 400 nm pore size membranes using a LIPEX 1.5 mL Extruder (Transferra Nanosciences Inc., Burnaby, B.C., Canada) at 50 °C to form vesicles of a more uniform size. Extruded vesicles with

dye inside were separated from free dye by using a PD-10 column. Samples were loaded into a black 96 well, flat bottom plate with a clear bottom (Corning, Corning, New York). Samples were read using a BMG Labtech FLUOstar Omega plate reader (fluorescence ex/em 485/520 nm). Fluorescent measurements of vesicles without treatment were read initially, treatments were added, and then read for 1 hr every 2 min. Finally, 1% Triton x-100 was added to obtain maximum fluorescence values. The initial readings were subtracted from each curve and the percent of max was calculated by taking each treatment reading and dividing it by the final, maximum Triton x-100 value. Data was plotted in GraphPad Prism using mean and standard deviation.

CyQUANT assay for bacterial membrane stability

Bacterial strains were maintained in 25% glycerol at $-80\,°C$. Bacteria were directly inoculated into 5 mL lysogeny broth (LB) media (10 g/liter NaCl, 5 g/L yeast extract, 12 g/L tryptone, 1 g/L glucose) and grown for ~16 hrs (overnight) with shaking at $30\,°C$. Bacterial overnight cultures were subcultured and grown to an approximate optical density at 600 nm of 0.1 to 0.3. The bacterial cells were washed twice by centrifugation, aspiration of supernatant, and suspension in zebrafish embryo media (pH=5.5); after the second aspiration step, bacterial cells were concentrated by suspending the pellet in a volume of sterile embryo media (pH=5.5) that was a quarter of the starting volume. CyQUANT Direct Red (Invitrogen; product no. C35014) 2X Reagent was prepared fresh as directed by the kit's protocol by diluting CyQUANT™ Direct Red nucleic acid stain (1:250) and CyQUANT Direct Red background suppressor (1:50) in embryo media (pH=5.5). In a sterile 96-well tissue culture-treated black flat-bottom microplate (Greiner Bio-One; product no. 655090), appropriate volumes of bacteria, or sterile embryo media (pH=5.5) in controls, were mixed in triplicate with 50 μL/well of CyQUANT 2X Reagent. Cells

and CyQUANT 2X Reagent were incubated in the dark at 30°C for one hour. A FLUOstar Omega microplate reader (BMG Labtech, Offenburg, Germany) was used to obtain baseline fluorescence measurements with 584 nm excitation and 650 nm emission filters before addition of treatment solutions. Treatments were added to each well to achieve desired final concentrations in a total volume of 100 μL/well. Final concentrations of treatments were as follows: 7 μM BefA, 3 μM Reg3α, and 0.08% SDS. Fluorescence measurements were then taken every 10 min for 400 min at 30°C, the plate was shaken briefly before each reading. Unfortunately, addition of BefA[99] in this assay caused dye precipitation which confounded our analysis. We also tried using BefA in a propidium iodide (PI) assay, another dye often used for assessing membrane permeability, and these efforts were also inconclusive due to natural autofluorescent in the presence of propidium iodide.

Bacterial imaging and quantification

Bacillus sp. (strain index # ZOR0058) a contaminant isolated from a GF zebrafish flask, was grown overnight at 30°C, shaking, from a glycerol stock. The culture was back-diluted in the morning (approximately 16 hrs later) at a 1:50 dilution and grown at 30°C shaking for approximately 1.5 hrs. 1 mL of bacterial culture was pelleted and gently washed into 500 μL of PBS. Purified mCherry-BefA or mCherry control protein was incubated with bacterial cells at 1 μM for 30 min at room temperature. SYBR Green nucleic acid stain (Thermo Fisher Scientific, Waltham, MA) was added to cells at 1X final concentration in PBS and incubated at room temperature for 20 min. 100 μL of cell mixture was loaded into multi-well microscope slides with a loose lid. Images in three channels, DIC, SYBR Green (ex/em 475/525 nm), and mCherry (ex/em 575/625 nm) were acquired using a GE DeltaVision Ultra fluorescent microscope using a 60x magnification, 1.42 numerical aperture, Plan Apo N, oil immersion objective. The DIC

channel was taken at 50 msec exposure with 10% power, the SYBR Green channel was taken at 100 msec exposure with 10% power and the mCherry channel was taken at 75 msec exposure with 32% power. 10-15 original images (1024 x 1024 px) were taken of each sample and processed identically within each channel using Fiji (ImageJ). Each channel was pseudocolored using the channels tool (DIC, gray; SYBR Green DNA, green; mCherry proteins, magenta) and the min and max displayed values set for the first two channels (DIC min/max 702/3934 and green 575/15605), while magenta/protein channel was background subtracted using a rolling ball radius of 50 px. A 20 x 20 μm square was cropped around representative cells, and then each image scaled identically in Adobe Illustrator CC 2018.

For quantification of BefA binding, all cell regions in the two treatments that were not overlapping with other cells were quantified for BefA binding using Fiji. In the DIC channel, the segmented line tool at 35 px wide was used to draw along the length of each cell, clicking at each segment. For every image, a line of roughly the same length as the average bacterial cell was drawn to capture the background intensity. The mean gray value and standard deviation was measured for each line using the Measure analysis tool. The mean background values were subtracted from each value taken along a cell and were plotted in GraphPad Prism.

Anti-BefA antibody preparation and testing

Two rabbit polyclonal antibodies were produced through Cocalico Biologicals (Stevens, Pennsylvania) using their standard 56-day protocol (three boosts), ending in exsanguination. Briefly, BefA with 6x His affinity tag was purified as previously described (Hill et al., 2016), separated from impurities by SDS PAGE and the BefA band excised. The gel slices containing 1-2 mg protein were mailed to Cocalico Biologicals. Two rabbits were chosen based on pre-bleeds showing little response to other antigens of similar molecular weight to BefA on western

blots. The rabbits were injected with BefA antigen and two bleeds (~6-8 mL each) plus the final

bleed (~30-50 mL each) were collected. The bleeds were analyzed via western blot for reactivity

against purified BefA (full-length, truncated BefA99 and fluorescently-tagged forms) as well as

endogenous BefA from bacterial cultures. 1 µg of purified protein or 25 µL of bacterial culture

was loaded on an SDS PAGE gel and transferred to a PVDF membrane. The membrane was

blocked in 5% milk in Tris-Base Saline Tween-20 (TBST) and then probed with our antibody

bleeds against BefA at a 1:500 concentration in 5% milk/TBST overnight at 4°C. To visualize

the resulting BefA bands, a secondary Anti-rabbit IgG HRP-linked antibody (Thermo Fisher

Scientific, Waltham, MA) was used at 1:1000 for 1 hour at room temperature. Additional

samples of bacteria cultures that lacked endogenous BefA were run simultaneously with our

other samples to confirm that the antibody is specific.

Gnotobiotic zebrafish husbandry and protein treatment

All zebrafish experiments were performed using protocols approved by the University of

Oregon Institutional Care and Use Committee (protocol number 14-14RR) and followed

standard protocols. Wild-type (ABC x Tu strain), *Tg(-1.0insulin:eGFP)* (referred to as *ins:gfp*)

(diIorio et al., 2002), *Tg(ins:mCherry)jh2* (Pisarath et al., 2007), *myd88*[-/-] (Burns et al., 2017),

and *sox9bfh313* (referred to as *sox9b*[-/-]) (Delous et al., 2012), npas4l (referred to as *cloche)*

(Reischauer et al., 2016; Stainier et al., 1995), and *Tg(nxk2.2:eGFP)* (Pauls et al., 2007)

zebrafish embryos were either raised conventionally or derived GF as previously described (Hill

et al., 2016; Bates et al., 2006). Following GF derivation, fish were mono-associated with

bacterial strains as previously described (Hill et al., 2016; Bates et al., 2006) or treated with a

given purified protein. For experiments involving the treatment of larvae with sterile purified

protein, the protein was either added directly to the embryo media at a final concentration of 500

ng/mL at 4 dpf, unless specified otherwise, or delivered using oral microgavage at the concentration indicated in the text with rigorous sterile technique. Larvae were incubated with the protein of interest for 48 hrs or until 6 dpf.

Quantification of zebrafish β-cells

Larvae were fixed in 4% PFA, antibody stained, and imaged using a confocal microscope as described previously (Hill et al., 2016) to quantify total β-cell numbers. For experiments utilizing *Tg(-1.0insulin:eGFP)* larvae, antibody staining was conducted as previously described (Hill et al., 2016). For experiments utilizing *myd88^{-/-}*, *sox9b^{-/-}*, and *cloche* larvae, they were first permeabilized using an acetone freeze step for 7 min at -20° C before staining with guinea-pig anti-insulin (Dako, Carpinteria, CA), Alexa Fluor 488 or 647 goat anti-guinea-pig (Thermo Fisher Scientific, Waltham, MA), and either TO-PRO-3-Iodide (Thermo Fisher Scientific, Waltham, MA) or DAPI (Thermo Fisher Scientific, Waltham, MA).

Larval zebrafish oral microgavage and cardiac valley injections

For oral microgavage, 4 or 6 dpf larval zebrafish were anesthetized, mounted in sterile 4% methylcellulose and gavaged as described previously (Cocchiaro et al., 2013). Proteins at the concentrations specified in the text were administered to each larval zebrafish with a 4.6 nL volume at an injection rate of 23 nL/sec using a Nanoject II Auto-Nanoliter Injector (Drummond Scientific Company, Broomall, PA). To visualize the global localization of BefA *in vivo*, mCh-BefA, mCh, mNG-BefA, or mNG were administered at a concentration of 1 mg/mL. All images were taken no longer than 2 hrs after gavage using the wide-field feature of a LEICA DM6 confocal microscope (Leica Microsystems Inc., Buffalo Grove, IL).

For cardiac valley injections, 4 dpf larval zebrafish were anesthetized and injected directly into the cardiac valley with 2 pg of BefA or negative control vehicle again using the

Nanoject II Auto-Nanoliter Injector (Drummond Scientific Company, Broomall, PA) as previously described (Wiles et al., 2009). Zebrafish were revived and housed in separate wells of a 24-well plate after the injection to monitor health. To ensure that gavaged or injected GF larvae were not contaminated with microbes during these procedures, a sample of embryo media and homogenized gastrointestinal tracts from some larvae were plated onto tryptic soy agar (TSA) or luria broth (LB) agar at the time of harvest, typically 48 hrs later.

BODIPY feeding and analysis

A BODIPY/egg yolk mixture was prepared by emulsifying sterile 5% chicken egg yolk in zebrafish embryo media via sonication for 3-5 min (*37*). BODIPY dissolved in DMSO (Thermo Fisher, Waltham, MA) was quickly added to the emulsified egg mixture at a final concentration of 6.4 μM and vigorously vortexed for 3-5 min. 1ml of the BODIPY/egg yolk mixture was added to 4dpf zebrafish in a sterile 12-well plate (10 zebrafish/well) for 5 hrs. Zebrafish were covered in foil during their exposure to the BODIPY/egg yolk mixture. To visualize the patency of the pancreatic ductal system, the confocal feature of a LEICA DM6 confocal microscope (Leica Microsystems Inc., Buffalo Grove, IL) was used for imaging live zebrafish administered BODIPY/egg yolk. With DIC, the pancreas was located and z-stack images throughout the entire head of the pancreas were acquired. The presence or absence of BODIPY within the pancreas was assessed by compiling the z-stacks taken of the pancreas into maximum projections.

Gnotobiotic mouse husbandry

SPF C57BL/6J and Swiss Webster mice were originally purchased from Jackson Labs and reared in the SPF facility at the University of Utah. Germfree mice were reared in sterile gnotobiotic chambers in the germfree facility at the University of Utah. Microbial sterility was

determined every 3 weeks and at sacrifice by plating and PCR. For *E. coli* Nissle mono-associations, breeder pairs were inoculated via oral gavage with approximately 10^6 CFU of a given strain. Mono-associations for each strain were performed in separate gnotobiotic chambers to prevent cross-contamination. Successful and continual Nissle colonization was confirmed by periodically plating feces to monitor the presence of a single colony morphology, and was further confirmed by 16S sequencing. Contamination was also ruled out in the same manner. Periodic PCR amplification of the *befA* gene was also used to monitor each mono-association experiment over time. All experiments were performed in accordance with federal regulations as well as the guidelines for animal use set forth by the University of Utah Institutional Animal Care and Use Committee under protocol numbers 17-04009 & 20-03006.

Oral gavage delivery of BefA to neonates was performed three times at p3, 5, and 7, as described (Francis et al., 2019). BefA was delivered at a final concentration of 500 µg/mL. The same concentration was also used to deliver BefA via neonatal IP injection at P3, 5 and 7. C57B6 mice were used for oral gavage experiments and assayed at P12 and Swiss Webster mice were used for IP injection experiments and assayed at P21.

Quantification of mouse beta-cells and alpha-cells

Entire wild-type litters of either C57BL/6J or Swiss Webster of the indicated gnotobiology were collected at neonatal or adult ages as indicated in the text, and immediately euthanized before harvesting the pancreas. Freshly dissected pancreata were weighed before fixing overnight in 10% buffered zinc formalin or Z-FIX (Anatech LTD, Battle Creek, MI) at RT. Fixed pancreata were rinsed 3 times in PBS before dehydrating in increasing concentrations of EtOH. Dehydrated samples were paraffin embedded and for each sample at least 3 longitudinal sections of 6 um thickness and 100 um apart (for neonates) or 200 um apart (for

adults) were cut and mounted onto glass slides for analysis. Slides were then deparaffinized, rehydrated, permeabilized, blocked in PBS with 5% goat serum (Jackson ImmunoResearch Laboratories, INC., West Grove, PA) at RT for one hour, and stained with a guinea pig anti-insulin antibody (Dako, Carpinteria, CA) at a final concentration of 1:200 overnight (ON) at 4°C. Primary antibody was washed in PBS with 0.1% Tween 20 (PBT) before applying a horseradish peroxidase (HRP) conjugated anti-guinea pig antibody (Jackson ImmunoResearch Laboratories, INC., West Grove, PA) at a final concentration of 1:250 for 2 hrs at RT or ON at at 4°C. After rinsing with PBT, HRP was detected using a DAB substrate kit (Abcam, Cambridge, MA). Counterstaining of nuclei was performed using Meyer's Hematoxylin (Millipore Sigma, St. Louis, MO). Slides were dehydrated before mounting with Permount (Fisher Scientific, Waltham, MA). An Axio Scan.Z1 slide scanner (Zeiss, White Plains, NY) brightfield microscope was used to image and subsequently stitch entire stained pancreatic cross sections into single images for analysis. The open access software Fiji was used to quantify both insulin-positive area and total cross-sectional area for each section. The ratio of insulin area:total pancreatic area was determined for each section and then averaged across three sections for each mouse. For adult mouse beta-cell quantifications, the same procedure was used, except that mice were harvested at 6-8 weeks of age and beta-cell mass was calculated by multiplying the final ratio of insulin area:total pancreatic area by pancreas weight.

For quantifying alpha-cell mass, immunofluorescence staining, also with the same guinea pig anti-insulin antibody, as well as a mouse anti-glucagon antibody (1:500, Sigma Aldrich, St. Louis, MO) to visualize beta-cell and alpha-cell content. Alexa-flour 647 goat anti-guinea pig (1:500, Thermo Fisher Scientific, Waltham, MA) and Alexa-flour 488 goat anti-mouse (1:500, Thermo Fisher Scientific, Waltham, MA) secondary antibodies were applied. Sections were

counterstained with DAPI (1:1500, Thermo Fisher Scientific, Waltham, MA) and mounted for imaging in Prolong Diamond Antifade Mountant (Thermo Fisher Scientific, Waltham, MA). Total insulin and glucagon positive area was calculated for each section using FIJI (Schindelin et al., 2012). The total area covered by pancreatic tissue was determined from DAPI staining, allowing for the percentage of α-cell and β-cell area to be calculated for each section. The ratio of glucagon positive area to either whole pancreatic area or insulin positive area for each cell type was determined in at least 3 cross sections for each mouse by averaging across each section.

Quantification of blood glucose and insulin levels from mice

Blood glucose readings were taken in the morning from nonfasted mice at the same time across treatment groups from tail vein samples using a Contour Next EZ glucometer and Contour Next Blood Glucose Test Strips. Serum insulin (also collected from tail vein blood) was measured using the Crystal Chem (Elk Grove Village, IL) Insulin ELISA Kit. For each mouse, two separate 5 uL serum samples were assayed and averaged for the final reported value.

Zebrafish primary islet dissection and treatment

Gut dissections were performed on 4 dpf GF *Tg(-1.0insulin:eGFP)* larval zebrafish as previously described (Bates et al., 2006). Since the pancreas is tightly associated with the gut, it came along readily with dissected gut tissue. The head of the pancreas was dissected away from the rest of the gastrointestinal organs, so all that remained was the primary islet and a small amount of surrounding exocrine tissue. We found that leaving this small amount of supporting tissue was necessary to keep the islet intact for accurate β-cell quantification. Dissections also likely included small amounts of extra-pancreatic tissue, which varied between dissections, but may have included cell types of the following: ducts, gallbladder, and intestine. Each dissection was transferred into a sterile PCR tube containing 50 uL of Leibovitz's L15 Media with

GlutaMax Supplement (L15) (Thermo Fisher Scientific, Waltham, MA). Protein treatments were added at a final concentration of 500 ng/mL and allowed to incubate at room temperature for 48 hrs. Plating to LB agar at the end of the treatment period allowed us to test sterility of the preparation. Any bacterial contamination was also obvious by color change of the pH indicator in the L15 media. Following treatment, single islets were mounted directly onto a slide with 5 μL of ProLong (Molecular Probes, Eugene, OR) and a coverslip. Each islet was imaged immediately after mounting on a confocal microscope and then analyzed as previously described for *in vivo* β-cells.

Preparation of primary zebrafish gastrointestinal cell culture

Gut dissections were performed, as described above, on 4-6 dpf GF Wild-type larval zebrafish, except that an effort was made to preserve other digestive tissues such as pancreas and liver. Dissected tissue of 45-60 larval zebrafish (per group/tube) was collected in 1mL Leibovitz's L15 media (Fisher Scientific, Waltham, MA) with Penicillin/Streptomycin and supplemented with 15% fetal bovine serum (Fisher Scientific, Waltham, MA). Tubes were gently spun down at 300xg for 3 min to remove the supernatant and the pellet was rinsed 2x with 1mL of 1X HBSS (Fisher Scientific, Waltham, MA). To gently dissociate the dissected tissue into individual cells or small cell clusters, 400 ul of 1X TrypLE (Thermo Fisher Scientific, Waltham, MA) warmed to 30°C was added and mixed with the pellet by pipetting every 5 min for 15 min. To stop the dissociation reaction, 1mL of cold 1% BSA/HBSS was added and the tubes were spun down at 300xg for 3min. The supernatant was removed and the pellet rinsed twice with 1% BSA/HBSS. The pellet of cells was then re-suspended in 200ul of 15% FBS in Leibovitz's L15 media with Penicillin/Streptomycin and aliquoted directly onto poly-D lysine coated coverslips (Corning, Corning, NY) in 24-well tissue culture plates.

BTC-6 Cell culture

βTC-6 cells (*43*) (ATCC, Manassas, VA) were cultured in low glucose DMEM (Dubelco's Modified Eagle's Medium) (Fisher Scientific, Waltham, MA) supplemented with 10% fetal bovine serum (FBS) (Fisher Scientific, Waltham, MA). Cells were incubated at 37° C and 5% CO_2. Media was changed every 24-48 hrs and cells were split every 4-7 days.

Protein treatments were added to the culture media of either βTC-6 or primary zebrafish cells at a final concentration of 0.22 μM. For imaging analysis, cells were split onto poly-D lysine coated coverslips (Corning, Corning, NY) in 24-well plates. Cells were given at least 24 hrs to attach to the coverslips before treating them with mNG-BefA[99] or mNG for several hrs. Following the incubation, the media was removed and each well gently rinsed twice with sterile PBS to wash away unbound protein before fixing with 4% PFA for 15 min at room temperature. Cells were permeabilized with 0.1% triton in PBS and then blocked in a solution of 3% BSA, 1% DMSO, and 0.01% Triton in PBS for 10 min. Cells were then antibody stained with guinea-pig anti insulin (1:500, Dako, Carpinteria, CA), Alexa Fluor 647 anti guinea-pig (1:1000 Thermo Fisher, Waltham, MA), and DAPI (1:1000, Thermo Fisher Scientific, Waltham, MA). Coverslips were mounted onto a microscope slide with ProLong Diamond mounting media (Molecular Probes, Eugene, OR) and imaged with a LEICA DM6 confocal microscope (Leica Microsystems Inc., Buffalo Grove, IL) to localize mNG. Images taken were randomized across each coverslip.

