Elucidating regulatory mechanisms shaping the transcriptome and proteome of the gut immune response in *Drosophila melanogaster*

Thèse N° 9723

Présentée le 27 septembre 2019

à la Faculté des sciences de la vie
Unité du Prof. Deplancke
Programme doctoral en biotechnologie et génie biologique

pour l'obtention du grade de Docteur ès Sciences

par

Michael Vincent FROCHAUX

Acceptée sur proposition du jury

Prof. A. L. A. Persat, président du jury
Prof. B. Deplancke, directeur de thèse
Prof. F. Leulier, rapporteur
Prof. M. C. Gambetta, rapporteuse
Prof. B. Lemaître, rapporteur

2019

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Version: 1st draft

Name: Michael Vincent Frochaux
N° SCIPER: 174580

Private defense : July 17th, 2019
Place : Lausanne, Switzerland

Public defense : to be determined
Place : Lausanne, Switzerland

University : EPFL
Faculty : SV-IBI
Laboratory : UPDE
Doctoral program : EDBB

Thesis committee

Thesis Director:	Prof. Bart DEPLANCKE	EPFL, Lausanne, Switzerland
President:	Prof. Alexandre PERSAT	EPFL, Lausanne, Switzerland
Experts	Prof. Bruno LEMAITRE	EPFL, Lausanne, Switzerland
	Prof. François LEULIER	CNRS, Lyon, France
	Prof. Maria Cristina GAMBETTA	UNIL, Lausanne, Switzerland

“Time is a drug. Too much of it kills you.”
— Terry Pratchett, Small Gods

“We make choices everyday, some of them good, some of them bad. And - if we are strong enough - we live with the consequences.”
— David Gemmell

“Do, or do not. There is no try.”
— Yoda

Summary

Work on *Drosophila melanogaster* paved the way to our current understanding of modern genetics. Since then, this model organism has contributed greatly to various fields such as neurobiology, development, and immunology. The discovery and analysis of the various pathways of the *Drosophila* immune response led the way to the in-depth characterization of the gut immune response. Furthermore, a genome-wide association study on a population of Drosophila identified resistance to infection by the entomopathogenic bacteria *Pseudomonas entomophila* (*P.e.*) as a complex phenotype, with highly resistant and susceptible lines, and highlighted genes and pathways mediating inter-individual differences. However, the molecular mechanisms mediating the resistance to enteric infection remain largely uncharacterized. This study uses the same Drosophila population to analyze what are the molecular mechanisms mediating the resistance to enteric infection. We analyzed the genetic determinants of gene expression variation among the most resistant and susceptible lines in response to infection using expression quantitative trait loci (eQTL) analysis. We also characterized the mode of action of these eQTLs into cis- and trans-acting. Furthermore, we identified the gene *nutcracker* (*ntc*) as the only differentially-expressed gene between resistance classes. We then demonstrate that loss of function *ntc* mutants are significantly more susceptible to infection and then identify a single nucleotide polymorphism (SNP) located in a transcription factor binding site (TFBS) which affects the binding affinity of the transcription factor Broad leading to allele-specific expression variation of the gene *ntc*.

If the transcriptomic response to infection has been thoroughly characterized, it is not the case for the proteomic response. We thus sought to characterize for the first time the proteomic response of the *Drosophila* to enteric infection. We performed mass spectrometry and RNA sequencing on dissected guts from flies infected by two Gram-negative bacteria, *Erwinia carotovora carotovora 15* (Ecc15) or *Pseudomonas entomophila,* 4h and 16h after infection. We found that a large portion of the measurable proteome (12%) varies after infection and that protein changes are strongly time- and infection-dependent. We showed the relatively poor correlation between gene expression and protein abundance in the *Drosophila* enteric immune response. Finally, we performed a screen of several proteins that were identified in our proteomics work but had previously not been found when assessing the gene expression response to infection using loss-of-function mutants and identifying 7 genes that modulate overall susceptibility to *P.e.* infection.

In summary, this work analyze the genetic determinants of gene expression variation in the DGRP population upon infection and characterize them according to their mode of action. Furthermore, we describe how a non-coding variant lowers resistance to infection by modulating *ntc* gene expression through altered Broad repressor binding.

Finally, it provides the first comprehensive characterization of the *Drosophila* gut proteome upon infection by Gram-negative bacteria which can be the basis of future proteomic analysis of the enteric immune response

Keywords

Systems Genetics – Genomics – Proteomics – QTL – Natural Variation – *Drosophila melanogaster* – DGRP– Immune response – Gut – Allele specific expression

Résumé

Les travaux sur *Drosophila melanogaster* ont ouvert la voie à notre compréhension actuelle de la génétique moderne. Depuis lors, cet organisme modèle a grandement contribué à divers domaines tels que la neurobiologie, le développement et l'immunologie. La découverte et l'analyse des différentes voies de régulation de la réponse immunitaire ont ouvert la voie à une caractérisation approfondie de la réponse immunitaire de l'intestin. En outre, une étude d'association pangénomique sur une population de Drosophila (DGRP) a identifié la résistance à l'infection par la bactérie entomopathogène *Pseudomonas entomophila* (*P.e.*) comme un phénotype complexe, avec des lignées hautement résistantes et sensibles, et a mis en évidence des gènes et des voies de transmission des différences interindividuelles. Cependant, les mécanismes moléculaires impliqués dans la résistance à l'infection entérique restent largement inconnus. Cette étude utilise la même population de drosophiles pour analyser quels sont les mécanismes moléculaires impliqués dans la résistance à l'infection entérique. Nous avons analysé les déterminants génétiques de la variation de l'expression des gènes parmi les lignées les plus résistantes et les plus sensibles en réponse à une infection, en utilisant une analyse des locus d'expression de caractère quantitatif (LeCQ). Nous avons également caractérisé le mode d'action, *cis*- ou *trans*-, de ces LeCQ. De plus, nous avons identifié le gène *nutcracker* (*ntc*) comme étant le seul gène exprimé de manière différentielle entre les classes de résistance. Nous démontrons ensuite que les mouches mutantes pour *ntc* sont significativement plus sensibles à l'infection, puis nous identifions un polymorphisme d'un seul nucléotide (PSN) situé dans un site de liaison au facteur de transcription (SLFT) qui affecte l'affinité de liaison du facteur de transcription Broad conduisant à une variation spécifique de chaque allèle du gène ntc.

Si la réponse transcriptomique à l'infection a été complètement caractérisée, ce n'est pas le cas pour la réponse protéinique. Nous avons donc cherché à caractériser pour la première fois la réponse protéinique de la drosophile à une infection entérique. Nous avons utilisé la spectrométrie de masse et le séquençage de l'ARN sur les intestins disséqués de mouches infectées par deux bactéries Gram-négative, *Erwinia carotovora carotovora 15* (Ecc15) or *Pseudomonas entomophila* 4h et 16h après l'infection. Nous avons constaté qu'une grande partie du protéome mesurable (12%) varie après l'infection et que les modifications protéiques dépendent fortement du temps et de l'infection. Nous avons montré la corrélation relativement faible entre l'expression des gènes et l'abondance des protéines dans la réponse immunitaire entérique de Drosophila. Enfin, nous avons effectué un criblage de plusieurs protéines identifiées dans notre travail de protéinique mais qui n'avaient pas été découvertes auparavant

lors de l’évaluation de la réponse de l’expression génique à une infection à l’aide de plusieurs mutantes pertes de fonction identifiant 7 gènes modulant la susceptibilité globale à l’infection par *P.e*.

En résumé, ces travaux analysent les déterminants génétiques de la variation de l'expression des gènes dans la population de DGRP après infection et les caractérisent en fonction de leur mode d'action. En outre, nous décrivons comment un variant non codant réduit la résistance à l'infection en modulant l'expression du gène *ntc* par le biais d'une liaison de répresseur large modifiée.

Enfin, il fournit la première caractérisation complète du protéome de l'intestin de Drosophila après infection par une bactérie à Gram négatif, qui peut constituer la base d'une future analyse protéinique de la réponse immunitaire entérique.

Mots-clefs
Génétique des systèmes – Génomique– Protéomique – LCQ – Variation naturel – *Drosophila melanogaster* – DGRP– réponse immunitaire – Gut – Expression spécifique à un allèle

Acknowledgments

I would first like to thank Pr. Bart Deplancke for his supervision, guidance and kind words throughout my thesis.

I wish to express my gratitude to the members of the jury, Prof. Alexandre Persat, Prof Bruno Lemaître, Prof. François Leulier and Prof. Maria Cristina Gambetta for their constructive criticism and discussion of my work.

I want also to warmly thank the Drosophila Crew of the lab, Dr. Roel Bevers, Virginie Braman, Masha Litovchenko, Dr. Brian Hollis, and Sakshi Sharda. A special thanks as well to lab alumni Dr. Maroun Bou Sleiman who has been extremely helpful and supportive during my work. We had a great collaboration during these years that lead to several papers co-authored and a lasting friendship. Special thanks goes to Dr. Roel Bevers. We started our PhD together, but you were faster. We shared a lot of experiences and coffee as well as many discussion work and not work related. A huge thanks as well for Virginie Braman who helped me a lot when I was doing large experiment and for her friendship throughout my PhD.
I want to acknowledge as well the help and support of the rest of the Deplancke lab, especially Dr. Vincent Gardeux for his help with everything bioinformatics related and for driving me to Yverdon, Julie Russeil who smoothly runs the lab, help us find common consumable in the most obvious place and laugh at my jokes. Magali Masson and Aurélia Tacchini for helping us organize ourselves. Special thanks to Riccardo Dainese, Marjan Biocanin, Johannes Bues and Jörn Pezoldt for sharing ideas, laugh, beers, memes and epic meal time. Thanks to Magda for her kind words and encouragement. Finally, huge thanks to the rest of the lab for the friendly atmosphere and to lab alumni, especially Alina, Carine and Jean-Daniel.

To my friends, in no special orders. Thank you Adrian for these roughly 20 years of unending friendship. We don’t see each other as often as I’d like, but it is always as if we just saw each other the day before. To Andrea, my partner in SG cafeteria coffee and gaming, board or video for his constant support in all aspects of life. To Grégoire Genolet, pour ces années passes à jouer au unihockey et regarder le hockey. To Grégoire Gapany pour les nombreuses sorties légèrement alcoolisées. Finally, a special thanks to Roel Bevers which whom we shared so much, from struggles in science to ERB and video games, you deserve a second mention. To Davide for accompanying me to several gaming events.

Finally, I want to thanks my family, my parents Catherine and Vincent for their unconditional support. My brother and sisters Cédric, Alice and Axelle who are always there for me when I need a lift from the station. Warm thanks to my grand-father, Jean-Paul, who always takes time to come see me in Lausanne for all the interesting conversations. Finally, a warm thank you to all my aunts and uncles, Anne-Michèle, Isabelle, Brigitte, Daniel, Nicolas and Frédéric and my cousins Laure, Florence, Mathieu, Roxanne, Marion and Pierre.

Contents

List of Figures

List of Supplementary figures

List of Tables

Chapter 1: General introduction

1.1 Quantitative traits

One of the key aims of biological research is to decipher the phenotype-genotype interaction to understand how variation in the genome affect the phenotype. The phenotype is the sum of all observable traits of an organism and is the combined result of its genotype, its hereditary material, and the environment (Dawkins, 1978). The complete phenotype of an individual is therfore composed of many discrete or quantifiable traits. While the phenotype indeed is the ensemble of traits of an individual, it should be highlighted that the word phenotype can also be used to describe the outcome of a specific trait. The phenotype-genotype relation is at the crux of biological research where researcher try to understand how variation in the hereditary material (genomic variation) affect a given phenotype. Some traits are classified as categorical, such as the ABO blood group system where someone can have one of the four existing blood types; , A, B, O and AB (Ferguson-Smith, Aitken, Turleau, & de Grouchy, 1976). The majority of traits, however, are quantitative traits, a good example of a quantitative trait would be height in humans (Visscher, McEvoy, & Yang, 2010). In that case, the phenotype is defined on a scale and cannot be separated in specific categories. Whereas categorical traits can sometimes be explained by an association with a single gene, or genetic locus (a stretch of DNA), complex traits are usually the result of a combination of multiple genes and/or loci interacting with each other. Moreover, we can note that complex traits do not usually follow the Mendelian laws of inheritance.

The first step undertaken to understand how a phenotype can be explained by the genotype is to research which loci are affecting the trait of interest. Then we can understand how these loci interact with each other and with the environment. We can then assess the strength of each loci, usually called the effect size. The effect size can be defined as “a measure of the magnitude of a phenomenon” (Kelley & Preacher, 2012) and is derived from statistical analysis. Taken together, these results form the genetic architecture of a trait (Hansen, 2006). Although some traits can be relatively easily deciphered, it can still take many years of research before the entirety of the molecular mechanisms underlying a complex trait are understood. Therefore, one can study intermediate phenotypes such as gene expression and proteins abundance that may explain in part some more complex phenotypes. Indeed, with the recent advances in technologies, it is possible to probe for these intermediate phenotypes on a large scale. A very popular method is to link variation in gene expression or another trait in a population to genomic variation. The result is that for a given trait, several loci in the genome

can be associated with the trait, forming a quantitative trait loci (QTL). A QTL can be linked to different type of trait, such as gene expression. In that case, the QTL is dubbed an expression QTL or eQTL, meaning that loci in the genome explain the variation in the expression of a gene. eQTL mapping is a popular type of analysis in which researchers measure the gene expression in as many individuals and correlate changes in gene expression with genetic molecular markers such as single-nucleotide polymorphism (SNP), insertion or deletions (indels), microsatellites or transposable elements (A. Y. K. Albert et al., 2008; F. W. Albert & Kruglyak, 2015). The researchers are then able to link variation in gene expression with genetic markers and phenotypes, explaining in this way how genomic variation affect the expression of genes. In general, these analyses are relying on large populations consisting of individuals with different genotypes and phenotypes. Because these analysis rely on large population of individuals (Nica & Dermitzakis, 2013), it is sometimes preferable to use model organism such as the *Drosophila melanogaster* to study a given trait (Gilad, Rifkin, & Pritchard, 2008).

1.2 *Drosophila melanogaster*

Drosophila melanogaster, commonly known as the fruit fly, was first reared at the beginning of the twentieth century by Charles W. Woodworth but only when it was picked up by Thomas H. Morgan around 1908 is when the fruit fly started its illustrious career in science as a model organism (Therese Ann Markow, 2015; Morgan, 1910) (**Fig. 1-1**). It was originally used to study the Charles Darwin theory of evolution by mutating the hereditary material of the fly using various mutagens. The goal was to find a heritable trait to test the heredity of this trait in subsequent generation. The observation of the *white* mutant, that change the color of the eye of the Drosophila from bright red to white led the way to our current knowledge in the modern genetic field. A direct consequence of this research was the dissemination of the *Drosophila* as a model organism throughout universities around the world where it started being commonly used for research other than evolution due to its ease of use. More than a century of *Drosophila* research later, there is a wealth of knowledge gathered about the fruit fly, ranging from development and physiology,

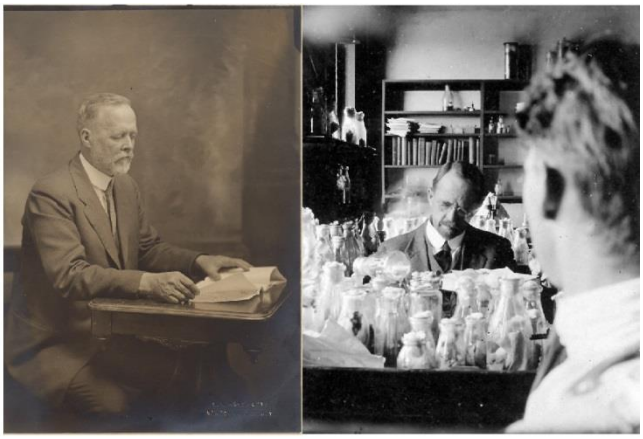


Figure 1-1 : The Drosophila Pioneers

Charles W. Woodworth (left) and Thomas H. Morgan in his fly room (right)

to sexual behavior (Miguel-Aliaga, Jasper, & Lemaitre, 2018; Tolwinski, 2017; D. Yamamoto, Jallon, & Komatsu, 1997). A by-product of this research is a staggering amount of characterized mutants and genetic tools available (Bellen, Tong, & Tsuda, 2010; Rubin & Lewis, 2000).

It is important to highlight several characteristics of this organisms to outline its relevance in science and for this work particularly. First, the fly is ideal to perform large scale experiments which require a lot of material due to its ease of manipulation. Indeed, the space needed to rear flies is small, it has a short generation time of 10 days at standard rearing conditions and flies generate a high numbers of offsprings in a short time period, meaning it is relatively easy to expand a population for large experiments (T. A. Markow, Beall, & Matzkin, 2009; Therese Ann Markow, 2011; St Johnston, 2002). Moreover, the creation of balancer chromosomes harboring dominant phenotypic mutants and recessive death-inducing genes allows easily traceable mutations (Kaufman, 2017). Balancer chromosome are one of the many tools developed, and are widely used, by the *Drosophila* community which greatly simplifies working with mutant strains. Another example of extremely useful tool is the GAL4/UAS system, a system which allows the expression of specific genetic constructs in a spatially controlled way. Furthermore, this system coupled with the GAL80 protein allow for temporal control of transgene expression (Harrison & Perrimon, 1993; Ma & Ptashne, 1987).

The *Drosophila* genome consists of 4 chromosomes and was sequenced in its entirety for the first time in 2000 (Adams et al., 2000). This unveiled a genome of lower complexity than the human one, with approximately 165’000’000 base pairs and roughly 15’000 genes, though this number increased as newer version of the genome were released. Currently, the latest assembly (assembly Release 6 plus ISO1 MT) contains 17738 genes (Hoskins et al., 2015; NCBI, 2015). Interestingly, there is relatively high similarity between the fruit fly and the human genome regarding the genes involved in disease development in humans. Indeed, up to 65% of human-disease related genes have a homolog in the fly (Chintapalli, Wang, & Dow, 2007; Millburn, Crosby, Gramates, & Tweedie, 2016; Ugur, Chen, & Bellen, 2016; S. Yamamoto et al., 2014). This similarity was leveraged to study several human disorders in the fly ranging from neurodegenerative disorders such as Alzheimer (Tan & Azzam, 2017) and Parkinson disease (Feany & Bender, 2000; Moyerbrailean et al., 2016; Xiong & Yu, 2018) as well as cancer progression/development (Enomoto & Siow, 2018).

Moreover, aside from classical genetics, the fruit fly has also been extensively studied in the field of population and quantitative genetics (Coyne & Orr, 1997; Roff D.A. & Mousseau T.A., 1987; Wright, 2006). As this field depends on the use of large populations, various populations

were developed. First flies collected from the wild were isogenized and formed the *Drosophila* Genetic Reference Panel (DGRP) (W. Huang et al., 2015; Mackay et al., 2012). Then, recombinant inbred lines were used to create the *Drosophila* Synthetic Population Resource (DSPR) (King, Sanderson, McNeil, Long, & Macdonald, 2014; Qu, Gurdziel, Pique-Regi, & Ruden, 2018; Stanley, Ng'oma, O'Day, & King, 2017). Both these resources allow for the measurement of various quantitative, or discrete, phenotypes which can be measured in relative high-throughput such as olfactory behavior (Arya et al., 2015), pigmentation (Dembeck et al., 2015) food intake (Garlapow, Huang, Yarboro, Peterson, & Mackay, 2015), lifespan and fecundity (Durham, Magwire, Stone, & Leips, 2014) or innate immunity (Bou Sleiman et al., 2015). Moreover some other flies collected from the wild were kept in the lab for several decades of experimental evolution aimed at understanding genetic determinants of complex phenotypes through isolation of rare variants (Burke & Rose, 2009) for traits such as adaptation to temperature variation (Klepsatel et al., 2013), egg-size variation (Jha et al., 2015) or adaptation to darkness (Izutsu, Toyoda, Fujiyama, Agata, & Fuse, 2016).

Furthermore, to collect and organize the wealth of information created, several online resources and database have been developed by the fly community, with the most prominent one being FlyBase and is freely accessible (for now). This website aims to integrate all knowledge garnered on the fly from as many sources as possible and is carefully curated to remain up to date, giving researchers access to highly detailed information on specific genes such as their (putative) function, expression patterns, mutant phenotypes, and whether mutant stocks are available and where (McQuilton, St Pierre, & FlyBase Consortium, 2012; Thurmond et al., 2018).

All these resources have helped keeping the *Drosophila melanogaster* as one of the main model organism in science. Indeed, insight from Drosophila have helped understand molecular mechanisms in the human. For example, the fly innate immunity which can also be considered as a quantitative trait, was extensively studied because, as in humans, where one person gets sick easily, another one may not. However, it is still largely unknown what drives this variability. Drosophila, with all advantages mentioned before, can thus be an exceptional model to study these aspects of immunity.

1.3 The *Drosophila* immune system

1.3.1 General introduction to the immune system

The immune system can be defined as a collective of mechanisms that an organism can muster to prevent infection by pathogenic agents and to defend itself against them. To prevent infection, organisms, possess a physical barrier that acts as an outer defense inhibiting the entry of pathogens. This can include the human skin, or, for the *Drosophila*, its cuticle. Both of these organs help keeping pathogens outside of the body. We often find a mucus layer that cover some of the internal exposed area, such as the digestive track or the respiratory tract, that act in complement to these physical barriers by entrapping pathogens (B. Alberts et al., 2002; Janeway, Travers, & Walport, 2001; Nochi & Kiyono., 2006). When a pathogen breaches through these barriers, the organism must mount a successful immune response to survive. First the organism must be able to recognize the pathogen before it can neutralize him using various means and, finally, restore the homeostasis of the organism.

The immune response (**Figure 1-2**) can be divided into two types of responses, first the innate immune response, and second, the adaptive immune response (B. Alberts et al., 2002; Fearon DT & Locksley RM, 1996). Whereas the adaptive immune response relies on specifically tailored responses to a particular pathogen and is able to ”memorize” this response for future infections, the innate immune response provides a rapid response against many pathogens, and is thus more general. Besides eliminating the pathogens, the elements of the innate immune response can also activate the adaptive immunity by presenting antigens for recognition. Then, the adaptive immune response will produce highly specific compounds (antibodies) and cells against the detected pathogen (B. Alberts et al., 2002; Janeway et al., 2001). However, the adaptive immune response is not found in all organisms, and has mostly been shown in vertebrates whereas *Drosophila* and other invertebrates lack any adaptive immune system.

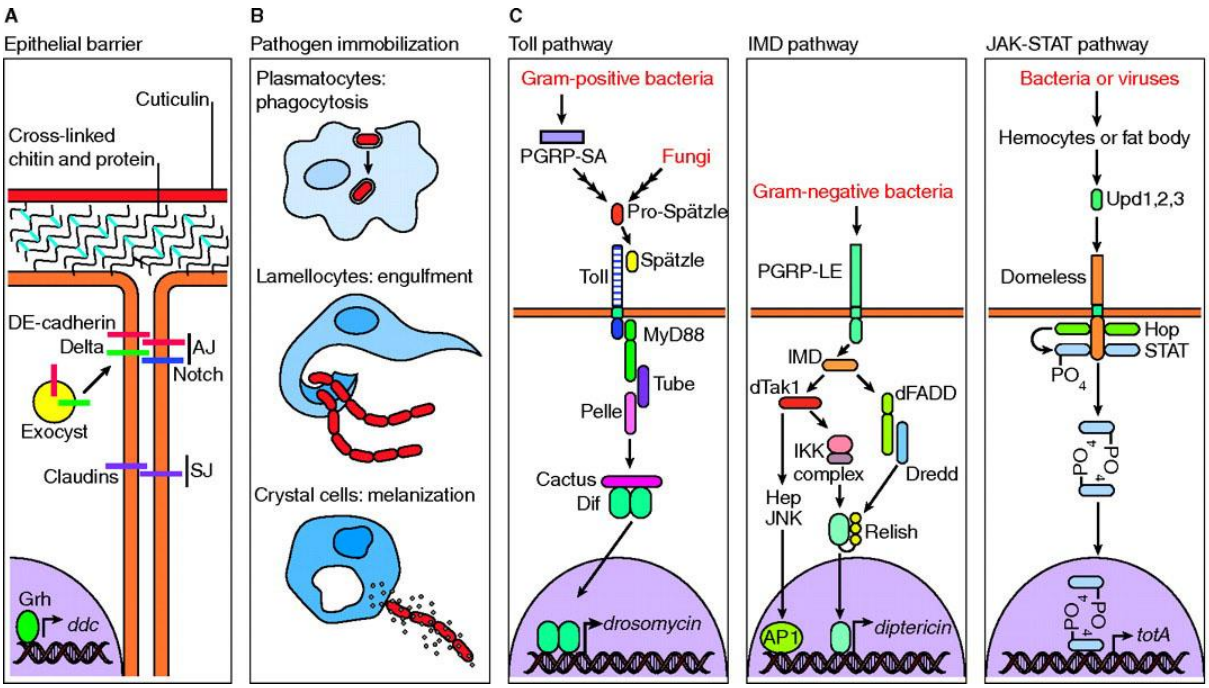


Figure 1-2 : Schematic of the Drosophila immune response

Adapted from (E. Bier & Guichard, 2012)

(a) The various physical barriers

(b) Cellular immune response

(c) Toll pathway - Imd pathway - JAK-STAT pathway

1.3.2 The *Drosophila* immune response

The *Drosophila* immune system has been extensively characterized and possess both a cellular response and a cell-free, or humoral, response. The cellular immune response encompasses mechanisms such as phagocytosis, cellular encapsulation, melanization and coagulation. The humoral response refer to the expression of anti-microbial peptides (AMPs) and reactive oxygen species (ROS) by cells triggered by the recognition of pathogens (Lemaitre & Hoffmann, 2007).

1.3.3 The cellular immune response

The cellular immune response effectors are hemocytes which comprises of plasmatocytes, crystal cells and lamellocytes. Plasmatocytes are cells involved in phagocytosis of pathogens. They recognize foreign or dead cells, to which they attach and finally, through cytoskeletal and internal reorganization, internalize the object. The foreign object is kept in a vesicle dubbed the phagosome, where it will eventually be destroyed through

acidification of the internal environment after the phagosome fuse with a lysosome (Kinchen & Ravichandran, 2008; Kounatidis & Ligoxygakis, 2012; Lemaitre & Hoffmann, 2007). Crystal cells are responsible for the melanization, a process by which the fly encapsulates invading microorganisms at the site of the injury by secreting melanin. It is a defense mechanism aimed at protecting the insect against breaching of the cuticle. Moreover, some of the byproducts of the chemical reaction needed for melanin synthesis, such as quinones, are toxic to the pathogens. This process is easily visible as it produces a dark patch on the fly at the site of injury (Tang, 2009). Next, lamellocytes encapsulate foreign objects that cannot be internalized due to their large size. It is a potent reaction to the detection of a large foreign object in the body, such as the egg of parasitoid wasps. Upon detection of the foreign object, the fly will induce the production of lamellocytes through differentiation of prohemocyte precursors. These lamellocytes will bind to each other around the foreign object in several layers, thus creating a capsule around the invader which will be eventually killed inside it (Kounatidis & Ligoxygakis, 2012; Lemaitre & Hoffmann, 2007; Vlisidou & Wood, 2015). Finally, coagulation or clotting is a mechanism that prevent the spread of microorganisms by using filaments to attach them together, thus creating a clot. Clotting is mediated by plasmatocytes and crystal cells and involves many pathways and secreted proteins. The most prominent proteins are Hemolectin, Fondue and Transglutaminase. These together create a matrix of filaments (Hemolectin) linked together by Fondue and connected to bacteria via Transglutaminases (Kounatidis & Ligoxygakis, 2012; Lemaitre & Hoffmann, 2007; Vlisidou & Wood, 2015). However, the cellular immune response alone is not capable of repelling all threats to the organisms and need the help of the humoral immune response to effectively fight some pathogens.

1.3.4 The humoral immune response

The humoral immune response is responsible for expressing and secreting proteins and other products to fight the pathogens. The most important ones are the antimicrobial peptides (AMPs) and the components to produce reactive oxygen species. The AMPs are generally small proteins with an anti-bacterial or anti-fungal activity.