To localize where BefA bound to βTC-6 cells, cell fractionation followed by Western blotting was used. Either mCh-BefA[99], or mCh alone, was administered at a final concentration of 0.89 μM and 1.85 μM, respectively, to the βTC-6 cultures and incubated overnight. For analysis, the cell culture media was collected and concentrated to 1 mL. The cells were rinsed 3 times with 1 mL of cold PBS. All the PBS rinses were collected together and concentrated to 500

μL. Lysis buffer containing 0.01% Triton X-100 and protease cocktail inhibitor (Sigma-Aldrich, St. Louis, MO) was added to each culture dish and put into the -20°C freezer for 5 min. After thawing, the cell remnants were scraped from the bottom of the dish and collected in a clean centrifuge tube. Tubes were spun at 14,000 rpm for 15 min to separate the cytosolic components freed upon lysis from the cell debris. The cytosol within the supernatant and the remaining cell debris pellet were collected separately. The protein concentration for all samples (culture media, PBS rinses, cytosol, and pellet) was measured using the Pierce BCA Protein Assay Kit (Thermo Scientific, Rockford IL) and 20 μg of protein from each sample was run on an SDS PAGE gel and transferred to a PVDF membrane. The membrane was blocked in 5% BSA and then probed with a mouse monoclonal antibody to mCherry (Novus Biologicals, Littleton CO) at 1:1000 at 4°C overnight. To visualize the resulting bands, a secondary Anti-mouse IgG HRP-linked antibody was used (Cell Signaling Technology, Dancers MA) at 1:1000 for 1 hr at room temperature.

Primary mouse islet collection and culture.

Primary islets were isolated from Swiss Webster pups aged P12 as described (Huang et al., 2017). Additional purification was done by multiple rounds of hand-picking. Islets were allowed to recover overnight in RPMI supplemented with 10% FBS (Fisher Scientific, Waltham, MA) at 37° C and 5% CO_2. Before segregating into separate treatment wells for treatments, islets were screened and unhealthy or dying islets were removed. mNG-BefA[99] and mNG were each added at a final concentration of 250 μg/mL. Glucose was added as a positive control (Vetere et al., 2012) for beta-cell proliferation at a final concentration of 10 μM. Control wells received no additional treatment besides basal RPMI media supplemented with FBS. Islets were incubated for 48 hours, and were supplemented with 10mM EdU (5-ethyl-2'-deoxyuridine) (ThermoFisher

Scientific, Rockford, IL) during the final 6 hours of incubation to mark proliferating cells. Islets were fixed and processed as per the Click-iT® EdU Imaging Kit Protocol before mounting in ProLong Diamond Anitfade Mountant with DAPI (ThermoFisher, Rockford, IL). Islets were imaged using an Olympus IX71 microscope and total EdU labeling was quantified.

Amino acid alignment of SYLF containing proteins across kingdoms

Amino acid sequences of SYLF containing proteins obtained from NCBI were aligned using Clustal Omega (Goujon et al., 2010; Williams et al., 2013; Sievers et al., 2011).

Phylogenetic analysis

The amino acid sequence of the SYLF domain of BefA from *Aeromonas HM21*, starting at residue 99 and continuing to residue 261, was used to screen for homologs across microbial species using NCBI's BLASTp (Altschul et al., 1997) function. Default search parameters were maintained except to increase the maximum number of aligned sequences displayed to be 20,000, and the Expect Threshold was changed to 1.0. All standard scoring parameters were maintained. Homologs were identified as those with an amino acid sequence identity equal to or greater than 30%, and a query coverage equal to or greater than 80%. For phylogenetic analysis, each bacterial taxa with a putative SYLF domain-containing protein was represented by the hit of highest percent identity to the SYLF domain of *Aeromonas HM21* among isolates within that taxa. Amino acid sequences of putative SYLF domain-containing proteins were aligned by MUSCLE through Genious Prime 2020.1.2 and the predicted SYLF domain from each sequence was used in a subsequent MEGA 7.0 (Kumar et al., 2016) Alignment by MUSCLE (Edgar 2004). A rooted phylogenetic tree was constructed using Dendroscope 3 (Huson et al., 2012). Taxa habitats were manually curated by cross referencing literature.

Analysis of SYLF domain-containing protein lengths

A broad search of "SYLF" was performed in NCBI using Geneious Prime 2020.1.2. All Eukarytoic and Prokaryotic sequence files contianing "SYLF" annotations were downloaded. All sequence length data from the broad NCBI search, along with sequence length data from was exported into a comma separated value file for further analysis. To determine the distribution of sizes of SYLF domain containing proteins, a Frequency Distibution analysis was performed using GraphPad Prism version 6.07 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com) on sequence lengths from Eukaryotes, Prokaryotes, and the BefA homologs using a bin width of 25 amino acids. The frequency of sequences that were equal to or greater than 1000 amino acids were summed for each group (Eukaryotes, Prokaryotes, BefA homologs) and were represented in one bin. The percent of sequences in each bin was determined by dividing the frequency by the total number of sequences within each group. These percentages were then plotted using GraphPad Prism version 6.07 for Windows.

Statistical analysis

Appropriate sample sizes for all experiments were estimated *a priori* using a power of 80% and a significance level of 0.05. From preliminary experiments we estimated variance and effect. For larval β-cell quantification, these parameters suggested using a sample size of 6 in order to detect significant changes between treatment groups. Each of our experiments contained about 10–15 biological replicates or individual fish per treatment group, although some larger experiments had fewer biological replicates due to limited material, making our studies highly powered. Furthermore, entire experiments or technical replicates were repeated multiple times, resulting in pooled data sets of about 20–50 biological replicates. For mouse beta-cell analysis, our power calculations suggested using a minimal sample size of 4, each mouse being one biological replicate, to detect significant changes. Each of our experiments contained at least 5

116

mice, except for one which was limited to 3 due to extreme technical difficulty. Technical replicates of mouse experiments were carried out 1 to 3 times. These data are represented in the figures as box and whisker plots, which display the data median (line within the box), first and third quartiles (top and bottom of the box), and max and min of the data (whiskers). Plot of mouse data display all biological replicates in the set as dots overlaying the box plot. In box plots of zebrafish data, which contain far more biological replicates, box plot whiskers represent the 95% confidence interval, and any data point falling outside the 95% confidence interval is represented as a solid dot. Pooled data from multiple replicate experiments were analyzed through either the statistical software RStudio (RStudioTeam 2020) or GraphPad Prism 8. For experiments comparing just two differentially treated populations, a Student's t-test with equal variance assumptions was used. For experiments measuring a single variable with multiple treatment groups, a single factor ANOVA with post hoc means testing (Tukey) was utilized. Tukey results are displayed as lower-case letters above each treatment group on every box plot. A p-value of less than 0.05 was required to reject the null hypothesis that no difference existed between groups of data. In rare cases, extreme outliers were removed from data sets using the ROUT method for Robust regression and outlier removal developed in GraphPad PRISM (Motulsky and Brown, 2006), using a Q score of 1%.

Supplemental Figures

Figure S3.1 Amino-acid alignment and phylogenetic analysis of diverse SYLF domain-containing proteins. (A) ClustalW alignment of amino acid sequences for the SYLF domains of four different SYLF domain-containing proteins: SH3YL1, YSC84, BefA, and the BefA homolog from *K. aerogenes*. The percentage values in the top left at the beginning of the sequences, indicate conserved amino acid identity to BefA across the sequence region shown. The corresponding secondary structure of BefA is shown across the top of the alignment, starting at H4 in teal. Pink triangles denote the predicted start site of the SYLF domain based on sequence alone. Purple boxes show residues involved in actin binding. Black underlines indicate partial conservation between BefA and key functional residues from YSC84. The numbers flanking the BefA SYLF sequence denote amino acid number. (B&C) 3D schematics of the bacterial BPSL1445 NMR structure (blue) (B) and the predicted structures for the SYLF domains of yeast protein YSC84 (gold), and human protein SH3YL1 (purple). Each one is overlayed with full length BefA structure (green). (D) A relatedness tree of SYLF homologs across bacterial genera. Homologs are grouped by genus and each branch represents the most closely related homolog to BefA within a genus. Bold taxa are annotated to a higher taxonomic rank than genus. Taxa are highlighted based on the environments where members within that clade have been found: green = genera found in aquatic hosts, grey = genera found in aquatic environments, yellow = genera found in terrestrial hosts, and pink = genera found in human hosts, the original BefA sequence from Aeromonas HM21 SYLF query and its close homolog from ZOR0001 are not highlighted, but instead have pink arrows for genera found in human hosts.(E) The distribution of sequence lengths of putative SYLF-domain containing proteins is shifted in Eukaryotes compared to Prokaryotes. A relative frequency distribution (in percentages) of sequence lengths of proteins that contain an annotated SYLF-domain from Eukaryotes, Prokaryotes, and predicted BefA homologs.

118

Figure S3.2 Exploration of the biochemical interactions of BefA with F-actin and membrane lipids. (A) Coomasie Brilliant Blue denaturing protein gels illustrating results of F-actin co-pelleting assay. Ladder in lane 1 with band sizes in kDa labelled left. Expected band sizes for: BefA (29 kDa), BefA[99] (18 kDa), cortactin (80 kDa) and F-actin (42 kDa). + indicates the corresponding protein in left-hand label was added to the reaction shown in the lane immediately above. S = supernatant, P = pellet. (B) Top row, Membrane lipid strip binding assay. From Left to right: legend, representative strip incubated with BefA shows positive binding to phosphatidylserine (PS) and cardiolipin (CL), negative control performed with immunomodulatory bacterial protein His-AimA, point mutants created within individual predicted lipid binding sites (amino acid change indicated by label). Middle row, PIP lipid strip binding assay. From left to right: legend, representative strip incubated with BefA shows positive binding to PS, positive control was performed with PI(4,5)P$_2$ Grip and bound to PI(4,5)P$_2$ as expected (*denotes positive control of HRP secondary antibody spotted directly onto strip above Sphingosine-1-phosphate to confirm visualizing reagents were working), negative control with His-AimA. Bottom row, Membrane lipid array assay was performed with BefA with or without 1 mM Ca. (C) Concentration of total fluorescent liposomes with separate quantifications of total single and multi-vesicle structures, illustrating untreated groups do not form multi-vesicle structures.

119
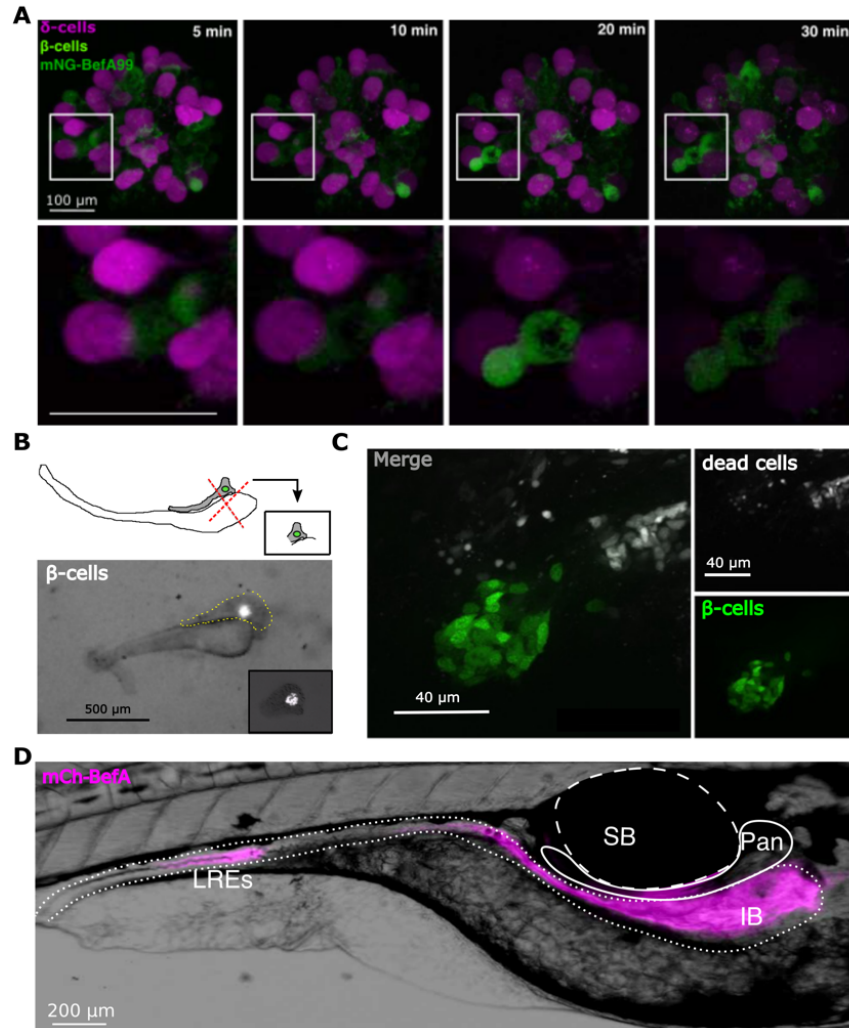
Figure S3.3 Tracking the activity of BefA in zebrafish. (A) Fluorescent images of a single primary zebrafish islet treated with mNG-BefA[99] over the course of 30 minutes. magenta = δ-cells, green = mNG and β-cells. Top row from left to right: whole islet images 5-,10-, 15-, 20- and 30-minutes post addition of mNG-BefA[99]. Scale bar = 100 μm. boxes outline inset of image at higher magnification in bottom row directly below. Scale bar = 100 μm. (B) Top: cartoon representation of a dissected larval intestine (white) with attached pancreas (grey) and primary islet indicated in green. Red dashed lines illustrate the approximate dissections to separate the head of the pancreas from the rest of the tissue. Bottom: greyscale image of an actual dissection. Pancreas (outlined in dashed yellow line) attached above the intestine, with the final resulting pancreatic head containing in-tact islet in lower right inset. Insulin promoter expressing β-cells in white. Scale bar = 500 μm. (C) Confocal image of β-cells in explant tissue after 48 hrs of ex-vivo culture. TOPRO staining denoting dead or dying cells in white (top left), insulin promoter expressing β-cells in green (bottom left). Merge = right. Scale bars = 40 μm. (D) DIC image of 4 dpf larval zebrafish following oral microgavage treatment with mCh-BefA. LREs = Lysozyme Rich Enterocytes of the distal intestine. For anatomical orientation: IB = intestinal bulb, SB = swim bladder, Pan = pancreas. Scale bar = 200 μm.
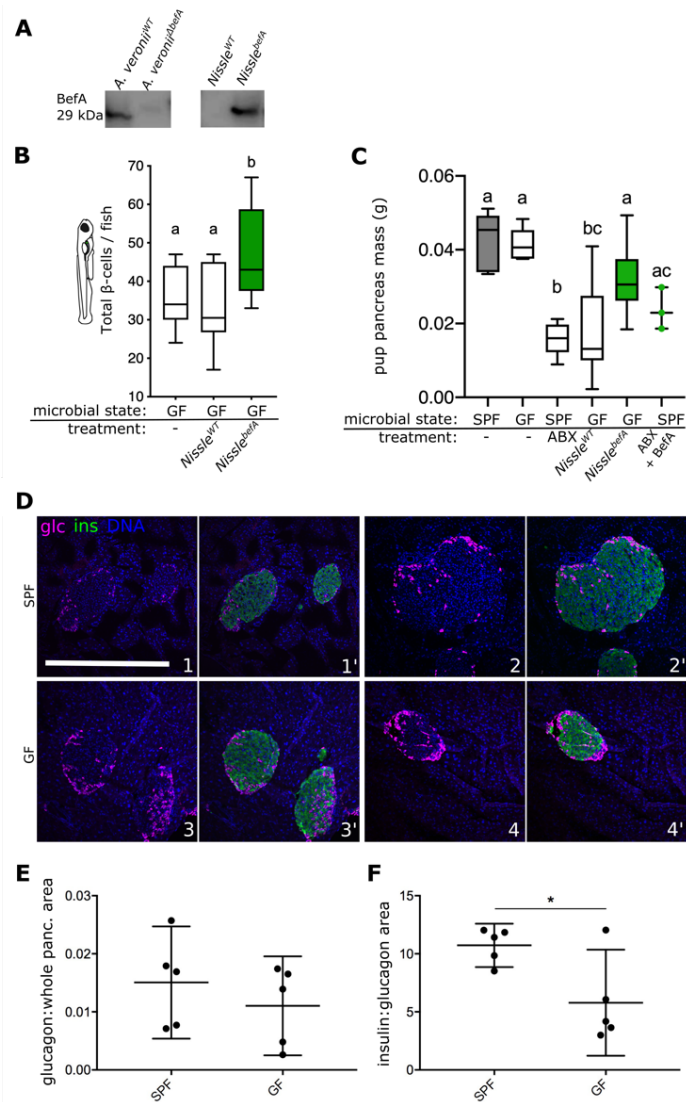
Figure S3.4 Activity of BefA is conserved in mice. (A) Western blots of anti-BefA antibody binding to CFS from indicated bacterial cultures. (B) Boxplots representing total beta-cell per GF larvae monoassociated with Nissle strains. (C) Boxplots illustrating total pancreas mass in grams (g) for neonatal P12 aged mice used in Figure 4A. (D) Representative images of immunofluorescence stained pancreatic tissue from four different Swiss Webster mice (1-4). Top row: SPF mice, Bottom row: GF mice. Magenta = Glucagon (glc) denoting alpha cells. Green = insulin (ins) denoting beta-cells (only in1'-4'), blue = DAPI denoting nuclei. Scale bar = 500 μm. (E) Average ratio of alpha-cell area to whole pancreas area per mouse. (F) Average ratio of alpha-cell area to beta-cell area per mouse.
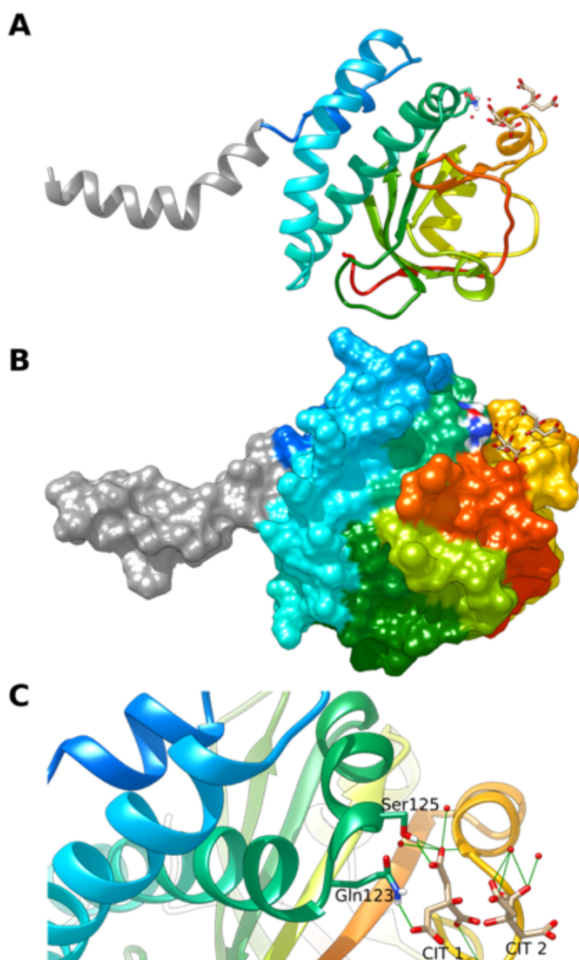
Figure S3.5 (Methods) Citric acid binding by BefA. Two citrate molecules are bound near Ser125 and Gln123. (A) Citrate was in the crystallization buffer at 100 mM and co-crystallized with BefA in the SYLF domain. Citrate molecules shown in tan and red sticks. Red is oxygen, tan is carbon. (B) Surface depiction of BefA bound to citrates. (C) Zoom of the two citrate molecules (CIT 1 & CIT 2) bound to BefA with potential hydrogen bonds to side chains, main chain and waters shown in green lines.

CHAPTER IV

CONCLUDING REMARKS

This chapter contains pre-published, co-authored material. Together, Dr. Karen Guillemin and I will use the ideas and content within this chapter to contribute a review piece for the *Annual Review of Genetics*, Volume 56. This review will expound on our experiments in chapter II and describe experimental data that is currently in preparation for publication that includes work from an undergraduate student I've mentored through the Knight Campus for Accelerating Scientific Impact at the University of Oregon, Robin Black. This data could also not have been collected without assistance and impressive team-work of the Lab Managers Elena Wall and Laura Taggart-Murphy from the Guillemin and Eisen Labs as well as the coordination and resourcefulness of Guillemin lab Zebrafish Wrangler Rose Sockol.

Introduction

Sometimes the most innovative discoveries in science come, not from trying to cure a disease, but instead from investigating how nature works in some of its smallest organisms. In the late 19th century, the 'father of innate immunity' Elie Metchnikoff earned a Nobel Prize from his observations of immune cell phagocytosis in the transparent larvae of starfish, later connecting these findings to mechanisms behind human host immunity. Recognized as the one of the greatest advances in therapeutic medicine, Alexander Fleming discovered the antibiotic properties of penicillin in the 1920s by observing the lack of *Staphylococcus* growth surrounding a mound of mold on his petri dishes, also earning a Nobel Prize. In the 20th century, genes discovered in bacteria would become the first targets for anti-tumor therapies for cancer. Further, in 2020 Drs. Jennifer Doudna and Emmanuelle Charpentier were awarded the Nobel Prize for

their discovery of CRISPR gene editing which has revolutionized basic science research as well as approaches to new cancer therapies and curing genetic diseases. In their research, Doudna and Charpentier did not outright seek to develop a pioneering gene editing technology, but instead were investigating how the bacterial immune system defends themselves against viruses. These examples illustrate how investigating basic mechanisms in microscopic life can illuminate our understanding of biology across animal species and lead to innovation.

At first glance, it may not seem like investigating the underlying biology of organisms as small as bacteria could also have implications in understanding human biology. But all animal life has evolved from microbial cells which populated the earth as the first living organisms billions of years ago. Multicellular organisms have always co-existed with microbes and in these intricate interactions, animals were required to evolve strategies to compete or tolerate microbial life, all based on the 'microbial-set-rules' of biology. Mounting studies are illustrating that the microbes that live on and within animals, ie their microbiota, are critical for host homeostasis beginning with their development into mature organisms and sustaining these relationships throughout life. Although the mechanisms behind the roles of the microbiota on animal biology are largely obscure, it is clear that our *conscious* lessons from microbes, have only just begun.

This review will focus on what is known about the important interactions between host organisms and their microbiota within the model vertebrate *Danio rerio*, the zebrafish. Larval zebrafish are an excellent model organism to study host-microbe interactions. One important advantage is their *ex utero* development which enables easy surface sterilization of the outer chorion for deriving germ-free zebrafish, that lack their microbiota (Melancon et al. 2017).

Associated isolates of the complex zebrafish microbiota are well-known, including their draft genome sequences allowing mechanistic gnotobiotic experimentation (Stephens et al. 2016). At the genetic level, zebrafish have high genome conservation with mammals including 70% of human genes and 84% of human disease-causing genes. Further, combining their small size with single cell RNA sequencing, larval zebrafish enable a whole-vertebrate-organism approach to understanding cell-type identity at the transcriptional level across their development (Farnsworth, Saunders, and Miller 2019) and examine which host cell types respond to the presence of the microbiota (Massaquoi et al., *submitted*). In addition to highlighting the pioneering work that explores the roles of the microbiota in zebrafish homeostasis *in vivo*, we will expound upon our insights gained by surveying which cell types respond to the presence of the microbiota and cataloguing how those cells responded at single cell resolution.

Synchrony of Microbiota Assembly & Host Development

The brilliant work accomplished in developmental biology has primarily focused on the intrinsic genetic controls and canonical signaling networks that drives animal morphogenesis into functionally mature organisms. However, in most cases the assembly of a host's microbiota occurs during development, redefining the organism from an individual to an ecosystem. The embryonic stages of zebrafish development take place within the sterile, translucent chorion. At ~3 days post fertilization (dpf), larvae hatch, marking their first opportunity for exposure to microbial cells from the outside the environment. The host intestine harbors the highest density and diversity of resident microbiota within zebrafish, like all vertebrates. From 52-76 hours post fertilization (hpf), the larval intestine establishes a continuous lumen with distinct compartmentalization of the digestive organ into proximal, mid, and distal regions established from 5-6 dpf (Ng et al. 2005; Lickwar et al. 2017). The timeline of intestinal development into

mature functioning epithelia coincides not only with the initial inoculation of the larvae's first resident microbes but also their dramatic microbial expansion in the lumen from 4-6 dpf (Bates et al. 2006; Jemielita et al. 2014; Stephens et al. 2016).

Establishment of the microbiota during critical windows of development can have life-long consequences for the host. In neonate mice, the coordination of hair follicle morphogenesis and colonization of commensal microbes induces the migration and activation of T cells in the hair follicle during a critical window of development (Scharschmidt et al. 2017). The recruitment of these regulatory T cells enables the host to discriminate between pathogenic versus commensal bacteria on the skin. Preventing the migration of these regulatory T cells during this critical window of development prevents tolerance to commensals, not rescued by bacterial colonization in adult life (Scharschmidt et al. 2015; Leech et al. 2019). Conventionalization of germ-free larvae with a complex microbiota at 7dpf fails to rescue zebrafish growth and several behavioral patterns measured at 14dpf (Bruckner et al., BioRx), implicating that interactions with the microbiota during early development can have irreversible consequences in the establishment of the gut-brain-axis and host metabolism. Further, germ-free zebrafish colonized at 8 dpf showed significantly higher blood glucose levels a year later in female adults even though all larvae were re-introduced to the main zebrafish facility with community-shared water (Fig4.1) (Massaquoi & Black, *unpublished*). Together, these findings show that interactions with microbes during early critical windows of development can have life-long consequences for the host. We predict that competency to the microbiota during early development, when host cells are most malleable in their cell-fate-decisions, could be decisive not only for training the host immune system, but also for host metabolic imprinting (Tremaroli and Bäckhed 2012; Reinhardt, Reigstad, and Bäckhed 2009) and long-term cell-type specific functionality. This section will
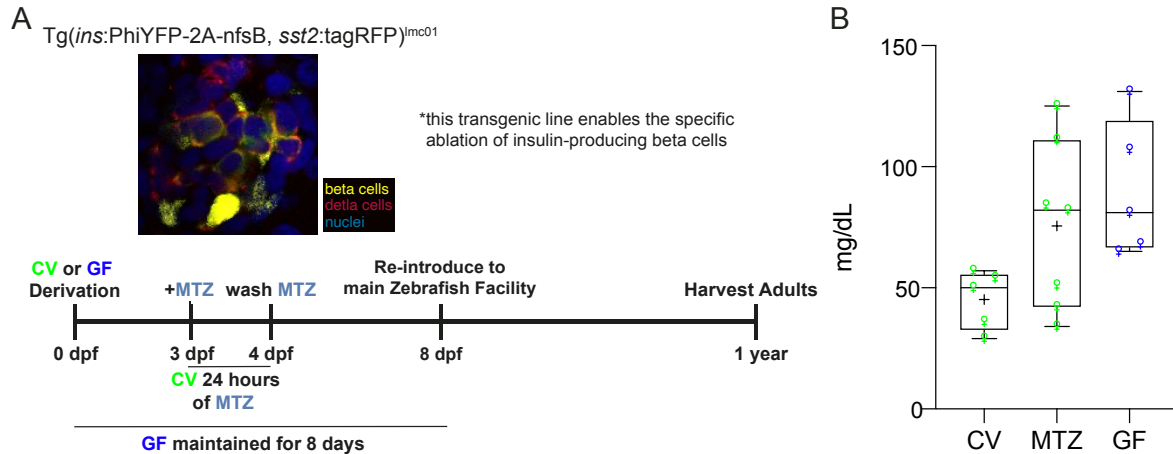
Figure 4.1 Ablation of beta cells and exposure to microbes early in life have long-term consequences in adulthood. A) The experimental design allows specific ablation of beta cells in transgenic zebrafish that use the insulin promoter to express nitro reductase (nfsB) specifically within beta cells. The administration of metronidazole (MTZ) is metabolized by nfsB into toxic metabolites that induce cell death. B) The boxplot shows the concentration of blood glucose within female adult zebrafish a year after exposure to MTZ for 24 hours during larval stages of development under conventional conditions (CV) or germ-free (GF) derivation that was maintained for 8 days post fertilization (dpf). In the boxplot of this figure, and following figures, the middle line represents the median, the top and bottom of the box representing the upper and lower quartile, the whiskers representing the min and max values, and the '+' symbol representing the mean. Each symbol represents an adult zebrafish.

review different strategies microbes employ to colonize the host zebrafish and what is known

about how the host perceives these different microbial cues during their development.

Attendance & Occupation: inevitable cues from the assembling microbiota

Although the world is brimming with microbes, only selective residents assemble within

animal hosts (Thompson et al. 2017), best observed by the exclusive colonization of *Vibrio*

within the squid light-organ (Mcfall-Ngai 2014). Isolates of *Aeromonas* experimentally evolved

to colonize the larval zebrafish illustrate that adaptive strategies of the bacterium to immigrate

from the local environment can play an important role in the establishment and maintenance of

the zebrafish microbiota (Robinson et al. 2018). Once they are within the intestine, microbes

employ different strategies to maintain their establishment by increasing their burden (Jemielita

et al. 2014), competing with other microbes for resources (Stecher and Hardt 2011) or adapting

to constraints of the host environment (Wiles et al. 2016; Stagaman et al. 2017; Rawls et al.

2006). Different challenges facing the microbiota within the intestine include intestinal

contractions and luminal pH under the control of the enteric nervous system (Rolig et al. 2017)

(Hamilton et al., 2021 *accepted*), maintenance of the intestinal barrier and mucus secretion, and

activity of the immune system (Murdoch and Rawls 2019). An important strategy, echoing the

selective pressure of immigration for host colonization, is the advantage of microbial motility

(Stephens et al. 2016). A proinflammatory *Vibrio* symbiont native to zebrafish uses swimming

motility and chemotaxis to counteract intestinal expulsion, promoting their persistence and

spatial organization within the intestine (Wiles et al. 2020). Another strategy that enables

retention within the intestinal lumen is through microbial-associated competitive activities

(MACAs), which includes a repertoire of largely un-characterized microbial-derived bioactive

molecules that enable competition with other microbes and/or host cells (Fischbach and Segre

2016; Wiles and Guillemin 2020). Examples of MACAs can include bacterial-secreted products

such as anti-microbial peptides or pore-forming proteins which have characteristic features of

non-discriminately compromising cell membrane integrity. Other examples of MACAs include

nutrient salvaging molecules like siderophores or degradative enzymes that breakdown essential

macromolecules (Kramer, Özkaya, and Kümmerli 2019). Host perception and response to

MACAs can indirectly shape the microbiota community structure. A zebrafish isolate of

Shewanella, or its secreted factors alone, potently suppresses host immune responses by

decreased intestinal influx of neutrophils, a type of white blood cell indicative of inflammation

(Rolig et al. 2015). In addition to bacterial motility and MACA strategies to compete in the

luminal space, successfully colonizing microbes are required to adjust to the nutritional

landscape within the intestine. Nutritional resources for the microbiota primarily come from the

host diet and host-produced mucins. Metabolites derived from microbial catabolism significantly

contribute to community attendance. The digestion of a complex carbon source into metabolites

from one type of bacteria can provide otherwise not accessible nutrients for other types of bacteria, permitting a habitable lumen for the former. Additionally, metabolites derived from microbial catabolism can impact the overall intestinal environment via the host. Bacterial-fermentation of dietary fiber into short chain fatty acids (SCFA) promotes intestinal barrier function (Koh et al. 2016). Zebrafish isolate *Edwardsiella tarda* produces indole derivatives that modulates the microbiota community by regulating host intestinal motility (Ye et al. 2020). All of these factors and strategies employed by residents of the microbiota to compete within the intestinal lumen illustrate a diversity of inevitable microbial cues that have the potential to communicate the context of the community and the outside environment to the host.

Perception & Reaction: signaling mechanisms that allow the host to discern microbial cues during development

The innate immune system refers to ancient defense programs employed by the host. The title 'innate' describes the host's inborn defense mechanisms that include primitive strategies such as physical protective barriers and secretion of mucus. Myeloid cells of the innate immune system are constantly monitoring the microbiota and are first responders to compromised host cells where they can mount local or systemic inflammatory responses and elicit phagocytosis of microbes or dying host cells. The presence of the microbiota is necessary for the recruitment of macrophages and neutrophils to the intestine (Rolig et al. 2015; Koch et al. 2019). The innate immune system also encompasses the broad expression of sensory receptors that have evolved to recognize microbial cues. Together, these host defense strategies included in the innate immune system are well poised to perceive and relay information about the microbiota and their cues during early host development.

Macrophages and neutrophils are both innate immune myeloid cells that have a repertoire of responses to microbial cues including chemotaxis/recruitment to compromised tissues, phagocytosis, and secretion of proinflammatory cytokines and digestive enzymes. In the context of some pathogenic infections, the expression of inflammatory signaling proteins at distant tissues are a result of indirect responses to bacterial products. For example, cytokines produced locally at the site of injury systemically activate immune responses. This demonstrates a mechanism that enables the perception and response to microbial cues in tissues and cells distant from the microbiota. Displaying that organs both local and distant to the microbiota are responsive to microbial cues, bacterial motility in zebrafish larvae induces the systemic expression of tumor necrosis factor alpha (*tnfa*), a proinflammatory cytokine, within gut-associated macrophages and the liver (Wiles et al. 2020). Wiles and colleagues report a dichotomy in the source of cells expressing *tnfa*: in the intestinal environment, the majority of macrophages expressed *tnfa* but within the liver the enriched expression of *tnfa* did not come from the resident macrophages or neutrophils. This data illustrates a holistic response to a single microbial cue where it is perceived by different host tissues and elicits similar responses through different cell types in different locations. In addition to the sensing of microbes or damaged host cells by appointed myeloid cells, innate immune sensing of microbial cues is broadly implemented across cell types including epithelia.

A well-characterized mechanism behind innate immune host sensing of microbial cues is through Toll-like family receptor (TLRs) signaling (Murdoch and Rawls 2019). TLRs are part of an ancient family of pattern recognition family of receptors and their activation by generic microbial products, including as lipopolysaccharide (LPS) and flagellin, regulates the immune response (Jault, Pichon, and Chluba 2004; Deguine and Barton 2014). LPS is a prominent

component of cell walls in Gram-negative bacteria and flagellin proteins assemble together for bacterial flagellum during bacterial motility. The examples of TLR activation by LPS or flagellin illustrates the competency of host cells to bacterial colonization strategies. The downstream response of TLR activation is partly dictated by the recruitment of intracellular adaptor proteins, including myeloid differentiation primary response 88 (MyD88) within zebrafish (Hall et al. 2009; Jault, Pichon, and Chluba 2004; Meijer et al. 2004). The zebrafish genome has a single copy of *myd88* that has conserved function of regulating innate immune function. Several studies have shown the necessity of Myd88 in larval zebrafish intestinal epithelial maturation and secretory cell-fate specification, demonstrating that the sensing of microbial cues by the innate immune system plays important roles in digestive system development (Bates et al. 2006, 2007; Cheesman et al., 2011). In the absence of microbial cues, the intestinal epithelia of germ-free larval zebrafish display identical phenotypes to Myd88-deficient larvae including a paucity of proliferative epithelia (Rawls, Samuel, and Gordon 2004), mucus-secreting goblet cells, and hormone-secreting enteroendocrine cells (Bates et al. 2006). Germ-free zebrafish also display a marked decrease in alkaline phosphatase (AP) expression, a marker of enterocyte maturation within the intestine (Rawls, Samuel, and Gordon 2004; Bates et al. 2007). The enzymatic activity of AP is dependent on the presence of Myd88 and promotes tolerance to the microbiota by dephosphorylation of LPS, rendering it less proinflammatory (Bates et al. 2007). Together, these works established a convergence of the innate immune system and host tissue development as their epistasis experiments revealed that TLR/Myd88 sensing of the microbiota is upstream of canonical developmental signaling pathways including activation of Wnt signaling for epithelial proliferation and inhibition of Notch signaling for secretory cell fate (Bates et al. 2006; Cheesman et al., 2011). Interestingly, mono-association of germ-free larvae rescued the paucity

of all epithelial phenotypes but the administration of LPS alone only rescued the intestinal

expression of AP (Bates et al. 2006; Cheesman et al., 2011). The differential responses to LPS

that contrasts the response to mono-associated bacteria (that include LPS in their cell walls)

suggest that multiple cues from the microbiota non-redundantly activate innate immune sensing

pathways and impact diverse aspects of host development. In addition to detecting cell structure

components of microbes, innate immune sensing can play a role in the detection of secreted

MACAs.

First coined in 2020, MACAs encompass the fleet of activities displayed by microbial

cells that enable them to compete with other microbes and the host (Wiles and Guillemin 2020).

One example of a MACA perceived by zebrafish larvae is the *Aeromonas*-secrete protein,

Aeromonas immune modulator A (AimA), which decreases the influx of neutrophils and

proinflammatory cytokine expression within the intestine (Rolig et al. 2018). Bacterial Aim

proteins have structural similarity with vertebrate immunomodulatory Lipocalin proteins (Lcns)

whose properties bind hydrophobic ligands such as prostaglandins and retinols. The *lcn* gene

family contains 13 members within zebrafish and its well characterized that mouse LCN2s are

hallmarks of inflammation that inhibit bacterial growth, create local hypoferremia and bind

bacterial siderophores that scavenge iron (Golonka, Yeoh, and Vijay-Kumar 2019; Garcia-Estañ

et al. 2021). Aim proteins do not bind siderophores directly, but they do bind and alter the

function of Lcns. Secretion of Aim proteins are considered MACAs because it confers a

colonization advantage to *Aeromonas*. This property is likely not limited to its modulation of

host inflammation, but also from other Aim protein activities such as those that counteract mouse

LCN2-mediated iron-starvation while simultaneously accentuating mouse LCN2's inhibition of

Escherichia coli growth (Rolig et al. 2018). Subsequently, host expression of Myd88 is required

for the AimA-induced attenuated neutrophil response illustrating innate immune sensing of MACAs. Although it's not fully understood which host cells directly perceive AimA and pacify the immune response, several *lcn* genes have been observed within larval neutrophils, liver and epithelia at single cell resolution (Massaquoi et al., *submitted*). Together, Aim proteins illuminate how MACAs permit colonization advantages while inevitably providing a cue to the host, that when perceived, promotes tolerance to the microbiota. The host also has the capacity to perceive different types of MACAs without innate immune sensing.