To recognize a pathogen, the innate immune system use pattern recognition receptors (PRRs) to detect pathogen-associated molecular patterns (PAMPs) (Suresh & Mosser, 2013). Successful detection of a PAMPs will trigger a signaling cascade which will induce the production of AMPs (Lemaitre & Hoffmann, 2007; Zasloff, 2002). The two main pathways regulating the expression of AMPs are the Toll and immune deficiency (Imd) pathways. Each of these pathways recognize different types of pathogens. For instance, the Toll pathway

recognizes yeast, fungi, and bacteria with Lysin-type peptidoglycan, which come from the cell wall of primarily gram-positive bacteria. On the other hand, the Imd pathway recognizes Diaminopimelic acid (DAP) type peptidoglycan, mostly present when an infection with gram-negative bacteria occurs. Both pathway cascades lead to the activation of various transcription factors of the NF-κB family (Toll: Dif and Dorsal; Imd: Relish) which then translocate to the nucleus and regulate gene expression, particularly AMPs (Kounatidis & Ligoxygakis, 2012; Lemaitre & Hoffmann, 2007). However, AMPs are not the only response as cells are also able to produce reactive oxygen species (ROS), which consist of a wide assortment of microbicidal compounds containing oxygen which are released upon infection. They damage pathogens through oxidative damage to biological molecules (Paiva & Bozza, 2014). A basal level of ROS production is nevertheless kept in the gut to control for the microbiota (S.-H. Kim & Lee, 2014; Lemaitre & Hoffmann, 2007).

The Toll pathway

The detection of pathogens leading to Toll activation is mediated by the Nerve Growth Factor-related cytokine Spz. Specific detection of infectious agents by the Toll pathway is mediated upstream of Spz by the recognition molecules that recognize different substrates. PGRP-SA, PGRP-SD and GNBP1 recognize Gram-positive bacteria, GNBP3 recognizes yeasts and Persephone recognizes fungi. Their signals are integrated by Spe that then activate Spz. Spz is linked to Toll and upon detection will then form a dimer, thus creating a dimerized version of Toll. This will create a complex to transmit the signal composed of MyD88, Tube and Pelle. This complex can then phosphorylate Cactus, a protein binding Dif and Dorsal and preventing them to reach the nucleus. The phosphorylation of Cactus leads to its degradation, freeing both NF-κB homologs to relocate to the nucleus to trigger expression of targets genes.

The Imd pathway

The Imd pathway is activated by the transmembrane receptor PGRP-LC and the intracellular receptor PGRP-LE. Upon detection of DAP-type peptidoglycan, both receptor recruit Imd, which itself recruit and interact with FADD which does the same with DREDD. DREDD unmasks another interaction domain of Imd which allow for the recruitment of dIAP-2. The latter will ubiquitinate Imd, allowing it to trigger a cascade involving TAK1 and IKK and finally leading to the cleavage of Relish which then translocate to the nucleus to trigger the expression of several genes.

The JAK/STAT pathway

The JAK/STAT pathway is composed of the ligands unpaired 1-3 (Upd1-3) which bind the transmembrane receptor Dome which in turn activates Hop (a human JAK2 homologue) and Stat92E, a transcription factor. Upon activation, Stat92E will translocate to the nucleus and alter the expression of several genes, among them *tep1*, a thioester-containing protein (TEP) with anti-microbial activity (Agaisse & Perrimon, 2004; Bou Aoun et al., 2011; Dostálová, Rommelaere, Poidevin, & Lemaitre, 2017). However, it was shown that the JAK/STAT pathway play an important role in the local immune response of the *Drosophila* gut (Nicolas Buchon, Broderick, Poidevin, Pradervand, & Lemaitre, 2009).

Negative regulation of the immune response

While the immune response is of course advantageous to the survival of an organism, there are downsides to it when it goes on uncontrolled. Indeed, control of the immune response is critical as it can be detrimental to the host over time. It was shown in *Drosophila* that uncontrolled immune response is detrimental to the fly (Paredes, Welchman, Poidevin, & Lemaitre, 2011). Several negative regulators of the Imd pathway have been identified with different modes of action. For example, PGRP-LB is a secreted molecule that cleaves DAP-type PG and leads to a dampening of the immune response (Zaidman-Rémy et al., 2006). In another case, Pirk was shown to limit the activation of the pathway by interacting with PGRP-LC. (Kleino et al., 2008). And for the Toll pathway, it was shown that WntD negatively regulates the pathway by preventing translocation of Dorsal to the nucleus (Gordon, Dionne, Schneider, & Nusse, 2005). These are just a set of examples which can prevent overregulation of the immune response which could ultimately harm the organism.

If the negative regulation of the immune system is critical for host survival, it has another important role in the intestine of the fly. Indeed, the *Drosophila* gut, like its human counterpart, is host to a thousands of bacteria forming the microbiota (N Buchon, Broderick, & Lemaitre, 2013; Walton, 2015) and must thus accommodate them in the gut lumen.

1.4 The Drosophila gut

One of the major evolution in the animal was the emergence of a gastrointestinal tract within the body cavity which allowed the transition from an intra- to an extracellular mode of digestion (Lemaitre & Miguel-Aliaga, 2013; Stainier, 2005). The *Drosophila* gut (**Figure 1-3**) is composed of a simple epithelium surrounded by trachea, visceral muscle and enteric nerves. It is relatively close to its human counterpart, and, like it, is constantly renewed during the lifespan of the fly. The gut is divided into three parts, the foregut, the midgut and the hindgut surrounded by visceral muscles, trachea and enteric nerves. Each of these parts have been shown to be highly compartmentalized with each compartment showing very different genes expression patterns as well as different cellular, chemical and physiological characteristics (Nicolas Buchon & Osman, 2015; Nicolas Buchon et al., 2013; Marianes & Spradling, 2013).
The gut epithelium is composed of four types of cells: large absorptive enterocytes (ECs), small secretory enteroendocrine cells (EECs), pluripotent intestinal stem cells (ISCs) and enteroblasts (EBs). Enterocytes are large polypoid cells and constitute the majority of the cells present in the gut. Their main role is to absorb nutrients, although they also have a secretory role. Enteroendocrine cells secrete small peptides that are thought to control the physiology of the lumen (Veenstra, Agricola, & Sellami, 2008). ISCs ensure the renewal of the gut by replenishing both the ECs and EECs (Micchelli & Perrimon, 2006; Ohlstein & Spradling, 2006). It can divide symmetrically or asymmetrically thus producing an ISC and an enteroblast. The enteroblast is a transient precursor to both enteroendocrine cells and enterocyte and signaling by Notch determines the fate of the enteroblast (Ohlstein & Spradling, 2007).

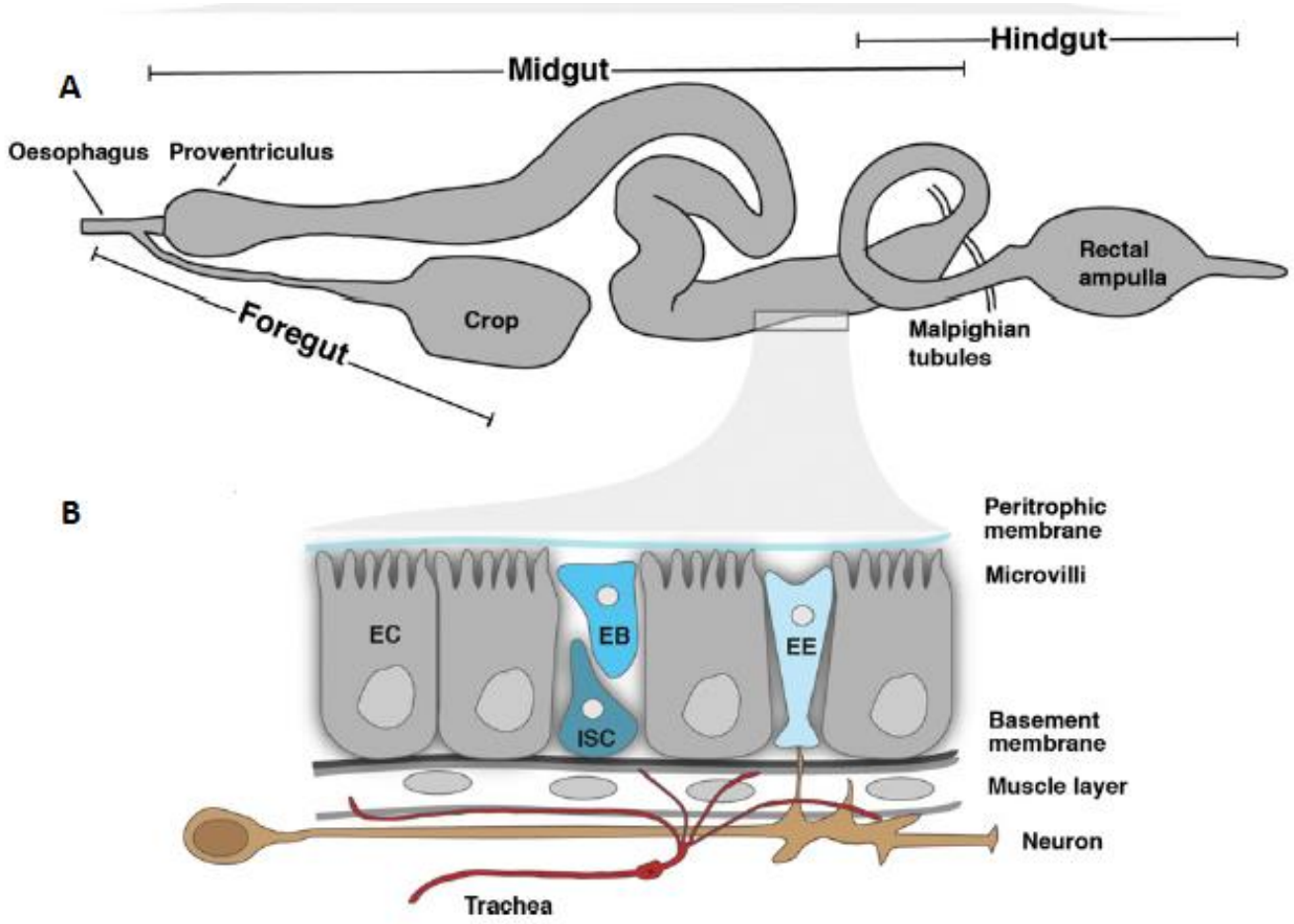


Figure 1-3 : The *Drosophila* gut

adapted from (Miguel-Aliaga et al., 2018)

(a) The main subdivision of the gut

(b) The different cell forming the gut

1.4.1 The *Drosophila* enteric immune response

Because the gut evolved to be exposed to the exterior environment, this also opened the opportunity for pathogens to exploit this route of infection. Therefore, Gut-bearing organisms evolved complex mechanisms to defend themselves against these agents, making them “gut immunocompetent”. However, the gut immune response must be tightly controlled to allow for the microbiota to survive. Interestingly, there are marked difference in the immune response of the fly depending on whether the infection is contained on the site of entry or if it spreads to the whole fly. In the case of a systemic infection, the immune response is under the control of the fat body, whereas the local immune response is mediated by the locally affected epithelia (Kounatidis & Ligoxygakis, 2012; Liehl, Blight, Vodovar, Boccard, & Lemaitre, 2006).

The gut is permanently exposed to the commensal bacteria (microbiota) and to the environment, making it one of the main paths of infection by external pathogens. It must though carefully regulate its immune response to accommodate the microbiota and fight pathogens at the same time. This tight balance is done through the secretion of AMPs as well as through the production of ROS. The production of AMPs in the gut is mediated through the Imd pathway and its effector, NF-κB-like transcription factor (TF), Relish.

Another important part of the gut immune response is the production of ROS which is mediated by the NADPH oxidase Duox (E.-M. Ha et al., 2009; E. M. Ha et al., 2005). The Duox pathway is activated by the detection of Uracil, a compound only produced by opportunistic bacteria, through the Gαq-phospholipase C-ß-Ca2+ pathway (K. A. Lee, Kim, Bhin, et al., 2015; K. A. Lee, Kim, You, & Lee, 2015). The detection of Uracil directs a constant expression of ROS that control microbiota level and is strongly induced when the gut is colonized by pathogenic bacteria due to an increase amount of Uracil. Moreover, the Imd pathway is activated upon infection and further increases *Duox* expression. DUOX activity is directly controlled by PLC who regulates both the expression and the activity of DUOX. To avoid damage due to excessive oxidative stress, the immune-regulated catalase (IRC) removes ROS in a dynamic way Duox transcription is regulated by the TF ATF2 downstream of the p38a-Mkk3-Mekk1 pathway (Sveta Chakrabarti, Poidevin, & Lemaitre, 2014).

1.5 Proteomics

The previous decade has seen a tremendous rise in the technologies available to probe large biological datasets. Along with them came sophisticated informatics tool needed to treat and analyze the amount of information generated. This led to more and more work aimed at understanding how complex biological systems react to changes. Indeed, with the rise of high-throughput sequencing technologies and the decrease in cost, experiments such as mRNA sequencing (RNA-seq) or chromatin immunoprecipitation followed by sequencing (ChIP-seq) allowed the researchers to probe genome wide changes. It is interesting to note that all of these technologies rely on the PCR reaction, first developed by Kary Mullis in 1985 (Saiki et al., 1985). However, these methods all measure changes happening at the nucleic acid level and not at the protein level. Interestingly, the development of mass spectrometry, one of the methods to investigate changes at the protein level, predates even the modern genetics (Thomson, 1913). It was pioneered at the beginning of the 20th century By Sir Joseph John Thomson and has since then been used in many fields, such as physics, chemistry, biology, etc. A mass spectrometer analyzes the mass-to-charge ratio (m/z) of elements. This allow to both qualify and quantify each element that is present in a sample. The element must first be ionized after which all subsequent steps are to be carried out in a vacuum. There have been many types of mass spectrometers and ionization methods developed, but all of them use electric and magnetic fields to separate the ions and measure their m/z ratio. Another critical technological development, and advancement, was the Fourier transformation that allowed the deconvolution of signals into individual signals, thus allowing for an improved analysis of the output by reaching a higher resolution. However, it is only recently that these techniques have matured enough and have become more cost-effective to be applied to large biological datasets, mainly by increasing the weight range of the molecules that can be ionized (Tanaka, 2003; Tanaka et al., 1988), thus opening the door to the analysis of large biological molecules (Maher, Jjunju, & Taylor, 2015).

The standard modus operandi of mass spectrometry analyses using biological samples include two critical steps before the data acquisition: a digestion step and a separation step (Zhang et al., 2010). During the digestion steps, researchers used different methods and enzymes to break proteins into smaller peptides (Gundry et al., 2009). Then, the goal of these separation steps is to divide the digested samples into different fractions. The goal is to separate the peptides to such extent that they can be injected in the mass spectrometer in small quantities to allow for an improved signal acquisition. One of the most technically challenging steps in mass spectrometry is the analysis of the results. Compared to a DNA or cDNA sequencing run which yields a relatively simple text file of DNA sequences, the output of a mass spectrometer

is a set of m/z spectra. Similarly to sequencing analysis, where one must match the DNA segment to a part of the genome, the m/z spectra generated must be matched to known spectra. Most of the proteomics software will thus generate spectra for each putative peptide. These software will generate these spectra not only based on the underlying genome specific to the source of each sample, but also based on the digestion methods that were used. Finally, the quantification is also technically complicated. It is indeed not possible to count the number of peptides, thus forcing the researchers to derive the number of peptides based on other quantification methods. We can here distinguish between two types of experiment, label-free quantification and quantification with label. The labelling methods require mixing of the samples with stable isotopes. Isotopes marked proteins then crate a specific mass tag which is easily recognizable by the spectrometer and is then used as an internal control to infer the amount measured (Bantscheff, Schirle, Sweetman, Rick, & Kuster, 2007). Label-free quantification relies on data-derived methods that assess the relative amount of peptides compared to other peptides (Asara, Christofk, Freimark, & Cantley, 2008; Bridges et al., 2007).

Despite the relative hurdles of mass spectrometry, it offers many advantages. The most obvious one is that it allows the observation of the proteome, which none other methods currently can provide. Although extremely instructive, the analysis of the transcriptome provides an imperfect view of the changes happening in an organism. Thus, connecting both transcriptome and proteome may therefore aid in our understanding of gene regulatory networks, or molecular mechanisms, underlying the various immune responses. While indeed many studies have found that the correlation between transcriptome and proteome is rather low (Bonaldi et al., 2008; Butter et al., 2013; Casas-Vila et al., 2017; Griffin et al., 2002; Grün et al., 2014; Li, Bickel, & Biggin, 2014; Schwanhäusser et al., 2011), combining such data can still provide us with a better understanding of specific molecular mechanisms. Indeed, because the proteins are generally the effectors of the cell, directly studying their variation could provide new insight into molecular mechanisms not detected by transcriptomics studies.

1.6 The scope of this thesis

In this thesis, I expand on previous work on the genetic determinants of the resistance to enteric infection in the DGRP after infection by *Pseudomonas entomophila* and further explore the enteric immune response by looking at the proteome of the gut after enteric infection.

In **Chapter 2**, I use RNA-sequencing to analyze the genetic determinants of gene expression variation in the DGRP population. Then, we use F1 offsprings to measure the extent of eQTLs-driven allele specific expression we performed a large in-depth characterization of gene-expression variation eQTLs identified in a subset of the DGRP population and discuss the characterization of their effects. Further analyses of differentially expressed genes between different resistance classes interestingly reveal only one gene, *Nutcracker* (*ntc)*. The effect of this gene on the resistance of the fly to enteric infection is then confirmed using other various mutants and molecular tools. Finally, I identify, and show, how an infection specific eQTL linked with *ntc* decreases the binding affinity of the transcription factor Broad, and ultimately leads to an increased expression of *ntc*.

In **Chapter 3**, we characterize the proteome of the *Drosophila* gut upon infection. I described how the proteome react to the ingestion of both *P.e.* and *Ecc15* and the differences between these two responses. I then highlight differences between transcriptome and proteome changes at the individual gene level. Finally, I select several candidates genes that may have been overlooked in previous transcriptome studies for further analysis of their impact on resistance to infection and perform survival analysis on mutants of these genes to identify potential new players in the gut enteric immune response.

Finally, I summarize my findings in **Chapter 4** and provides future directions and outlooks.

Chapter 2: *cis*-regulatory variation modulates susceptibility to enteric infection in the *Drosophila* Genetic Reference Panel

Michael V. Frochaux[1]*, Maroun Bou Sleiman[2]*, Vincent R. J. Gardeux[3], Riccardo Dainese[4], Brian Hollis[5], Maria Litovchenko[6], Virginie S. Braman[7], Dani Osman[8], Bart Deplancke[9+]

1. Michael Vincent Frochaux, Laboratory of System Biology and Genetics, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, Lausanne, Switzerland, michael.frochaux@epfl.ch
2. Maroun Bou Sleiman, Laboratory of Metabolic Signaling, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, maroun.bousleiman@epfl.ch
3. Vincent Roland Julien Gardeux, Laboratory of System Biology and Genetics, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, Lausanne, Switzerland, vincent.gardeux@epfl.ch
4. Riccardo Dainese, Laboratory of System Biology and Genetics, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, Lausanne, Switzerland, riccardo.dainese@epfl.ch
5. Brian Hollis, Laboratory of System Biology and Genetics, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, Lausanne, Switzerland, brian.hollis@epfl.ch
6. Maria Litovchenko, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, Lausanne, Switzerland, maria.litovchenko@epfl.ch
7. Virginie Braman, Laboratory of System Biology and Genetics, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, virginie.braman@epfl.ch
8. Dani Osman, Faculty of Sciences III and Azm Center for Research in Biotechnology and its Applications, LBA3B, EDST, Lebanese University, 1300, Tripoli, Lebanon, daniosmanlb@gmail.com

9. Bart Deplancke, Laboratory of System Biology and Genetics, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, bart.deplancke@epfl.ch

*: These people contributed equally to the paper

+: Corresponding author

Note

This chapter is a draft submitted to genome biology April 15th, 2019.

Author contributions: Michael Frochaux, Maroun Bou Sleiman and Bart Deplancke designed the study with the help of Dani Osman for the RNA-seq. Maroun Bou Sleiman and Dani Osman prepared the RNA-sequencing samples. Maroun Bou Sleiman performed the statistical and computational analyses on the RNA-seq data with assistance from Tommaso Andreani and Maria Litovchenko. Michael Frochaux and Maroun Bou Sleiman performed infection and RT-qPCR experiments. Michael Frochaux, Maroun Bou Sleiman and Virginie Braman performed the round robin experiment. Michael Frochaux validated the eQTLs with assistance from Vincent Gardeux and Brian Hollis. Riccardo Dainese performed the MITOMI experiment.

2.1 Abstract

2.1.1 Background

Resistance to enteric pathogens is a complex trait at the crossroads of multiple biological processes. We have previously shown in the *Drosophila* Genetic Reference Panel (DGRP) that resistance to infection is highly heritable, but our understanding of how the effects of genetic variants are channeled through distinct molecular layers to determine gut immunocompetence is still limited.

2.1.2 Results

To address this, we performed a systems genetics analysis of the gut transcriptomes from 38 DGRP lines that were orally infected with *Pseudomonas entomophila* (*P.e.*), identifying a large number of condition-specific expression quantitative trait loci (*local*-eQTLs) that were enriched for the Atf-2 and Disco transcription factor motifs. By assessing the allelic imbalance in the transcriptomes of 19 F1 hybrid lines from a large round-robin design, we could independently attribute a robust *cis*-regulatory effect to only 10% of these detected *local*-eQTLs. However, additional analyses indicated that many *local*-eQTLs (~10%) may act in *trans* instead. Comparison of the transcriptomes of DGRP lines that were either susceptible or resistant to *P.e.* infection revealed *Nutcracker* (*ntc*) as the only differentially expressed gene. Interestingly, we found that *ntc* is linked to infection-specific eQTLs that not only correlate with its expression level, but also to enteric infection susceptibility. This is consistent with our findings that *ntc* expression is induced upon infection, whereas loss of *ntc* confers an overall greater susceptibility to infection. Further regulatory analysis revealed one particular *ntc* eQTL that significantly decreases the binding affinity for the repressor Broad, driving differential allele-specific *ntc* expression.

2.1.3 Conclusion

Our collective findings point to a large number of infection-specific *cis*- and *trans*-acting eQTLs in the DGRP, including one common non-coding variant that lowers enteric infection susceptibility by modulating *ntc* gene regulation through altered Broad repressor binding.

2.2 Background

Deciphering the relationship between genomic and phenotypic variation is a central question in genetics. Genome Wide Association Studies (GWAS) have been extensively used to address this question by looking for variants that could explain a certain fraction of the genetic variance of phenotypes (Manolio, 2010; Visscher et al., 2017). More often than not, those variants are located in non-coding regions of the genome, rendering the inference of their putative function difficult (F. W. Albert & Kruglyak, 2015; Gan, Pro, Sewell, & Fuxman Bass, 2018; Hindorff et al., 2009; Rojano, Seoane, Ranea, & Perkins, 2018). Therefore, the study of intermediate molecular traits, such as gene expression levels, and how they are affected by genomic variation is a powerful complementary approach to linking geno- to phenotype (Deplancke, Alpern, & Gardeux, 2016; Nica & Dermitzakis, 2013).

Ever since the first expression quantitative trait locus (eQTL) report in yeast (Brem, Yvert et al. 2002), it became clear that eQTLs could account for a substantial proportion of variability in gene expression following a cellular or organismal response to external stimuli. These eQTLs in turn advanced our understanding of the genetic basis of disease susceptibility. Indeed, eQTL studies in both mouse and human using monocytes, macrophages, dendritic cells or other immune cells have been useful to better understand how genetic regulatory effects affect auto-immune disease (R. Alberts et al., 2011; Kim-Hellmuth et al., 2017; Raj et al., 2014), inflammatory bowel disease (Peters et al., 2016), resistance to *Salmonella* (Gilchrist et al., 2015) and the molecular response to an infection stimulus (Fairfax et al., 2014, 2012; M. N. Lee et al., 2014; Orozco et al., 2012). These advances motivated the establishment of even larger-scale projects such as DICE (Database of Immune Cell Expression, eQTL and Epigenomics) to characterize gene expression in all human immune cell types and to study how genetic variants affect these immune cell-related transcriptomes (Schmiedel et al., 2018). However, eQTL-related studies aimed at better understanding the genetic and molecular basis underlying gut immunocompetence have been lacking for practical and ethical reasons. Indeed, human intestine eQTL studies have to our knowledge so far been restricted to inflammatory bowel disease (Di Narzo et al., 2016; Hulur et al., 2015; Kabakchiev & Silverberg, 2013; Peters et al., 2016; Singh et al., 2015).

A valuable alternative model to uncover the genetic and molecular mechanisms underlying variation in gut immunocompetence is *Drosophila melanogaster* given that this organism is by now widely used to study the biological processes mediating the response to enteric infection (Galenza & Foley, 2019; Gupta, Stewart, Rund, Monteith, & Vale, 2017; Lemaitre & Hoffmann, 2007; Mistry, Kounatidis, & Ligoxygakis, 2015; Mondotte et al., 2018). Moreover, previous work

including ours has shown that gut immunocompetence is a highly variable and heritable trait, not only in human (Gundogdu & Nalbantoglu, 2017) and mouse (Davenport et al., 2015), but also in *Drosophila* (Bou Sleiman et al., 2015; Early, Shanmugarajah, Buchon, & Clark, 2017). Consequently, population resources such as the *Drosophila* Genetic Reference Panel (DGRP) can be effectively used to study the molecular nature of enteric infection-induced gene expression variation. In this study, we therefore explored the effect of genetic variation on gene expression and organismal phenotypes in the context of *in vivo* enteric infection in the DGRP. Despite several valuable eQTL studies in *Drosophila* involving the DGRP (Cannavò et al., 2016; W. Huang et al., 2015; W. Huang, Massouras, & Inoue, 2014; Massouras et al., 2012; Zichner et al., 2012) and the *Drosophila* Synthetic Population Resource (DSPR) (King et al., 2014; Qu et al., 2018; Stanley et al., 2017), none have so far focused on the response to infection.

To do so, we generated a large set of *Drosophila* control and *P.e.*-infected gut transcriptomes to systematically investigate the link between gut gene expression levels and genetic variation. We showed that genotype is a major determinant of global gene expression levels, revealing a large number of both shared and condition-specific *local*-eQTLs (Hasin-Brumshtein et al., 2014; Khansefid et al., 2018; Rockman & Kruglyak, 2006). We then validated and catalogued these *local*-eQTLs into *cis* and *trans*-acting eQTLs using allele-specific expression on a set of F1 siblings from crosses between isogenic DGRP lines. Importantly, we identified *nutcracker (ntc*) as a gene that is differentially expressed between susceptible and resistant DGRP lines. Through classical genetic analyses, we found that it affects the immunodeficiency (Imd)-dependent enteric immune response through the induction of the major effector *Diptericin (Dipt)*. We also identified and *in vivo* validated a *cis*-regulatory variant in a predicted transcription factor (TF) binding site responsible for the difference in *ntc* expression between the resistance classes and validated the effect of the SNP on allele-specific gene expression *in vivo*. In this study, we thus leveraged the genetic tractability of the fruitfly, the ability to easily replicate experiments on the same genetic backgrounds and the study at the whole organism level to characterize in depth the genetic and molecular mechanisms that contribute to gut immunocompetence variation in *Drosophila*.

2.3 Results

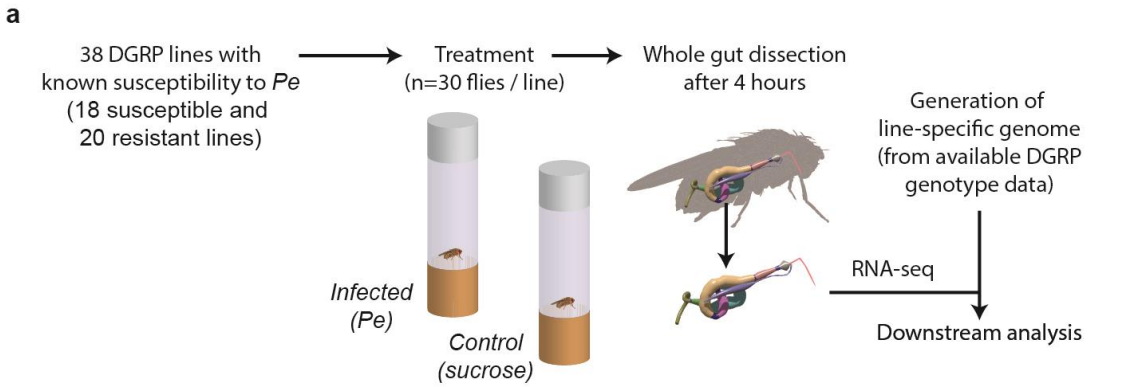
2.3.1 Genetic analysis reveals pervasive, condition-specific gene expression variation

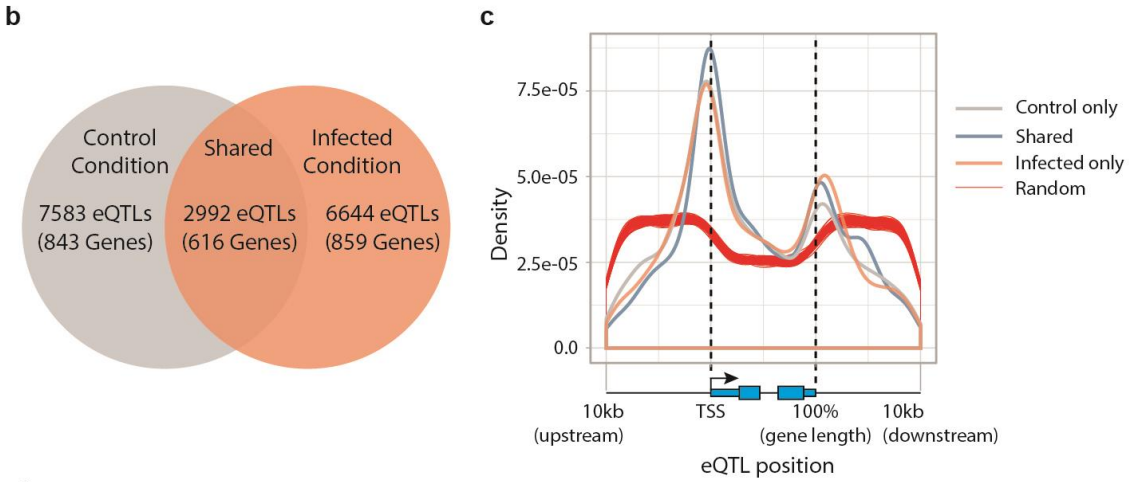
To study global gene expression variation between two enteric infection resistance classes, we selected 38 DGRP lines from the phenotypic extremes from our previous study (Bou Sleiman et al., 2015) with 20 being highly susceptible and 18 being highly resistant to enteric infection by *Pseudomonas entomophila* (*P.e.*) (**Fig. 1a**). Adult female flies were infected and mRNA sequencing (mRNA-seq) performed on dissected guts 4 hours post infection. In parallel, for each line, we also sequenced guts of sucrose-fed flies as controls. Each genotype and condition were replicated once. Since the DGRP lines are highly polymorphic, we opted for analyses on individualized genomes. For that, we used the available genotype data (W. Huang et al., 2014), including single nucleotide variants as well as indels and structural variations, to generate individualized genomes and gene annotations (see **Methods**) which we used throughout the analyses. Seven of the lines were already included in our previous study (Bou Sleiman et al., 2015), which allowed us to assess the biological reproducibility of the mRNA-seq experiment. After combining the expression count data from the two experiments and performing normalization and removal of batch effects, we performed conventional hierarchical clustering (**Supplementary fig. 2-1a**). This revealed that the samples from the same line and condition always cluster together, indicating that genotypic differences mediate expression-level differences and that batch effects are weaker than the infection or genotype effects. Principal Component Analysis (PCA) on the same data also supported this observation (**Supplementary fig. 2-1b-c**).

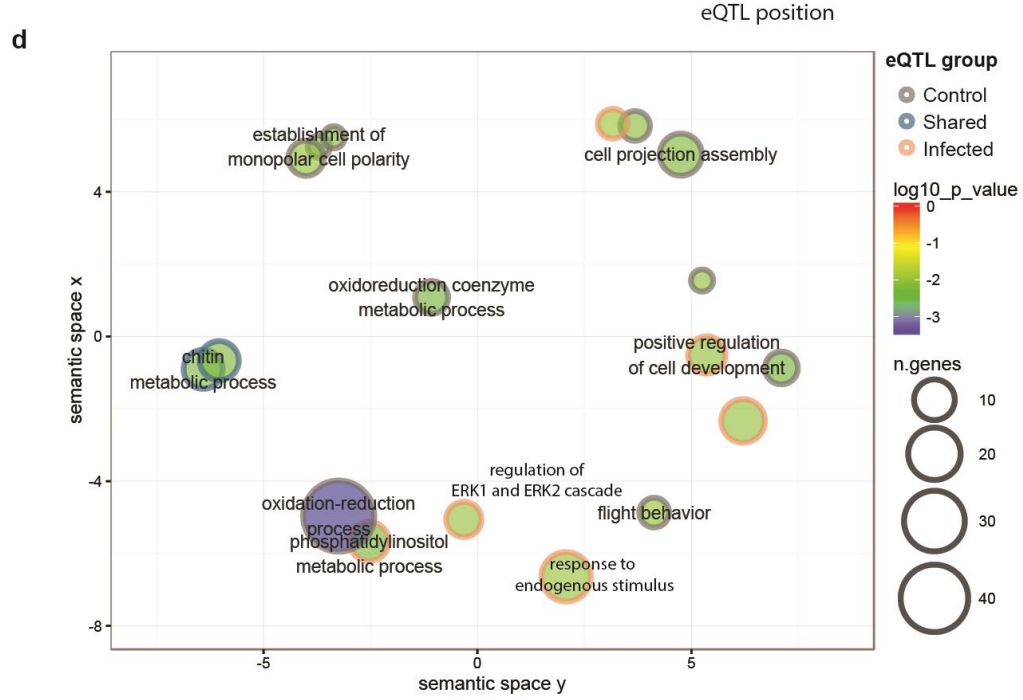
We then sought to catalogue the effect of genetic variation on gene expression levels for the two treatment conditions. To do so, we used Matrix-eQTL (Shabalin, 2012) to identify *local*-expression Quantitative Trait Loci (eQTLs) (i.e. within a window of 10 kb up- and downstream of genes) whose alleles correlate with the expression levels of nearby genes. We performed the analysis separately for the control and infected conditions and identified 7583 and 6644 *local*-eQTLs (p-value < 0.05) for 1459 and 1475 genes in the control and infected condition respectively (**Fig. 2-1b**). Interestingly, while 26% of *local*-eQTL-associated genes were shared between the two treatment conditions, the majority of detected *local*-eQTLs were condition-specific, emphasizing the substantial contribution of cryptic genetic variants to gene expression variation, especially in the presence of a strong transcriptome-altering stimulus such as infection. However, we found that the meta-distribution of detected *local*-eQTLs around the respective transcription start sites (TSSs) was similar between the two conditions. The

distribution also followed the expected pattern in that their density was highest around the TSS with a peak immediately downstream of the TSS, also involving the most significant associations (**Fig. 2-1c**). 26% of the genes expressed in the gut could also be linked to at least one eQTL, reflecting pervasive genomic variation-mediated gene expression differences.

To uncover pathways affected by genetic variation, we performed Gene Ontology analyses on the control, infected, and shared set of *local*-eQTL genes. This analysis revealed that the shared *local*-eQTL-associated genes are most enriched for the chitin metabolic process, possibly suggesting that genetic variation could be mediating differences in genes that are essential in maintaining the intestinal epithelium (**Fig. 2-1d**). Genes linked to control-specific *local*-eQTLs tended to be involved in the establishment of cell polarity, while infection-specific terms included a "response to endogenous stimuli" and the regulation of the ERK1/ERK2 cascade. This suggests that genetic regulatory variation in the infected condition might be affecting distinct biological processes. To provide an additional layer of characterization, we next explored whether infection-specific *local*-eQTLs are preferentially located in the proximity of specific transcription factor (TF) binding sites. To do so, we considered a region of 200bp around each eQTL and used AME from the MEME suite (Bailey et al., 2009; McLeay & Bailey, 2010) to test for TF motif enrichment in infection-specific compared to control-specific regions. These analyses revealed only two TF motifs, for Atf-2 and Disco (**Fig. 2-1e**), that passed the 5% FDR threshold. Interestingly, Atf-2 is a known downstream effector of the p38 pathway that has already been identified as a major player controlling gut immunity (Sveta Chakrabarti et al., 2014; E.-M. Ha et al., 2009; Sano et al., 2005), suggesting that infection-specific gene expression variation is in part driven by inter-individual Atf2 DNA binding differences. For Disco, no link with gut immunity has so far been reported. Taken together, our analyses catalogued a large set of genomic loci that affect gene expression levels only in the infected condition, collectively rendering them interesting candidates for examining their role in influencing the overall susceptibility of *Drosophila* to infection.







e

Logos	Database	Name	FDR
	Fly Factor Survey	Atf-2	7.03×10^{-3}
	Fly Factor Survey	Disco	8.58×10^{-3}

Figure 2-1 : Expression profiling of phenotypic extremes does not reveal consistently differentially expressed genes between classes

(a) Study design: 30 adult female flies from two phenotypic extremes (18 resistant and 20 susceptible) of the DGRP were infected orally with Pe or fed sucrose. Whole guts of ~30 flies were dissected per condition and line, then RNA-sequencing was performed. Sequencing reads were mapped to individualized genomes, and the number of reads was counted per gene.

(b) Infection leads to the differential expression of around 2400 genes (BH-corrected p-value < 0.05, fold change > 2).

(c) Metaplot of locations of cis-eQTLs with respect to their associated genes' transcription start sites (TSS). Solid grey line and dashed orange line are for the control and infected conditions respectively.

(d) Graphical representation of enriched biological process gene ontology terms based on the lists of genes with significant cis-eQTL associations. The GO analysis was performed using the GOstats (Falcon & Gentleman, 2007) R package (Hypergeometric test p-value < 0.005), and REVIGO (Supek, Bošnjak, Škunca, & Šmuc, 2011) was used to reduce redundancy in the ontology groups and plot them by semantic similarity (allowed similarity = 0.7). The size of the circle indicates the number of genes belonging to a certain GO category, and the color indicates enrichment significance.

(e) TF motifs that are enriched within local-eQTLs in the infected compared to the control condition.

2.3.2 Large-scale *in vivo local*-eQTL characterization via allele-specific expression

We have so far uncovered many shared and condition-specific *local*-eQTLs, but our analyses did not inform on the validity of the identified eQTLs nor whether these are *cis*- or *trans*-acting. eQTL studies have often focused on validating the effect of a particular variant on relevant genes using molecular biology techniques such as chromatin immunoprecipitation and small-scale reporter assays to measure the effect of an eQTL (Jin, Jung, DebRoy, & Davuluri, 2016; Lawrenson et al., 2015). While the recent emergence of Massively Parallel Reporter Assays allows for a much more systematic analysis of the regulatory effect of variants in transcriptional elements (Inoue & Ahituv, 2015; Maricque, Chaudhari, & Cohen, 2019; Tewhey et al., 2016), these assays are still unable to consider the complex interaction between genetic variation and gene expression. We therefore decided to exploit our experimental setting to thoroughly validate the detected *local*-eQTLs and explore their *cis*-regulatory nature by investigating their effect in a different genetic background. Specifically, by implementing a large-scale allele-specific expression analysis, we aimed at examining whether *local*-eQTLs induce the expected imbalance in expression between maternal and paternal alleles in an F1 cross (Hu, Sun, Tzeng, & Perou, 2015; Zou et al., 2018).

To achieve this, we selected 19 DGRP lines and crossed them in a round robin scheme (**Fig. 2-2a and Supplementary fig. 2-2a**) to maximize the number of F1 offspring that feature heterozygous genotypes for our set of predicted *local*- eQTLs such that we could assess allele-specific gene expression and infer *cis*-regulatory effects. Using the F1 individuals, we infected 2-3 days old adult females for 4 hours and extracted RNA from their dissected guts. As a control, a similar number of female adults were fed sucrose and processed in similar fashion. We replicated this experiment a second time to obtain two biological replicates and subsequently used BRB-seq, the high-throughput transcriptomics approach developed by our lab (Alpern, Gardeux, Russeil, & Deplancke, 2018) to derive gene expression profiles for each of the processed samples (see **Methods**). We assessed the quality of the replicas by performing PCA and correlation analysis on the gene count matrix which revealed no major batch effects between replicate experiments and strong separation between infected and control samples (**Supplementary fig. 2-2b – 2-2e**).

We selected a subset of 19,121 eQTLs from the control condition and 19,166 ones from the infected condition among our detected *local*-eQTLs, with an FDR < 0.1 for further validation. To detect differential allele expression as driven by an eQTL-linked variant, we required at least one cross whose offspring would be heterozygous for the selected variant to assess

allelic expression imbalance. To identify those crosses, we used the DGRP freeze 2 genomic data resource (W. Huang et al., 2014). Even though we used only 19 DGRP lines, our design allowed us to comprehensively interrogate the majority of detected *local*-eQTLs on the subset of 38 lines. For instance, only 67 and 66 *local*-eQTLs from the control and infected conditions (0.35% of *local*-eQTLs in both conditions) could not be tested due to the absence of any F1 that is heterozygous at these loci. The average number of heterozygous crosses per eQTL variant was 6.5 and 6.4 for the control and infected conditions respectively (**Fig. 2-2b**). Although one eQTL is linked to one gene, it is possible that one gene may be affected by multiple eQTLs. The distribution of the number of eQTLs linked to each gene revealed that most genes are linked to one or two variants, with a maximum of 123 eQTLs linked to one gene. Moreover, we did not detect any difference in the distribution of eQTLs per gene between control and infected condition-linked eQTLs (**Supplementary fig. 2-2f**).

To detect *cis*-eQTL variant-driven allele-specific expression (ASE) over several different genetic backgrounds, we applied a generalized linear mixed model (GLMM) with the response modelled by a binomial test of maternal *vs* paternal reads and crosses as random effect. The binomial test has been widely used to detect allelic imbalance (J. Chen et al., 2016; Degner et al., 2009; Kukurba et al., 2014; McManus et al., 2010) and by adding the genetic background as a random effect, we can detect consistent allelic imbalance over multiple crosses. Thus, variants validated by our model are able to drive allelic imbalance across several genetic backgrounds. We applied strict cutoff parameters to the samples that were passed to the GLMM which eliminated approximately 16% and 18% of the *local*-eQTLs from the control and infected conditions respectively because those variants did not have sufficient reads to be considered in the analysis (**Fig. 2-2c and Supplementary fig. 2-2g – 2-2h)**. At the end, our model allowed us to uncover 11.8% of the control (2023 *local*-eQTLs with FDR < 0.1) and 11.5% of the infected (1990 *local*-eQTLs with FDR < 0.1) condition-linked *local*-eQTLs across all tested genetic backgrounds as *cis*-acting eQTLs. We next assessed if an increased number of F1 hybrids would result in a higher probability for a *local*-eQTL to be validated, but found no evidence for this (**Fig. 2-2b**). Interestingly, when we compared the adjusted p-values computed by Matrix-eQTL for the *local*-eQTLs to the adjusted p-values from the F1 data, we observed only a very modest correlation, indicating that a low p-value for a *local*-eQTL is not necessarily a good predictor of an actual *cis*-effect across mixed genetic backgrounds (**Fig. 2d and 2e**). Furthermore, we found no correlation between the computed effect size and the measured effect size in both control and infected conditions (**Supplementary fig. 2-3a and 2-3d**). However, we observed that when a *local*-eQTL is found to act in *cis*, there is a high probability that the computed beta accurately predicts the direction of the measured effect (**Supplementary fig. 2-3b – 2-3c and 2-3e – 2-3f**). We subsequently tested for a difference in

the meta-distribution around the TSS between *local* and *cis*-eQTLs but found again no striking difference when all variant positions were compared (**Supplementary fig. 2-3g and 2-3j**). To test if more variants linked to one gene would lead to variants situated further away from the TSS, we separated eQTLs into “single eQTL per gene” and “multiple eQTLs per gene” categories and tested whether these two categories where distributed differently, but found no difference in either case (**Supplementary fig. 2-3h – 2-3i and 2-3k – 2-3l,** Kolmogorov-Smirnov p-value = 1 and 0.1 in Control condition for unique and multiple eQTLs respectively and 0.35 and 0.55 in Infected condition).

We then tested if *local*-eQTLs that were not characterized as *cis* could have a measurable *trans*-effect instead. To do so, we applied a linear mixed model to the crosses that were homozygous for each variant, using the crosses as a random effect (see **Methods**). We were able to detect a *trans* effect for 7.5% of control and 11.9% infected condition non-*cis local*-eQTLs (1190 and 1833 *trans*-eQTLs with FDR < 0.1 in control and infected conditions respectively) (**Fig. 3f**). However, due to the restricted number of available, homozygous crosses, we could only test 62% and 61% of the non-*cis local*-eQTLs in control and infected conditions respectively, while also being relatively underpowered. In summary, we detected a large number of *local*-eQTLs across conditions, but the majority of those cannot be defined as *cis*-eQTLs in a mixed heterozygous background. Rather, we found that, even within a conservative and underpowered analytical framework for *trans*-effect analysis, already a non-negligible portion of these non-*cis local*-eQTLs feature a robust, measurable *trans*-effect.

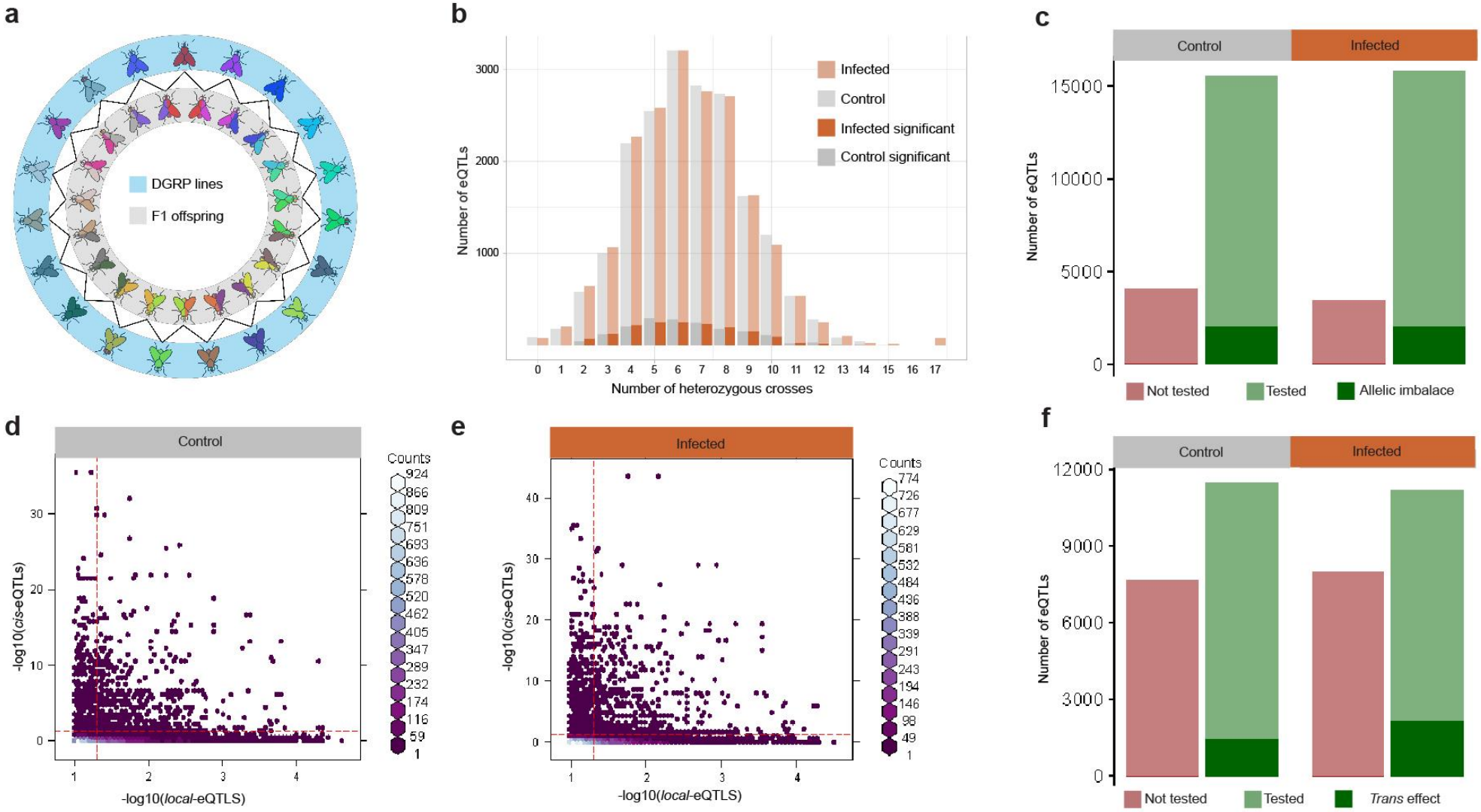

a
DGRP lines
F1 offspring
b
Infected
Control
Infected significant
Control significant
Number of eQTLs
Number of heterozygous crosses
c
Control
Infected
Number of eQTLs
Not tested
Tested
Allelic imbalace
d
Control
Counts
-log10(cis-eQTLs)
-log10(local-eQTLS)
e
Infected
Counts
-log10(cis-eQTLs)
-log10(local-eQTLS)
f
Control
Infected
Number of eQTLs
Not tested
Tested
Trans effect

Figure 2-2 *:* Predicted *cis*-eQTL validation by allele-specific gene expression

(a) Schematic of the round robin design: isogenic parental lines (blue) were crossed to two different lines and heterozygote F1 female offspring (grey) were used for infection and further processing.

(b) Distribution of the number of heterozygous crosses per eQTL in control (grey) and infected (orange) conditions. The distribution of *cis*-eQTLs (dark grey and dark orange) are not affected by the number of heterozygous crosses that are available to perform the calculations.

(c) Number of eQTLs passing the data cutoff for *cis*- and *trans*-characterization (light green) and rejected (red) in control (16%) and infected (17%) conditions. *Cis*-eQTLs are indicated in dark green with 6.8% and 6% of *local*-eQTLs in infected and control conditions respectively.

(d) And (e) Correlation between *local*-eQTL p-values (x-axis, -log10(Benjamini-Hochberg adjusted p-value)) compared to *cis*-eQTL calculated p-values (y-axis, -log10(Benjamini-Hochberg adjusted p-value)). Vertical and horizontal lines represent the 0.05 cutoff in control and infected Conditions

(f) Number of eQTLs passing the data cutoff for *cis*- and *trans*- haracterization (light green) and rejected (red) in control (16%) and infected (18%) conditions. *cis*-eQTLs are indicated in dark green with 11.8% and 11.5% of *local*-eQTLs in infected and control conditions respectively.

2.3.3 Few genes are significantly differently expressed between resistance classes

In a next step, we aimed to investigate how genetic variation influences the molecular and phenotypic differences between resistance classes. To first gain an unbiased, overall insight into the relatedness of the transcriptomes of the homozygous lines, we performed PCA on gene expression levels (**Fig. 2-3a and Supplementary fig. 2-1b**). While the infection effect is obvious and recapitulated by the first principal component (PC), lines from different resistance classes did not show any clear separation on the first two PCs. This is in contrast to our previous study, where we were able to see a modest separation on the second PC (Bou Sleiman et al., 2015). Furthermore, performing PCA on the expression levels within conditions yielded a similar result, with no obvious separation of the resistance classes on the first two principal components. A rationale for the disappearance of any separation compared to our previous study may include i) our expansion of the number of lines (from 8 to 20 per extreme), therefore reducing the phenotypic spread, or ii) the fact that the separation observed with the eight lines in our previous study may have been dominated by genotypic rather than treatment effects. Taken together, our findings suggest that, while the molecular impact of infection is similar among all tested lines and while the phenotypic differences are striking between the two resistance classes, the underlying transcriptomic differences are neither evident at the single gene- nor transcriptome-wide levels. This is in line with our previous findings that higher-level modules related to specific biological processes such as stress response, ROS metabolism and intestinal homeostasis (Bou Sleiman et al., 2015) could explain differences between resistance classes.

Using standard gene-based differential expression analysis, we identified around 2400 genes that are either up- or down-regulated 4 hours post *P.e.* infection (FDR<0.05, log fold change > 2, **Fig. 2-3b**). This is consistent with previous RNA sequencing and microarray results (Bou Sleiman et al., 2015; S Chakrabarti, Liehl, Buchon, & Lemaitre, 2012). Next, we explored gene expression differences between the resistance classes in the two experimental conditions. In our previous study, we had only found five and 34 mostly uncharacterized, differentially expressed genes in the control and infected conditions respectively. We reasoned that this low number may reflect either the underpowered nature of our previous study, involving only four lines from each resistance class, or that there are effectively few consistent differences between the resistance classes at the single gene level. Strikingly, when considering 38 lines, we found again no differentially expressed genes in the control condition, and only one gene, *nutcracker* (*ntc*), in the infected condition (**Fig. 2-3c**). This observation supports the notion that the differences between the classes, while being overt at the physiological level (i.e. being

alive vs dead), cannot be fully explained at the single gene level using standard differential expression approaches.

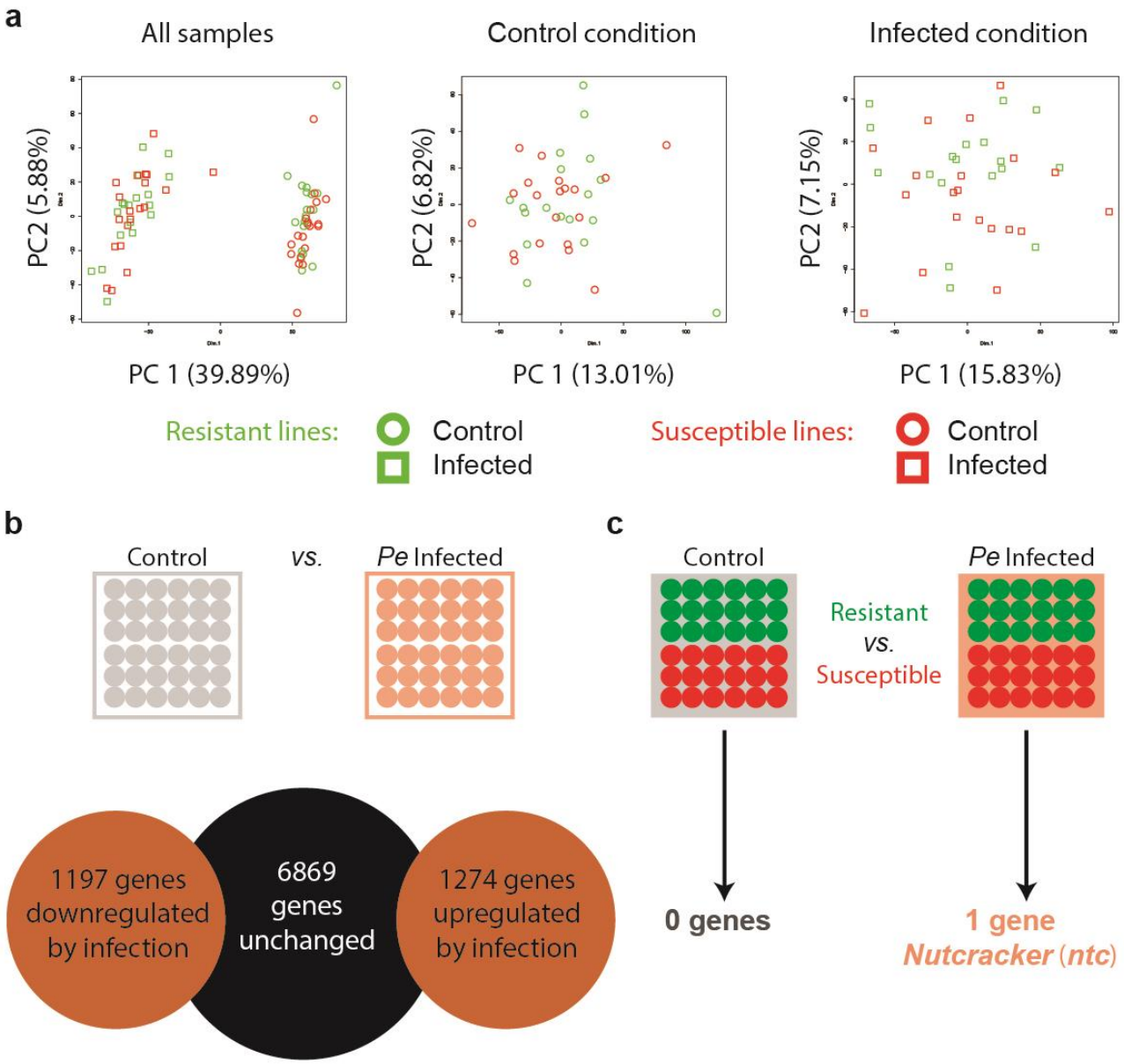

a
All samples
Control condition
Infected condition
PC2 (5.88%)
PC2 (6.82%)
PC2 (7.15%)
PC 1 (39.89%)
PC 1 (13.01%)
PC 1 (15.83%)
Resistant lines:
Control
Infected
Susceptible lines:
Control
Infected
b
Control
vs.
Pe Infected
1197 genes downregulated by infection
6869 genes unchanged
1274 genes upregulated by infection
c
Control
Pe Infected
Resistant
vs.
Susceptible
0 genes
1 gene
Nutcracker (ntc)

Figure 2-3 *:* local-eQTL analysis links natural variation to gene expression levels.

(a) Principal component analysis plots of all the samples (left), the control condition (middle), and the infected condition alone (right). The R package FactomineR was used to obtain the coordinates of each sample in the first two components, as well as the variance explained by each component (in parentheses).