Another strategy included in MACAs are the secretion of proteins or peptides that have anti-microbial properties. A trademark feature of anti-microbials is their inherent property of disrupting cell membranes, in both pro- and eukaryotes. Host cells have evolved many strategies to combat perturbations of their cell membranes from toxins or pore-forming units (Rühl et al. 2018; Reiss and Bhakdi 2012; Swanson, Deng, and Ting 2019). Recent work has discovered that the membrane perturbing property of bacterial-derived protein, beta cell expansion factor A (BefA), is necessary for inducing the proliferation of beta cells in germ-free mouse neo-nates (Hill et al., 2022 *in review*). Initially discovered as a secreted product from *Aeromonas*, BefA is necessary and sufficient to rescue the expansion of insulin-producing beta cells in developing germ-free zebrafish and can act directly on pancreas tissue (Hill et al. 2016, 2022 *in review*). The ability of BefA to rescue beta cell expansion does not require Myd88, suggesting that host competency to this MACA is independent of innate immune sensing. Similar to BefA, the administration of pore-forming regenerating (Reg) proteins, rescues the development of beta cells in germ-free larvae and neo-nates (Hill et al., 2022 *in review*). Reg proteins were first discovered by their abundance in regenerating pancreatic tissue but were later characterized as pore-forming anti-microbials secreted by host enterocytes (Watanabe et al. 1994; Vaishnava et

al. 2008; Hooper, Littman, and Macpherson 2012). The common host response of beta cell proliferation from treatment with either BefA or Reg proteins shows that generic perturbations of the host cell membrane can play important roles in host development, even in tissues distant to the intestine. Exploring how BefA disseminates within the zebrafish larvae illustrates that MACAs have the potential to interact with many cell types throughout the body, not limited to host tissues directly at the host-microbe interfaces (Hill et al., 2022 *in review*). Concurrent with the secretion of different MACAs within the developing host intestine, luminal digestion of proteins is delayed before mammalian weaning, allowing for the persistence of maternal antibodies and antigens to help train the developing immune system (Kulkarni and Newberry 2019). Suckling mice and larval zebrafish have specialized epithelia within the mid-intestine that promiscuously absorb luminal proteins for intracellular digestion (Park et al. 2019). The mid-intestine is distinguished with highly vacuolated lysosomal-rich enterocytes (LREs) and protein absorption from the LREs is critical for vertebrate growth and survival (Park et al. 2019). Interestingly, fluorescent proteins absorbed from the LREs were also located in distant larval organs including the kidney and liver signifying that if MACA proteins are also absorbed by LREs, they could interact with diverse cell types via the circulatory system. Although its anticipated that BefA naturally accesses the pancreas through retrograde transport via the pancreatic duct, it is also strongly absorbed by the LREs and administration of BefA into the circulatory system is sufficient to induce beta cell expansion (Hill et al., 2022 *in review*). Direct cellular interactions with MACAs that have membrane perturbing properties is an example of an inclusive microbial cue that virtually all host cells can respond too. In addition to membrane perturbation, host cells can perceive microbial cues with their transmembrane signaling receptors and ion channels.

Found only in eukaryotic cells, G protein-coupled receptors (Gpcr) are cell surface receptors that detect a diversity of ligands and stimulate cellular activation. Metabolites from microbial digestion can act as Gpcr ligands. The best characterized microbial-derived metabolites are SCFAs from the fermentation of dietary fiber, that include acetate, propionate and butyrate. Through direct activation of Gpcr, SCFAs impact a breadth of widespread physiological processes within the host ranging from maintenance of intestinal barrier function to nervous system activity (Koh et al. 2016). The intestinal microbiota of adult zebrafish include residents that can synthesize the main three SCFAs and host perception of butyrate has conserved anti-inflammatory effects during wound repair (Cholan et al. 2020). These experiments demonstrated that the immersion of larvae in butyrate attenuated neutrophil and macrophage influx to the wound site and decreased macrophage *tnfa* expression but the G protein-coupled receptor, hydroxycarboxylic acid receptor 1 (*hcar1*), was necessary for neutrophil reduction. Further, Enriched levels of SCFAs within zebrafish larvae is associated with an up-regulation of genes related to glucose level reduction (Falcinelli et al. 2016), illustrating connections between microbial-derived metabolites and host metabolism.

SCFAs also activate G protein coupled receptors on enteroendocrine cells (EECs) (Lu, Gribble, and Reimann 2018). The EECs of the intestine are specialized epithelia that detect stimuli from the lumen and respond by the secretion of hormones or neurotransmitters. EECs can regulate the local intestinal environment or they can communicate with tissues distant to the intestine in the case of systemic hormone secretion and activation of the enteric nervous system (ENS) (Furness et al. 2014). The ENS is part of the Peripheral Nervous System (PNS) and regulates many aspects of intestinal homeostasis including motility, by stimulating the contractions of the smooth muscles that surround the intestine. Recent work has established

direct connections between EECs and enteric neurons (Bohórquez et al. 2014, 2015; Kaelberer et al. 2018). Further, the Vagus nerve of the PNS facilitates the coordination between EECs, the ENS, and neurons of the central nervous system (CNS) enabling a bi-directional route of communication between the luminal microbiota and the brain. In addition to detecting SCFA stimuli, the catabolism of tryptophan by *Edwardsiella tarda* produces indole derivatives that specifically activate transient receptor potential ankyrin A1 (Trpa1) on subsets of EECs within the larval intestinal epithelia (Ye et al., 2020). Trpa1 is an excitatory calcium-permeable cation channel involved in the perception of inflammatory pain (Bautista et al. 2006; Lapointe and Altier 2011). Activation of Trpa1 signaling by indole derivatives fosters intestinal motility by activation of cholinergic enteric neurons and in the absence of *trpa1* conventional zebrafish larvae showed an overgrowth of *E. tarda* (Ye et al., 2020). These experiments demonstrate that direct perception of microbial metabolites by host sensory epithelia facilitate clearance of specific types of bacteria. Further, microbial stimulation of Trpa1 on EECs activates vagal sensory ganglia, suggesting that microbial derived metabolites from tryptophan catabolism may be a cue for communication within the gut-brain-axis (Ye et al., 2020).

The different strategies that microbes elicit to colonize and compete within the host has driven the evolution of host mechanisms to detect and respond to those microbial cues. Although the gross morphological features are consistent between conventional and germ-free zebrafish (Rawls, Samuel, and Gordon 2004), competency to the presence of the microbiota and their MACAs can have profound impacts in cell type specification and function at the single cell level, not limited to host cells directly at the host-microbe interface. The works highlighted in this section demonstrate different mechanisms by which the host perceives and responds to the

microbiota as we continue to learn which host cell types monitor different microbial cues. The next section of this review expands on our previous work that compares the transcriptional profiles of gnotobiotic zebrafish larvae at single cell resolution to understanding which cell types across the entire organism perceive the microbiota and describe the ways in which those cells respond.

Transcriptional Responses to the Microbiota Across Zebrafish Cell Types

The small size of zebrafish larvae in combination with their gnotobiotic tractability allows an unparalleled opportunity to understand the role of the microbiota on host development using single cell RNA transcriptomics. We have built a whole organism gnotobiotic atlas resource that allows researchers to explore the transcriptional differences between conventionalized and germ-free larvae across the whole organism (Massaquoi et al., *submitted*). A diversity of cell types respond to the presence of the microbiota and the responses are diverse in tissue-type specific ways. There are two main themes across the host cellular responses to microbes that we will discuss further; first, the microbiota induced the enriched expression of genes associated with adenosine triple phosphaote (ATP)/nucleotide metabolism in a diversity of cell types. Second, the cells from germ-free derived zebrafish have a ubiquitous increase in the expression of *crystallin* genes across all cells.

ATP as a universal language between hosts and microbes

ATP is an ancient molecule that has propelled the evolution of early biology. The most famous example of host-microbe co-evolution is in the Endosymbiotic Theory of mitochondria but even before this strategic alliance that led to mass production of ATP within eukaryotes, bacteria evolved ATP-degrading enzymes and release mechanisms. The first receptors evolved to detect extracellular ATP were in single cell eukaryotic protozoa (Verkhratsky and Burnstock

137

2014). At single cell resolution, our data illustrate that a diversity of cells respond to the presence of the microbiota by an enrichment of genes involved in ATP/nucleotide metabolism ranging from cells of the digestive system to neurons of the CNS (Massaquoi et al., *submitted*). Although different cell types have specific functions, the energy gained by de-phosphorylation of ATP to ADP is essential for most cellular enzymatic reactions. Therefore, the consistent enrichment of genes involved in ATP/nucleotide metabolism across diverse cell types likely reflects the tissue-specific status or function of a given cell, which was observed for different cell types by gene ontology analysis. Re-analyzing cells from conventionalized larvae based solely on ~200 ATP/nucleotide metabolism genes found within the data showed that host cells largely maintained their original tissue-type specific segregation, suggesting that the microbiota induce a similar suite of genes that host cells employ for specific responses (Massaquoi et al., *submitted*). Subsequently, measuring the total concentration of ATP within whole larvae displayed consistent levels of ATP between conventional and germ-free larvae but the production and turnover of ATP was significantly increased within the intestines of conventional zebrafish (Massaquoi., et al submitted). These experiments suggest discreet differences in ATP/nucleotide metabolism depending on the context of the tissue type and that the microbiota play a role in homeostatic levels of ATP concentration within the lumen.

Cellular energetic demands aside, ATP also acts as a signaling molecule and most cell types have the capacity to detect extracellular ATP (Junger, 2011). Enabling bi-directional communication between neurons and glia, ATP is co-released with neurotransmitters (Verderio and Matteoli 2011; Huang, Otrokocsi, and Sperlágh 2019). The production of ATP and nucleotides is used for paracrine signaling in many tissues including the liver and pancreas, illustrating that perception of extracellular ATP plays important roles in tissue specific functions

138

and host metabolism (Burnstock, Vaughn, and Robson, 2014; Bours et al. 2006; Burnstock and

Knight 2004; Novak 2003; Sørensen and Novak 2001). Between host and microbial cells, ATP is

a potent facilitator of communication within intestinal lumen (Inami, Kiyono, and Kurashima

2018) and fluxes of luminal ATP concentrations can act as a microbial cue for host perception.

Bacteria release ATP during their growth phases (Mempin et al. 2013; Hironaka et al. 2013) and

from perturbations, such as via MACAs, that compromise the bacterial-cell membrane.

Intracellular ATP concentration is several orders of magnitude higher than within the

extracellular space and subtle increases of extracellular ATP acts as a distress signal from

damaged or dying cells to the immune system (Bours et al. 2006; Chen et al. 2006; Mempin et al.

2013; Akopova et al. 2012; Roberts et al. 2012; Spari and Beldi 2020). At single cell resolution,

the data indicate that neutrophils transcriptionally respond to the microbiota by an enrichment of

genes involved in ATP metabolism as well as transcriptional changes involved in cell shape

changes for migration and chemotaxis (Massaquoi et al., submitted). These transcriptional

responses within neutrophils of conventionalized larvae is consistent with their well-

characterized mechanisms of activation where these cells detect a stimulus, such as extracellular

ATP, induce autocrine ATP release to guide their chemotaxis, and produce endogenous ATP to

sustain their functional response (Chen et al. 2006; Bao et al. 2014). To understand the extent to

which extracellular ATP within the lumen acts as a microbial cue perceived by the developing

host, we challenged gnotobiotic zebrafish larvae with exogenous $ATP_\gamma S$, which has decreased

rates of hydrolyzation (Bagshaw 2020; Chen et al. 2006), and quantified the number of

neutrophils recruited to the intestine 4 hours later (Fig4.2A; $F_{(7, 289)}$=4.891; p<0.0001).

Reproducible with previous work (Rolig et al. 2015) CV zebrafish (9.3 neutrophils) on average

recruited significantly more neutrophils to the intestine compared to both GF (4.5 neutrophils)

and GF+vehicle (VEH) (4.2 neutrophils) larvae (p<0.0001). Although not statistically different compared to VEH, all gavage doses: 10mM (6 neutrophils), 25mM (5.6 neutrophils), 40mM (5.3 neutrophils), 50mM (6.1 neutrophils) of ATP$_\gamma$S showed a consistent increase in neutrophil recruitment with 50mM ATP$_\gamma$S indicating a partial rescue not statistically different from CV. Although the changes in neutrophil recruitment in response to ATP$_\gamma$S are subtle, this trend was consistent across numerous blinded replicate experiments despite the natural high variability of the data and these trends held by adding ATP$_\gamma$S directly to the flask water (data not shown). The data illustrates that developing zebrafish larvae can perceive extremely subtle changes of extracellular ATP within the environment, suggesting an additional mechanism by which the host perceives and responds to microbial-induced changes in the intestinal lumen (Fig4.2B).
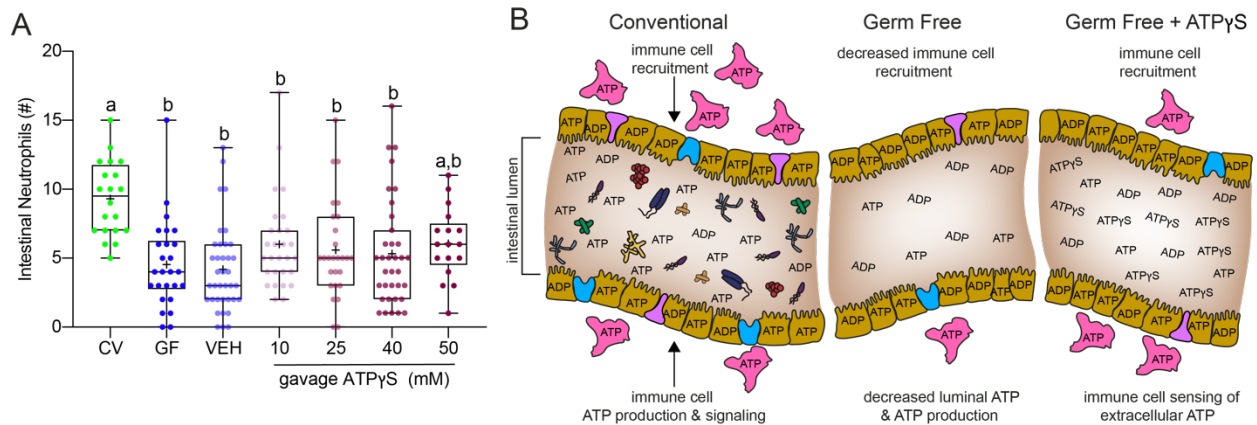
Figure 4.2 Neutrophils respond to subtle changes in luminal ATP levels. A) The boxplot shows the number of neutrophils recruited to the intestine in conventional (CV), germ free (GF) derived zebrafish and larvae micro gavaged with control (VEH) or doses of ATP$_\gamma$S. A One Way Analysis of Variance (ANOVA) was used to determine if variation within the data could be attributed to the experimental groups (p < 0.001) and statistical differences between the groups were tested with Tukey-Kramer Post Hoc analyses. Non-matching letters above the groups demonstrates statistically different means (p <0.05). B) The schematic illustrates that in a conventional zebrafish larvae intestine, neutrophils are recruited to the epithelia in which the lumen contains microbes and extracellular ATP. In a germ free intestine, the lack of microbes results in lower turnover of ATP and decreased recruitment of neutrophils. The recruitment of neutrophils in germ free intestines can be partially rescued with the administration of ATP$_\gamma$S.

Together, the transcriptional changes in conventionalized larvae at single cell resolution and *in vivo* experiments support the hypothesis that despite their proximity to host-microbe interfaces, several host cell types perceive and respond to the microbiota. Not only is ATP

critical for its energy currency but can also mediate diverse messages of communication between host and microbial cells. Further, microbiota-induced enrichment of genes involved in ATP/nucleotide metabolism across diverse cell types illustrates how similar genes can be employed for different host cell-type specific functions and how the microbiota globally influences host metabolism. In addition to an enrichment of genes involved in ATP metabolism across cell types in conventionalized zebrafish, the gene ontology analysis based on enriched transcripts also demonstrated that the microbiota promote plasticity in tissue function and development versus germ-free larvae, best described within the exocrine pancreas (Massaquoi et al., *submitted*). Proper metabolism is critical during host development as cells are required to grow and function simultaneously, a balance severely impacted at the transcriptional level by the lack of microbial cues in germ-free derived larvae.

Cells from germ-free larvae are in a synthetic and metabolically inactive state

The use of gnotobiotic systems have significantly contributed to our understanding of host-microbe interactions from discovering the atomic crystal structure of novel MACA proteins (Rolig et al. 2018, Hill et al., 2022 *in review*) to understanding microbiota population dynamics within the host (Stephens et al. 2016; Burns et al. 2016). However, it's important to be mindful that germ-free derived organisms represent a synthetic biological state. In conventional organisms the microbiota plays invaluable roles in host metabolism including the assistance in food digestion, providing exogenous nutrition, regulating intestinal absorption and regulating fatty acid metabolism (Magnúsdóttir et al. 2015; Walther et al. 2013; Orlando et al. 2014; Semova et al. 2012). We suspect that the lack of metabolic input by the microbiota led to the surprising observation of *crystallin* transcript and protein expression across the germ-free larval body (Massaquoi et al., *submitted*). Crystallin constitute a heterogenous and highly stable family

of proteins, with the zebrafish genome containing alpha, beta and gamma *crystallin* genes

(Farnsworth et al., 2021). Although crystallin are well-known as water-soluble proteins that

make up the vertebrate lens, related *crystallin* genes were present in early vertebrates before the

evolution of the lens suggesting that they have other cellular functions (Zigler and Sinha 2015).

Interestingly, many 'taxon-specific' crystallin have been discovered to be identical to metabolic

enzymes which maintained their ancestral roles within several cell types while being recruited to

serve as crystallin (Slingsby and Wistow 2014). Alpha crystallin was evolutionarily recruited

from stress proteins, where they function as chaperones to prevent aggregation during

environmental stress (Sprague-Piercy et al. 2021). Expression of alpha and beta crystallin in non-

lens tissues has been reported in various rodents and zebrafish studies. (Piri, Kwong, and

Caprioli 2013; Sprague-Piercy et al. 2021; Zigler and Sinha 2015; Andley 2007). Experiments

using gamma crystallin deficient mice show that in the lens fiber it's important for the

organization of actin filaments, suggesting that the general properties of crystallin can stabilize

and prevent aggregation of cytoskeletal filaments (Fan et al. 2012). Although beta and gamma

crystallin are generally classified as providing structural properties, their N-terminus contains

serine protease activity under certain conditions (Andley 2007; Gupta, Chen, and Srivastava

2010; Slingsby and Wistow 2014). We speculate that the expression of crystallin in the artificial

state of germ-free larvae is a response to decreased metabolic function or other cellular stresses

experienced by zebrafish developing without microbial cues. This hypothesis is consistent with

the decreased expression of genes involved in ATP/nucleotide metabolism and decreased

consumption ATP within the intestinal lumen of germ-free larvae (Massaquoi et al., submitted).

The generation of ATP and metabolism of nutrients is essential during host development to fuel

growth and survival. Insufficient cellular levels of ATP and other nutrients can lead to

dysfunctional states and oxidative stress as observed in aging post-mitotic cells (Schütt et al. 2012; Terman et al. 2010). The remainder of this section will discuss the possible interpretations of crystallin expression in germ-free zebrafish.

Context and location in the expression and function of crystallin

Across organisms, Crystallin proteins have shown to be protective against a range of stresses. Consistent with neuroprotection, there is a transient upregulation of all representative members of the crystallin gene family within the embryonic brain of mice from several models of maternal immune activation (Garbette et all 2012). In the retina and other neuronal tissues, alpha crystallin have anti-apoptotic and neuroprotective functions while both alpha and beta crystallin in the retinal epithelium are protective against lysosomal dysfunction by stabilizing the vacuolar ATPase (Phadte, Sluzala, and Fort 2021; Zhu and Reiser 2018; Cui et al. 2020; Valapala et al. 2016). AlphaB crystallin protects against cardiac stress induced by cortisol or overcrowding in zebrafish (Mishra et al. 2018). Nematodes in the *dauer* state and insects and brime shrimp in diapuase express small heat shock proteins, similar to alpha crystallin, consistent with the expression of stable proteins functioning to maintain cells during low metabolic states (Fu et al. 2021; King and Macrae 2015). The lack of microbial cues in germ-free derived zebrafish likely disengages multifaceted cellular dynamics involved in tuning host metabolism, which could induce cellular stress globally across the body. This is apparent in germ-free zebrafish, which have a paucity of pancreatic insulin-producing beta cells (Hill et al., 2016). Modulation of insulin signaling plays a key role in organ cross-talk and regulating metabolic homeostasis (Romero and Eckel 2021). Beta cells are the sole producers of insulin in vertebrates, and the binding to insulin receptors activates complex intracellular signaling networks that not only facilitate glucose uptake but also pathways involved in proliferation, growth and survival.