(b) Variants with a minor allele frequency greater than 5 in the 38 lines and that are within a 10kb window of each expressed gene were tested for their association with gene expression levels. Results of two local-eQTL analyses (one for each infected condition) using Matrix eQTL (43) are presented in a Venn diagram (FDR < 0.05). The number of genes with significant associations is indicated in parentheses.

(c) When lines of the two resistance classes are compared within condition, no genes are significantly differentially expressed in the control condition, and only one gene in the infected condition.

2.3.4 The gene *nutcracker* is induced in resistant lines, has a validated *cis*-eQTL, and is involved in the gut immune response

As reported above, *ntc* was the only differentially expressed gene between the resistant and susceptible lines (**Fig. 2-3c**). Mining of our *local*-eQTL data revealed five infected condition-specific eQTLs belonging to two linkage groups, one group consisting of two eQTLs 7.6kb upstream and the other group composed of three 4.5kb downstream of its TSS (**Fig. 2-4a**), raising the question whether the *cis*-regulatory variation of *ntc* expression is one of the likely several mechanisms that contribute to resistance class stratification. This is further supported by our observations that *ntc* expression is induced in the gut after infection and that resistant lines tend to have greater *ntc* expression than susceptible ones (**Fig. 2-4b**)

We therefore first explored whether *ntc* affects gut immunocompetence given that its only described role so far is in sperm differentiation (Bader, Arama, & Steller, 2010). To do so, we used a null mutant line that harbors a point mutation in the F-box domain of *ntc*, ntc^{ms771} and tested its susceptibility to *P.e.* infection. Because flies homozygous for ntc^{ms771} are fragile and have a short lifespan (**Supplementary fig. 2-4a and 2-4b**) in both wildtype and infected conditions, we backcrossed them to their background line (*bw;st*). We assessed the survival of F1 offspring compared to their control, i.e., we compared the survival of *bw;st,+/TM6b* to $bw;st,ntc^{ms771}/TM6b$ and *bw;st,+/+* to $bw;st,+/ntc^{ms771}$. We observed decreased survival in all offspring flies harboring the *ntc* mutant allele. However, this decrease was stronger in the balancer line compared to the one without a balancer chromosome, suggesting that the *ntc* effect is stronger in weaker lines (**Fig. 2-4c**, $p < 0{,}0001$ and $p=0.081$, log-rank test). Furthermore, we performed RT-qPCR on dissected guts from the same lines and found that *ntc* expression is strongly reduced in mutant allele lines compared to control. Concurrently, the expression of the anti-microbial peptide *Dipt* was severely reduced in flies harboring the *ntc* mutant allele compared to controls (**Fig. 2-4d and 2-4e**). We replicated these findings using two lines harboring P-element-induced mutations, ntc^{f03797} and ntc^{f07259}, in or around the *ntc* locus (**Supplementary fig. 2-4c and 2-4f**). Interestingly, we also found that *ntc* is not expressed in the Rel^{E20} mutant line, a *Relish* loss of function line that disrupt the Imd pathway, upon infection (**Supplementary fig. 2-4f**). Together, these results suggest that loss of or decreased *ntc* expression negatively influences the enteric immune response through downregulation of the Imd pathway effectors upon *P.e.* infection.

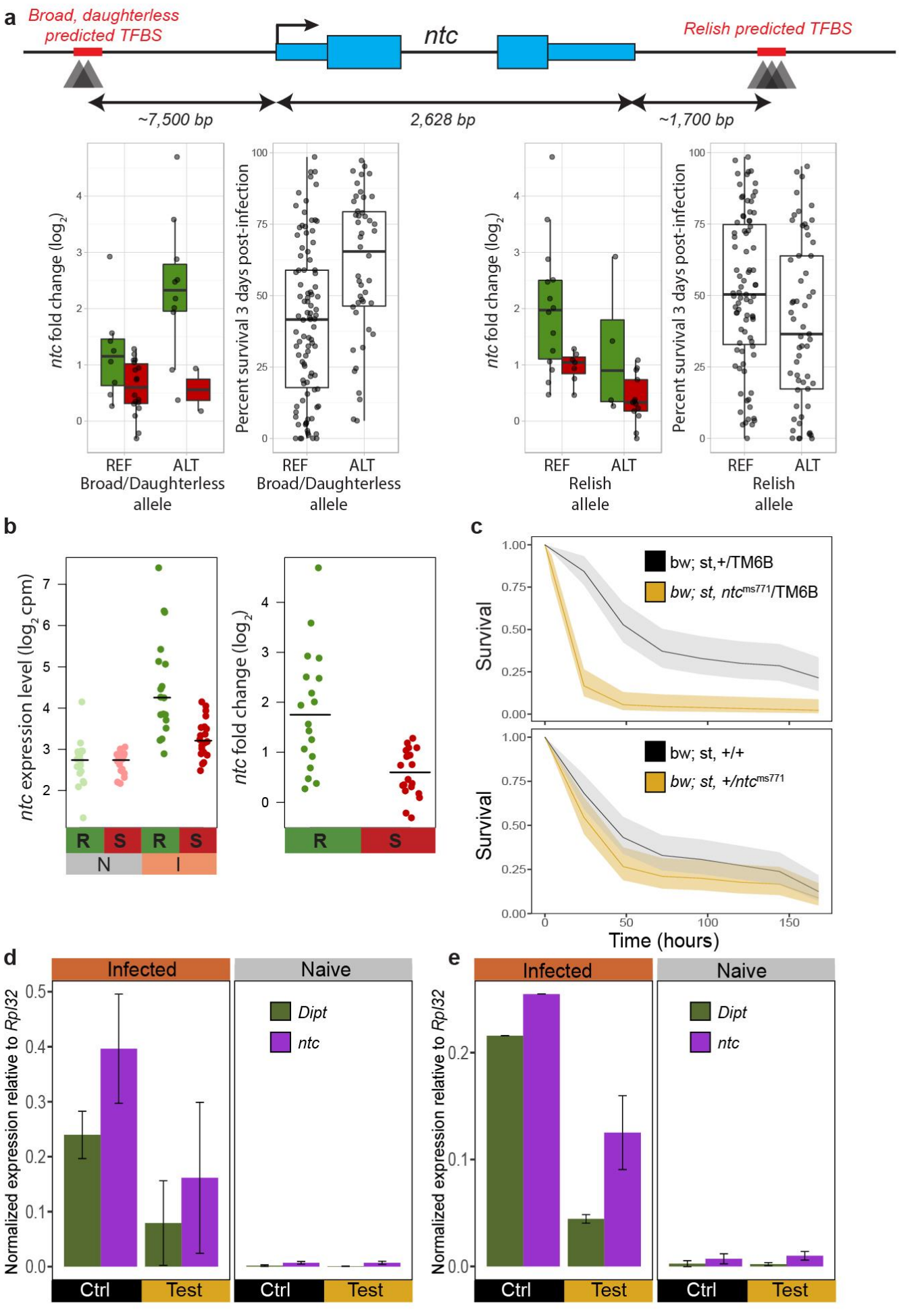

a
Broad, daughterless predicted TFBS
ntc
Relish predicted TFBS
~7,500 bp
2,628 bp
~1,700 bp
ntc fold change (log2)
Percent survival 3 days post-infection
REF
ALT
Broad/Daughterless allele
Relish allele
b
ntc expression level (log2 cpm)
ntc fold change (log2)
R
S
N
I
c
Survival
bw; st,+/TM6B
bw; st, ntcms771/TM6B
bw; st, +/+
bw; st, +/ntcms771
Time (hours)
d
e
Infected
Naive
Dipt
ntc
Normalized expression relative to Rpl32
Ctrl
Test

Figure 2-4 *:* The gene nutcracker is induced in resistant lines, has cis-eQTLs, and is involved in the gut response.

(a) Top panel: schematic of the ntc gene with specific annotations: cis-eQTLs around the ntc locus, and their overlap with predicted transcription factor binding sites (TFBS). TFBS prediction was done using FIMO (86) and motifs from the Fly Factor Survey (87) and OnTheFly (88) databases. The expression fold change by resistance class and two of those alleles (termed the broad/daughterless allele (right panel), and the relish allele(left panel)) is plotted, as well as the percentage death of 140 DGRP lines (30).

(b) Left panel: Expression level (in log2(cpm)) of the ntc gene by resistance class and infected condition. Right panel: Fold change of ntc by resistance class.

(c) Survival of lines harboring a null mutant allele ntc^{ms771} with (top panel) and without (bottom panel) TM6B balancer upon Pe infection compared to control. Log ranked test $p < 0{,}0001$ and $p=0.081$ for lines with and without balancer chromosome respectively. Shaded area represent 95% confidence interval.

(d) Gene expression of *ntc* and *Dipt* as measured by qPCR, normalized to *Rpl32*. Control: *bw;st,+/TM6B* and Test: $bw;st,ntc^{ms771}/TM6B$.

(e) Gene expression of ntc and Dipt as measured by qPCR, normalized to Rpl32. Control: *bw;st,+/+* and Test: $bw;st,+/ntc^{ms771}$.

(f) Data presented in (c), (d) and (e) are based on at least three biological replicates.

2.3.5 Determining the *cis*-regulatory mechanism underlying differential *ntc* expression among resistant and susceptible DGRP lines

Given the potential phenotypic consequences of differential *ntc* expression between the two resistance classes, we next set out to better understand the genetic and molecular mechanisms underlying differential expression of the gene. A transcription factor motif scanning analysis of the *ntc* locus revealed several potential binding sites that overlapped with the *ntc*-linked *local*-eQTLs, including Broad Complex and Daughterless TF binding sites for the upstream *local*-eQTLs, and a Relish/NF-kB one for a downstream *local*-eQTL. The alleles at both sites showed a high correlation with *ntc* expression for the studied 38 lines, but when associated with enteric infection susceptibility variation among the 140 DGRP lines, the allele at the Broad/Daughterless site was more significant than the Relish/NF-κB binding site one (GWAS p-value of 6.1*10-5 vs. 0.024 respectively). However, both failed to pass the stringent, implemented nominal 1*10-5 p-value (Bou Sleiman et al., 2015), although the Broad allele was obviously very close. In addition, since the gene *IntS10* is physically closer to these variants than *ntc*, we would not intuitively have linked these variants to *ntc*.

Next, we investigated the impact of the *local*-eQTL variant on the binding activity of four different TFs predicted to bind the sites overlapping *ntc*-linked *local*-eQTL sites: Broad, Daughterless, Sage, and Relish. To do so and given the difficulty in performing line-specific ChIP on these TFs, we used our in-house MITOMI setup (Maerkl & Quake, 2007) to measure *in vitro* the binding affinity of the selected TFs to double-stranded 20-mers that encompassed the respective binding site and that represented either the reference or alternate alleles. These analyses showed that among all four tested TFs, only Broad, a protein able to act as both a repressor and an activator (Bayer, Holley, & Fristrom, 1996; Mugat et al., 2000), exhibited a differential binding activity (**Fig. 2-5a** and **Supplementary fig. 2-5,** Welch's t-test p-value = 0.0063), showing substantially reduced binding to the alternate compared to the reference binding site allele.

Because the increase in *ntc* expression upon infection is substantially higher in DGRP lines harboring the alternate Broad binding site allele and because the alternate allele has a weaker affinity for Broad, we hypothesized that Broad, in our study, acts as a repressor on *ntc*. Consequently, a decrease in Broad binding affinity would lead to less repression and thus increased *ntc* expression. To verify this hypothesis *in vivo*, we again turned to the round-robin BRB-seq data to measure the ASE at each variant, with 14 F1 lines being heterozygous for the variant (**Fig. 2-5b**). In the control condition, we detected no significant difference between the ratio of alternate read count over reference read count (**Fig. 2-5c,** t-test p-value = 0.21) in

the few samples in which *ntc* expression was detected, which was expected due to the low expression of *ntc* in wildtype flies. However, we found that, in the infected condition, the ratio was significantly skewed towards the alternate allele (**Fig. 2-5c**, t-test p-value = 0.04), supporting our hypothesis that the variant in the Broad TF binding site is a *cis*-acting eQTL that affects *ntc* expression. Together, these results present a compelling mechanism explaining how a variant located in a TF binding site contributes to variation in gut immunocompetence by altering the expression level of a particular gene that itself influences an organism's resistance to infection.

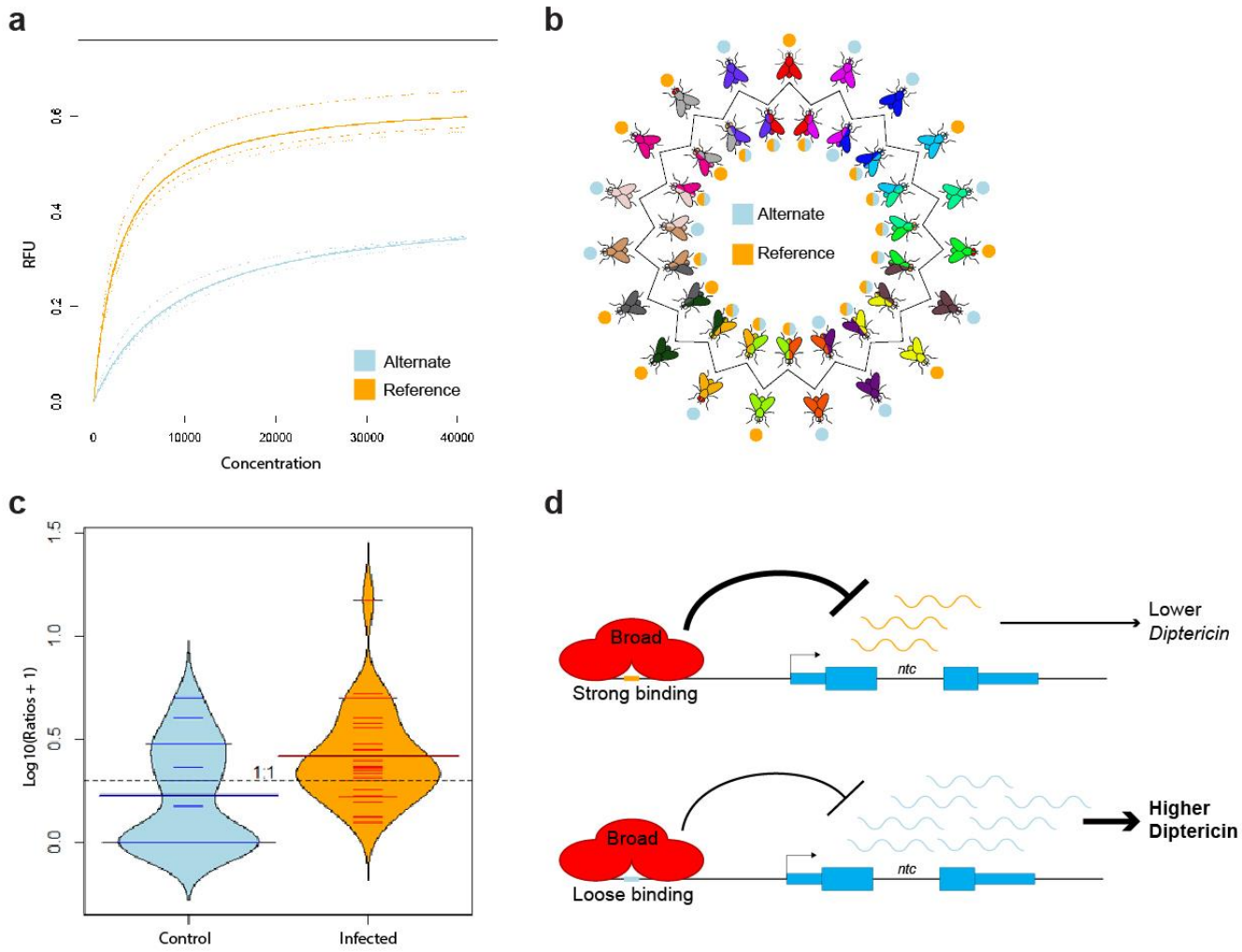


Figure 2-5 *:* Broad affinity and ntc expression is lowered for the alternate allele

(a) Measure of the binding affinity between Broad and the reference or alternate allele as measured by MITOMI in three different replicates.

(b) Repartition of lines in the round robin scheme based on reference or alternate Broad TFBS alleles.

(c) Ratios of read count mapping to the alternate over the reference allele reveal no difference in control condition (T-test, p-value = 0.21) but is significant in infected condition (T-test p-value = 0.04).

(d) Proposed model of ntc-mediated variation in gut immunocompetence: an enteric immune challenge increases ntc expression, while Broad acts as a repressor of ntc expression. The SNP in the Broad binding site decreases the binding affinity for Broad and thus the extent of ntc repression, resulting in greater ntc expression, which in turn increases Dipt expression.

2.4 Discussion

This study aimed at elucidating the effect of genetic variation on gene expression and organismal phenotypes in the context of *in vivo* enteric infection in the DGRP. One of the major findings that emerged is that DGRP lines with diametrically opposite resistance to infection all have a similar response after ingestion of a pathogenic bacterium, at least at an early timepoint after infection (**Fig. 2-1**). We show that this is not due to our inability to detect genotype-specific differences, since lines of the same genotype cluster together at the transcriptional level (**Supplementary fig. 1**). It is therefore clear that genomic variation imparts line-specific systemic differences on the transcriptome, yet only a small subset of those differences appears to be relevant in determining resistance.

To directly assess the effect of genomic variation on gene expression levels, we catalogued the possible *local*-eQTLs around all expressed genes. We found that in both the control and infected conditions, around a third of all associations are unchanged, confirming that genotypic differences indeed drive gene expression differences. However, the majority of *local*-eQTLs proved to be condition-specific, including the *local*-eQTLs at the *ntc* locus*,* suggesting an important contribution of cryptic variation to infection resistance (Gibson & Dworkin, 2004; Gibson, Powell, & Marigorta, 2015). Furthermore, our study allowed us to acquire unique insights into the regulatory nature of detected *local*-eQTLs. Most notable is that our study, to our knowledge the most comprehensive and systematic *in vivo local*-eQTL characterization effort to date, indicates that we tend to vastly overestimate the frequency of *cis*-eQTLs. This conclusion is in line with a previous study on mice in which only 17% of *local*-eQTLs could be defined as *cis*-eQTLs (Hasin-Brumshtein et al., 2014). Moreover, while we were able to still classify many *local*-eQTLs as *trans*, the majority of *local*-eQTLs remained unvalidated in variable genetic backgrounds*.* Of course, it is possible that the *cis* effect of a *local*-eQTL may be masked by other *trans*-acting eQTLs affecting the same gene (F. W. Albert & Kruglyak, 2015). Indeed, when several eQTLs were predicted to affect one gene, we were not able to disentangle their effects. In addition, a single polymorphism may drive differential expression and the other eQTLs may be merely in linkage disequilibrium (LD) with the effector SNP. It is also possible that a given variant is able to affect a gene only in a small set of genetic backgrounds and thus even more crosses would be required to increase the number of testable heterozygous genomic sites. Several confounding factors may also influence these validation numbers, including the fact i) that some variants may affect different target genes that are located farther away (e.g. in the case of intergenic variants) or that are even separated from the variant by other genes and ii) that some variants only affect a gene in combination with other variants (Y.-T. Huang, VanderWeele, & Lin, 2014). Importantly though, even if only

considering the validated *cis*-eQTLs*,* our earlier statement of pervasive, condition-specific gene expression variation between genotypes remains intact, since 10% of the validated *cis*-eQTLs were condition-specific. Interestingly, we found that highly significant *local*-eQTLs were not necessarily more likely to act in *cis*. However, when a variant was characterized as a *cis*-eQTL, then the *local*-eQTL measured directionality effect size was a good indicator of the *cis*-eQTL measured effect size.

Strikingly, we found only one gene differentially expressed between the resistant and susceptible lines, *nutcracker* (*ntc*). This gene was initially identified in a screen for mutants that failed to undergo sperm individualization due to their inability to activate caspases (Bader et al., 2010). Through its F-box domain, *ntc* interacts with other partners to form an SCF (Skp, Cullin, F-box) ubiquitin ligase (E3) complex that controls caspase activity in *Drosophila* (Bader et al., 2011). Caspases play important roles in insect immunity and homeostasis through both apoptotic and non-apoptotic pathways. For instance, Dredd, the homolog of human Caspase-8 is required for Relish cleavage and activation (Leulier, Rodriguez, Khush, Abrams, & Lemaitre, 2000). Furthermore, activation of the IKK complex is dependent on ubiquitination (R. Zhou et al., 2005), and studies in mammals have shown that commensal bacteria can affect ROS levels, leading to modification of the activity of the SCF complex, thus affecting NF-κB signaling (Kumar et al., 2007). While there are therefore several possible functional scenarios, the exact function of *ntc* in the gut and specifically enteric infection remains unclear and should be the subject of a more mechanistic, follow-up study. However, we were able to demonstrate that impaired *ntc* expression and null mutants of *ntc* negatively impact the survival of flies harboring these mutations. This effect is stronger in flies with a genetic background that is itself more susceptible to infection, such as lines harboring balancer chromosomes. This could be interpreted as a result of the sum of several different factors that are, when taken individually, not impactful, but lead to an increase in susceptibility when combined, as is suggested by our RNA-seq results not displaying strong separation between resistant and susceptible lines. Moreover, we were able to show that *Dipt* expression is severely reduced in the absence of *ntc*, showing a direct impact of *ntc* expression on potent immune response effectors.

We thereby uncovered how a SNP in a TF binding site proximal to *ntc* may impact its expression upon enteric infection. It is by now well-established that variants in TF binding sites can impact binding affinity and in turn the expression of the respective target gene (Deplancke et al., 2016; Haldane, Manhart, & Morozov, 2014). Here, we found that only one mutated binding site out of two possible *local*-eQTL sites displays variable binding affinity to a TF, namely Broad. Furthermore, allele-specific expression of F1 hybrids carrying the two alleles showed that the two copies of *ntc* are being induced differently, demonstrating a *cis* effect of

the SNP on the expression of *ntc*. These results establish a causal relationship between the binding site variant and variable *ntc* expression through differential binding of the TF *Broad*, constituting to our knowledge a rare example of an eQTL that modifies an ecologically-relevant complex trait through its effect on binding of a specific TF in a particular environmental condition. That said, it is unlikely that the extreme phenotype observed for *ntc* mutants reflect all of the underlying molecular mechanisms differentiating the resistant and susceptible DGRP lines since the difference in *ntc* expression between susceptible and resistant lines is not as severe as those measured in the mutants.

Together, these observations support the following model regarding how the *ntc* locus mediates variation in enteric infection susceptibility (**Fig. 2-5d**): upon infection, the expression of *ntc* is increased, together with that of *Broad* as well as several other immune response genes, as inferred from (S Chakrabarti et al., 2012; De Gregorio, Spellman, Rubin, & Lemaitre, 2002). Given Broad's role as a repressor in metamorphosis (Karim, Guild, & Thummel, 1993), we hypothesize that this TF may also act as a negative (feedback) regulator of *ntc* expression. Consequently, in flies harboring the alternate allele showing diminished affinity for Broad binding, *ntc* repression is reduced, resulting in greater *ntc* expression. This in turn positively affects the expression of *Dipt* through an as yet unknown mechanism, resulting in greater infection resistance compared to susceptible lines.

2.5 Conclusions

Our study shows the advantage of allele-specific experiments as a complement to standard eQTL approaches to identify causal variants as well as the power of systems genetics to assign novel roles to genes in biological processes unrelated to their originally discovered roles. During our research, we did not take into account the fact that the gut is a highly regionalized organ (Nicolas Buchon & Osman, 2015; Marianes & Spradling, 2013) that consists of multiple cell types (Dutta et al., 2015). It is possible that some eQTLs could therefore be restricted to a certain cell-type or environment, which cannot be detected using our current strategy, but could be investigated in a follow-up study.

2.6 Material and Methods

2.6.1 Fly Stocks

DGRP lines were obtained from the Bloomington stock center and reared at room temperature on a standard fly medium with 12 hours light dark cycle. The fly medium we used is composed of (for 1L water): 6.2g Agar powder (ACROS N. 400400050), 58.8g Farigel wheat (Westhove N. FMZH1), 58.8g yeast (Springaline BA10), 100ml grape juice; 4.9ml Propionic acid (Sigma N. P1386), 26.5 ml of Methyl 4-hydroxybenzoate (VWR N. ALFAA14289.0) solution (400g/l) in 95% ethanol, 1L Water. We used w^{1118} and *bw;st* flies as wildtype. Various DGRP lines, ntc^{f03797} and ntc^{f07259} stocks were obtained from the Bloomington Stock Center. The *bw;st,*ntc^{ms771}*/TM6B* mutant stock was a kind gift from the Hermann Steller lab.

2.6.2 RNA sequencing

RNA extraction

RNA extraction was performed using Trizol Reagent (Invitrogen) following the standard protocol.

Library preparation and sequencing

Standard Illumina Truseq libraries were prepared from 1ng total RNA as measured by a Nanodrop 1000 device (Thermo Scientific) by the Lausanne Genomic Technologies Facility. Single end sequencing was performed for 100 cycles. Initially, 80 samples from 40 lines were sequenced but we excluded 4 samples from two lines. One of the lines was contaminated, as its reads were derived from two genotypes and another DGRP line had a smaller library size in one condition, with led to its elimination from the analysis.

Mapping to individualized genome

For each DGRP line, we generated an individualized fasta genome sequence based on homozygous variants in the published Freeze 2 DGRP genotypes and the Release 5 reference genome. We also generated individualized gene annotations by applying the offsetGTF tool included in the mmseq package (Turro et al., 2011) on the Ensembl BDGP5.25.

For each sample, reads were mapped to the respective genome using STAR aligner. Reads for each gene were counted using HTseq-count.

Normalization and differential expression

We used the edgeR package to perform TMM normalization, followed by conversion to Counts Per Million Voom with quantile normalization. When we combined samples from this study and the previous study, we used the same approach, starting from combined gene counts, with the addiction of the removeBatchEffect function in the limma package. Differential expression was performed in limma using the weights obtained by voom while adjusting for intra-line correlations using the duplicate correlation function with the DGRP lines as the blocking factor. The following model was used: y = treatment + class + treatment:class. For each predictor variable, genes having a fold change of 2 and a Benjamini-Hochberg corrected adjusted p-value of 0.05 were deemed differentially expressed.

Principal component analyses

The FactoMineR package was used to perform the principal component analyses with scaling and centering.

cis-eQTL analysis

We performed separate analyses for each infected condition using Matrix-eQTL (Shabalin, 2012). Variants that are within 10kb of an expressed gene and whose minor allele frequency is greater than 5 in the 38 tested lines were used. *Cis*-eQTL associations with an FDR corrected p-value that is less than 0.05 were considered significant. Metaplots were plotted in R. The GO analysis was performed using the GOstats (Falcon & Gentleman, 2007) R package (Hypergeometric test p-value < 0.005), and REVIGO (Supek, Bošnjak, Škunca, & Šmuc, 2011) was used to reduce redundancy in the ontology groups and plot them by semantic similarity (allowed similarity = 0.7)
All analysis were performed in R version 3.5.0

2.6.3 Round Robin BRB-seq

RNA extraction

Flies were killed in cold 70% Ethanol, the Ethanol was wiped and replaced with cold RNAse free 1x PBS supplemented with 0.02% Tween-20. 10 guts were dissected for each sample and placed in a screw cap Eppendorf tube containing 350 uL Trizol and 10 uL plastic beads. Samples were homogenized in a Precellys 24 Tissue Homogenizer at 6000 rpm for 30 seconds. Samples were then transferred to liquid nitrogen for flash freezing and stored at –80°C. For RNA extraction, tubes were thawed on ice, supplemented with 350 uL of 100% Ethanol before homogenizing again with the same parameters. We then used the Direct-zol™ RNA Miniprep R2056 Kit, with the following modifications: we did not perform DNAse I treatment, we added another 2 min centrifugation into an empty column after the RNA Wash step, finally elution was performed by adding 10 uL of RNAse-free water to the column, incubation at room temperature for 2 min and then centrifugation for 2 min. RNA was transferred to a low-binding 96 well plate and stored at -80°C.

BRB-seq Library preparation

RNA quantity was assessed using picogreen. Samples were then diluted to an equal concentration in 96 well plates. RNA was then used for gene expression profiling using the bulk RNA barcoding and sequencing (BRB-seq) approach recently developed by our lab (see Alpern et al., BiorXiv, 2018, for experimental details). This protocol is able to provide high-quality 3’ transcriptomic data by implementing an early multiplexing scheme as in single-cell protocols and at a fraction of the cost of its competitors (e.g. 10-fold lower than Illumina Truseq Stranded mRNA-seq). In short, the BRB-seq protocol starts with oligo-dT barcoding, without TSO for the first-strand synthesis (reverse transcription), performed on each sample separately. Then all samples are pooled together after which the second-strand is synthesized using DNA PolII Nick translation. The sequencing library is then prepared using cDNA tagmented by an in-house produced Tn5 transposase preloaded with the same adapters (Tn5-B/B) and further enriched by limited-cycle PCR with Illumina compatible adapters. Libraries are then size-selected (200 - 1000 bp), profiled using a High Sensitivity NGS Fragment Analysis Kit (Advanced Analytical, #DNF-474), and measured using a Qubit dsDNA HS Assay Kit (Invitrogen, #Q32851). Finally, 6-8 pg of libraries were sequenced twice with Illumina NextSeq 500 with 21 cycles for read 1 (R1) and 101 cycles for read 2 (R2), only for the second sequencing.