Several studies suggest that insulin signaling also reinforces mitochondrial electron transport chain integrity (Cheng, Tseng, and White 2010). The paucity of insulin-producing beta cells and the global enrichment of crystallin within germ-free zebrafish signifying a metabolically stressed organism is consistent with work that analyzed the transcriptome of the entire retina in diabetic mice and showed a dramatic increase in beta and gamma crystallin (Kandpal et al. 2012; Hill et al. 2016). Energy gained from cellular metabolism is vital for cells to function, including cellular transport and the removal of cellular debris. We reason that the decreased metabolic function within germ-free larvae contributes to our observations that several cell types appear to be fully differentiated but in an inactive state. Within the exocrine pancreas, we show that the microbiota promotes tissue development and function. The pancreases of germ-free larvae exhibit stunted tissue growth and appear to be composed of fully differentiated, yet metabolically and functionally inert acinar (Massaquoi et al., *submitted*). Such a state may render the necessity of cytoprotective Cry proteins to maintain. For example, a common feature among cell types that characteristically express crystallin such as lens fiber, retinal pigment cells and many types of neurons, is their post-mitotic status or low rates of cell turnover. Over time these cells accumulate deposits that cannot be discarded through normal 'house-keeping' mechanisms, inducing cellular stress. With the potential roles of impeding the aggregation of complex macromolecules, crystallin could have evolved to add stability to these ageing cells (Slingsby and Wistow 2014). Not only have germ-free zebrafish demonstrated a lack of proliferation in several different cell types but also in the case of the exocrine pancreas, the dramatic increase of amylase expression within secretory granules versus conventional larvae echoes the need for stabilizing aggregates of macromolecules.
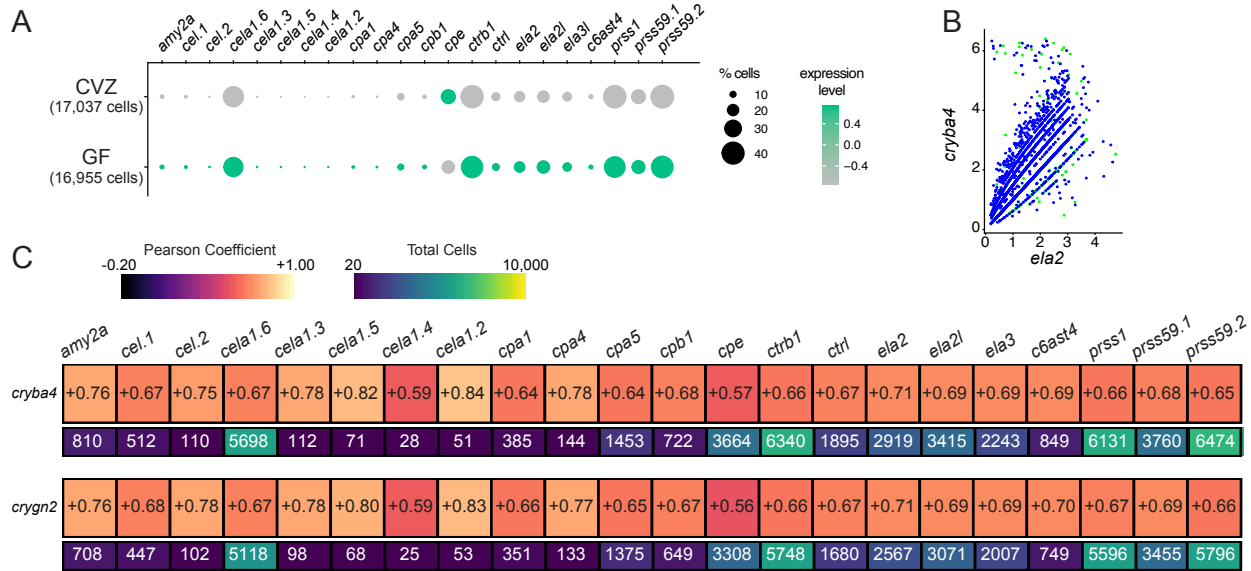
Figure 4.3 Digestive enzymes and crystallin genes are co-expressed in cells from germ free larvae by single cell transcriptomic profiling. A) The dotplot displays the percentage of individual cells across the entire body of conventionalized (CVZ) or germ free (GF) 6dpf zebrafish expressing different digestive enzyme transcripts and the level of their expression. Data is from single cell transcriptomic profiling of whole gnotobiotic zebrafish, as described in Chapter II. B) The scatterplot demonstrates the expression of crystallin *cryba4* and digestive enzyme *ela2* positively correlate among CVZ (green) and GF (blue) cells. These data do not include cells from the exocrine pancreas which produces high levels of digestive enzymes for its canonical function. C) The heat plot demonstrates the Pearson Correlation Coefficient of cryba4 and crygn2 expression with respect to different digestive enzymes and the number of cells positive for both transcripts in the comparison.

To further explore the possible connections between crystallin and cellular stressors we analyzed the expression pattern of *crystallin* with respect to different transcripts involved in inflammation, ATP metabolism and digestion. We did not observe expression of the canonical suite of transcripts anticipated with inflammation, nor strong correlations with genes involved in ATP metabolism but did recognized expression of several digestive enzymes that positively correlated with *crystallin* (Fig4.3). Consistent with our transcriptional findings, the protein families of both peptidases/proteases and crystallin are included in the 'Top 25' protein families list of zebrafish-specific responses to various physiological, chemical or biological stressors (Groh and J-F Suter 2015). Digestive enzymes play a critical role in host metabolism by their enzymatic activity of breaking down macromolecules. They are produced and secreted by exocrine glands of the digestive tract, including the intestine and pancreas. Relative to

conventionalized zebrafish, diverse cell types from germ-free larvae have exogenous

transcriptional expression of digestive enzymes, predominantly serine proteases, which cleave

peptide bonds in proteins. We speculate that the dual expression of crystallin and digestive

enzymes could be involved in a genetic program responding to cellular debris buildup due to

metabolic stressors of the artificial germ-free environment. In the case of stunted tissues with

cellular debris buildup from lack of functionality, the expression of crystallin would chaperone

protein aggregates, while serine proteases assist in debris degradation. In wild speculation, this

dual function is consistent with human beta crystallin possessing serine-type protease activity

(Gupta, Chen, and Srivastava 2010). However, serine proteases perform various biological

functions. With their broad substrate specificity, neutrophils deploy serine proteases (ie

elastases) for multiple reactions in response to infection (Stapels, Geisbrecht, and Rooijakkers

2015) and Factor B is a serine protease functioning as a key component in the activation of the

complement system (Seeger, Mayer, and Klein 1996). Interestingly, Alpha crystallin from

bovine lens can act as an inhibitor to elastase (Ortwerth and Olesen 1992). In addition to

performing immune responses, a serine protease homologue in *Drosophila* plays a role in the

perception of nutrient availability and modulates the formation of the blood brain barrier

(Contreras et al., 2021). On the theme of cell barriers, serine protease activity plays a role in the

regulation epithelial tissue integrity and cell-adhesion maintenance, implicating an important role

in the morphology of tissues during organogenesis (Carney et al. 2007; Ohler and Becker-Pauly

2011). Another interpretation of the expression of crystallin and serine proteases within germ-

free larvae is as a genetic program to maintain cellular and tissue architecture in response to the

metabolically deficient state. Beta and gamma crystallin are the most prevalently expressed

within the germ-free zebrafish and these crystallin share tertiary protein structure with microbial

proteins spherulin 3a of the eukaryotic slime mold *Physarum polycephalum* involved in cell-cell adhesion and spore coat Protein S of *Myxococcus xanthus* (Clout et al. 2001; Wistow, Summers, and Blundell 1985). Across the literature the definitive role of crystallin outside of the vertebrate lens is only speculated, however the lens is clear on the diversity of important roles the microbiota play in host metabolism and development shown by years of *in vivo* experiments with gnotobiotic zebrafish and with transcriptional single cell resolution.

REFERENCES CITED

Adams *et al*., a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213-221 (2010).

Adolph, Timon E., Lisa Mayr, Felix Grabherr, Julian Schwärzler, and Herbert Tilg. 2019. "Pancreas-Microbiota Cross Talk in Health and Disease." *Annual Review of Nutrition* 39 (August). NLM (Medline): 249–66. doi:10.1146/annurev-nutr-082018-124306.

Aguayo-Mazzucato, S. Bonner-Weir, Pancreatic beta Cell Regeneration as a Possible Therapy for Diabetes. *Cell Metab* **27**, 57-67 (2018).

Akira, S. (2003) 'Toll-like receptor signaling', *Journal of Biological Chemistry*, pp. 38105–38108. doi: 10.1074/jbc.R300028200.

Altschul *et al*., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402 (1997).

Andley, Usha P. 2007. "Crystallins in the Eye: Function and Pathology." *Progress in Retinal and Eye Research* 26 (1). Pergamon: 78–98. doi:10.1016/J.PRETEYERES.2006.10.003.

Angelova M.I., Méléard P., Faucon F., Bothorel P., in *Trends in Colloid and Interface Science VI. Progress in Colloid & Polymer Science*, L. M. Helm C., Möhwald H, Ed. (Steinkopff, 1992), vol. 89.

Arentsen, T, Y Qian, S Gkotzis, T Femenia, T Wang, K Udekwu, H Forssberg, and R Diaz Heijtz. 2016. "The Bacterial Peptidoglycan-Sensing Molecule Pglyrp2 Modulates Brain Development and Behavior." doi:10.1038/mp.2016.182.

Atarashi, Koji, Junichi Nishimura, Tatsuichiro Shima, Yoshinori Umesaki, Masahiro Yamamoto, Masaharu Onoue, Hideo Yagita, et al. 2008. "ATP Drives Lamina Propria TH17 Cell Differentiation." *Nature 2008 455:7214* 455 (7214). Nature Publishing Group: 808–12. doi:10.1038/nature07240.

Ayres, J.S. (2016). Cooperative Microbial Tolerance Behaviors in Host-Microbiota Mutualism. Cell *165*, 1323-1331.

Bates *et al*., Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Developmental Biology* **297**, 374-386 (2006).

Bates, Jennifer M. *et al*. (2007) 'Intestinal Alkaline Phosphatase Detoxifies Lipopolysaccharide and Prevents Inflammation in Zebrafish in Response to the Gut Microbiota', *Cell Host and Microbe*, 2(6), pp. 371–382. doi: 10.1016/j.chom.2007.10.010.

Bayles, Bacterial programmed cell death: making sense of a paradox. *Nat Rev Microbiol* **12**, 63-69 (2014).

Bacon *et al.*, Serum levels of pancreatic stone protein (PSP)/reg1A as an indicator of beta-cell apoptosis suggest an increased apoptosis rate in hepatocyte nuclear factor 1 alpha (HNF1A-MODY) carriers from the third decade of life onward. *BMC Endocr Disord* **12**, 13 (2012).

Bagshaw, Clive R. 2020. "ATP Analogues at a Glance." Accessed December 30. www.biologists.com/jcs.

Bao, Yi, Carola Ledderose, Thomas Seier, Amelie F. Graf, Bianca Brix, Eritza Chong, and Wolfgang G. Junger. 2014. "Mitochondria Regulate Neutrophil Activation by Generating ATP for Autocrine Purinergic Signaling." *Journal of Biological Chemistry* 289 (39): 26794–803. doi:10.1074/jbc.M114.572495.

Bartoli, N. Baeza, C. Figarella, I. Pellegrini, D. Figarella-Branger, Expression of peptide-23/pancreatitis-associated protein and Reg genes in human pituitary and adenomas: comparison with other fetal and adult human tissues. *J Clin Endocrinol Metab* **83**, 4041-4046 (1998).

Bautista, Diana M., Sven Eric Jordt, Tetsuro Nikai, Pamela R. Tsuruda, Andrew J. Read, Jeannie Poblete, Ebenezer N. Yamoah, Allan I. Basbaum, and David Julius. 2006. "TRPA1 Mediates the Inflammatory Actions of Environmental Irritants and Proalgesic Agents." *Cell* 124 (6). Elsevier B.V.: 1269–82..

Belkaid, Y. and Harrison, O. J. (2017) 'Homeostatic Immunity and the Microbiota', *Immunity*. Cell Press, pp. 562–576. doi: 10.1016/j.immuni.2017.04.008.

Berger, M. *et al.* (2015) 'Gαi/o-coupled receptor signaling restricts pancreatic β-cell expansion', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 112(9), pp. 2888–2893. doi: 10.1073/pnas.1319378112.

Biemar, F, F Argenton, R Schmidtke, S Epperlein, B Peers, and W Driever. 2001. "Pancreas Development in Zebrafish: Early Dispersed Appearance of Endocrine Hormone Expressing Cells and Their Convergence to Form the Definitive Islet." *Developmental Biology* 230 (2): 189–203. doi:10.1006/dbio.2000.0103.

Bohórquez, Diego V., Leigh A. Samsa, Andrew Roholt, Satish Medicetty, Rashmi Chandra, and Rodger A. Liddle. 2014. "An Enteroendocrine Cell – Enteric Glia Connection Revealed by 3D Electron Microscopy." *PLOS ONE* 9 (2). Public Library of Science: e89881. doi:10.1371/JOURNAL.PONE.0089881.

Bohórquez, Diego V., Rafiq A. Shahid, Alan Erdmann, Alex M. Kreger, Yu Wang, Nicole Calakos, Fan Wang, and Rodger A. Liddle. 2015. "Neuroepithelial Circuit Formed by Innervation of Sensory Enteroendocrine Cells." *The Journal of Clinical Investigation* 125 (2). J Clin Invest: 782–86. doi:10.1172/JCI78361.

Bokulich *et al.*, Antibiotics, birth mode, and diet shape microbiome maturation during early life. *Sci Transl Med* **8**, 343ra382 (2016).

Bours, M. J.L., E. L.R. Swennen, F. Di Virgilio, B. N. Cronstein, and P. C. Dagnelie. 2006. "Adenosine 5′-Triphosphate and Adenosine as Endogenous Signaling Molecules in Immunity and Inflammation." *Pharmacology and Therapeutics*. Pergamon. doi:10.1016/j.pharmthera.2005.04.013.

Brito, D. Cabanes, F. Sarmento Mesquita, S. Sousa, Mechanisms protecting host cells against bacterial pore-forming toxins. *Cell Mol Life Sci* **76**, 1319-1339 (2019).

Burns *et al.*, Interhost dispersal alters microbiome assembly and can overwhelm host innate immunity in an experimental zebrafish model. *Proc Natl Acad Sci U S A* **114**, 11181-11186 (2017).

Burns, A. R. and Guillemin, K. (2017) 'The scales of the zebrafish: host–microbiota interactions from proteins to populations', *Current Opinion in Microbiology*, 38, pp. 137–141. doi: 10.1016/j.mib.2017.05.011.

Burns, Adam R, W Zac Stephens, Keaton Stagaman, Sandi Wong, John F Rawls, Karen Guillemin, and Brendan JM Bohannan. 2016. "Contribution of Neutral Processes to the Assembly of Gut Microbial Communities in the Zebrafish over Host Development." *The ISME Journal* 10 (3). Nature Publishing Group: 655–64. doi:10.1038/ismej.2015.142.

Burnstock, Geoffrey, and Gillian E. Knight. 2004. "Cellular Distribution and Functions of P2 Receptor Subtypes in Different Systems." *International Review of Cytology* 240 (SPEC.ISS.): 31–304. doi:10.1016/S0074-7696(04)40002-3.

Burnstock, Geoffrey, Byron Vaughn, and Simon C Robson. n.d. "Purinergic Signalling in the Liver in Health and Disease." doi:10.1007/s11302-013-9398-8.

Cash, C. V. Whitham, C. L. Behrendt, L. V. Hooper, Symbiotic bacteria direct expression of an intestinal bactericidal lectin. *Science* **313**, 1126-1130 (2006).

Carlos, A.R., Weis, S., and Soares, M.P. (2018). Cross-Talk Between Iron and Glucose Metabolism in the Establishment of Disease Tolerance. Front Immunol *9*, 2498.

Carmona, Santiago J., Sarah A. Teichmann, Lauren Ferreira, Iain C. MacAulay, Michael J.T. Stubbington, Ana Cvejic, and David Gfeller. 2017. "Single-Cell Transcriptome Analysis of Fish Immune Cells Provides Insight into the Evolution of Vertebrate Immune Cell Types." *Genome Research* 27 (3). Cold Spring Harbor Laboratory Press: 451–61. doi:10.1101/gr.207704.116.

Carney, Thomas J., Sophia von der Hardt, Carmen Sonntag, Adam Amsterdam, Jacek Topczewski, Nancy Hopkins, and Matthias Hammerschmidt. 2007. "Inactivation of Serine Protease Matriptase1a by Its Inhibitor Hai1 Is Required for Epithelial Integrity of the Zebrafish Epidermis." *Development* 134 (19). The Company of Biologists: 3461–71. doi:10.1242/DEV.004556.

Carten, M. K. Bradford, S. A. Farber, Visualizing digestive organ morphology and function using differential fatty acid metabolism in live zebrafish. *Dev Biol* **360**, 276-285 (2011).

Cash, C. V. Whitham, L. V. Hooper, Refolding, purification, and characterization of human and murine RegIII proteins expressed in Escherichia coli. *Protein Expr Purif* **48**, 151-159 (2006).

Cheesman, S. E. *et al*. (2011) 'Epithelial cell proliferation in the developing zebrafish intestine is regulated by the Wnt pathway and microbial signaling via Myd88'. doi: 10.1073/pnas.1000072107.

Chen, S. Downing, E. S. Tzanakakis, Four Decades After the Discovery of Regenerating Islet-Derived (Reg) Proteins: Current Understanding and Challenges. *Front Cell Dev Biol* **7**, 235 (2019).

Chen, Yu, Ross Corriden, Yoshiaki Inoue, Linda Yip, Naoyuki Hashiguchi, Annelies Zinkernagel, Victor Nizet, Paul A. Insel, and Wolfgang G. Junger. 2006. "ATP Release Guides Neutrophil Chemotaxis via P2Y2 and A3 Receptors." *Science* 314 (5806): 1792–95. doi:10.1126/science.1132559.

Cheng, Zhiyong, Yolanda Tseng, and Morris F. White. 2010. "Insulin Signaling Meets Mitochondria in Metabolism." *Trends in Endocrinology & Metabolism* 21 (10). Elsevier: 589–98. doi:10.1016/J.TEM.2010.06.005.

Cholan, Pradeep Manuneedhi, Alvin Han, Brad R. Woodie, Maxinne Watchon, Angela R.M. Kurz, Angela S. Laird, Warwick J. Britton, et al. 2020. "Conserved Anti-Inflammatory Effects and Sensing of Butyrate in Zebrafish." *Gut Microbes* 12 (1). Bellwether Publishing, Ltd.: 1–11. doi:10.1080/19490976.2020.1824563/SUPPL_FILE/KGMI_A_1824563_SM3703.ZIP.

Chu, H., and Mazmanian, S.K. (2013). Innate immune recognition of the microbiota promotes host-microbial symbiosis. Nat Immunol *14*, 668-675.

Clout, Naomi J., Michael Kretschmar, Rainer Jaenicke, and Christine Slingsby. 2001. "Crystal Structure of the Calcium-Loaded Spherulin 3a Dimer Sheds Light on the Evolution of the Eye Lens Betagamma-Crystallin Domain Fold." *Structure (London, England : 1993)* 9 (2). Structure: 115–24. doi:10.1016/S0969-2126(01)00573-1.

Cocchiaro, Jordan L., and John F. Rawls. 2013. "Microgavage of Zebrafish Larvae." *Journal of Visualized Experiments*, no. 72 (February). Journal of Visualized Experiments. doi:10.3791/4434.

Collaborative Computational Project, The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D Biological Crystallography* **50**, 760-763 (1994).

Contreras, Esteban G, Álvaro Glavic, Andrea H Brand, and Jimena Sierralta. n.d. "The Serine Protease Homologue, Scarface, Is Sensitive to 1 Nutrient Availability and Modulates the Development of the 2 Drosophila Blood Brain Barrier."

Crosnier, C. *et al*. (2005) 'Delta-Notch signalling controls commitment to a secretory fate in the zebrafish intestine', *Development*, 132(5), pp. 1093–1104. doi: 10.1242/dev.01644.

Cui, Xiukun, Ruiping Feng, Jungai Wang, Chunxiao Du, Xiahui Pi, Danling Chen, Jing Li, et al. 2020. "Heat Shock Factor 4 Regulates Lysosome Activity by Modulating the AB-Crystallin-ATP6V1A-MTOR Complex in Ocular Lens." *Biochimica et Biophysica Acta (BBA) - General Subjects* 1864 (3). Elsevier: 129496. doi:10.1016/J.BBAGEN.2019.129496.

Dedeine, F. *et al*. (no date) *Removing symbiotic Wolbachia bacteria specifically inhibits oogenesis in a parasitic wasp*. Available at: www.pnas.orgcgidoi10.1073pnas.101304298 (Accessed: 4 March 2019).

Deguine, J. and Barton, G. M. (1000) 'MyD88: a central player in innate immune signaling'. doi: 10.12703/P6-97.

Deguine, Jacques, and Gregory M Barton. 1000. "MyD88: A Central Player in Innate Immune Signaling." doi:10.12703/P6-97.

del Castillo *et al*., The Microbiomes of Pancreatic and Duodenum Tissue Overlap and Are Highly Subject Specific but Differ between Pancreatic Cancer and Noncancer Subjects. *Cancer Epidemiology Biomarkers & Prevention* **28**, 370-383 (2019).

Delous *et al*., Sox9b is a key regulator of pancreaticobiliary ductal system development. *PLoS Genet* **8**, e1002754 (2012).

Desai, M. S. *et al*. (2016) 'A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility', *Cell*. Cell Press, 167(5), pp. 1339-1353.e21. doi: 10.1016/j.cell.2016.10.043.

Deverman, Benjamin E., and Paul H. Patterson. 2009. "Cytokines and CNS Development." *Neuron*. Cell Press. doi:10.1016/j.neuron.2009.09.002.

Dewar *et al*., Novel proteins linking the actin cytoskeleton to the endocytic machinery in Saccharomyces cerevisiae. *Mol Biol Cell* **13**, 3646-3661 (2002).

diIorio, J. B. Moss, J. L. Sbrogna, R. O. Karlstrom, L. G. Moss, Sonic hedgehog is required early in pancreatic islet development. *Developmental Biology* **244**, 75-84 (2002).

Donohoe, Dallas R., Nikhil Garge, Xinxin Zhang, Wei Sun, Thomas M. O'Connell, Maureen K. Bunger, and Scott J. Bultman. 2011. "The Microbiome and Butyrate Regulate Energy Metabolism and Autophagy in the Mammalian Colon." *Cell Metabolism* 13 (5). Elsevier: 517–26. doi:10.1016/J.CMET.2011.02.018/ATTACHMENT/47419042-0759-4B45-AADD-3942F67E2AE4/MMC3.XLS.

E. Astorri *et al*., Circulating Reg1alpha proteins and autoantibodies to Reg1alpha proteins as biomarkers of beta-cell regeneration and damage in type 1 diabetes. *Horm Metab Res* **42**, 955-960 (2010).

Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113-113 (2004).

Emsley, K. Cowtan, *Coot* : model-building tools for molecular graphics. *Acta Crystallographica Section D Biological Crystallography* **60**, 2126-2132 (2004).

Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).

Falcinelli, Silvia, Ana Rodiles, Suraj Unniappan, Simona Picchietti, Giorgia Gioacchini, Daniel Lee Merrifield, and Oliana Carnevali. 2016. "Probiotic Treatment Reduces Appetite and Glucose Level in the Zebrafish Model." *Scientific Reports 2016 6:1* 6 (1). Nature Publishing Group: 1–13. doi:10.1038/srep18061.

Falk, P., Roth, K. A. and Gordon, J. I. (1994) 'Lectins are sensitive tools for defining the differentiation programs of mouse gut epithelial cell lineages', *American Journal of Physiology - Gastrointestinal and Liver Physiology*, 266(6 29-6). doi: 10.1152/ajpgi.1994.266.6.g987.

Fan, Jianguo, Lijin Dong, Sanghamitra Mishra, Yingwei Chen, Paul Fitzgerald, and Graeme Wistow. 2012. "A Role for ΓS-Crystallin in the Organization of Actin and Fiber Cell Maturation in the Mouse Lens." *The FEBS Journal* 279 (16). John Wiley & Sons, Ltd: 2892–2904. doi:10.1111/J.1742-4658.2012.08669.X.

Farnsworth, Dylan R., Lauren M. Saunders, and Adam C. Miller. 2019. "A Single-Cell Transcriptome Atlas for Zebrafish Development." *Developmental Biology*, November. Elsevier BV. doi:10.1016/j.ydbio.2019.11.008.

Farnsworth, Dylan R., Mason Posner, and Adam C. Miller. 2021. "Single Cell Transcriptomics of the Developing Zebrafish Lens and Identification of Putative Controllers of Lens Development." *Experimental Eye Research* 206 (May). Exp Eye Res. doi:10.1016/J.EXER.2021.108535.

Field, Holly A., P. D. Si Dong, Dimitris Beis, and Didier Y.R. Stainier. 2003. "Formation of the Digestive System in Zebrafish. Ii. Pancreas Morphogenesis☆." *Developmental Biology* 261 (1). Academic Press: 197–208. doi:10.1016/S0012-1606(03)00308-7.

Fischbach, Michael A, and Julia A Segre. 2016. "Leading Edge Review Signaling in Host-Associated Microbial Communities Human-Associated Microbiota Form and Stabilize Communities Based on Interspecies Interactions. We Review How These Microbe-Microbe and Microbe-Host Interactions." *Cell* 164: 1288–1300. doi:10.1016/j.cell.2016.02.037.

Francis *et al*., Probiotic Studies in Neonatal Mice Using Gavage. *J Vis Exp*, (2019).

Frost, Fabian, Tim Kacprowski, Malte Rühlemann, Robin Bülow, Jens Peter Kühn, Andre Franke, Femke Anouska Heinsen, et al. 2019. "Impaired Exocrine Pancreatic Function Associates With Changes in Intestinal Microbiota Composition and Diversity." *Gastroenterology* 156 (4). W.B. Saunders: 1010–15. doi:10.1053/j.gastro.2018.10.047.

Fu, Xinmiao, Anastasia N. Ezemaduka, Xinping Lu, and Zengyi Chang. 2021. "The Caenorhabditis Elegans 12-KDa Small Heat Shock Proteins with Little in Vitro Chaperone Activity Play Crucial Roles for Its Dauer Formation, Longevity, and Reproduction." *Protein Science* 30 (10). John Wiley & Sons, Ltd: 2170–82. doi:10.1002/PRO.4160.

Furness, John B, Brid P Callaghan, Leni R Rivera, Hyun-Jung Cho, J B Furness, B P Callaghan, • L R Rivera, H.-J Cho, M Lyte, and J F Cryan. 2014. "The Enteric Nervous System and Gastrointestinal Innervation: Integrated Local and Central Control." doi:10.1007/978-1-4939-0897-4_3.

Garay-Malpartida *et al*., Toll-like receptor 4 (TLR4) expression in human and murine pancreatic beta-cells affects cell viability and insulin homeostasis. *BMC Immunol* **12**, 18 (2011).

Garcia-Estañ, Joaquin, Kiyoshi Mori, Pablo Vivas-Mejia, University of Puerto Rico, Puerto Rico Ginette Santiago-Sanchez, Diego Sanchez, Sergio Diez-Hermano, Maria D Ganfornina, Arne Skerra, and Gabriel Gutiérrez. 2021. "An Evolutionary Perspective of the Lipocalin Protein Family." doi:10.3389/fphys.2021.718983.

Gnainsky, Yulia, Nofar Zfanya, Michael Elgart, Eman Omri, Alexander Brandis, Tevie Mehlman, Maxim Itkin, Sergey Malitsky, Jerzy Adamski, and Yoav Soen. 2021. "Systemic Regulation of Host Energy and Oogenesis by Microbiome-Derived Mitochondrial Coenzymes." *Cell Reports* 34 (1). ElsevierCompany. doi:10.1016/j.celrep.2020.108583.

Golonka, Rachel, Beng San Yeoh, and Matam Vijay-Kumar. 2019. "The Iron Tug-of-War between Bacterial Siderophores and Innate Immunity." *Journal of Innate Immunity* 11 (3). Karger Publishers: 249–62. doi:10.1159/000494627.

Goujon *et al*., A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Res* **38**, W695-699 (2010).

Granato, T. A. Meiller-Legrand, K. R. Foster, The Evolution and Ecology of Bacterial Warfare. *Curr Biol* **29**, R521-R537 (2019).

Groh, Ksenia J, and Marc J-F Suter. 2015. "Stressor-Induced Proteome Alterations in Zebrafish: A Meta-Analysis of Response Patterns." *Aquatic Toxicology* 159: 1–12. doi:10.1016/j.aquatox.2014.11.013.

Grunwald, D. J. and Eisen, J. S. (2002) 'Headwaters of the zebrafish - emergence of a new model vertebrate', *Nature Reviews Genetics*, pp. 717–724. doi: 10.1038/nrg892.

Guha, J. Ghimire, E. Wu, W. C. Wimley, Mechanistic Landscape of Membrane-Permeabilizing Peptides. *Chem Rev* **119**, 6040-6085 (2019).

Gupta, R, J Chen, and O P Srivastava. 2010. "A Serine-Type Protease Activity of Human Lens BA3-Crystallin Is Responsible for Its Autodegradation." http://www.molvis.org/molvis/v16/a240.

Hall, C. *et al*. (2009) 'Transgenic zebrafish reporter lines reveal conserved Toll-like receptor signaling potential in embryonic myeloid leukocytes and adult immune cell lineages.', *Journal of leukocyte biology*, 85(5), pp. 751–65. doi: 10.1189/jlb.0708405.

Hall, Chris, Maria Vega Flores, Annie Chien, Alan Davidson, Kathryn Crosier, and Phil Crosier. 2009. "Transgenic Zebrafish Reporter Lines Reveal Conserved Toll-like Receptor Signaling Potential in Embryonic Myeloid Leukocytes and Adult Immune Cell Lineages." *Journal of Leukocyte Biology* 85 (5): 751–65. doi:10.1189/jlb.0708405.

Hancock, E. F. Haney, E. E. Gill, The immunology of host defence peptides: beyond antimicrobial activity. *Nat Rev Immunol* **16**, 321-334 (2016).

Hasegawa *et al*., SH3YL1 regulates dorsal ruffle formation by a novel phosphoinositide-binding domain. *The Journal of Cell Biology* **193**, 901-916 (2011).

Henderson, S. Hazra, A. M. Dunkle, M. E. Salvucci, R. M. Wachter, Biophysical characterization of higher plant Rubisco activase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1834**, 87-97 (2013).

Heppert, Jennifer K., James M. Davison, Cecelia Kelly, Gilberto Padilla Mercado, Colin R. Lickwar, and John F. Rawls. 2021. "Transcriptional Programmes Underlying Cellular Identity and Microbial Responsiveness in the Intestinal Epithelium." *Nature Reviews Gastroenterology and Hepatology*. Nature Research. doi:10.1038/s41575-020-00357-6.

Hill, J.H., Franzosa, E.A., Huttenhower, C., and Guillemin, K. (2016). A conserved bacterial protein induces pancreatic beta cell expansion during zebrafish development. Elife *5*.

Hironaka, Ippei, Tadayuki Iwase, Shinya Sugimoto, Ken ichi Okuda, Akiko Tajima, Katsuhiko Yanaga, and Yoshimitsu Mizunoe. 2013. "Glucose Triggers ATP Secretion from Bacteria in a Growth-Phase-Dependent Manner." *Applied and Environmental Microbiology* 79 (7). American Society for Microbiology: 2328–35. doi:10.1128/AEM.03871-12.

Holm En Larsson, Jessica M, Hasse Karlsson, Jessica Gråberg Crespo, Malin E V Johansson, Lisbeth Eklund, Henrik Sjövall, and Gunnar C Hansson. 2011. "Altered O-Glycosylation Profile of MUC2 Mucin Occurs in Active Ulcerative Colitis and Is Associated with Increased Inflammation." doi:10.1002/ibd.21625.

Hooper, Lora V., Dan R. Littman, and Andrew J. Macpherson. 2012. "Interactions between the Microbiota and the Immune System." *Science*. American Association for the Advancement of Science. doi:10.1126/science.1223490.

Horn, J. K. Jaiswal, Cellular mechanisms and signals that coordinate plasma membrane repair. *Cell Mol Life Sci* **75**, 3751-3770 (2018).

Horton, Z. L. Cai, S. N. Ho, L. R. Pease, Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *BioTechniques* **8**, 528-535 (1990).

Howe, Douglas G., Yvonne M. Bradford, Tom Conlin, Anne E. Eagle, David Fashena, Ken Frazer, Jonathan Knight, et al. 2013. "ZFIN, the Zebrafish Model Organism Database: Increased Support for Mutants and Transgenics." *Nucleic Acids Research* 41 (D1): 854–60. doi:10.1093/nar/gks938.

Hsiao, Elaine Y., Sara W. McBride, Sophia Hsien, Gil Sharon, Embriette R. Hyde, Tyler McCue, Julian A. Codelli, et al. 2013. "Microbiota Modulate Behavioral and Physiological Abnormalities Associated with Neurodevelopmental Disorders." *Cell* 155 (7). Elsevier Inc.: 1451–63. doi:10.1016/j.cell.2013.11.024.

Huang, G. Gu, Effective Isolation of Functional Islets from Neonatal Mouse Pancreas. *J Vis Exp*, (2017).

Huang, Lumei, Lilla Otrokocsi, and Beáta Sperlágh. 2019. "Role of P2 Receptors in Normal Brain Development and in Neurodevelopmental Psychiatric Disorders." *Brain Research Bulletin*. Elsevier Inc. doi:10.1016/j.brainresbull.2019.01.030.

Huson, C. Scornavacca, Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* **61**, 1061-1067 (2012).

Inami, Akie, Hiroshi Kiyono, and Yosuke Kurashima. 2018. "ATP as a Pathophysiologic Mediator of Bacteria-Host Crosstalk in the Gastrointestinal Tract." *International Journal of Molecular Sciences* 19 (8): 7–9. doi:10.3390/ijms19082371.

Iwanami, Norimasa, Divine Fondzenyuy Lawir, Katarzyna Sikora, Connor O´Meara, Kohei Takeshita, Michael Schorpp, and Thomas Boehm. 2020. "Transgenerational Inheritance of Impaired Larval T Cell Development in Zebrafish." *Nature Communications* 11 (1). Nat Commun. doi:10.1038/S41467-020-18289-9.

Jameson, K. G., C. A. Olson, S. A. Kazmi, and E. Y. Hsiao. 2020. "Toward Understanding Microbiome-Neuronal Signaling." *Molecular Cell* 78 (4). Mol Cell: 577–83. doi:10.1016/J.MOLCEL.2020.03.006.

Jault, Cyril, Laurent Pichon, and Johanna Chluba. 2004. "Toll-like Receptor Gene Family and TIR-Domain Adapters in Danio Rerio." *Molecular Immunology* 40 (11). Elsevier Ltd: 759–71. doi:10.1016/j.molimm.2003.10.001.

Jemielita, Matthew, Michael J. Taormina, Adam R. Burns, Jennifer S. Hampton, Annah S. Rolig, Karen Guillemin, and Raghuveer Parthasarathy. 2014. "Spatial and Temporal Features of the Growth of a Bacterial Species Colonizing the Zebrafish Gut." *MBio* 5 (6). American Society for Microbiology. doi:10.1128/mBio.01751-14.

Jevtov, Irena, Tore Samuelsson, Grace Yao, Adam Amsterdam, and Katharina Ribbeck. 2014. "Zebrafish as a Model to Study Live Mucus Physiology OPEN." doi:10.1038/srep06653.

Ji *et al*., Toll-like receptors TLR2 and TLR4 block the replication of pancreatic beta cells in diet-induced obesity. *Nat Immunol* **20**, 677-686 (2019).

Jimah, P. H. Schlesinger, N. H. Tolia, Liposome Disruption Assay to Examine Lytic Properties of Biomolecules. *Bio Protoc* **7**, (2017).

Johansson, M. E. V. and Hansson, G. C. (2016) 'Immunological aspects of intestinal mucus and mucins', *Nature Reviews Immunology*. Nature Publishing Group, 16(10), pp. 639–649. doi: 10.1038/nri.2016.88.

Jone, T.A., and Guillemin, K. (2018). Racing to Stay Put: How Resident Microbiota Stimulate Intestinal Epithelial Cell Proliferation. Current Pathobiology Reports *6*, 23-28.

Junger, Wolfgang G. n.d. "Immune Cell Regulation by Autocrine Purinergic Signalling." doi:10.1038/nri2938.

Kaelberer, Melanie Maya, Kelly L. Buchanan, Marguerita E. Klein, Bradley B. Barth, Marcia M. Montoya, Xiling Shen, and Diego V. Bohórquez. 2018. "A Gut-Brain Neural Circuit for Nutrient Sensory Transduction." *Science* 361 (6408). American Association for the Advancement of Science. doi:10.1126/SCIENCE.AAT5236/SUPPL_FILE/AAT5236S2.MP4.

Kandpal, Raj P, Harsha K Rajasimha, Matthew J Brooks, Jacob Nellissery, Jun Wan, Jiang Qian, Timothy S Kern, and Anand Swaroop. 2012. "Transcriptome Analysis Using next Generation Sequencing Reveals Molecular Signatures of Diabetic Retinopathy and Efficacy of Candidate Drugs." *Molecular Vision* 18: 1123–46.

Kanther, Michelle, Sarah Tomkovich, Sun Xiaolun, Melinda R. Grosser, Jaseol Koo, Edward J. Flynn, Christian Jobin, and John F. Rawls. 2014. "Commensal Microbiota Stimulate Systemic Neutrophil Migration through Induction of Serum Amyloid A." *Cellular Microbiology* 16 (7). Blackwell Publishing Ltd: 1053–67. doi:10.1111/CMI.12257/SUPPINFO.

Karplus, K. Diederichs, Linking Crystallographic Model and Data Quality. *Science* **336**, 1030-1033 (2012).

Kashyap, P. C. *et al*. (2013) 'Genetically dictated change in host mucus carbohydrate landscape exerts a diet-dependent effect on the gut microbiota', *Proceedings of the National Academy of Sciences of the United States of America*, 110(42), pp. 17059–17064. doi: 10.1073/pnas.1306070110.

Khodajou-Masouleh, Hamidreza, S. Shirin Shahangian, Farnoosh Attar, Reza H. Sajedi, and Behnam Rasti. 2021. "Characteristics, Dynamics and Mechanisms of Actions of Some Major Stress-Induced Biomacromolecules; Addressing Artemia as an Excellent Biological Model." *Journal of Biomolecular Structure & Dynamics* 39 (15). J Biomol Struct Dyn: 5619–37. doi:10.1080/07391102.2020.1796793.

Kimmel, R. A. and Meyer, D. (2010) *Molecular regulation of pancreas development in zebrafish*. Third Edit, *Methods in Cell Biology*. Third Edit. Elsevier Inc. doi: 10.1016/B978-0-12-384892-5.00010-4.

King, Allison M., and Thomas H. Macrae. 2015. "Insect Heat Shock Proteins during Stress and Diapause." *Annual Review of Entomology* 60 (January). Annu Rev Entomol: 59–75. doi:10.1146/ANNUREV-ENTO-011613-162107.

Koch, Bjørn E V, Shuxin Yang, Gerda Lamers, Jens Stougaard, and Herman P Spaink. 2019. "Intestinal Microbiome Adjusts the Innate Immune Setpoint during Colonization through Negative Regulation of MyD88." Accessed November 11. doi:10.1038/s41467-018-06658-4.

Koh, Ara, Filipe De Vadder, Petia Kovatcheva-Datchary, and Fredrik Bäckhed. 2016. "From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites." *Cell* 165 (6). Elsevier: 1332–45. doi:10.1016/J.CELL.2016.05.041.

Kramer, Jos, Özhan Özkaya, and Rolf Kümmerli. 2019. "Bacterial Siderophores in Community and Host Interactions." *Nature Reviews Microbiology 2019 18:3* 18 (3). Nature Publishing Group: 152–63. doi:10.1038/s41579-019-0284-4.

Krissinel, K. Henrick, Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **60**, 2256-2268 (2004).

Krissinel, and Henrick, K., in *Computational Life Sciences*, G. R. C. R. Berthold M., Diederichs K., Kohlbacher O., Fischer I., Ed. (Springer, Berlin, Heidelberg, 2005), vol. 3695, pp. 67-78.

Ku *et al.*, Predicting melting temperature directly from protein sequences. *Comput Biol Chem* **33**, 445-450 (2009).