2.6.4 Allele specific expression analysis

Alignment

We first aligned the two libraries, only the R2 file, to the *Drosophila* reference genome release 3 and the BDGP5.25 release annotation using STAR 2.5.3a (Dobin et al., 2013) with the following relevant parameters --twopassMode Basic --outFilterMultimapNmax 1 --outSAMmapqUnique 60. Then we used an in-house built software (https://github.com/DeplanckeLab/BRB-seqTools) to annotate the two aligned BAM files with the R1 info (Barcode and UMI if the latter exists), generating read groups for each libraryXsample. Then the two BAM files were merged into a unique BAM file that was further sorted. Picard was then used to remove the duplicates using the read group information and the barcode tag (options BARCODE_TAG=BC READ_ONE_BARCODE_TAG=BX). One of the samples failed due to a very low amount of reads and was removed from further analysis (**Supplementary Fig. 2b**). We then used PicardTools (http://broadinstitute.github.io/picard) to add read groups, sort, index and remove duplicates using the UMI information (parameter BARCODE_TAG=BC READ_ONE_BARCODE_TAG=BX). We then used GATK (Depristo et al., 2011) to split N cigars reads and realign the reads following the GATK best practices (Depristo et al., 2011). Finally, we used an in-house built software that assigns the reads to the maternal or paternal lines based on the variants present in the read, using the DGRP Freeze 2.0 VCF file (W. Huang et al., 2014).

Allelic imbalance measurement

For each *local*-eQTL and its linked gene, we used the variant information to select only crosses that were heterozygous for the respective variant. We then applied a generalized linear mixed model (GLMM, R package lme4::glmer, binomial(alternate read count, reference read count) ~ (1|cross)) with the response modelled by a binomial distribution with the crosses as random effects and no fixed effect. For each *local*-eQTL, we only selected samples with a minimum number of reads superior to the maximum value between 6 or the 25^{th} quantile of the total of reads assigned to the lineage lines in each sample. The obtained P-values were then adjusted using the Benjamini-Hochberg method. The effect-size was computed as the inverse logit of the estimated intercept computed by the GLMM function.

Trans effect measurement

For each *local*-eQTL and its linked gene, we used the variant information to select only crosses that were homozygous for the variant. We used the log2 count per million of total read count normalized using voom and assigned them as alternate or reference variant. We then applied a linear mixed model (GLMM, R package lme4::lmer, log2(cpm) ~ variant + (1|cross)) using the normalized count as a response and modelled by the allele (reference or alternate) and the crosses as random effects. For each *local*-eQTL, we only selected samples with at least 2 homozygous crosses for each variant. The obtained P-values were then adjusted using the Benjamini-Hochberg method.
All analysis were performed in R version 3.5.1

2.6.5 Oral infection, Survival and qPCR

Oral infection was performed as previously described (Neyen, Bretscher, Binggeli, & Lemaitre, 2014). Briefly, 1-day old females were transferred to 29°C rearing conditions. When the female flies were 2-3 days old, they were starved for 2 hours and then transferred to a tube containing bacteria and allowed to feed on the bacteria for a maximum of 24 hours. To prepare the *P.e.* bacterial pellet, bacteria were plated from glycerol stocks on a standard LB-agar plate supplemented with 1% milk and grown overnight at room temperature. Two days prior to infection, one single colony was transferred to a 50 ml Erlenmeyer with 12.5 ml LB and incubated for 8 hours at 29°C with 180 rpm shaking. The pre-culture was then transferred to a 1L Erlenmeyer with 200 ml LB and the culture was incubated overnight using the same conditions as the pre-culture. The culture was then centrifuged at 2500 g at 4°c for 20 min. The remaining LB was discarded and the pellet was resuspended by pipetting up and down. The OD600 was measured using a CO8000 Cell density meter. The pellet was then diluted to a final OD600 of 100 with distilled water and supplemented with Sucrose to a final volume/volume of 1.25%. A control solution contained only Sucrose at the same concentration. A disc of whatman paper was layered on top of the food and 225 µl of the bacterial or control solution was added to the paper.

Survival

Flies were infected as described previously. 4h after infection, surviving flies were scored. After 24 hours of feeding on bacteria, flies were transferred to fresh tubes and survivors were scored. Then, every 24 hours, survivors were scored and flies were transferred to fresh

tubes every 48 hours. The R package Survival was used to compute the log-rank test to assess statistical differences between genotypes. The analysis was performed in R 3.5.1

qPCR

RNA was extracted using the same method as for the BRB-seq library preparation described above. cDNA was synthesized from 500 ng total RNA using *SuperScript II* enzyme (ThermoFisher 18064014). qPCR experiments were performed on a StepOnePlus Real-Time PCR system (Applied Biosystems) using the Power SYBR® Green PCR Master Mix (Applied Biosystems). Gene expression relative to the housekeeping gene *RpL32* was calculated separately for each biological replica. Primers used were: ntc_F GATCAGGTGGGGAAAAAGCAG and ntc_R : GTTGTTCGCTCAGGATTCGC

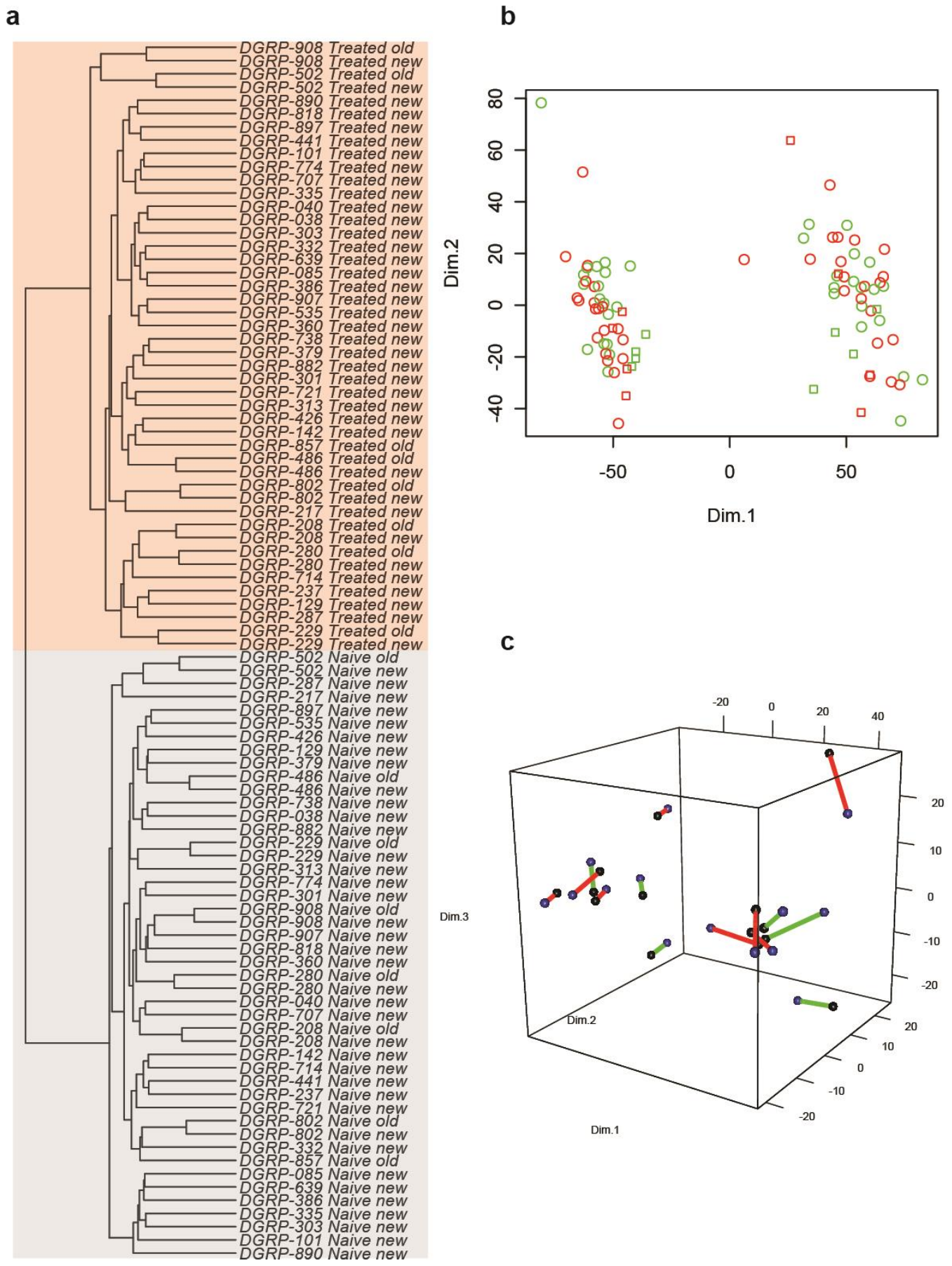
2.6.6 MITOMI

All target DNA fragments were obtained as single-strand oligonucleotides from IDT. These oligonucleotides were subsequently used to generate labeled double stranded oligonucleotides as described previously (Maerkl & Quake, 2007). TFs were expressed *in vitro* using the TnT SP6 High-Yield Wheat Germ protein expression system (Promega) with a C-terminal eGFP tag. The surface chemistry, MITOMI and image acquisition were performed as described previously (Isakova et al., 2017; Maerkl & Quake, 2007). We quantified the amount of each mutated sequence that is bound to the respective TF at the equilibrium state by means of fluorescence in a range of six input DNA concentrations. The obtained kinetic binding curves for each sequence were then fitted with the non-linear regression function according to the Michaelis-Menten law.

2.7 Acknowledgements

The *ntc*^{ms771} stock was a kind gift from Professor Hermann Steller. The Pseudomonas entomophila strain is a kind gift from Professor Gruno Lemaître. The authors would like to thanks Roel P. J. Bevers and Jörn Pezoldt for reviewing the manuscript. Sequencing was performed at the University of Lausanne Genomic Technologies Facility and at the EPFL GECF. Computational analyses were performed at the Vital-IT (http://www.vital-it.ch) Center for high-performance computing of the Swiss Institute of —Bioinformatics.

2.8 Supplementary material

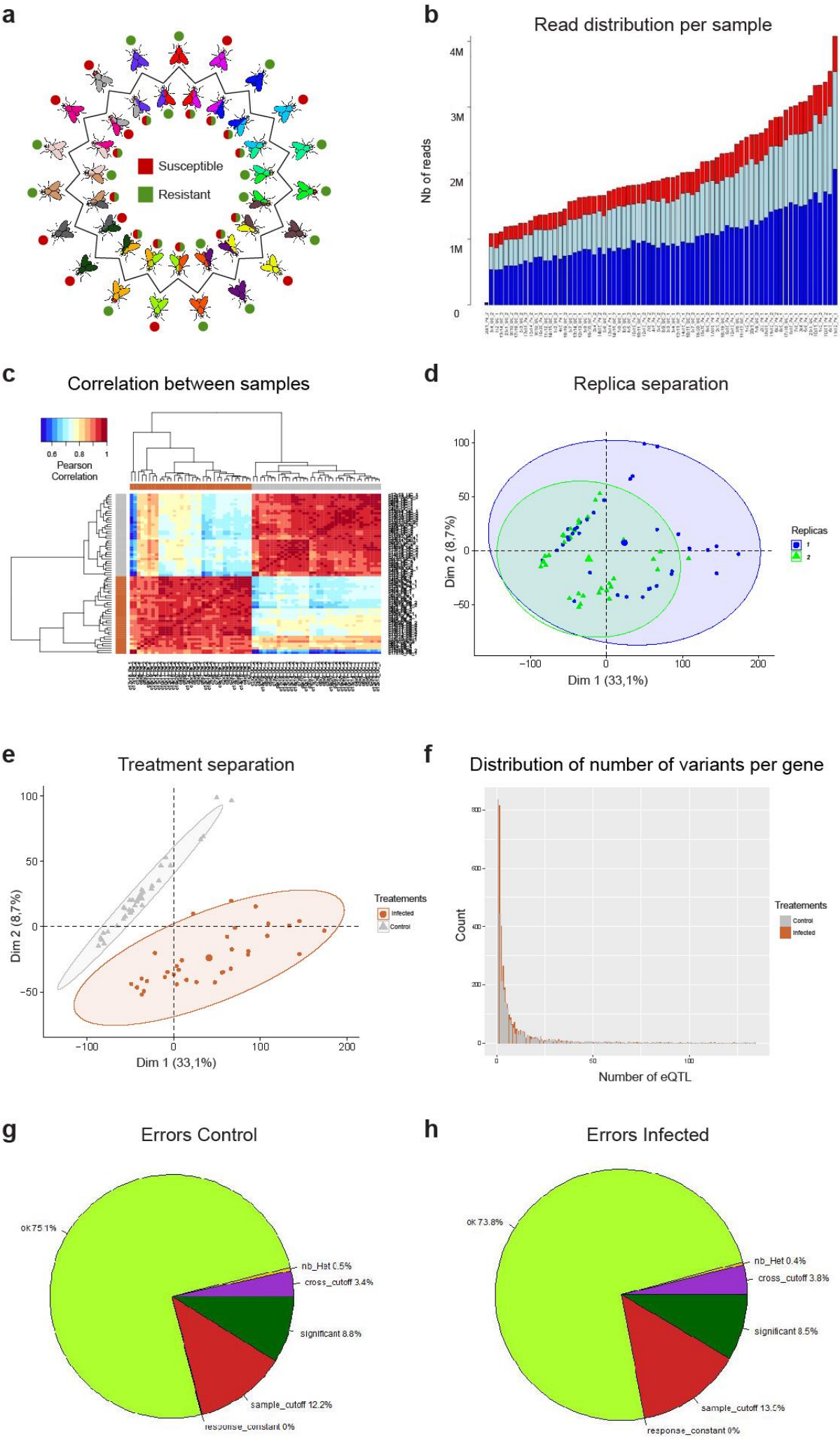

a
DGRP-908 Treated old
DGRP-908 Treated new
DGRP-502 Treated old
DGRP-502 Treated new
DGRP-890 Treated new
DGRP-818 Treated new
DGRP-897 Treated new
DGRP-441 Treated new
DGRP-101 Treated new
DGRP-774 Treated new
DGRP-707 Treated new
DGRP-335 Treated new
DGRP-040 Treated new
DGRP-038 Treated new
DGRP-303 Treated new
DGRP-332 Treated new
DGRP-639 Treated new
DGRP-085 Treated new
DGRP-386 Treated new
DGRP-907 Treated new
DGRP-535 Treated new
DGRP-360 Treated new
DGRP-738 Treated new
DGRP-379 Treated new
DGRP-882 Treated new
DGRP-301 Treated new
DGRP-721 Treated new
DGRP-313 Treated new
DGRP-426 Treated new
DGRP-142 Treated new
DGRP-857 Treated old
DGRP-486 Treated old
DGRP-486 Treated new
DGRP-802 Treated old
DGRP-802 Treated new
DGRP-217 Treated new
DGRP-208 Treated old
DGRP-208 Treated new
DGRP-280 Treated old
DGRP-280 Treated new
DGRP-714 Treated new
DGRP-237 Treated new
DGRP-129 Treated new
DGRP-287 Treated new
DGRP-229 Treated old
DGRP-229 Treated new
DGRP-502 Naive old
DGRP-502 Naive new
DGRP-287 Naive new
DGRP-217 Naive new
DGRP-897 Naive new
DGRP-535 Naive new
DGRP-426 Naive new
DGRP-129 Naive new
DGRP-379 Naive new
DGRP-486 Naive old
DGRP-486 Naive new
DGRP-738 Naive new
DGRP-038 Naive new
DGRP-882 Naive new
DGRP-229 Naive old
DGRP-229 Naive new
DGRP-313 Naive new
DGRP-774 Naive new
DGRP-301 Naive new
DGRP-908 Naive old
DGRP-908 Naive new
DGRP-907 Naive new
DGRP-818 Naive new
DGRP-360 Naive new
DGRP-280 Naive old
DGRP-280 Naive new
DGRP-040 Naive new
DGRP-707 Naive new
DGRP-208 Naive old
DGRP-208 Naive new
DGRP-142 Naive new
DGRP-714 Naive new
DGRP-441 Naive new
DGRP-237 Naive new
DGRP-721 Naive new
DGRP-802 Naive old
DGRP-802 Naive new
DGRP-332 Naive new
DGRP-857 Naive old
DGRP-085 Naive new
DGRP-639 Naive new
DGRP-386 Naive new
DGRP-335 Naive new
DGRP-303 Naive new
DGRP-101 Naive new
DGRP-890 Naive new
b
Dim.2
80
60
40
20
0
-20
-40
-50
0
50
Dim.1
c
-20
0
20
40
Dim.3
20
10
0
-10
-20
Dim.2
20
10
0
-10
-20
Dim.1

Supplementary Figure 2-1 : **Reproducibility of line-specific transcriptomes**

(a) Hierarchical clustering of the combined samples from this study and the previous one (Bou Sleiman et al., 2015). Hclust was used on the Euclidean distance matrix in R.

(b) Principal component analysis based on the gene expression profiles of the combined samples. Samples from the new and old study are represented as circles and squares, respectively.

(c) Three-dimensional representation of the first three principal components based only on the samples that belong to lines replicated between the two studies. Corresponding samples are connected by a segment that is colored based on susceptibility group. The sphere color indicates the batch (blue is new, black is old).


a
Susceptible
Resistant
b
Read distribution per sample
Nb of reads
4M
3M
2M
1M
0
c
Correlation between samples
Pearson Correlation
0.6
0.8
1
d
Replica separation
Dim 2 (8,7%)
Dim 1 (33,1%)
Replicas
1
2
e
Treatment separation
Dim 2 (8,7%)
Dim 1 (33,1%)
Treatements
Infected
Control
f
Distribution of number of variants per gene
Count
Number of eQTL
Treatements
Control
Infected
g
Errors Control
ok 75.1%
nb_Het 0.5%
cross_cutoff 3.4%
significant 8.8%
sample_cutoff 12.2%
response_constant 0%
h
Errors Infected
ok 73.8%
nb_Het 0.4%
cross_cutoff 3.8%
significant 8.5%
sample_cutoff 13.5%
response_constant 0%

Supplementary Figure 2-2 : **Quality control of BRB-seq libraries**

(a) Repartition of resistant and susceptible lines in the round robin scheme.

(b) Distribution of reads in each library. The left-most library was subsequently dropped from the analysis.

(c) Correlation matrix between samples computed on the count matrices displays good separation between treatment.

(d) Principal component analysis of gene expression from F1 lines, showing no major batch effect. Treatment-based PCA showing separation between different conditions.

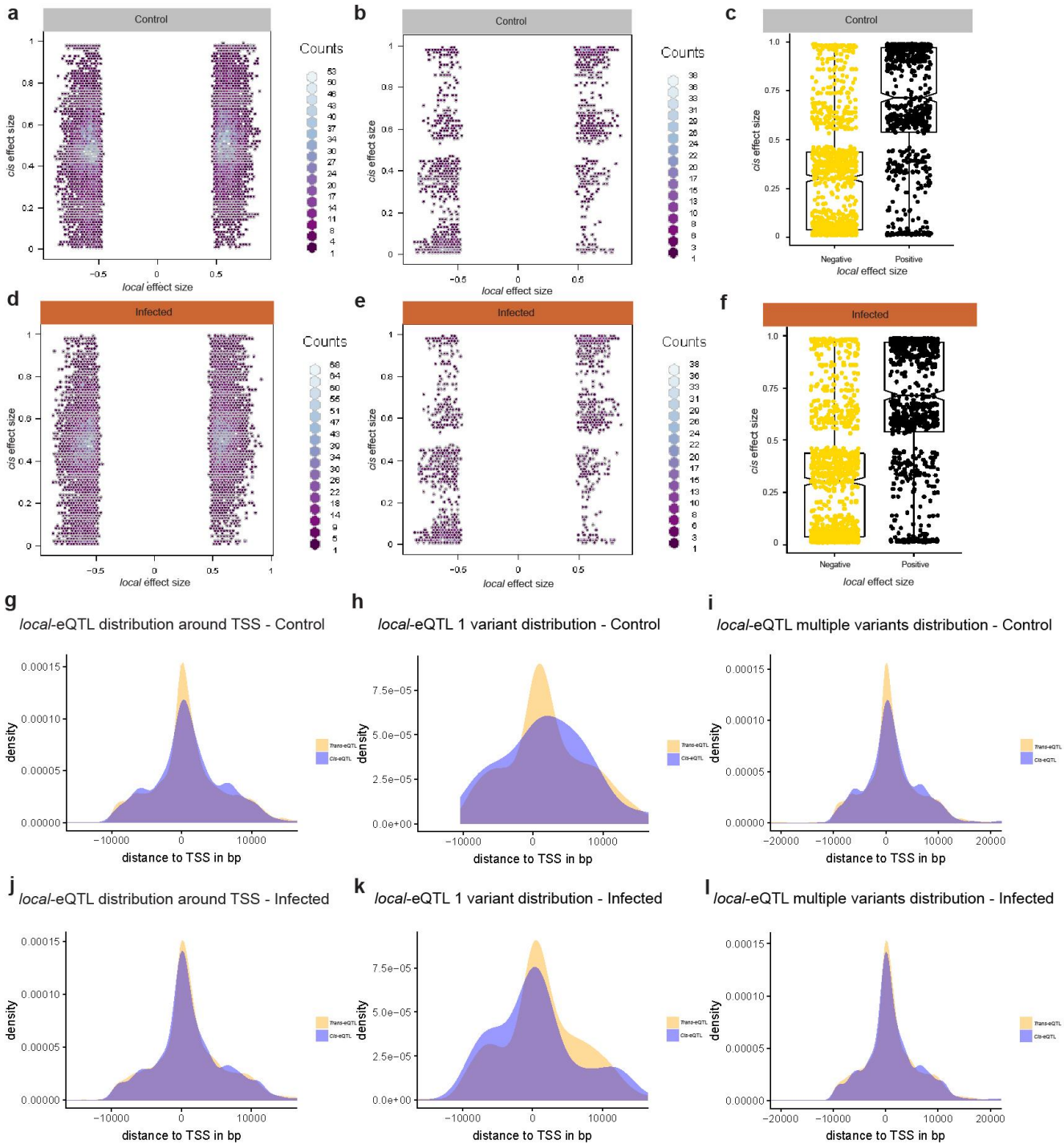
(e) Principal component analysis of gene expression from F1 lines, showing a separation between control and infected conditions.

(f) Distribution of number of variants per gene.

(g) Details of errors for tested *local*-eQTLs in control conditions: Nb_Het: not enough heterozygous crosses, cross_cutoff: not enough crosses pass the minimum requirement of assigned reads, sample_cutoff: not enough samples pass the minimum requirement of assigned reads.

(h) Details of errors for tested *local*-eQTLs in infected conditions.

.


a
Control
cis effect size
local effect size
Counts
b
Control
cis effect size
local effect size
Counts
c
Control
cis effect size
Negative
Positive
local effect size
d
Infected
cis effect size
local effect size
Counts
e
Infected
cis effect size
local effect size
Counts
f
Infected
cis effect size
Negative
Positive
local effect size
g
local-eQTL distribution around TSS - Control
density
distance to TSS in bp
h
local-eQTL 1 variant distribution - Control
density
distance to TSS in bp
i
local-eQTL multiple variants distribution - Control
density
distance to TSS in bp
j
local-eQTL distribution around TSS - Infected
density
distance to TSS in bp
k
local-eQTL 1 variant distribution - Infected
density
distance to TSS in bp
l
local-eQTL multiple variants distribution - Infected
density
distance to TSS in bp

Supplementary Figure 2-3 : **Comparison of predicted effect size and validated effect size**

(a) Correlation between *local*-eQTL effect size (x-axis) and *cis*-eQTL (y-axis) effect size showing poor overall correlation in the control condition.

(b) and (c) Correlation between *local*-eQTL effect size (x-axis) and *cis*-eQTL (y-axis) effect size showing good correlation in the control condition.

(d) Correlation between *local*-eQTL effect size (x-axis) and *cis*-eQTL (y-axis) effect size showing poor overall correlation in the infected condition.

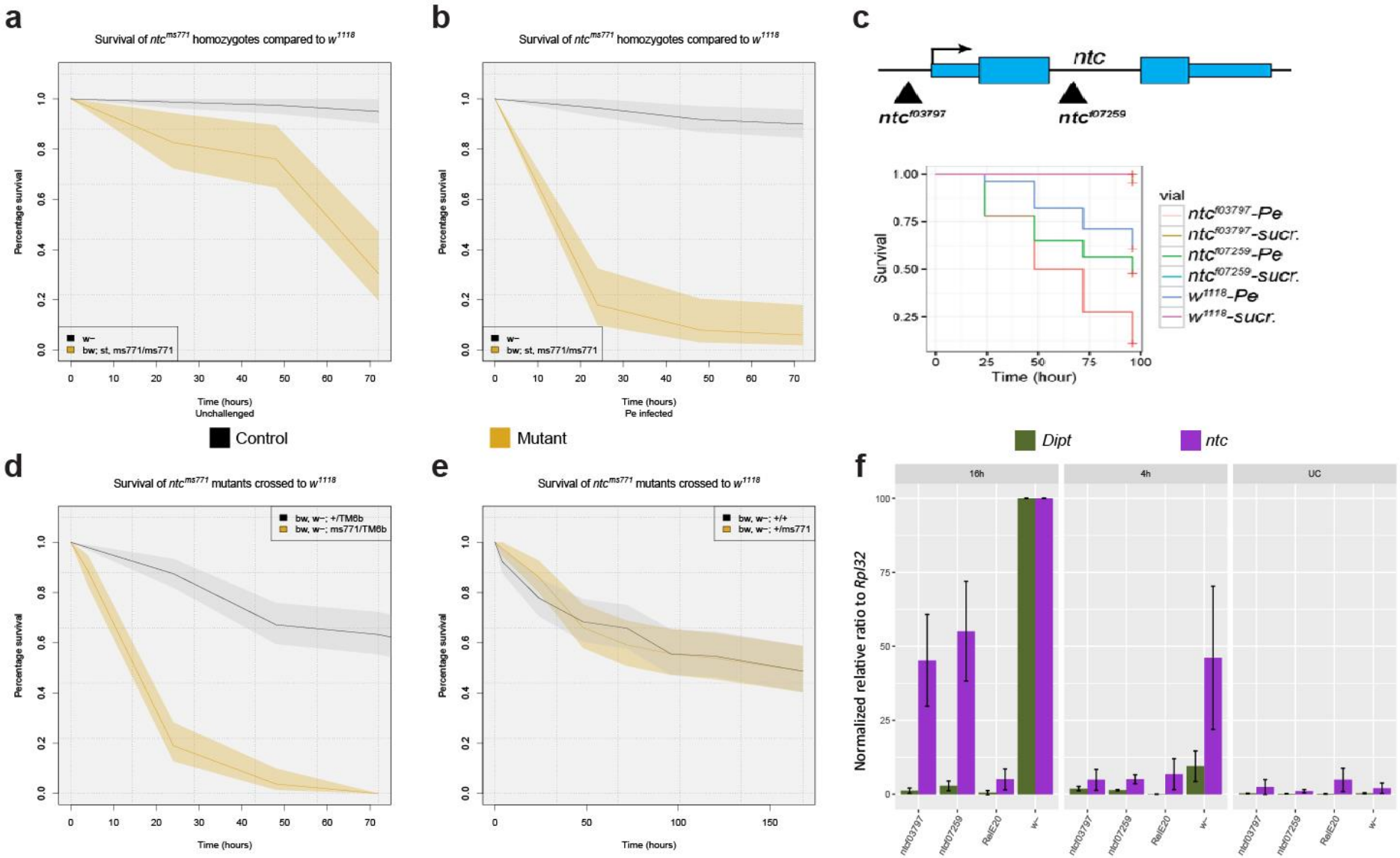
(e) and (f) Correlation between *local*-eQTL effect size (x-axis) and *cis*-eQTL (y-axis) effect size showing a good correlation in the infected condition.

(g) Distribution of *trans*- and *cis*-eQTLs around the TSS for the control condition.

(h) and (i) Distribution of *trans*- and *cis*- eQTLs around TSS based on the number of eQTLs linked to a gene (1 (e) or more (f)) in the control condition.

(j) Distribution of *trans*- and *cis*- eQTLs around the TSS from the infected condition.