Kulkarni, Devesha H., and Rodney D. Newberry. 2019. "Intestinal Macromolecular Transport Supporting Adaptive Immunity." *Cellular and Molecular Gastroenterology and Hepatology* 7 (4): 729–37. doi:10.1016/j.jcmgh.2019.01.003.

Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* **33**, 1870-1874 (2016).

Kurup, Naina, Yunbo Li, Alexandr Goncharov, and Yishi Jin. 2018. "Intermediate Filament Accumulation Can Stabilize Microtubules in Caenorhabditis Elegans Motor Neurons." *Proceedings of the National Academy of Sciences of the United States of America* 115 (12). National Academy of Sciences: 3114–19. doi:10.1073/PNAS.1721930115/VIDEO-4.

Lapointe, Tamia K., and Christophe Altier. 2011. "The Role of TRPA1 in Visceral Inflammation and Pain." *Channels* 5 (6). Taylor & Francis: 525. doi:10.4161/CHAN.5.6.18016.

Larsen, A. Grapin-Botton, The molecular and morphogenetic basis of pancreas organogenesis. *Semin Cell Dev Biol* **66**, 51-68 (2017).

Larsson, E. *et al.* (2012) 'Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88', *Gut*. BMJ Publishing Group, 61(8), pp. 1124–1131. doi: 10.1136/GUTJNL-2011-301104.

Leech, John M., Miqdad O. Dhariwala, Margaret M. Lowe, Kevin Chu, Geil R. Merana, Clémence Cornuot, Antonin Weckel, et al. 2019. "Toxin-Triggered Interleukin-1 Receptor Signaling Enables Early-Life Discrimination of Pathogenic versus Commensal Skin Bacteria." *Cell Host and Microbe* 26 (6). Cell Press: 795-809.e5. doi:10.1016/j.chom.2019.10.007.

Levy, C. A. Thaiss, E. Elinav, Metabolites: messengers between the microbiota and the immune system. *Genes Dev* **30**, 1589-1597 (2016).

Ley, R. E. *et al.* (2008) 'Worlds within worlds: Evolution of the vertebrate gut microbiota', *Nature Reviews Microbiology*, 6(10), pp. 776–788. doi: 10.1038/nrmicro1978.

Lickwar, Colin R, J Gray Camp, Matthew Weiser, Jordan L Cocchiaro, M Kingsley, Terrence S Furey, Shehzad Z Sheikh, and John F Rawls. 2017. *Genomic Dissection of Conserved Transcriptional Regulation in Intestinal Epithelial Cells*. doi:https://doi.org/10.1371/journal.pbio.2002054 August.

Liebschner *et al.*, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861-877 (2019).

Lopez, C.A., and Skaar, E.P. (2018). The Impact of Dietary Transition Metals on Host-Bacterial Interactions. Cell Host Microbe *23*, 737-748.

Lu, Van B., Fiona M. Gribble, and Frank Reimann. 2018. "Free Fatty Acid Receptors in Enteroendocrine Cells." *Endocrinology* 159 (7). Endocrinology: 2826–35. doi:10.1210/EN.2018-00261.

Mally, T. Otonkoski, A. D. Lopez, A. Hayek, Developmental gene expression in the human fetal pancreas. *Pediatr Res* **36**, 537-544 (1994).

Manfroid *et al.*, Zebrafish sox9b is crucial for hepatopancreatic duct development and pancreatic endocrine cell regeneration. *Dev Biol* **366**, 268-278 (2012).

MacVicar, Brian A., Leigh Wicki-Stordeur, and Louis Philippe Bernier. 2017. "The Cost of Communication in the Brain." *ELife*. NLM (Medline). doi:10.7554/eLife.27894.

Magnúsdóttir, Stefanía, Dmitry Ravcheev, Valérie de Crécy-Lagard, and Ines Thiele. 2015. "Systematic Genome Assessment of B-Vitamin Biosynthesis Suggests Co-Operation among Gut Microbes." *Frontiers in Genetics* 6 (April). Frontiers: 148. doi:10.3389/fgene.2015.00148.

Marchler-Bauer *et al.*, CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res* **45**, D200-D203 (2017).

Massaquoi, M. S. and Guillemin, K. (2018) 'Evolving in a Microbial Soup: You Are What They Eat', *Developmental Cell*. Cell Press, 47(6), pp. 682–683. doi: 10.1016/J.DEVCEL.2018.11.045.

Mcfall-Ngai, Margaret J. 2014. "MI68CH10-McFall-Ngai The Importance of Microbes in Animal Development: Lessons from the Squid-Vibrio Symbiosis." doi:10.1146/annurev-micro-091313-103654.

McLaughlin, R.N., Jr., and Malik, H.S. (2017). Genetic conflicts: the usual suspects and beyond. J Exp Biol *220*, 6-17.

McWilliam *et al.*, Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Res* **41**, W597-600 (2013).

Meijer, Annemarie H, S F Gabby Krens, Indira A Medina Rodriguez, Shuning He, Wilbert Bitter, B Ewa Snaar-Jagalska, and Herman P Spaink. 2004. "Expression Analysis of the Toll-like Receptor and TIR Domain Adaptor Families of Zebrafish." *Molecular Immunology* 40 (11): 773–83. doi:10.1016/j.molimm.2003.10.003.

Meisel, Jacquelyn S, Georgia Sfyroera, Casey Bartow-Mckenney, Ciara Gimblet, Julia Bugayev, Joseph Horwinski, Brian Kim, et al. n.d. "Commensal Microbiota Modulate Gene Expression in the Skin." doi:10.1186/s40168-018-0404-9.

Melancon, E., S. Gomez De La Torre Canny, S. Sichel, M. Kelly, T.J. Wiles, J.F. Rawls, J.S. Eisen, and K. Guillemin. 2017. "Best Practices for Germ-Free Derivation and Gnotobiotic Zebrafish Husbandry." In , 61–100. doi:10.1016/bs.mcb.2016.11.005.

Mempin, Roberto, Helen Tran, Connie Chen, Hao Gong, Katharina Kim Ho, and Sangwei Lu. 2013. "Release of Extracellular ATP by Bacteria during Growth." *BMC Microbiology* 13 (1). BioMed Central: 301. doi:10.1186/1471-2180-13-301.

Miller *et al*., Islet Formation during the Neonatal Development in Mice. *PLoS ONE* **4**, e7739 (2009).

Mishra, Sanjay, Shu Yu Wu, Alexandra W. Fuller, Zhen Wang, Kristie L. Rose, Kevin L. Schey, and Hassane S. Mchaourab. 2018. "Loss of Ab-Crystallin Function in Zebrafish Reveals Critical Roles in the Development of the Lens and Stress Resistance of the Heart." *Journal of Biological Chemistry* 293 (2). American Society for Biochemistry and Molecular Biology Inc.: 740–53.

Moreno-Hagelsieb, B. Vitug, A. Medrano-Soto, M. H. Saier, Jr., The Membrane Attack Complex/Perforin Superfamily. *J Mol Microbiol Biotechnol* **27**, 252-267 (2017).

Motulsky, R. E. Brown, Detecting outliers when fitting data with nonlinear regression - a new method based on robust nonlinear regression and the false discovery rate. *BMC Bioinformatics* **7**, 123 (2006).

Mukherjee *et al*., Antibacterial membrane attack by a pore-forming intestinal C-type lectin. *Nature* **505**, 103-107 (2014).

Mukherjee, L. V. Hooper, Antimicrobial defense of the intestine. *Immunity* **42**, 28-39 (2015).

Murdoch, Caitlin C., and John F. Rawls. 2019. "Commensal Microbiota Regulate Vertebrate Innate Immunity-Insights From the Zebrafish." *Frontiers in Immunology*. Frontiers Media S.A. doi:10.3389/fimmu.2019.02100.

Murtaugh, Pancreas and beta-cell development: from the actual to the possible. *Development* **134**, 427-438 (2007).

Ng, Annie N.Y., Tanya A. de Jong-Curtain, David J. Mawdsley, Sara J. White, Jimann Shin, Bruce Appel, P. Duc Si Dong, Didier Y.R. Stainier, and Joan K. Heath. 2005. "Formation of the Digestive System in Zebrafish: III. Intestinal Epithelium Morphogenesis." *Developmental Biology* 286 (1): 114–35. doi:10.1016/j.ydbio.2005.07.013.

Nishiyama, Hiroki, Tomoyuki Nagai, Masatoshi Kudo, Yoshihisa Okazaki, Yoshinao Azuma, Tomohiro Watanabe, Susumu Goto, Hiroyuki Ogata, and Toshiharu Sakurai. 2018. "Supplementation of Pancreatic Digestive Enzymes Alters the Composition of Intestinal Microbiota in Mice." *Biochemical and Biophysical Research Communications* 495 (1). Elsevier B.V.: 273–79. doi:10.1016/j.bbrc.2017.10.130.

Noguchi, M. O. Huising, Integrating the inputs that shape pancreatic islet hormone release. *Nat Metab* **1**, 1189-1201 (2019).

Novak, Ivana. 2003. "ATP as a Signaling Molecule: The Exocrine Focus." *News in Physiological Sciences* 18 (1): 12–17. doi:10.1152/nips.01409.2002.

Ohler, Anke, and Christoph Becker-Pauly. 2011. "Morpholino Knockdown of the Ubiquitously Expressed Transmembrane Serine Protease TMPRSS4a in Zebrafish Embryos Exhibits Severe Defects in Organogenesis and Cell Adhesion." *Biological Chemistry* 392 (7). De Gruyter: 653–64.

Okamoto, The Reg gene family and Reg proteins: with special attention to the regeneration of pancreatic beta-cells. *J Hepatobiliary Pancreat Surg* **6**, 254-262 (1999).

Orlando, Antonella, Michele Linsalata, Maria Notarnicola, Valeria Tutino, and Francesco Russo. 2014. "Lactobacillus GG Restoration of the Gliadin Induced Epithelial Barrier Disruption: The Role of Cellular Polyamines." *BMC Microbiology* 14 (1). BioMed Central: 19. doi:10.1186/1471-2180-14-19.

Ortwerth, B. J., and P. R. Olesen. 1992. "Characterization of the Elastase Inhibitor Properties of α-Crystallin and the Water-Insoluble Fraction from Bovine Lens." *Experimental Eye Research* 54 (1). Academic Press: 103–11. doi:10.1016/0014-4835(92)90074-3.

Otwinowski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).

Pais, R., Gribble, F. M. and Reimann, F. (2016) 'Stimulation of incretin secreting cells', *Therapeutic Advances in Endocrinology and Metabolism*. SAGE Publications Ltd, pp. 24–42. doi: 10.1177/2042018815618177.

Park, Jieun, Daniel S. Levic, Kaelyn D. Sumigray, Jennifer Bagwell, Oznur Eroglu, Carina L. Block, Cagla Eroglu, et al. 2019. "Lysosome-Rich Enterocytes Mediate Protein Absorption in the Vertebrate Gut." *Developmental Cell*. doi:10.1016/j.devcel.2019.08.001.

Pasquina-Lemonche *et al.*, The architecture of the Gram-positive bacterial cell wall. *Nature* **582**, 294-297 (2020).

Pauls, E. Zecchin, N. Tiso, M. Bortolussi, F. Argenton, Function and regulation of zebrafish nkx2.2a during development of pancreatic islet and ducts. *Dev Biol* **304**, 875-890 (2007).

Perfetti *et al.*, Regenerating (reg) and insulin genes are expressed in prepancreatic mouse embryos. *J Mol Endocrinol* **17**, 79-88 (1996).

Phadte, Ashutosh S., Zachary B. Sluzala, and Patrice E. Fort. 2021. "Therapeutic Potential of α-Crystallins in Retinal Neurodegenerative Diseases." *Antioxidants 2021, Vol. 10, Page 1001* 10 (7). Multidisciplinary Digital Publishing Institute: 1001. doi:10.3390/ANTIOX10071001.

Pietzner, Maik, Kathrin Budde, Malte Rühlemann, Henry Völzke, Georg Homuth, Frank U Weiss, Markus M Lerch, and Fabian Frost. 2021. "Exocrine Pancreatic Function Modulates Plasma Metabolites Through Changes in Gut Microbiota Composition." *The Journal of Clinical Endocrinology & Metabolism* XX (January). The Endocrine Society: 1–9. doi:10.1210/clinem/dgaa961.

Piri, Natik, Jacky M. K. Kwong, and Joseph Caprioli. 2013. "Crystallins in Retinal Ganglion Cell Survival and Regeneration." *Molecular Neurobiology* 48 (3). Springer US: 819–28. doi:10.1007/s12035-013-8470-2.

Piri, Natik, Jacky M. K. Kwong, and Joseph Caprioli. 2013. "Crystallins in Retinal Ganglion Cell Survival and Regeneration." *Molecular Neurobiology* 48 (3). Springer US: 819–28. doi:10.1007/s12035-013-8470-2.

Poitout *et al.*, Morphological and functional characterization of beta TC-6 cells--an insulin-secreting cell line derived from transgenic mice. *Diabetes* **44**, 306-313 (1995).

Pound *et al.*, Cathelicidin Antimicrobial Peptide: A Novel Regulator of Islet Function, Islet Regeneration, and Selected Gut Bacteria. *Diabetes* **64**, 4135-4147 (2015).

Pushalkar *et al.*, The Pancreatic Cancer Microbiome Promotes Oncogenesis by Induction of Innate and Adaptive Immune Suppression. *Cancer Discov* **8**, 403-416 (2018).

Quilici *et al.*, Solution Structure of the BPSL1445 Protein of Burkholderia pseudomallei Reveals the SYLF Domain Three-Dimensional Fold. *ACS Chem Biol*, (2021).

Rawls, John F, Buck S Samuel, and Jeffrey I Gordon. 2004. "Gnotobiotic Zebrafish Reveal Evolutionarily Conserved Responses to the Gut Microbiota." www.pnas.orgcgidoi10.1073pnas.0400706101.

Rawls, John F., Michael A. Mahowald, Ruth E. Ley, and Jeffrey I. Gordon. 2006. "Reciprocal Gut Microbiota Transplants from Zebrafish and Mice to Germ-Free Recipients Reveal Host Habitat Selection." *Cell* 127 (2): 423–33. doi:10.1016/j.cell.2006.08.043.

Reinhardt, Christoph, Christopher S Reigstad, and Fredrik Bäckhed. 2009. "Intestinal Microbiota during Infancy and Its Implications for Obesity." *Journal of Pediatric Gastroenterology and Nutrition* 48 (3): 249–56. doi:10.1097/mpg.0b013e318183187c.

Reischauer *et al*., Cloche is a bHLH-PAS transcription factor that drives haemato-vascular specification. *Nature* **535**, 294-298 (2016).

Reiss, Karina, and Sucharit Bhakdi. 2012. "Pore-Forming Bacterial Toxins and Antimicrobial Peptides as Modulators of ADAM Function." *Medical Microbiology and Immunology 2012 201:4* 201 (4). Springer: 419–26. doi:10.1007/S00430-012-0260-3.

Riquelme *et al*., Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes. *Cell* **178**, 795-806.e712 (2019).

Robertson *et al*., The WASP homologue Las17 activates the novel actin-regulatory activity of Ysc84 to promote endocytosis in yeast. *Mol Biol Cell* **20**, 1618-1628 (2009).

Robinson, Catherine D., Helena S. Klein, Kyleah D. Murphy, Raghuveer Parthasarathy, Karen Guillemin, and Brendan J. M. Bohannan. 2018. "Experimental Bacterial Adaptation to the Zebrafish Gut Reveals a Primary Role for Immigration." Edited by Jeff Gore. *PLOS Biology* 16 (12). Public Library of Science: e2006893. doi:10.1371/journal.pbio.2006893.

Rolig, Annah S., Emily Goers Sweeney, Lila E. Kaye, Michael D. Desantis, Arden Perkins, Allison V. Banse, M. Kristina Hamilton, and Karen Guillemin. 2018. "A Bacterial Immunomodulatory Protein with Lipocalin-like Domains Facilitates Host–Bacteria Mutualism in Larval Zebrafish." *ELife* 7 (November). eLife Sciences Publications Ltd. doi:10.7554/eLife.37172.

Rolig, Annah S., Erika K. Mittge, Julia Ganz, Josh V. Troll, Ellie Melancon, Travis J. Wiles, Kristin Alligood, W. Zac Stephens, Judith S. Eisen, and Karen Guillemin. 2017. "The Enteric Nervous System Promotes Intestinal Health by Constraining Microbiota Composition." Edited by Jeff Gore. *PLOS Biology* 15 (2): e2000689. doi:10.1371/journal.pbio.2000689.

Rolig, Annah S., Raghuveer Parthasarathy, Adam R. Burns, Brendan J.M. Bohannan, and Karen Guillemin. 2015. "Individual Members of the Microbiota Disproportionately Modulate Host Innate Immune Responses." *Cell Host and Microbe* 18 (5). Cell Press: 613–20. doi:10.1016/j.chom.2015.10.009.

Romero, Alejandra, and Juergen Eckel. 2021. "Cells Organ Crosstalk and the Modulation of Insulin Signaling." doi:10.3390/cells10082082.

Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* **5**, 725-738 (2010).

RStudioTeam. (Boston, MA, 2020).

Rühl, Sebastian, Kateryna Shkarina, Benjamin Demarco, Rosalie Heilig, José Carlos Santos, and Petr Broz. 2018. "ESCRT-Dependent Membrane Repair Negatively Regulates Pyroptosis Downstream of GSDMD Activation." *Science* 362 (6417). American Association for the Advancement of Science: 956–60. doi:10.1126/SCIENCE.AAR7607/SUPPL_FILE/AAR7607_S8.MP4.

Sala *et al*., Type 2 Diabetes Remission After Roux-en-Y Gastric Bypass: Evidence for Increased Expression of Jejunal Genes Encoding Regenerating Pancreatic Islet-Derived Proteins as a Potential Mechanism. *Obes Surg* **27**, 1123-1127 (2017).

Saliba, Antoine Emmanuel, Lei Li, Alexander J. Westermann, Silke Appenzeller, Daphne A.C. Stapels, Leon N. Schulte, Sophie Helaine, and Jörg Vogel. 2016. "Single-Cell RNA-Seq Ties Macrophage Polarization to Growth Rate of Intracellular Salmonella." *Nature Microbiology* 2 (2). Nature Publishing Group: 1–8. doi:10.1038/nmicrobiol.2016.206.

Sampson, Timothy R., and Sarkis K. Mazmanian. 2015. "Control of Brain Development, Function, and Behavior by the Microbiome." *Cell Host and Microbe*. doi:10.1016/j.chom.2015.04.011.

Sanchez, K.K., Chen, G.Y., Schieber, A.M.P., Redford, S.E., Shokhirev, M.N., Leblanc, M., Lee, Y.M., and Ayres, J.S. (2018). Cooperative Metabolic Adaptations in the Host Can Favor Asymptomatic Infection and Select for Attenuated Virulence in an Enteric Pathogen. Cell *175*, 146-158 e115.

Sancho, R., Cremona, C. A. and Behrens, A. (2015) 'Stem cell and progenitor fate in the mammalian intestine: Notch and lateral inhibition in homeostasis and disease', *EMBO reports*. EMBO, 16(5), pp. 571–581. doi: 10.15252/embr.201540188.

Scharschmidt, Tiffany C., Kimberly S. Vasquez, Hong An Truong, Sofia V. Gearty, Mariela L. Pauli, Audrey Nosbaum, Iris K. Gratz, et al. 2015. "A Wave of Regulatory T Cells into Neonatal Skin Mediates Tolerance to Commensal Microbes." *Immunity* 43 (5). Cell Press: 1011–21. doi:10.1016/j.immuni.2015.10.016.

Scharschmidt, Tiffany C., Kimberly S. Vasquez, Mariela L. Pauli, Elizabeth G. Leitner, Kevin Chu, Hong An Truong, Margaret M. Lowe, et al. 2017. "Commensal Microbes and Hair Follicle Morphogenesis Coordinately Drive Treg Migration into Neonatal Skin." *Cell Host and Microbe* 21 (4). Cell Press: 467-477.e5. doi:10.1016/j.chom.2017.03.001.

Schlomann, B. H. *et al*. (2018) 'Bacterial cohesion predicts spatial distribution in the larval zebrafish intestine', *bioRxiv*. Cold Spring Harbor Laboratory, p. 392316. doi: 10.1101/392316.

Schroeder, B. O. and Bäckhed, F. (2016) 'Signals from the gut microbiota to distant organs in physiology and disease', *Nature Medicine*. Nature Publishing Group, pp. 1079–1089. doi: 10.1038/nm.4185.

Schütt, Florian, Sebastian Aretz, Gerd U. Auffarth, and Jürgen Kopitz. 2012. "Moderately Reduced ATP Levels Promote Oxidative Stress and Debilitate Autophagic and Phagocytic Capacities in Human RPE Cells." *Investigative Ophthalmology and Visual Science* 53 (9): 5354–61. doi:10.1167/iovs.12-9845.

Seeger, Angela, Werner E. Mayer, and Jan Klein. 1996. "A Complement Factor B-like CDNA Clone from the Zebrafish (Brachydanio Rerio)." *Molecular Immunology* 33 (6). Pergamon: 511–20. doi:10.1016/0161-5890(96)00002-8.

Seleme, Maria C., Kate Kosmac, Stipan Jonjic, and William J. Britt. 2017. "Tumor Necrosis Factor Alpha-Induced Recruitment of Inflammatory Mononuclear Cells Leads to Inflammation and Altered Brain Development in Murine Cytomegalovirus-Infected Newborn Mice." *Journal of Virology* 91 (8). American Society for Microbiology. doi:10.1128/jvi.01983-16.

Semova, Ivana, Juliana D. Carten, Jesse Stombaugh, Lantz C. MacKey, Rob Knight, Steven A. Farber, and John F. Rawls. 2012. "Microbiota Regulate Intestinal Absorption and Metabolism of Fatty Acids in the Zebrafish." *Cell Host and Microbe* 12 (3). Cell Press: 277–88. doi:10.1016/j.chom.2012.08.003.

Shepherd, Iain, and Judith Eisen. n.d. "Development of the Zebrafish Enteric Nervous System." doi:10.1016/B978-0-12-387036-0.00006-2.

Shih, Li Jane, Yu Fen Lu, Yi Hua Chen, Chia Chi Lin, Jian An Chen, and Sheng Ping L. Hwang. 2007. "Characterization of the Agr2 Gene, a Homologue of X. Laevis Anterior Gradient 2, from the Zebrafish, Danio Rerio." *Gene Expression Patterns* 7 (4). Elsevier: 452–60. doi:10.1016/j.modgep.2006.11.003.

Sievers *et al.*, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

Sikander, A., Rana, S. V. and Prasad, K. K. (2009) 'Role of serotonin in gastrointestinal motility and irritable bowel syndrome', *Clinica Chimica Acta*, 403(1–2), pp. 47–55. doi: 10.1016/j.cca.2009.01.028.

Sjö, K. *et al*. (2012) 'The Gut Microbiota Regulates Bone Mass in Mice'. doi: 10.1002/jbmr.1588.

Slingsby, Christine, and Graeme J. Wistow. 2014. "Functions of Crystallins in and out of Lens: Roles in Elongated and Post-Mitotic Cells." *Progress in Biophysics and Molecular Biology* 115 (1). Pergamon: 52–67. doi:10.1016/J.PBIOMOLBIO.2014.02.006.

Sørensen, Christiane E, and Ivana Novak. 2001. "Visualization of ATP Release in Pancreatic Acini in Response to Cholinergic Stimulus USE OF FLUORESCENT PROBES AND CONFOCAL MICROSCOPY*." doi:10.1074/jbc.M103313200.

Spencer, Nick J., and Hongzhen Hu. 2020. "Enteric Nervous System: Sensory Transduction, Neural Circuits and Gastrointestinal Motility." *Nature Reviews*. *Gastroenterology & Hepatology* 17 (6). Nat Rev Gastroenterol Hepatol: 338–51. doi:10.1038/S41575-020-0271-2.

Sprague-Piercy, Marc A., Megan A. Rocha, Ashley O. Kwok, and Rachel W. Martin. 2021. "α-Crystallins in the Vertebrate Eye Lens: Complex Oligomers and Molecular Chaperones." *Annual Review of Physical Chemistry* 72 (April). Annu Rev Phys Chem: 143–63. doi:10.1146/ANNUREV-PHYSCHEM-090419-121428.

Stagaman, Keaton, Adam R Burns, Karen Guillemin, and Brendan Jm Bohannan. 2017. "The Role of Adaptive Immunity as an Ecological Filter on the Gut Microbiota in Zebrafish." *The ISME Journal Advance Online Publication* 17 (10). doi:10.1038/ismej.2017.28.

Stainier, B. M. Weinstein, H. W. Detrich, 3rd, L. I. Zon, M. C. Fishman, Cloche, an early acting zebrafish gene, is required by both the endothelial and hematopoietic lineages. *Development* **121**, 3141-3150 (1995).

Stapels, Daphne A.C., Brian V. Geisbrecht, and Suzan H.M. Rooijakkers. 2015. "Neutrophil Serine Proteases in Antibacterial Defense." *Current Opinion in Microbiology* 23. Elsevier Ltd: 42–48. doi:10.1016/j.mib.2014.11.002.

Stappenbeck, T. S., Hooper, L. V and Gordon, J. I. (no date) *Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells*. Available at: www.pnas.orgcgidoi10.1073pnas.202604299 (Accessed: 4 March 2019).

Stecher, Bärbel, and Wolf Dietrich Hardt. 2011. "Mechanisms Controlling Pathogen Colonization of the Gut." *Current Opinion in Microbiology* 14 (1). Elsevier Ltd: 82–91. doi:10.1016/j.mib.2010.10.003.

Stephens, W. Z. *et al*. (2015) 'Identification of population bottlenecks and colonization factors during assembly of bacterial communities within the zebrafish intestine', *mBio*. American Society for Microbiology, 6(6). doi: 10.1128/mBio.01163-15.

Stephens, W. Zac, Adam R. Burns, Keaton Stagaman, Sandi Wong, John F. Rawls, Karen Guillemin, and Brendan J.M. Bohannan. 2016. "The Composition of the Zebrafish Intestinal Microbial Community Varies across Development." *ISME Journal* 10 (3). Nature Publishing Group: 644–54. doi:10.1038/ismej.2015.140.

Strachan, Hay fever, hygiene, and household size. *BMJ* **299**, 1259-1260 (1989).

Straley, M. E., K. L. Togher, A. M. Nolan, L. C. Kenny, and G. W. O'Keeffe. 2014. "LPS Alters Placental Inflammatory and Endocrine Mediators and Inhibits Fetal Neurite Growth in Affected Offspring during Late Gestation." *Placenta* 35 (8). W.B. Saunders Ltd: 533–38. doi:10.1016/j.placenta.2014.06.001.

Stuart, Tim, Andrew Butler, Paul Hoffman, Christoph Hafemeister, Efthymia Papalexi, William Mauck, Marlon Stoeckius, Peter Smibert, and Rahul Satija. 2018. "Comprehensive Integration of Single Cell Data." *BioRxiv*, 460147. doi:10.1101/460147.

Stubbington, Michael J T, Orit Rozenblatt-Rosen, Aviv Regev, and Sarah A Teichmann. 2021. "Single-Cell Transcriptomics to Explore the Immune System in Health and Disease." Accessed February 2. http://science.sciencemag.org/.

Sutipatanasomboon *et al.*, Disruption of the plant-specific CFS1 gene impairs autophagosome turnover and triggers EDS1-dependent cell death. *Sci Rep* **7**, 8677 (2017).

Swanson, Karen V., Meng Deng, and Jenny P.Y. Ting. 2019. "The NLRP3 Inflammasome: Molecular Activation and Regulation to Therapeutics." *Nature Reviews Immunology 2019 19:8* 19 (8). Nature Publishing Group: 477–89. doi:10.1038/s41577-019-0165-0.

Taylor, Charlotte R., William A. Montagne, Judith S. Eisen, and Julia Ganz. 2016. "Molecular Fingerprinting Delineates Progenitor Populations in the Developing Zebrafish Enteric Nervous System." *Developmental Dynamics* 245 (11): 1081–96. doi:10.1002/dvdy.24438.

Terazono K, Yamamoto H, Takasawa S, Shiga K, Yonemura Y, Tochino Y, Okamoto H. A novel gene activated in regenerating islets. J Biol Chem. 1988 Feb 15;263(5):2111-4. PMID: 2963000.

Terman, Alexei, Tino Kurz, Marian Navratil, Edgar A. Arriaga, and Ulf T. Brunk. 2010. "Mitochondrial Turnover and Aging of Long-Lived Postmitotic Cells: The Mitochondrial-Lysosomal Axis Theory of Aging." *Antioxidants & Redox Signaling* 12 (4). Antioxid Redox Signal: 503–35. doi:10.1089/ARS.2009.2598.

Thompson, Luke, Jon G Sanders, Daniel mcDonald, Amnon Amir, Joshua Ladau, Kenneth J Locey, robert J Prill, et al. 2017. "Donna Berg-Lyons 19 , Valerie McKenzie 20 , Noah Fierer 20,21 , Jed A. Fuhrman 22 , Aaron Clauset 19,23 , Rick L. Stevens 24,25 , Ashley Shade 26" 18: 28. doi:10.1038/nature24621.

Tilg, P. D. Cani, E. A. Mayer, Gut microbiome and liver diseases. *Gut* **65**, 2035-2044 (2016).

Tolhurst, G., Reimann, F. and Gribble, F. M. (2012) 'Intestinal sensing of nutrients', *Handbook of Experimental Pharmacology*, 209, pp. 309–335. doi: 10.1007/978-3-642-24716-3_14.

Tonikian *et al.*, Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. *PLoS Biol* **7**, e1000218 (2009).

Tremaroli, Valentina, and Fredrik Bäckhed. 2012. "Functional Interactions between the Gut Microbiota and Host Metabolism." *Nature*. doi:10.1038/nature11552.

Troll, J.V., Hamilton, M.K., Abel, M.L., Ganz, J., Bates, J.M., Stephens, W.Z., Melancon, E., van der Vaart, M., Meijer, A.H., Distel, M., *et al*. (2018). Microbiota promote secretory cell determination in the intestinal epithelium by modulating host Notch signaling. Development *145*.

Udayangani, R. M.C., S. H.S. Dananjaya, Chamilani Nikapitiya, Gang Joon Heo, Jehee Lee, and Mahanama De Zoysa. 2017. "Metagenomics Analysis of Gut Microbiota and Immune Modulation in Zebrafish (Danio Rerio) Fed Chitosan Silver Nanocomposites." *Fish and Shellfish Immunology* 66 (July). Academic Press: 173–84. doi:10.1016/j.fsi.2017.05.018.

Unno *et al*., Production and characterization of Reg knockout mice: reduced proliferation of pancreatic beta-cells in Reg knockout mice. *Diabetes* **51 Suppl 3**, S478-483 (2002).

Urbanek, R. Chan, K. R. Ayscough, Function and interactions of the Ysc84/SH3yl1 family of actin- and lipid-binding proteins. *Biochem Soc Trans* **43**, 111-116 (2015).

Vaishnava, S., C. L. Behrendt, A. S. Ismail, L. Eckmann, and L. V. Hooper. 2008. "Paneth Cells Directly Sense Gut Commensals and Maintain Homeostasis at the Intestinal Host-Microbial Interface." *Proceedings of the National Academy of Sciences* 105 (52): 20858–63. doi:10.1073/pnas.0808723105.

Valapala, M, · S Hose, · J S Zigler, · D Sinha, Y Sergeev, E Wawrousek, S Hose, et al. 2016. "Modulation of V-ATPase by BA3/A1-Crystallin in Retinal Pigment Epithelial Cells." *Advances in Experimental Medicine and Biology* 854 (October). Springer, Cham: 779–84. doi:10.1007/978-3-319-17121-0_104.

van der Sar, A. M. *et al*. (2006) 'MyD88 innate immune function in a zebrafish embryo infection model.', *Infection and immunity*, 74(4), pp. 2436–41. doi: 10.1128/IAI.74.4.2436-2441.2006.

Van Duyne, R. F. Standaert, P. A. Karplus, S. L. Schreiber, J. Clardy, Atomic Structures of the Human Immunophilin FKBP-12 Complexes with FK506 and Rapamycin. *Journal of Molecular Biology* **229**, 105-124 (1993).

Veatch, Electro-formation and fluorescence microscopy of giant vesicles with coexisting liquid phases. *Methods Mol Biol* **398**, 59-72 (2007).

Verderio, Claudia, and Michela Matteoli. 2011. "ATP in Neuron-Glia Bidirectional Signalling." *Brain Research Reviews*. Elsevier. doi:10.1016/j.brainresrev.2010.04.007.

Verkhratsky, Alexei, and Geoffrey Burnstock. 2014. "Biology of Purinergic Signalling: Its Ancient Evolutionary Roots, Its Omnipresence and Its Multiple Functional Significance." *BioEssays* 36 (7). John Wiley and Sons Inc.: 697–705. doi:10.1002/bies.201400024.

Vetere, B. K. Wagner, Chemical methods to induce Beta-cell proliferation. *Int J Endocrinol* **2012**, 925143 (2012).

Vincent, K. M., Sharp, J. W. and Raybould, H. E. (2011) 'Intestinal glucose-induced calcium-calmodulin kinase signaling in the gut-brain axis in awake rats', *Neurogastroenterology and Motility*, 23(7). doi: 10.1111/j.1365-2982.2011.01673.x.

Vives-Pi *et al.*, Evidence of expression of endotoxin receptors CD14, toll-like receptors TLR4 and TLR2 and associated molecule MD-2 and of sensitivity to endotoxin (LPS) in islet beta cells. *Clin Exp Immunol* **133**, 208-218 (2003).

Walther, Barbara, J. Philip Karl, Sarah L. Booth, and Patrick Boyaval. 2013. "Menaquinones, Bacteria, and the Food Supply: The Relevance of Dairy and Fermented Food Products to Vitamin K Requirements." *Advances in Nutrition* 4 (4). Oxford University Press: 463–73. doi:10.3945/an.113.003855.

Wan, Haiyan, Svitlana Korzh, Zhen Li, Sudha Puttur Mudumana, Vladimir Korzh, Yun Jin Jiang, Shuo Lin, and Zhiyuan Gong. 2006. "Analyses of Pancreas Development by Generation of Gfp Transgenic Zebrafish Using an Exocrine Pancreas-Specific ElastaseA Gene Promoter." *Experimental Cell Research* 312 (9). Academic Press: 1526–39. doi:10.1016/J.YEXCR.2006.01.016.

Wang *et al.*, Intestinal REG3 Lectins Protect against Alcoholic Steatohepatitis by Reducing Mucosa-Associated Microbiota and Preventing Bacterial Translocation. *Cell Host Microbe* **19**, 227-239 (2016).

Watanabe, Takuo, Yutaka Yonemura, Hideto Yonekura, Yoshihisa Suzuki, Hikari Miyashita, Kazuo Sugiyama, Shigeki Moriizumi, et al. 1994. "Pancreatic Beta-Cell Replication and Amelioration of Surgical Diabetes by Reg Protein." *Proceedings of the National Academy of Sciences* 91 (9). National Academy of Sciences: 3589–92. doi:10.1073/PNAS.91.9.3589.

Weiss, L.C., Albada, B., Becker, S.M., Meckelmann, S.W., Klein, J., Meyer, M., Schmitz, O.J., Sommer, U., Leo, M., Zagermann, J., *et al*. (2018). Identification of Chaoborus kairomone chemicals that induce defences in Daphnia. Nat Chem Biol *14*, 1133-1139.

Wen, Jia, Gilberto Padilla Mercado, Alyssa Volland, Heidi L. Doden, Colin R. Lickwar, Taylor Crooks, Genta Kakiyama, et al. 2021. "Fxr Signaling and Microbial Metabolism of Bile Salts in the Zebrafish Intestine." *Science Advances* 7 (30). American Association for the Advancement of Science. doi:10.1126/SCIADV.ABG1371.

Wiles, J. M. Bower, M. J. Redd, M. A. Mulvey, Use of Zebrafish to Probe the Divergent Virulence Potentials and Toxin Requirements of Extraintestinal Pathogenic Escherichia coli. *PLoS Pathogens* **5**, e1000697 (2009).

Wiles, Travis J, Matthew Jemielita, Ryan P Baker, Brandon H Schlomann, Savannah L Logan, Julia Ganz, Ellie Melancon, Judith S Eisen, Karen Guillemin, and Raghuveer Parthasarathy. 2016. "Host Gut Motility Promotes Competitive Exclusion within a Model Intestinal Microbiota." Edited by Jeff Gore. *PLOS Biology* 14 (7). Public Library of Science: e1002517. doi:10.1371/journal.pbio.1002517.

Wiles, T. J. *et al*. (2018) 'Modernized tools for streamlined genetic manipulation and comparative study of wild and diverse proteobacterial lineages', *mBio*. American Society for Microbiology, 9(5). doi: 10.1128/mBio.01877-18.

Wiles, Travis J., and Karen Guillemin. 2020. "Patterns of Partnership: Surveillance and Mimicry in Host-Microbiota Mutualisms." *Current Opinion in Microbiology*. Elsevier Ltd. doi:10.1016/j.mib.2020.01.012.

Wiles, Travis J., Brandon H. Schlomann, Elena S. Wall, Reina Betancourt, Raghuveer Parthasarathy, and Karen Guillemin. 2020. "Swimming Motility of a Gut Bacterial Symbiont Promotes Resistance to Intestinal Expulsion and Enhances Inflammation." *PLoS Biology* 18 (3). Public Library of Science. doi:10.1371/journal.pbio.3000661.

Wistow, Graeme, Lesley Summers, and Tom Blundell. 1985. "Myxococcus Xanthus Spore Coat Protein S May Have a Similar Structure to Vertebrate Lens Bγ-Crystallins." *Nature* 315 (6022): 771–73. doi:10.1038/315771A0.

Wu, A. M. *et al*. (2009) 'Lectins as tools in glycoconjugate research.', *Glycoconjugate journal*, 26(8), pp. 899–913. doi: 10.1007/s10719-008-9119-7.

Wywial, S. M. Singh, Identification and structural characterization of FYVE domain-containing proteins of Arabidopsis thaliana. *BMC Plant Biol* **10**, 157 (2010).

Yang *et al*., Pancreatic stone protein/regenerating protein (PSP/reg): a novel secreted protein up-regulated in type 2 diabetes mellitus. *Endocrine* **48**, 856-862 (2015).

Yang *et al*., The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7-8 (2015).

Yassour *et al*., Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host Microbe* **24**, 146-154 e144 (2018).

Ye, Lihua, Munhyung Bae, Chelsi Cassilly, Sairam Jabba, Daniel Thorpe, Alyce Martin, Hsiu-Yi Lu, et al. 2020. "Enteroendocrine Cells Sense Bacterial Tryptophan Catabolites to Activate Enteric and Vagal Neuronal Pathways." *Cell Host and Microbe*. Elsevier Inc., 1–18. doi:10.1101/2020.06.09.142133.

Zhang, I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).

Zhang *et al*., Intestinal lysozyme liberates Nod1 ligands from microbes to direct insulin trafficking in pancreatic beta cells. *Cell Res* **29**, 516-532 (2019).

Zhang *et al*., Maternal cecal microbiota transfer rescues early-life antibiotic-induced enhancement of type 1 diabetes in mice. *Cell Host Microbe* **29**, 1249-1265 e1249 (2021)

Zhu, Xiao Hong, Hongyan Qiao, Fei Du, Qiang Xiong, Xiao Liu, Xiaoliang Zhang, Kamil Ugurbil, and Wei Chen. 2012. "Quantitative Imaging of Energy Expenditure in Human Brain." *NeuroImage* 60 (4). NIH Public Access: 2107–17. doi:10.1016/j.neuroimage.2012.02.013.

Zhu, Zhihui, and Georg Reiser. 2018. "The Small Heat Shock Proteins, Especially HspB4 and HspB5 Are Promising Protectants in Neurodegenerative Diseases." *Neurochemistry International* 115 (May). Pergamon: 69–79. doi:10.1016/J.NEUINT.2018.02.006.

Zigler, J. Samuel, and Debasish Sinha. 2015. "BA3/A1-Crystallin: More than a Lens Protein." *Progress in Retinal and Eye Research* 44 (January). Prog Retin Eye Res: 62–85. doi:10.1016/J.PRETEYERES.2014.11.002