(k) and (l) Distribution of *trans*- and *cis*- eQTLs around the TSS based on the number of eQTLs linked to a gene (1 (g) or more (i)) in the infected condition.


a
Survival of ntc^ms771 homozygotes compared to w^1118
Percentage survival
w–
bw; st, ms771/ms771
Time (hours)
Unchallenged
Control
b
Survival of ntc^ms771 homozygotes compared to w^1118
Percentage survival
w–
bw; st, ms771/ms771
Time (hours)
Pe infected
Mutant
c
ntc
ntc^f03797
ntc^f07259
Survival
Time (hour)
vial
ntc^f03797-Pe
ntc^f03797-sucr.
ntc^f07259-Pe
ntc^f07259-sucr.
w^1118-Pe
w^1118-sucr.
d
Survival of ntc^ms771 mutants crossed to w^1118
Percentage survival
bw, w–; +/TM6b
bw, w–; ms771/TM6b
Time (hours)
e
Survival of ntc^ms771 mutants crossed to w^1118
Percentage survival
bw, w–; +/+
bw, w–; +/ms771
Time (hours)
f
Dipt
ntc
16h
4h
UC
Normalized relative ratio to Rpl32
ntcf03797
ntcf07259
RelE20
w–

Supplementary Figure 2-4 : **Analysis of several *ntc* mutants**

(a) Survival of *ntc*^{ms771} homozygous flies in the control condition.

(b) Survival of *ntc*^{ms771} homozygous flies in the infected condition.

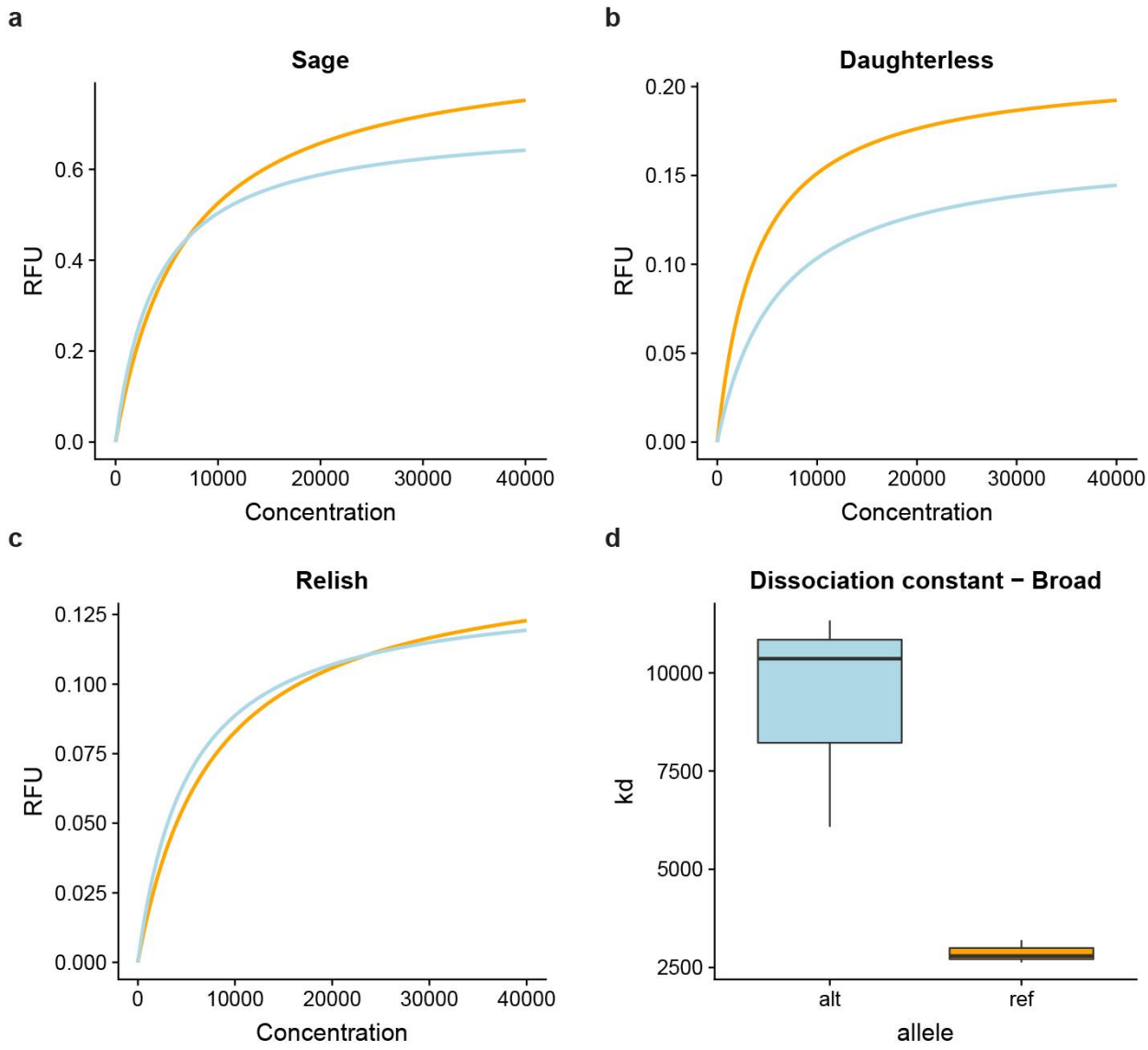
(c) Top panel: location of p-element insertions; bottom panel: survival of P-element lines upon *Pe* infection.

(d) Survival of *ntc*^{ms771} mutants crossed with *w*^{1118} with TM6B balancer chromosome.

(e) Survival of *ntc*^{ms771} mutants crossed with *w*^{1118} without TM6B balancer chromosome.

(f) qPCR-based expression of *ntc* and *Dipt* normalized to *Rpl32* in P-element lines, *w*^{1118} and *Rel*^{E20} in control and infected conditions. Data from at least three biological replicates.

Data presented in (a), (b), (c), (d), (e) and (f) are based on at least three biological replicates.



Supplementary Figure 2-5 : **MITOMI analysis of distinct TFs associated with the *ntc* locus**

(a) Measure of the binding affinity between Sage and the reference or alternate allele in one replica as measured by MITOMI (Isakova et al., 2017; Maerkl & Quake, 2007).

(b) Measure of the binding affinity between Daughterless and the reference or alternate allele in one replica as measured by MITOMI (Isakova et al., 2017; Maerkl & Quake, 2007).

(c) Measure of the binding affinity between Relish and the reference or alternate allele in one replica as measured by MITOMI (Isakova et al., 2017; Maerkl & Quake, 2007).

(d) Measure of the binding affinity between Sage and the reference or alternate allele in one replica as measured by MITOMI (Isakova et al., 2017; Maerkl & Quake, 2007).

Chapter 3: Proteomics of the *Drosophila melanogaster* enteric immune response reveals strong time- and pathogen-dependent protein signatures

Michael V. Frochaux[1], Maroun Bou Sleiman[2], Romain Hamelin3, Florence Armand3, Bart Deplancke[5+]

1. Michael Vincent Frochaux, Laboratory of System Biology and Genetics, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL) and Swiss Institute of Bioinformatics, Lausanne, Switzerland, michael.frochaux@epfl.ch
2. Maroun Bou Sleiman, Laboratory of Metabolic Signaling, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, maroun.bousleiman@epfl.ch
3. Romain Hamelin, Proteomics Core Facility, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, romain.hamelin@epfl.ch
4. Florence Armand, Proteomics Core Facility, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, florence.armand@epfl.ch
5. Bart Deplancke, Laboratory of System Biology and Genetics, Institute of Bioengineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, bart.deplancke@epfl.ch

+: Corresponding author

Note

This chapter describe the proteome of the Drosophila melanogaster gut after infection by Gram-negative bacteria. At the moment of the writing, this chapter is the latest draft of this work.

Author contributions: Michael Frochaux, Maroun Bou Sleiman and Bart Deplancke designed the study. Michael Frochaux and Maroun Bou Sleiman prepared the samples. Romain Hamelin performed the mass-spectrometry experiment. Florence Armand analyzed the raw data. Maroun Bou Sleiman performed the statistical and computational analyses on the RNA-seq

data. Michael Frochaux performed the statistical and computational analyses on the proteomics data and on the comparison between RNA and proteomics data. Michael Frochaux performed the loss of function screen.

3.1 Abstract

The enteric immune response in *Drosophila melanogaster* is a complex, multilayered process, which has been extensively characterized by several genome-wide gene expression studies. Surprisingly however, no study has to date investigated the proteomic response, even though discordances between gene and protein expression levels are well recognized. To address this, we performed mass-spectrometry and RNA sequencing on the guts of flies infected with one of two different bacteria, *Erwinia carotovora carotovora 15* (*Ecc15*) or *Pseudomonas entomophila* (*P.e.*) 4h and 16h post-infection. We found that a large portion of the measurable proteome (12%) varies after infection, often related to metabolism and immunity-related processes, and that protein changes are strongly time- and infection-dependent. We confirmed the relatively poor correlation between gene expression and protein abundance variation since up to one third of proteins with varying abundance did not show altered gene expression levels in the gut immune response. We then analyzed the potential role of several of these proteins that may have been overlooked in transcriptomics studies using a small loss-of-function screen. We found that out of 19 selected genes, 7 modulated the overall susceptibility to *P.e.* infection. Together, our study provides the first comprehensive characterization of the *Drosophila* gut proteome upon infection by Gram-negative bacteria, revealing widespread discordance between gene and protein expression levels as well as uncovering several novel gut immune response factors.

3.2 Introduction

Drosophila melanogaster has been used extensively as a model organism for immunity. Indeed, its immune system has been the subject of many important studies and seminal discoveries. Historically, the molecular mechanisms and pathways regulating the immune response were first thoroughly characterized. With the subsequent development of transcriptomic approaches, further molecular understanding of especially the fly systemic immune response was achieved (Boutros, Agaisse, & Perrimon, 2002; De Gregorio et al., 2002; Irving et al., 2001), revealing extensive variation at the RNA level and sequential activation of different molecular pathways. In these studies, the systemic immune response was typically provoked by a bacterial invasion of the fly's internal compartment, which raised the question whether other infection routes would induce a differential response.

Further research was therefore performed focusing on the response to oral infection, in which either larvae (Basset et al., 2000; Vodovar et al., 2005), or subsequently adult flies (Liehl et al., 2006) were fed infectious bacteria such as the mild pathogen *Erwinia carotovora carotovora 15* (*Ecc15*) (Basset et al., 2000) and the extremely potent entomopathogenic *Pseudomonas entomophila* (*P.e.*). Transcriptomic analysis of the adult fly gut after infection by either of these two Gram-negative bacteria revealed that the oral response does not rely on the Toll pathway and highlighted the importance of gut repair to maintain homeostasis as a critical step for survival (Nicolas Buchon et al., 2009; S Chakrabarti et al., 2012). Indeed, gut repair, acting through stimulation of intestinal stem cell division via the JAK/STAT and insulin receptor signaling pathways, was shown to be highly stimulated upon infection (Chatterjee & Ip, 2009). These studies thus shaped our understanding of the gut immune response by profiling the genome-wide gene expression changes during oral infection.

However, several studies in different model organisms including *Drosophila* have shown that mRNA level is, at best, an average predictor of protein abundance (Bonaldi et al., 2008; Butter et al., 2013; Casas-Vila et al., 2017; Griffin et al., 2002; Grün et al., 2014; Li et al., 2014; Schwanhäusser et al., 2011). Indeed, transcriptomic read-outs do not account for post-transcriptional processes such as regulation of mRNA transcription, miRNA regulation or mRNA stability / half-life (Barrett, Fletcher, & Wilton, 2012; Fabian, Sonenberg, & Filipowicz, n.d.; Lugowski, Nicholson, & Rissland, 2018). Furthermore, there are many protein-specific molecular mechanisms affecting its abundance, such as protein decay and export. Interestingly, variation in protein abundance was shown to be heritable in humans and *Drosophila* alike (Okada, Ebhardt, Vonesch, Aebersold, & Hafen, 2016; Ruffieux et al., 2019; Wu et al., 2013). Surprisingly however, despite the obvious need for dissecting complex

processes such as the *Drosophila* gut immune response at the proteome level, no such studies have to date been performed. In this study, we therefore set out to perform a comprehensive analysis of protein abundance changes in the fly gut facing oral infection by *Ecc15* or *P.e.*, aiming to uncover novel molecular mechanisms underlying the gut immune response.

3.3 Results

3.3.1 Proteomics of the fly gut after oral infection reveals extensive time- and infection-specific signatures

We harvested guts from *Ecc15* and *P.e.*-infected flies after 4h and 16h as well as control, sucrose fed flies at the same timepoints, using 4 replicates per condition. We subjected the guts to SDS-mediated protein extraction followed by trypsin digestion, liquid chromatography and mass spectrometry. The resulting spectra were compared to spectra generated from the Release 5 reference genome in MaxQuant (Cox & Mann, 2008). For each sample, we also extracted total RNA from dissected guts and performed RNA sequencing (RNA-seq) (**Figure 3-1a**). We first examined the quality of the mass-spectrometry derived dataset, looking at depth and reproducibility between samples. The average number of peptides identified and matched to the *Drosophila* proteome was around 45,000 (**Supplementary 3-1a**). This allowed us to reliably quantify an average of around 4000 proteins per sample with a total of 4503 proteins identified (**Supplementary 3-1b**), a number similar to other fly proteomic datasets from whole animals and guts (Casas-Vila et al., 2017; Tain et al., 2017). It is worth noting that in all *Ecc15*-infected samples, less peptides and thus proteins were identified and quantified compared to other conditions. As expected, peptides and proteins matching each bacterial species were found in the majority of corresponding samples.

To control for missing data, we performed imputation (Välikangas, Suomi, & Elo, 2017; R. Wei et al., 2018) following a normal distribution (**Supplementary 3-1c**, width = 0.3 and down-shift = 1.8). We then examined the correlation between samples at each timepoint. We found that the correlation is very high between replicas, but also between treatments (**Figure 3-1b**). Although the high correlation values obtained between replicas could be expected, the high correlation between treatments was surprising. This is also visible in the principal component analysis which failed to stratify samples upon consideration of all detected proteins (**Figure 3-1c**). These results suggest that most of the detected proteins are not affected by the immune response.

We then used Perseus (Tyanova et al., 2016) to identify proteins that exhibited differential expression (DE) after infection. We applied a weighted t-test between infected and control samples using the median of the variance and corrected for multiple testing using a permutation-based false discovery rate (250 permutations performed in Perseus). We found that the percentage of DE proteins depended on the harvesting time and type of infection (i.e. the type of pathogen) (**Figure 3-2** and **Table 3-1**). For example, 4h after infection, only 1.6%

and 1.8% of the detected proteins were differentially expressed for *Ecc15* and *P.e.* respectively. However, 16h after infection, 4.7% and 8.3% of the proteins were DE in *Ecc15* and *P.e.* infected flies respectively, indicating that proteome alterations were greater 16h compared to 4h after infection. Moreover, the impact of *P.e.* infection, based on the number of DE proteins, was stronger than that of *Ecc15* 16h post-infection, but not yet after 4h. We then performed PCA analysis using only DE proteins in at least one of the infection conditions (**Figure 3-3a**). This revealed a strong time- and infection-dependent sample stratification, except for the 4h and 16h control samples which clustered together. However, two *P.e.* infected samples did not cluster with the other ones. These two *P.e.*-infected samples were both harvested during our first biological replica. However, the uninfected and *Ecc15*-infected samples from the same replica are clustering accordingly with the same conditions of the other biological replicas. Thus, we ruled out any technical biases during the sample preparation or measurement phases. We therefore considered the pathogenicity of the utilized *P.e.* sample as a possible source of interference, revealing that it was indeed lower for the first replica as measured by the total amount of surviving flies three days post-infection (**Supplementary 3-2a**). This finding confirms the existence of a protein signature linked to pathogenicity or type of pathogen given the differences in samples reflecting low and high pathogenicity.

We then performed GO ontology analysis on all DE proteins for each condition (**Figure 3-3b**). For all four conditions, we observed an enrichment for immune response and biogenesis processes. Interestingly, DE proteins linked to *P.e.* infection were linked to ROS production whereas DE proteins for *Ecc15*-(16h) infected samples reflected a downregulation of the innate immune response. Furthermore, we more generally observed an enrichment for metabolism-related processes, and in particular mannose metabolism appeared to be downregulated after infection (**Figure 3-3c**).

Finally, we performed clustering on the DE proteins using k-means clustering (**Supplementary 3-3a**). To estimate the number of clusters, we used the within group sum of square and determined that 8 clusters would correctly separate the clusters. We observed that the correlation between clusters is relatively low with a few clusters (Cluster 2 and Cluster 3) being relatively close (**Supplementary 3-3b**). However, clustering supported the notion of strong infection- and time-specific proteomic signatures. For example, Clusters 2 and 3 harbored *Ecc15* specific proteins and Cluster 6 was *P.e.* specific. Cluster 4 contained proteins that were upregulated only at 16h whereas Cluster 8 had only proteins upregulated at 4h. We also uncovered a very specific time and pathogen-dependent cluster such as Cluster 7 containing proteins that were downregulated 16h after *Ecc15* infection and Cluster 5 which showed the opposite. Interestingly, Cluster 1 harbored anti-correlated proteins between *Ecc15* 4h and *P.e.*

16h, meaning that these proteins were downregulated 4h after *Ecc15* infection, not expressed for both *Ecc15* 16h and *P.e.* 4h and upregulated 16h after *P.e.* infection (**Supplementary 3-3c**).

Together, these results suggest that the proteomic gut response to oral infection is highly time- and pathogen-specific with various protein sets acting very differently. Moreover, metabolism-related processes emerged as a strong component of the gut immune response.

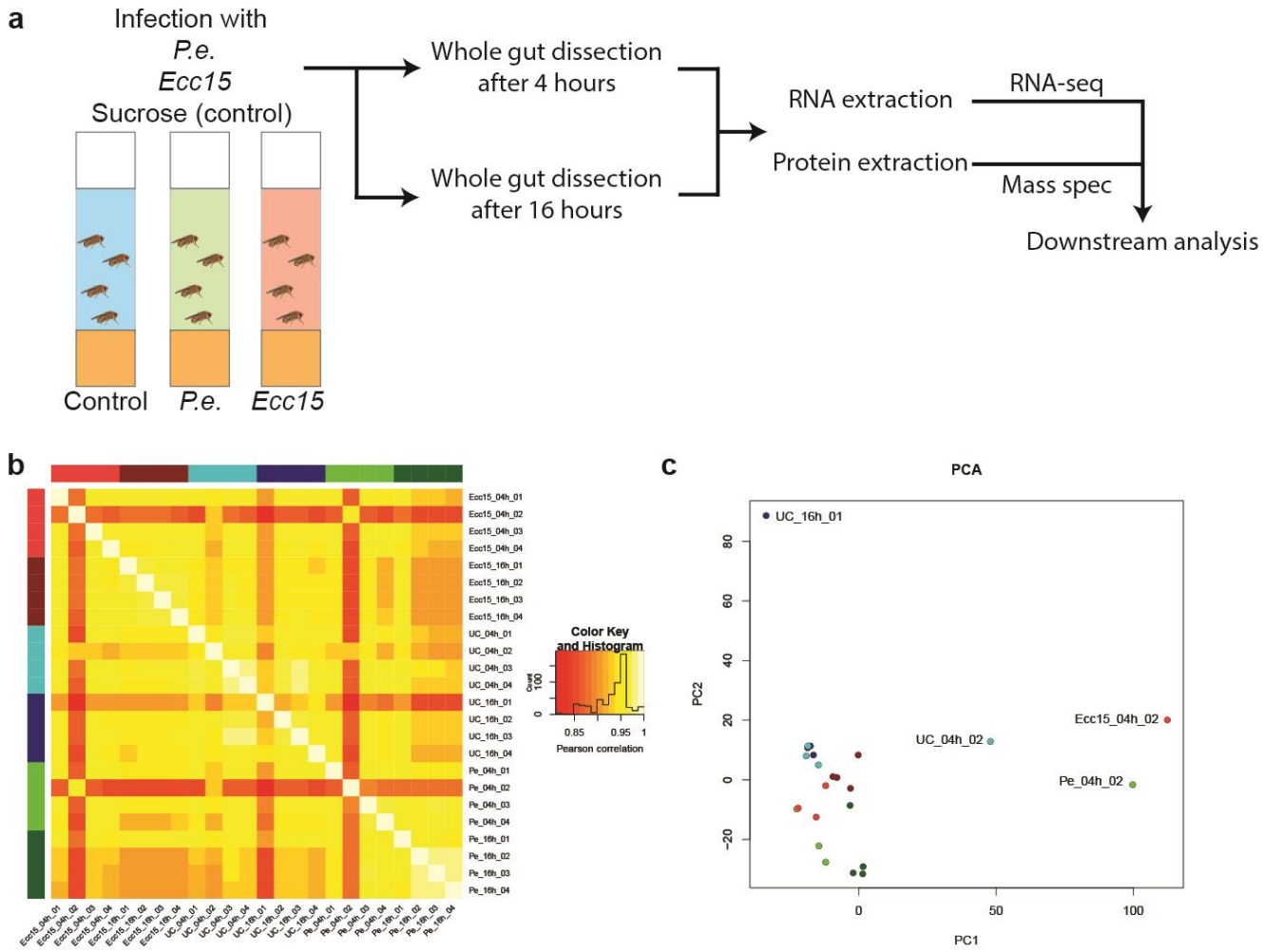


Figure 3-1: experimental procedure and data quality

(a) Schematic of the experimental protocol.

(b) Heatmap of the Pearson correlation on LFQ intensities between samples showing that the majority of the samples are highly correlated regardless of their treatment or time.

(c) PCA analysis of the samples showing that a few of the are separate from the core. Possible explanation may be technical variation during mass spectrometry.

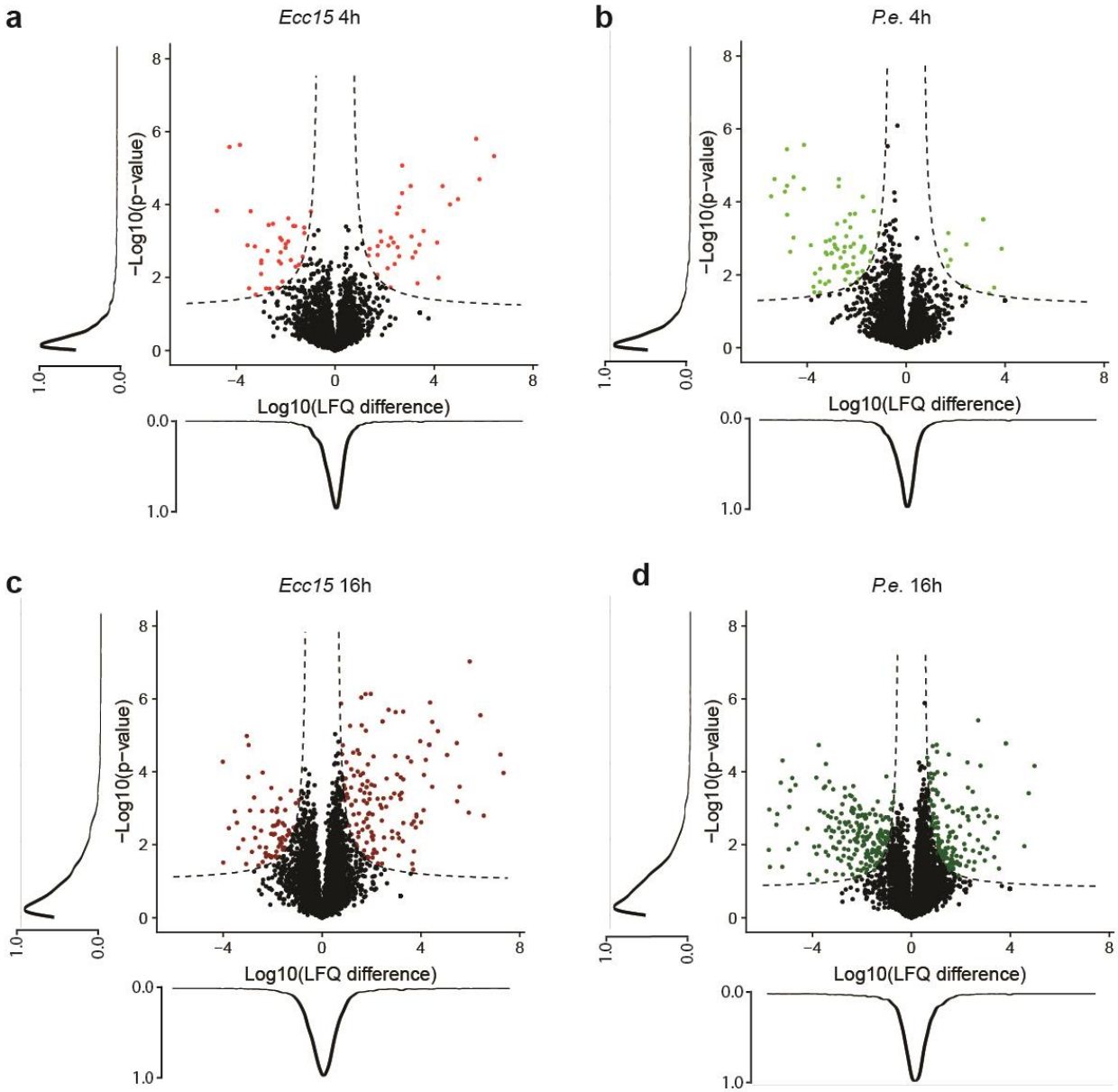


Figure 3-2 : Differentially abundant proteins

All plots display density curves of LFQ difference (bottom) and p-value (left). Dashed line represent the boundaries for a weighted t-test with FDR < 0.05 determined by random permutation (n=250)

(a) DE proteins 4 hours after infection with Ecc15.

(b) DE proteins 16 hours after infection with Ecc15.

(c) DE proteins 4 hours after infection with P.e.

(d) DE proteins 16 hours after infection with P.e.

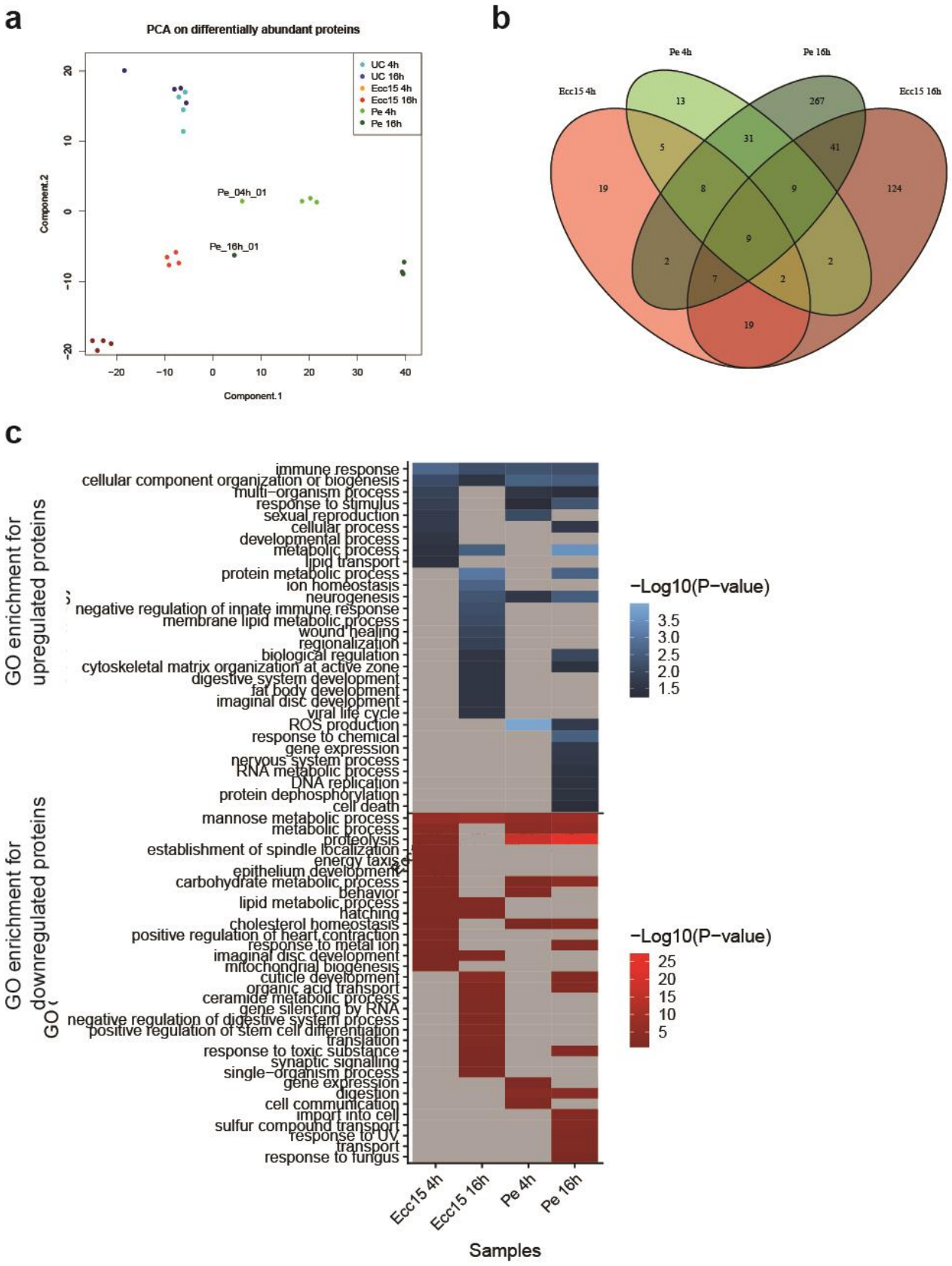

a
PCA on differentially abundant proteins
UC 4h
UC 16h
Ecc15 4h
Ecc15 16h
Pe 4h
Pe 16h
Pe_04h_01
Pe_16h_01
Component.2
Component.1
b
Pe 4h
Pe 16h
Ecc15 4h
Ecc15 16h
13
267
31
5
41
19
8
9
124
9
2
2
7
2
19
c
GO enrichment for upregulated proteins
immune response
cellular component organization or biogenesis
multi−organism process
response to stimulus
sexual reproduction
cellular process
developmental process
metabolic process
lipid transport
protein metabolic process
ion homeostasis
neurogenesis
negative regulation of innate immune response
membrane lipid metabolic process
wound healing
regionalization
biological regulation
cytoskeletal matrix organization at active zone
digestive system development
fat body development
imaginal disc development
viral life cycle
ROS production
response to chemical
gene expression
nervous system process
RNA metabolic process
DNA replication
protein dephosphorylation
cell death
−Log10(P−value)
3.5
3.0
2.5
2.0
1.5
GO enrichment for downregulated proteins
mannose metabolic process
metabolic process
proteolysis
establishment of spindle localization
energy taxis
epithelium development
carbohydrate metabolic process
behavior
lipid metabolic process
hatching
cholesterol homeostasis
positive regulation of heart contraction
response to metal ion
imaginal disc development
mitochondrial biogenesis
cuticle development
organic acid transport
ceramide metabolic process
gene silencing by RNA
negative regulation of digestive system process
positive regulation of stem cell differentiation
translation
response to toxic substance
synaptic signalling
single−organism process
gene expression
digestion
cell communication
import into cell
sulfur compound transport
response to UV
transport
response to fungus
−Log10(P−value)
25
20
15
10
5
Ecc15 4h
Ecc15 16h
Pe 4h
Pe 16h
Samples

Figure 3-3 *:* Differentially abundant proteins repartition

(a) PCA using only differentially abundant proteins display tight clustering of samples.

(b) Repartition of differentially abundant proteins by time and infection.

(c) GO terms enriched in upregulated (blue) and downregulate (red) for each time- and infection-specific samples.

3.3.2 Comparison between transcriptomic and proteomic gut immune response changes highlight an overall poor correlation

To explore the extent to which gene and protein expression changes are correlated during the gut immune response, we performed differential expression analysis using limma on RNA-seq data retrieved from the same samples that were used to assess protein levels. For each gene, its mRNA and protein expression can both go up, down, not change, differentially change or one (mRNA or protein) may not be detected. We first used these qualitative changes to compare variation in both mRNA and protein abundance after infection (**Figure 3-4** and **Table 3-2**). We observed a large number of proteins for which we did not detect any abundance change (i.e. "stable proteins") but whose mRNA levels were upregulated or downregulated. In *Ecc15*-infected samples, 16% and 21% of stable proteins were differentially expressed after 4h and 16h respectively. Interestingly, these proportions were even higher in *P.e.*-infected samples with 34% and 31% after 4h and 16h respectively, suggesting that oral infection induces many transcriptional changes without an obvious impact on respective protein levels. In addition, for about 4% of the stable proteins, we did not detect the corresponding mRNA. We then performed an in-depth investigation of the mRNA dynamics for all DE proteins. In both *P.e.*-infected samples, we found around 6.5% of DE proteins with an anti-correlated gene expression profile. Furthermore, 35% and 32% DE proteins exhibited no detectable change in gene expression in *P.e.*-infected samples at 4h and 16h respectively. Finally, 2% of the DE proteins in *P.e.* samples had no detectable mRNA levels. We found similar, but more dampened trends for *Ecc15*-infected samples. For example, the proportion of anti-correlated proteins was 4.2% and 2.8% for 4h and 16h respectively. However, we found more DE proteins in *Ecc15* samples with no detectable mRNA (4%) compared to *P.e.* samples.

We then investigated the extent of correlation between mRNA and protein levels (**Figure 3-5**). As expected based on the results discussed above, the correlation was poor between protein abundance and mRNA fold change (Spearman pair-wise correlation: *Ecc15* 4h: 0.23, *Ecc15* 16h: 0.35, *P.e.* 4h: 0.16, *P.e.* 16h: 0.19). Interestingly, we observed that the correlation is overall lower for *P.e.* compared to *Ecc15*-infected samples, supporting a previous report of *P.e.*-mediated translation inhibition (S Chakrabarti et al., 2012). We then investigated the same extent of correlation but now only involving DE proteins. Strikingly, the correlation was much higher for downregulated proteins in P.e. infected samples (Spearman pair-wise correlation downregulated proteins: *P.e.* 4h: 0.09, *P.e.* 16h: 0.36) compared to upregulated proteins (Spearman pair-wise correlation downregulated proteins: *P.e.* 4h: -0.03, *P.e.* 16h: 0.14) indicating negative regulation of translation on upregulated proteins. Interestingly, *P.e.* 4h showed an anti-correlation correlation whereas the 16h timepoint was positive but low.

However, by considering the abundance of bacterial proteins, we found that *P.e.* but not *Ecc15* is cleared from the gut after 16h (**Supplementary 3-4**). Thus, we can assume that the *P.e.* induce translational blockage has been effectively removed from the cell.

Non-coding RNA mediated translation control is an important control mechanism (Catalanotto, Cogoni, & Zardo, 2016). It was shown to be relevant in the fly immune response with several miRNAs being expressed in a time- and infection-type dependent manner (Atilano, Glittenberg, Monteiro, Copley, & Ligoxygakis, 2017; G. Wei et al., 2018). We thus sought to determine how miRNA regulation could affect DE proteins. We used the microRNA.org resource (Betel, Wilson, Gabow, Marks, & Sander, 2008) of miRNA predicted binding and the miRNA expression data (Atilano et al., 2017) to identify putative miRNA-regulated mRNAs. Specifically, we explored whether DE proteins have a greater likelihood of containing miRNA binding sites, and thus of possible stronger post-translational miRNA mediated regulation or whether an opposite scenario could occur. To do so, we calculated the enrichment of mRNA targets in our samples compared to all mRNAs. We did not find any difference when considering all DE proteins (Fisher's exact test p-value = 1). However, when we separated the DE proteins by condition, we found that the probability of a mRNA being less likely to be targeted by a miRNA to be higher in *P.e.*-infected compared to *Ecc15*-infected samples (Fisher's exact test: *Ecc15* 4h: 0.22, *Ecc15* 16h: 0.27, *P.e.* 4h: 0.06, *P.e.* 16h: 0.007)(**Table 3-3**). These results suggest that the gut immune response to *P.e.* infection has an overall lower level of miRNA regulation than that to *Ecc15*.

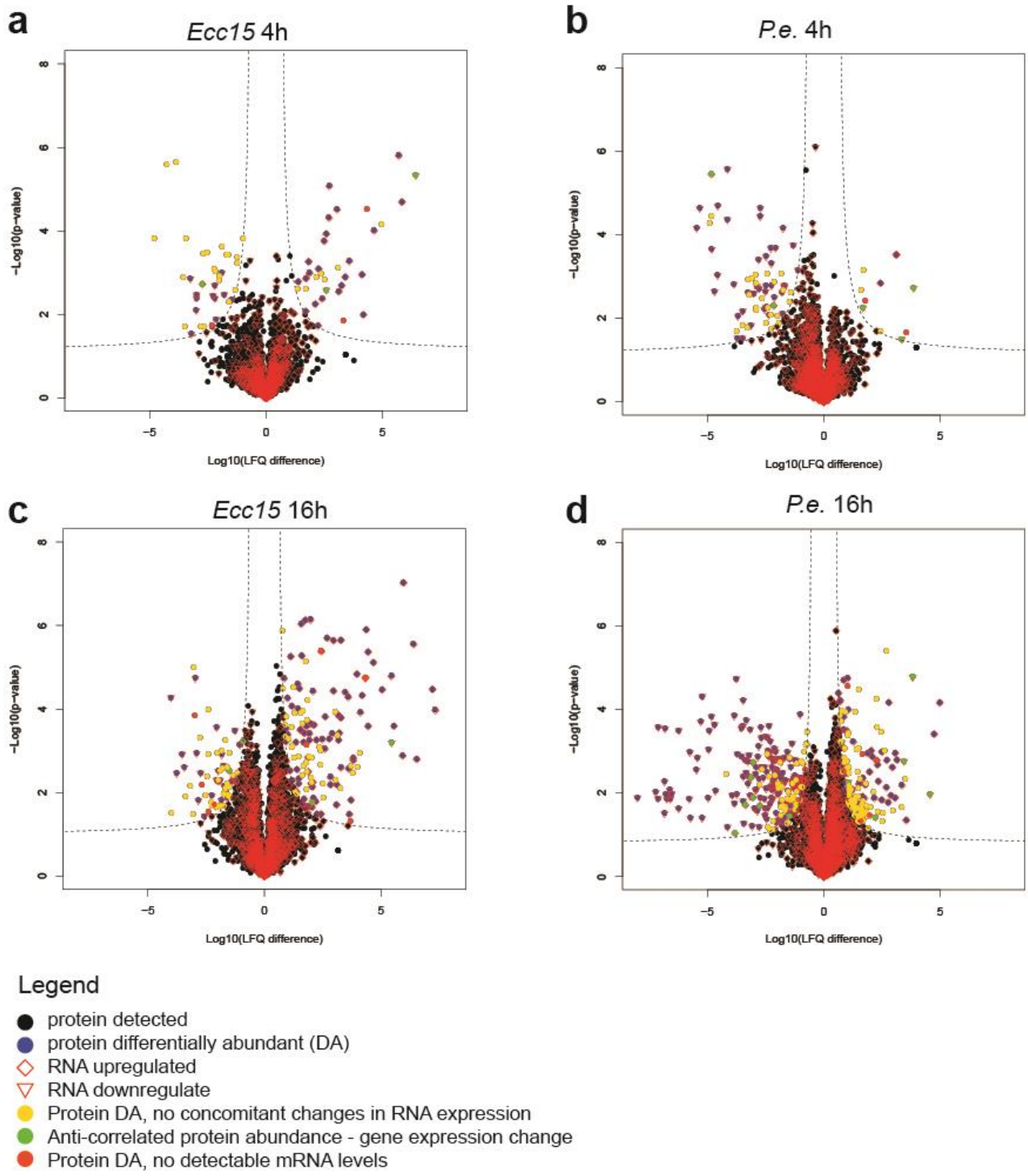


Figure 3-4 : Differentially abundant proteins with associated gene expression change

(a) DE proteins 4 hours after infection with Ecc15 with RNA direction and relationship between RNA direction and protein abundance.

(b) DE proteins 16 hours after infection with Ecc15.

(c) DE proteins 4 hours after infection with P.e.

(d) DE proteins 16 hours after infection with P.e.

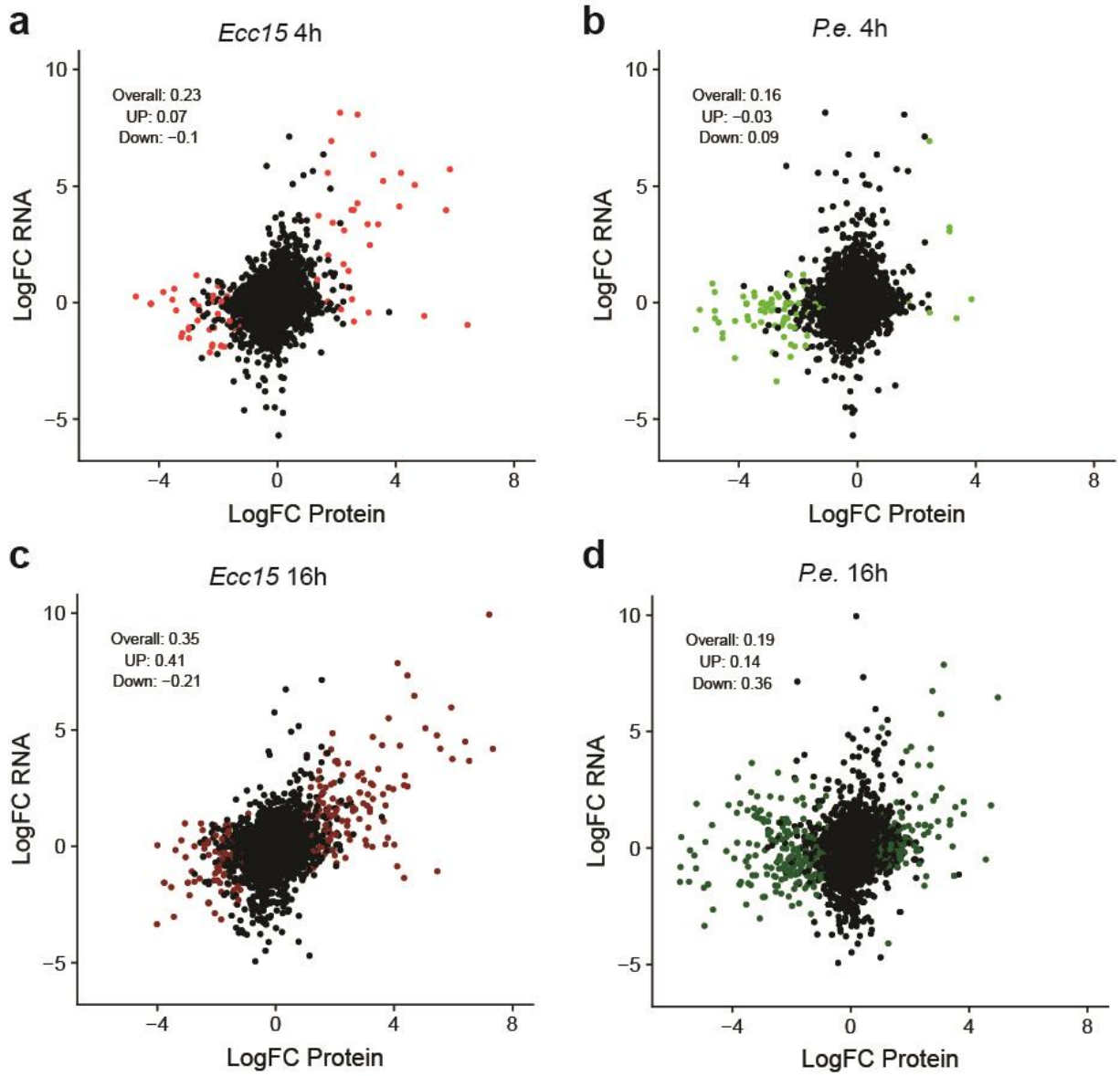


Figure 3-5 : Fold change comparison between mRNA and proteins abundance

All plots display the Spearman correlation for (from top to bottom):

- all the proteins
- the upregulated proteins
- the downregulated proteins

(a) DE proteins 4 hours after infection with Ecc15, spearman pair-wise correlation.

(b) DE proteins 16 hours after infection with Ecc15, spearman pair-wise correlation.

(c) DE proteins 4 hours after infection with P.e., spearman pair-wise correlation.

(d) DE proteins 16 hours after infection with P.e., spearman pair-wise correlation.

3.3.3 Survival of mutant flies highlight new genes mediating resistance to enteric infection

From our cross proteomics and transcriptomics analysis, we selected 19 genes whose protein abundance was significantly increased or decreased after infection but whose mRNAs were either not detected, downregulated or did not change upon infection. For each gene, we ordered a mutant line and we assessed its survival upon *P.e.* infection compared to the original control line. We used the log-odds ratio to determine if the test line was significantly more susceptible or resistant than the control one. Each survival dataset thereby reflects three biological replica with at least 30 flies per replica (**Supplementary 3-5**). Unfortunately, our control lines harboring the y^1 genotype were extremely susceptible to infection, thus we could not trust our results when this control was involved, which was the case for 7 proteins which effect on resistance could not be assessed. A summary of the genes and results is available in **Table 3-4**. We found that three mutants (CG12576, CG13531 and Mpc1) have severe phenotypes. Interestingly, flies harboring mutants for *Dab* started to die dramatically 10 days after infection. In total, we were able to confirm that 7 lines were significantly more susceptible to infection (37%). However, we cannot rule out at this point that the P-element insertion may alter the fly's infection susceptibility in other ways than by the sole disruption of the target genes. Thus, further analysis with different types of mutants or RNA interference lines may be required to validate these results. We can nevertheless conclude that several of these genes likely play an important role in the gut immune response.

3.4 Discussion

Response to an external stimulus such as enteric infection involves many processes from detecting the invasion, responding to it, and repairing putative damage. The underlying molecular mechanisms that mediate these processes act by remodeling the proteome through the activation of gene expression and production of different immune effectors to combat the infection. While most genes follow the canonical flow from transcription of their mRNA to translation of these mRNAs in proteins, there are several complementary mechanisms in place that control overall mRNA and protein levels. These include the targeted degradation of mRNA or protein, alternative splicing, or miRNA-mediated translation blockage. Thus, read-out of the proteome provides the most accurate picture of the molecular nature of the immune response against pathogens since it characterizes the final outcome of all regulatory processes.

This well-accepted notion motivated the current study, since no comprehensive proteomic analysis of the gut immune response had so far been undertaken. As expected, our analyses revealed a large number of differentially expressed (DE) proteins after infection. This number was however greater 16h compared to 4h after infection. This may be due to several factors: including a likely delay between transcription and translation (Gedeon & Bokes, 2012), and the fact that survival and repair mechanisms including pathways regulating stem cell proliferation and differentiation may be more protein-intensive than the immediate immune response (Nicolas Buchon et al., 2009; S Chakrabarti et al., 2012).

We further remarked not only a strong time- but also pathogen-dependent proteome signature, indicating that the general immune response is highly variable between conditions. We also observed that the pathogenicity of the bacterial sample strongly influences the immune response proteome. However, several pathways and proteins were differentially regulated in all conditions. Collectively, these core protein sets were especially enriched for immune, biogenesis and metabolism-regulatory processes. While the first of these is rather obvious, the two others may be less intuitive. Biogenesis encompasses the synthesis of a biological constituent. Thus, this process may directly reflect the attempt of the host to restore gut homeostasis and integrity. In terms of metabolism, we observed primarily an overall downregulation, which could be a way for cells and gut cells in particular to spare resources to respond efficiently to the infection. This hypothesis is in line with previous studies in which metabolism and infection were tightly co-regulated. In general, prolonged immune response can lead to disruption of metabolic stores and malnutrition can increase the susceptibility to disease (Dionne, 2014). Furthermore, Vibrio cholera can induce the death of flies by inducing a loss of intestinal acetate that results in lipid accumulation in enterocytes and host death

(Hang et al., 2014). Lipid catabolism was also shown to be involved in DUOX regulation in *Ecc15* infection (K.-A. Lee et al., 2018).

It is widely accepted that the correlation between mRNA and protein levels tends to be poor. It thus came not as surprise to observe that many stable proteins exhibited differential mRNA expression or vice versa. Many molecular mechanisms can explain these observations, such as miRNA translational blockage, regulated ribosome occupancy, elongation speed and translational pausing to list just a few. However, we must consider the design of our experiment as an important factor as well. Because there is a delay between gene expression and mRNA translation, it is reasonable to assume that a snapshot of the transcriptome will be shifted from that of the proteome. Furthermore, protein translation dynamics may vary from one mRNA to another. For example, mRNA could have a very short half-life while exhibiting a strong burst expression dynamic. In this scenario, the mRNA would be highly expressed for a short amount of time, immediately translated and then degraded such that protein levels would clearly rise while mRNA would be perceived as downregulated or even undetectable. Another scenario would be a gene that is expressed at a stable rate. This gene would not be identified as being differentially expressed, but it is possible that protein levels could increase through a change in ribosome occupancy. In this regard, it may be of interest to perform ribosome profiling, a technique allowing inference of ribosome occupancy on an mRNA strand, on control and infected *Drosophila* guts to identify which genes may be regulated through this mechanism.
Our survival-based validation of putative interesting proteins that had so far been missed given their stable mRNA but DE protein signature showed that a high percentage of these targets displayed reduced resistance when mutated. One could expect high percentages because of the intrinsic nature of the molecule assessed. Indeed, most of the functions of the cells are mediated via proteins. Thus, measuring the molecules that are present in virtually all molecular processes in the cells makes it more likely to identify a crucial cog in such a process. Conversely, variation in gene expression may not indicate whether the associated protein will be present. Furthermore, the lack of correlation between gene and protein expression levels implies that a gene that may have been overlooked in transcriptomics-based studies is more likely to play an important role. Indeed, a proteome analysis is more likely to find key proteins because the costs of protein synthesis is higher than that of mRNA metabolism (Bier, 1999). Thus, because the cost associated with the production, regulation and disposal of proteins is higher than for mRNA, a cell with limited resources will likely be more wary of its protein synthesis as to not waste too much resources.

3.5 Conclusion

Our study is, to our knowledge, the most in-depth characterization of the *Drosophila* gut proteome upon infection by Gram-negative bacteria. It paves the way for future studies on the link between genetic and protein abundance variation in a population using our dataset as a basis. It also identified the regulation of metabolism-related processes as a major target of regulation. Finally, we identified putative new proteins involved in the enteric immune response whose roles and mechanisms could be subject of a more detailed, mechanistic investigation in future studies.

3.6 Materials and Methods

3.6.1 Fly Stocks

We obtained w^{1118} and all mutants (see **Table 3-4**) from the Bloomington stock center. All flies were reared at room temperature on a standard fly medium with 12 hours light dark cycle. The fly medium composition for 1L of water is as follow: 6.2g Agar powder (ACROS N. 400400050), 58.8g Farigel wheat (Westhove N. FMZH1), 58.8g yeast (Springaline BA10), 100ml grape juice; 4.9ml Propionic acid (Sigma N. P1386), 26.5 ml of Methyl 4-hydroxybenzoate (VWR N. ALFAA14289.0) solution (400g/l) in 95% ethanol, 1L Water.

3.6.2 Infection

Oral infection was performed as previously described (Neyen et al., 2014). Briefly, 1-day old females were transferred to 29°C rearing conditions. When the female flies were 2-3 days old, they were starved for 2 hours and then transferred to a tube containing bacteria and allowed to feed on the bacteria for a maximum of 24 hours. To prepare the *P.e.* bacterial pellet, bacteria were plated from glycerol stocks on a standard LB-agar plate supplemented with 1% milk and grown overnight at room temperature. Two days prior to infection, one single colony was transferred to a 50 ml Erlenmeyer with 12.5 ml LB and incubated for 8 hours at 29°C with 180 rpm shaking. The pre-culture was then transferred to a 1L Erlenmeyer with 200 ml LB and the culture was incubated overnight using the same conditions as the pre-culture. The culture was then centrifuged at 2500 g at 4°c for 20 min. The remaining LB was discarded and the pellet was resuspended by pipetting up and down. The OD600 was measured using a CO8000 Cell density meter. The pellet was then diluted to a final OD600 of 100 with distilled water and supplemented with Sucrose to a final volume/volume of 1.25%. A control solution contained only Sucrose at the same concentration. A disc of whatman paper was layered on top of the food and 225 µl of the bacterial or control solution was added to the paper.

3.6.3 Protein extraction

Flies were killed by submersion in 70% ice-cold ethanol. The flies were then transferred to ice-cold 1x PBS and 30 guts were dissected and transferred to an Eppendorf containing 150 µl of extraction buffer (100 mM Tris-Cl pH 8, 2% SDS, protease and phosphatase inhibitor). The samples were incubated at 95°C for 2 min, then homogenized in a Precellys 24 Tissue Homogenizer at 6000 rpm for 30 seconds. Samples were then centrifuged for 10 min at max speed and 4°C. We aliquoted the supernatant in three separate tubes (1/5, 2/5 and 2/5 of total

volume of supernatant per tube) and stored at -80°C. We quantified the protein using a BCA assay and coomassie gel staining.

3.6.4 Mass Spectrometry

Samples were digested using trypsin and FASP protocols, then purified by desalting and fractionated during 24 hours. 6 fractions were collected and injected in an Orbitrap analyser. MaxQuant software was used to identify spectra.

3.6.5 RNA extraction

Flies were killed in cold 70% Ethanol then transferred to ice-cold RNAse free 1x PBS supplemented with 0.02% Tween-20. 10 guts were dissected for each sample and placed in a screw cap Eppendorf tube containing 350 uL Trizol and 10 uL plastic beads then homogenized in a Precellys 24 Tissue Homogenizer at 6000 rpm for 30 seconds. Samples were then flash frozen in liquid nitrogen and stored at –80°C. Once all samples were collected, tubes were thawed on ice, supplemented with 350 uL of 100% Ethanol and homogenized a second time with the same parameters as above. We then used the Direct-zol™ RNA Miniprep R2056 Kit, with the following modifications: we did not perform DNAse I treatment, we added another 2 min centrifugation into an empty column after the RNA Wash step, finally elution was performed by adding 20 uL of RNAse-free water to the column, incubation at room temperature for 2 min and then centrifugation for 2 min. RNA was transferred to a low-binding 96 well plate and stored at -80°C.

3.6.6 Library preparation and sequencing

Total RNA amount was assessed using picogreen, then standard Illumina Truseq libraries were prepared from 1ng total RNA. Single end sequencing was performed for 100 cycles.

3.6.7 Survival

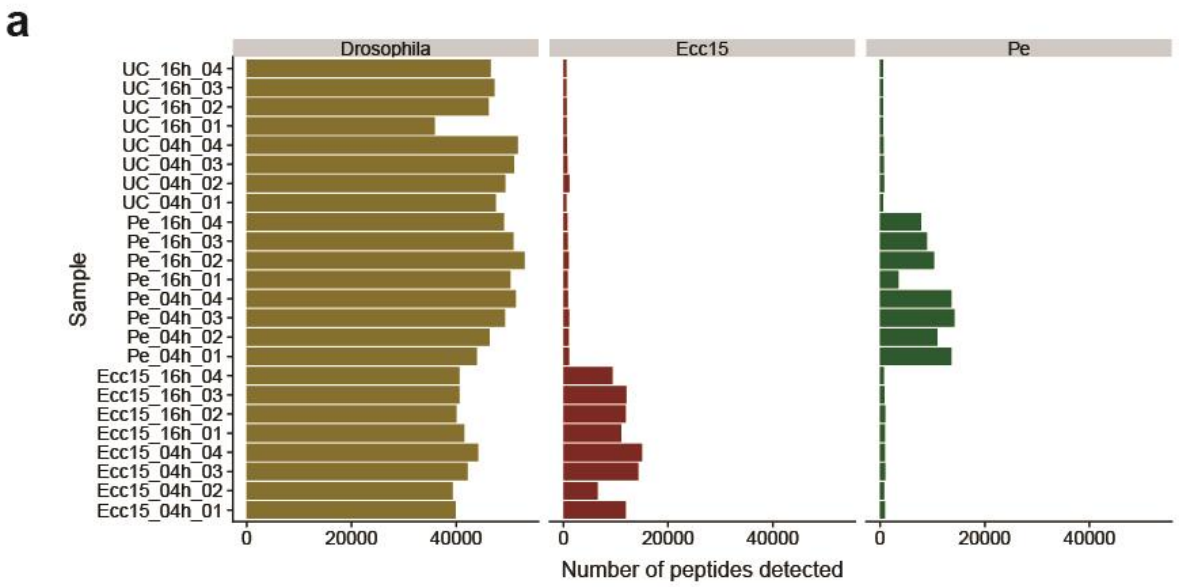
Flies survival was scored 4h after infection, then, after 24 hours flies were transferred to fresh tubes without bacteria and survivors were scored. Then, scoring occurred every 24 hours and flies were transferred to fresh tubes every 48 hours. The R version 3.5.1 and the package Survival was used to compute the log-rank test to test for effect compared to control.

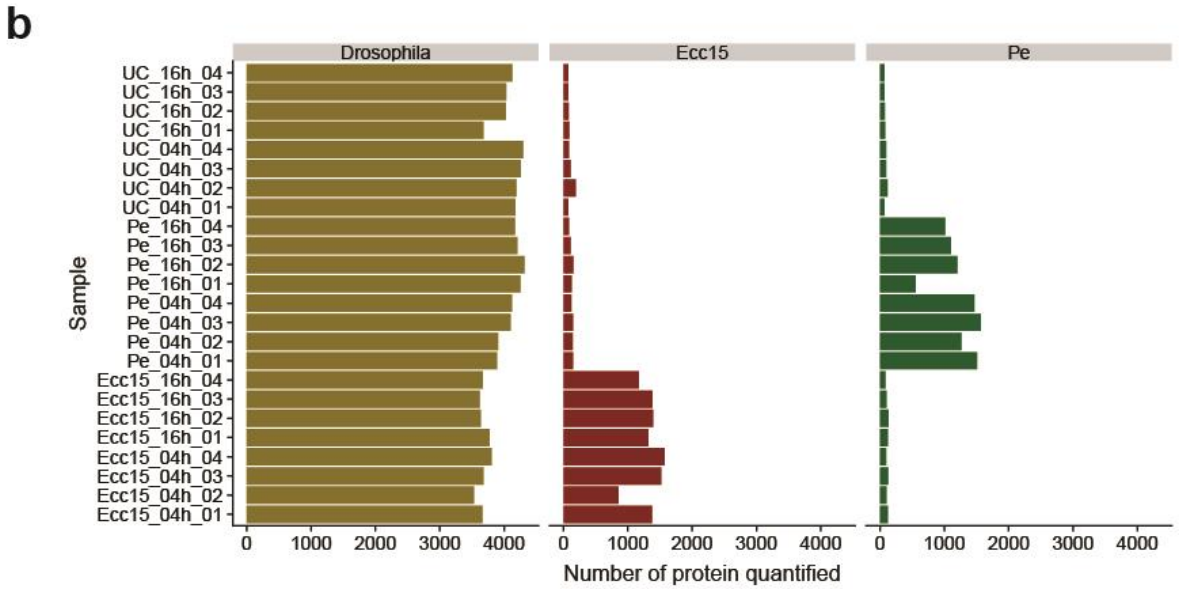
3.6.8 GO analysis

The GO analysis was performed using the GOstats (Falcon & Gentleman, 2007) R package (Hypergeometric test p-value < 0.05), and REVIGO (Supek et al., 2011) was used to reduce redundancy in the ontology groups and plot them by semantic similarity (allowed similarity = 0.5). We then manually curated the resulting list.

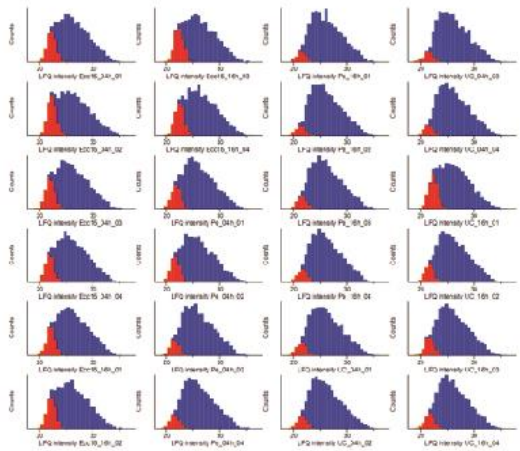
All analysis were performed in R version 3.5.0

3.7 Supplementary Material





c



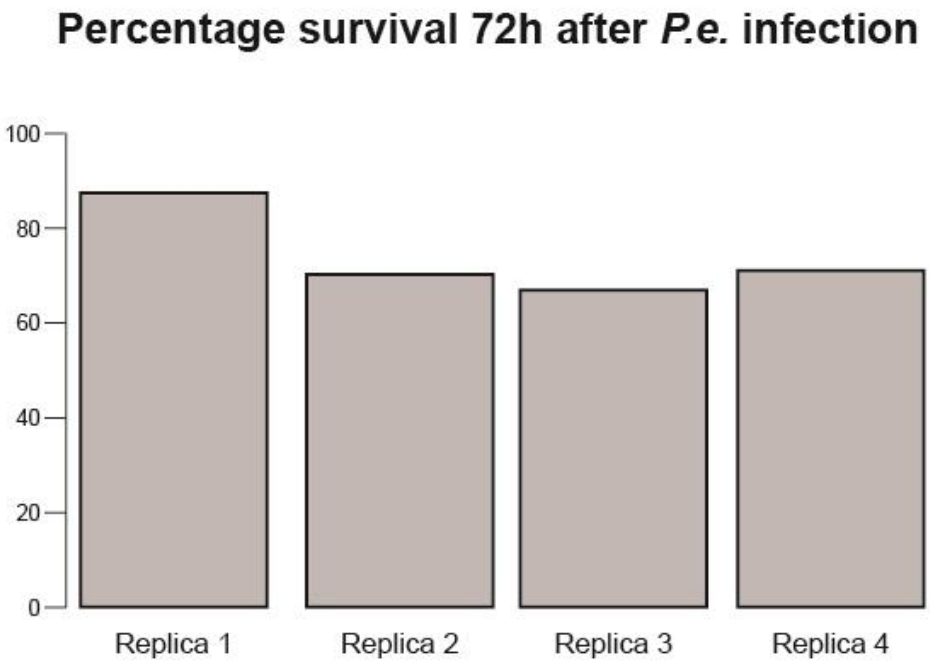
Supplementary Figure 3-1 : Peptides and proteins detected

(a) Number of peptides identified by mass spectrometry in each samples, separated by origins. (brown: *Drosophila* peptides, red: *Ecc15* peptides, green: *P.e*. peptides)

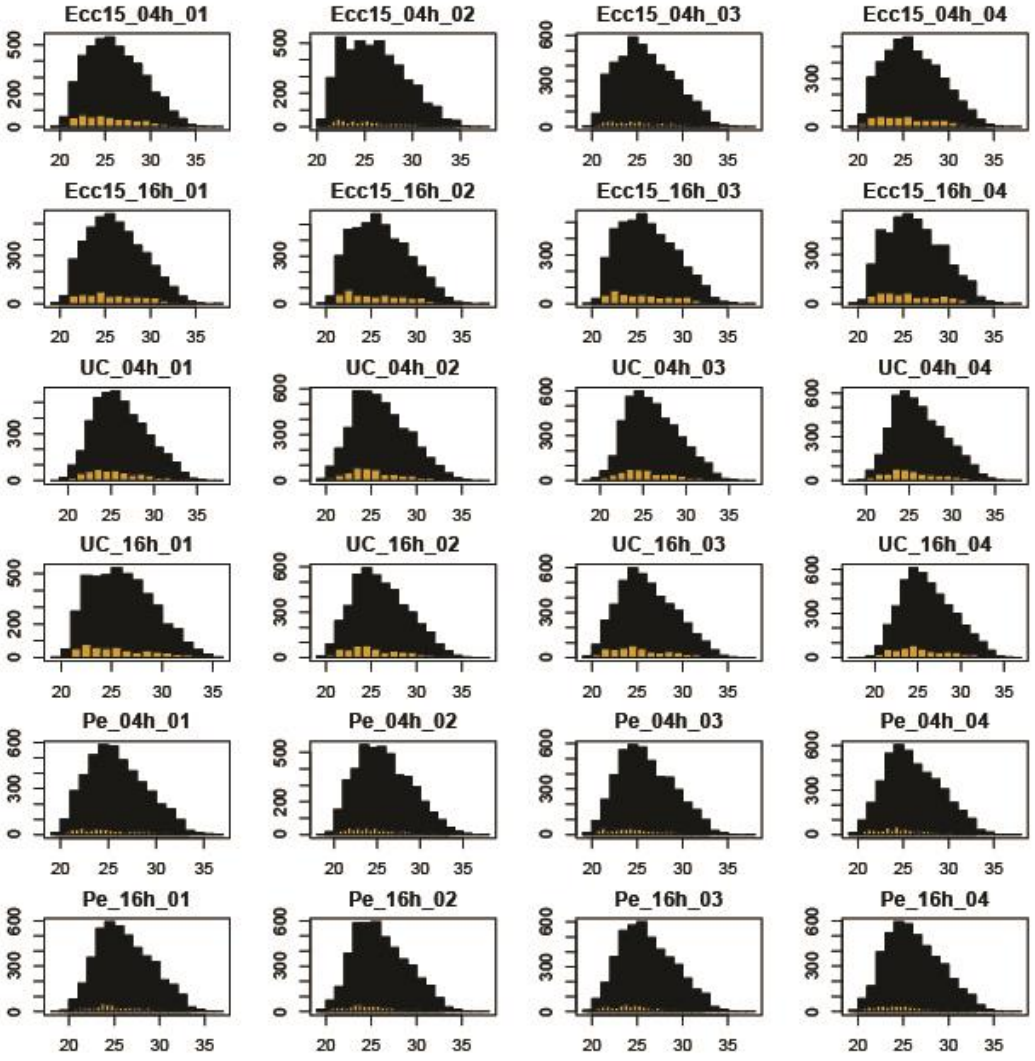
(b) Number of proteins quantified by mass spectrometry in each samples, separated by origins. (brown: *Drosophila* proteins, red: *Ecc15* proteins, green: *P.e*. proteins)

(c) Imputation showing that missing values imputed (red) are situated at the low expression level compared to all proteins (blue).

a


Percentage survival 72h after P.e. infection
100
80
60
40
20
0
Replica 1
Replica 2
Replica 3
Replica 4

b

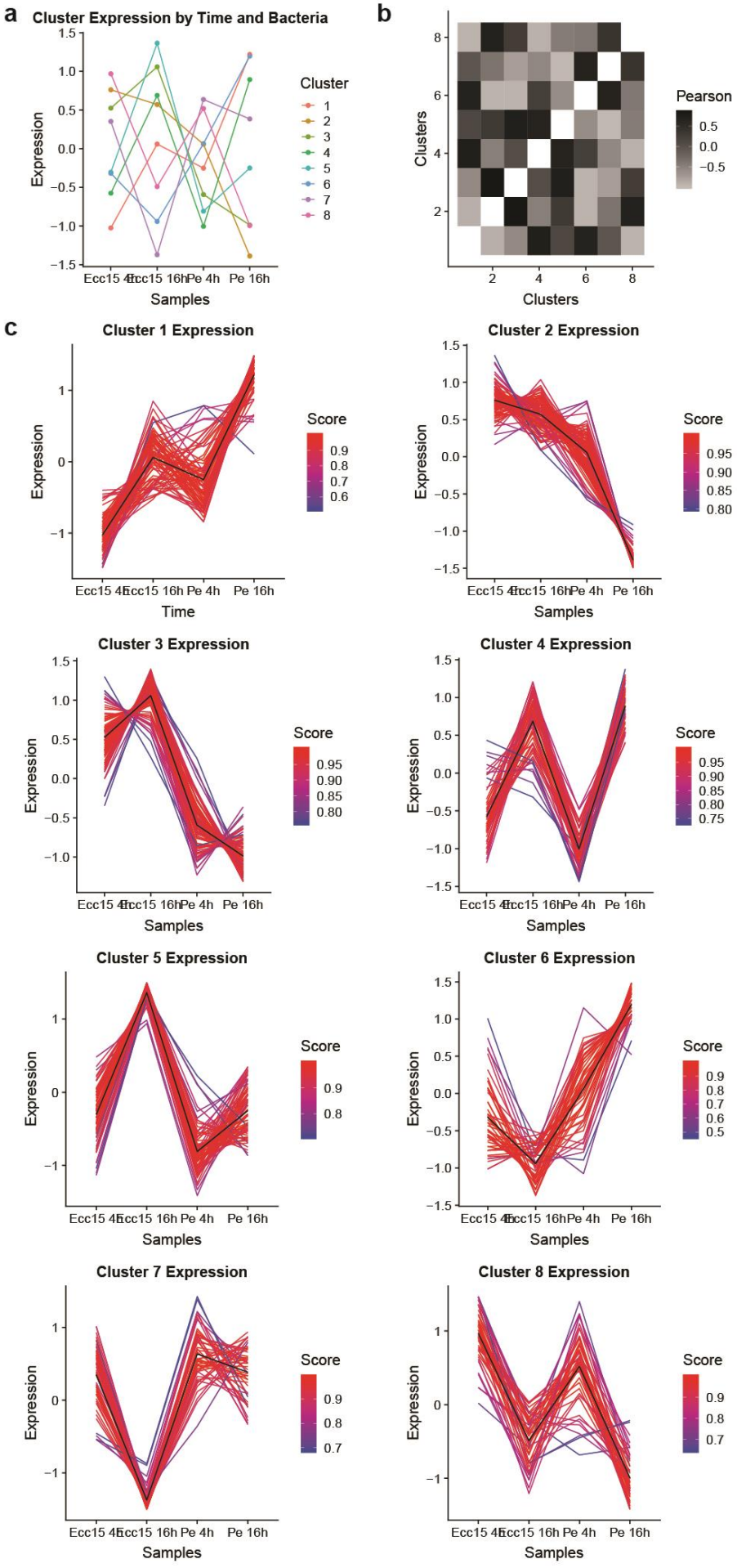

Ecc15_04h_01
Ecc15_04h_02
Ecc15_04h_03
Ecc15_04h_04
Ecc15_16h_01
Ecc15_16h_02
Ecc15_16h_03
Ecc15_16h_04
UC_04h_01
UC_04h_02
UC_04h_03
UC_04h_04
UC_16h_01
UC_16h_02
UC_16h_03
UC_16h_04
Pe_04h_01
Pe_04h_02
Pe_04h_03
Pe_04h_04
Pe_16h_01
Pe_16h_02
Pe_16h_03
Pe_16h_04

Supplementary Figure 3-2 : Lethality of baterial pellet and DE protein distribution

(a) Survival of flies after P.e. infection measured by counting the percentage of flies alive 3 days after infection showing that the first replica was fed a less lethal P.e. pellet than the 3 others.

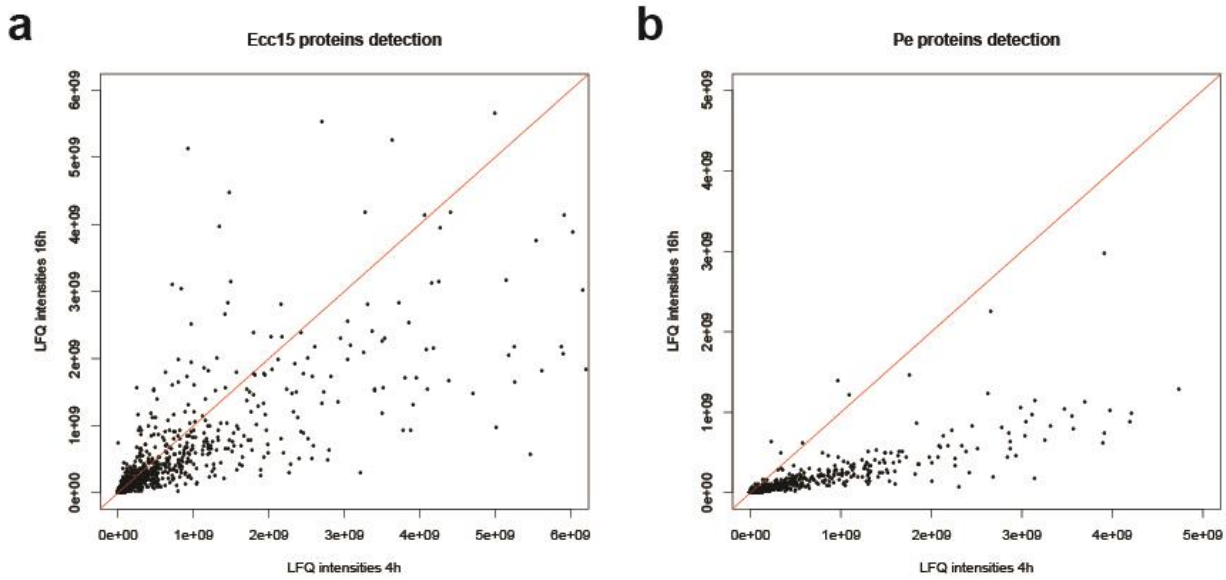
(b) Distribution of LFQ intensities, yellow bars represent LFQ intensities from DE proteins. Note that the distribution is mostly centered towards low to medium abundant proteins.

Supplementary 3:


a
Cluster Expression by Time and Bacteria
Expression
Samples
Ecc15 4h
Ecc15 16h
Pe 4h
Pe 16h
Cluster
1
2
3
4
5
6
7
8
b
Clusters
Pearson
c
Cluster 1 Expression
Time
Score
Cluster 2 Expression
Cluster 3 Expression
Cluster 4 Expression
Cluster 5 Expression
Cluster 6 Expression
Cluster 7 Expression
Cluster 8 Expression

Supplementary Figure 3-3 : Clustering of differentially expressed proteins

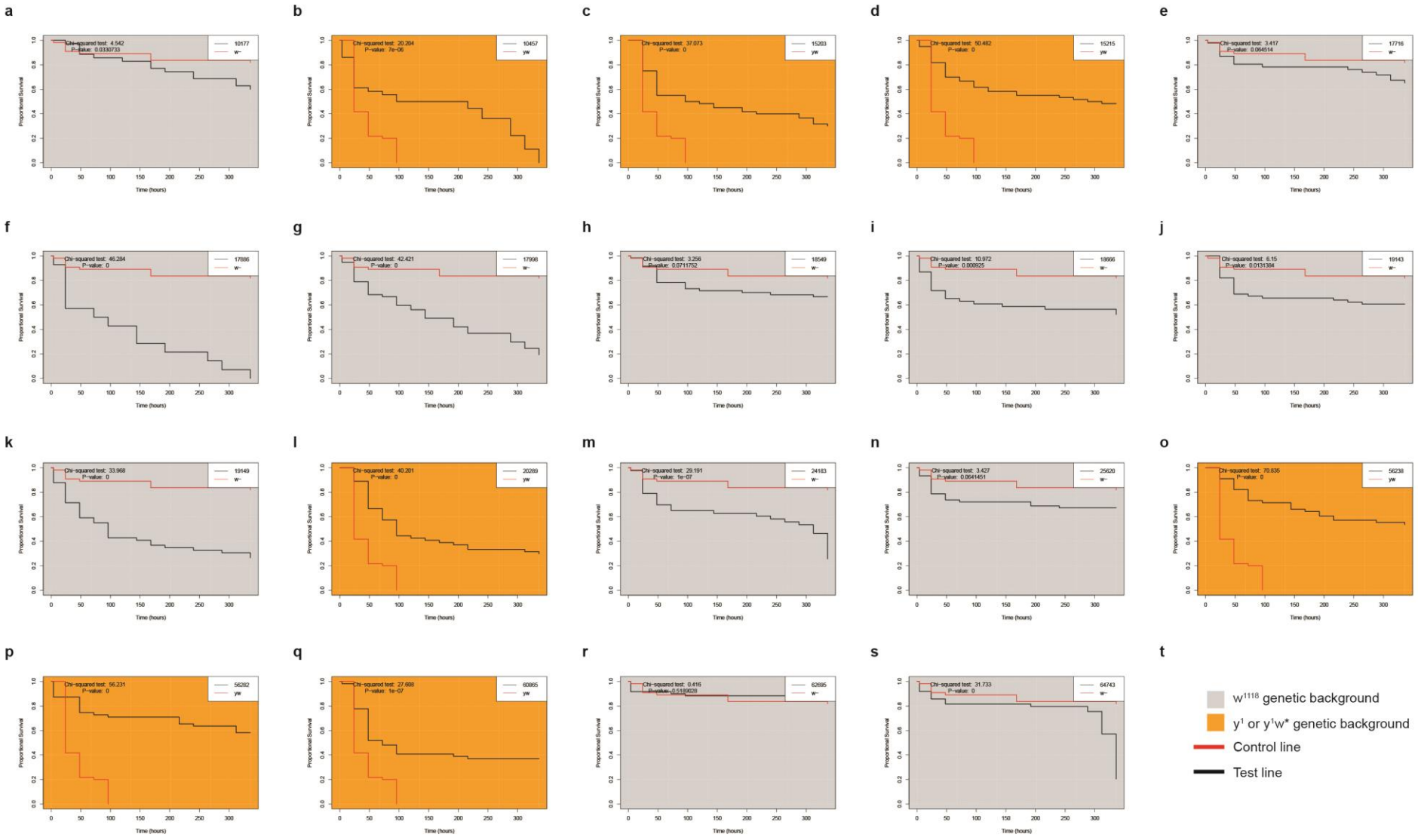
(a) Overview of all clusters behavior.
(b) Pearson coefficient of clusters correlation shows a few cluster closely correlated though the majority are loosely correlated.
(c) All identified clusters. The black line is the core of the cluster. Each line represent a gene which correlation with the core is color coded.



Supplementary Figure 3-4 : Bacterial proteins abundance

(a) Comparison of LFQ intensities of Ecc15 proteins at 4h and 16h shows slight reduction in abundance.

(b) Comparison of LFQ intensities of P.e. proteins at 4h and 16h shows complete clearance of the bacteria at 16h.


a
Chi-squared test: 4.542
P-value: 0.0330733
10177
w-
b
Chi-squared test: 20.204
P-value: 7e-06
10457
yw
c
Chi-squared test: 37.073
P-value: 0
15203
yw
d
Chi-squared test: 50.482
P-value: 0
15215
yw
e
Chi-squared test: 3.417
P-value: 0.064514
17716
w-
f
Chi-squared test: 46.284
P-value: 0
17886
w-
g
Chi-squared test: 42.421
P-value: 0
17998
w-
h
Chi-squared test: 3.256
P-value: 0.0711752
18549
w-
i
Chi-squared test: 10.972
P-value: 0.000925
18666
w-
j
Chi-squared test: 6.15
P-value: 0.0131384
19143
w-
k
Chi-squared test: 33.968
P-value: 0
19149
w-
l
Chi-squared test: 40.201
P-value: 0
20289
yw
m
Chi-squared test: 29.191
P-value: 1e-07
24183
w-
n
Chi-squared test: 3.427
P-value: 0.0641451
25620
w-
o
Chi-squared test: 70.835
P-value: 0
56238
yw
p
Chi-squared test: 56.231
P-value: 0
56282
yw
q
Chi-squared test: 27.688
P-value: 1e-07
60865
yw
r
Chi-squared test: 0.416
P-value: 0.518928
62695
w-
s
Chi-squared test: 31.733
P-value: 0
64743
w-
Proportional Survival
Time (hours)
t
w1118 genetic background
y1 or y1w* genetic background
Control line
Test line

Supplementary Figure 3-5 : Survival of mutants flies

The color of the background indicate the genetic background of the lines, y^1: orange background, w^{1118}: grey background. Log odds ratio p-values are given in parenthesis.

(a) *Pex1* (0.033)
(b) *RpS21* (<0.001)
(c) *RpL22* (<0.001)
(d) *CG42239* (<0.001)
(e) *CG4022* (0.065)
(f) *CG12576* (<0.001)
(g) *CG13531* (<0.001)
(h) *CG3342* (0.07)
(i) *Dnah3* (<0.001)
(j) *CG17083* (0.13)
(k) *Mpc1* (<0.001)
(l) *Zip48C* (<0.001)
(m) *Prx2540-2* (<0.001)
(n) *Tsp42Ec* (0.064)
(o) *ndl* (<0.001)
(p) *CG32113* (<0.001)
(q) *CG32850* (<0.001)
(r) *Lmpt* (0.52)
(s) *Dab* (<0.001)
(t) Legend.

Table 3-1 : Differentially expressed proteins after infection

Pathogen	Time	No change	Down	Up
Ecc15	4h	4432	37	34
Ecc15	16h	4290	72	141
P.e.	4h	4424	68	11
P.e.	16h	4129	223	151

Table 3-2 : Differentially expressed proteins and their associated mRNA behaviour

		mRNA			
	Ecc15 4h	Down	Up	No change	Not detected
Protein	No change	337	377	3538	180
	Down	13	1	22	1
	Up	2	23	7	2

		mRNA			
	Ecc15 16h	Down	Up	No change	Not detected
Protein	No change	508	425	3184	173
	Down	30	3	35	4
	Up	3	89	43	6

		mRNA			
	P.e. 4h	Down	Up	No change	Not detected
Protein	No change	867	651	2725	181
	Down	41	2	25	0
	Up	3	3	3	2

		mRNA			
	P.e. 16h	Down	Up	No change	Not detected
Protein	No change	646	641	2667	175
	Down	165	10	46	2
	Up	14	54	77	6

Table 3-3 : Fisher's exact test for overrepresentation of miRNA targets among mRNA of proteins differentially expressed

Samples	Fishers exact test	
Hypothesis	greater	less
Ecc15 4h	0.857	0.2151
Ecc15 16h	0.7766	0.2732
P.e. 4h	0.9689	0.05577
P.e . 16h	0.9952	0.006777

Table 3-4 : List of differentially expressed proteins selected for screening

gene name	Protein FC Ecc15 4h	Ecc15 16h	Pe 4h	Pe 16h	RNA direction Ecc15 4h	Ecc15 16h	Pe 4h	Pe 16h	gene name	mutant genotype	background	p-value
Pex1	2.008505	1.263032	3.34849	0.170368	nc	nc	-	-	Pex1	w[1118]; P{lacW}Pex1s48687/TM3, P{ftz/lacC}SC1, ryRK Sb1 Ser1	w[1118]	0.033
RpS21	0.202433	0.722968	0.291705	1.563637	nc	nc	nc	-	RpS21	y[1] w[67c23]; P{lacW}RpS21k16804a/CyO	y[1] w[67c23]	7.00E-06
RpL22	0.157558	0.679516	0.401796	1.059535	nc	nc	-	-	RpL22	y[1]; P{SUPor-P}RpL22KG09650/FM7c, sn+	y[1]	< 0.001
CG42239	0.241853	2.05046	1.86846	3.441985	nc	nc	nc	-	CG42239	y[1]; P{SUPor-P}CG42239KG09816; ry506	y[1]	< 0.001
CG4022	4.964305	4.105725	0.782682	-0.24005	nc	nc	nc	nc	CG4022	w[1118]; PBac{PB}CG4022c05627	w[1118]	0.065
CG12576	0.029052	2.056182	0.409032	1.033537	-	-	-	-	CG12576	w[1118] PBac{RB}CG12576e00577	w[1118]	< 0.001
CG13531	3.77467	4.00891	1.146057	0.689752	nc	nc	-	-	CG13531	w[1118]; PBac{RB}CG13531e01893	w[1118]	< 0.001
CG3342	0.646453	0.41185	1.703015	0.845707	nc	nc	nc	nc	CG3342	w[1118] PBac{WH}CG3342f02379	w[1118]	0.07
Dnah3	3.08434	1.17245	2.442032	0.427425	nc	+	nc	+	Dnah3	w[1118]; PBac{WH}Dnah3f03522	w[1118]	< 0.001
CG17083	4.340665	1.571351	3.554738	1.58765	NA	NA	NA	NA	CG17083	w[1118]; P{XP}CG17083d00403	w[1118]	0.13
Mpc1	-0.373345	0.490965	0.85306	2.21709	nc	nc	-	-	Mpc1	w[1118]; P{XP}Mpc1d00809	w[1118]	< 0.001
Zip48C	-0.428048	0.478882	0.336091	1.55562	nc	-	-	-	Zip48C	y[1] w[67c23]; P{EPgy2}Zip48CEy[1]1307 MCPH1Ey[1]1307	y[1] w[67c23]	< 0.001
Prx2540-2	0.17708	0.766967	3.11387	3.815894	+	+	+	-	Prx2540-2	w[1118]; Mi{ET1}CAPMB04531 Mi{ET1}Prx2540-2MB04531	w[1118]	< 0.001
Tsp42Ec	-3.851713	-0.534042	-4.821445	-4.199853	nc	nc	+	nc	Tsp42Ec	w[1118]; Mi{ET1}Tsp42EcMB06423	w[1118]	0.064
ndl	3.328858	3.684437	-0.47899	-0.08025	NA	NA	NA	NA	ndl	y[1] w*; Mi{MIC}ndlMI10748	y[1]w*	< 0.001
CG32113	2.585378	5.460623	0.237063	0.254228	-	-	-	-	CG32113	y[1] w*; Mi{MIC}CG32113MI11101	y[1]w*	< 0.001
CG32850	0.731995	0.762332	0.69144	1.293517	nc	-	-	-	CG32850	y[1]; Mi{MIC}CG32850MI14327	y[1]	< 0.001
Lmpt	-1.031998	-0.231789	-1.08867	3.492775	nc	-	nc	nc	Lmpt	w[1118]; PBac{wl+mC]=lT.GAL4}Lmpt[0200-G4]	w[1118]	0.52
Dab	6.423627	1.29409	-0.167462	0.150797	-	-	-	-	Dab	w[1118]; PBac{lT.GAL4}Dab0764-G4	w[1118]	< 0.001

Chapter 4: Overall discussion and conclusion

4.1 Cryptic variants are the main driver behind gene expression variation in the gut enteric immune response

To identify the genetic determinants of gene expression variation in the fly we first performed RNA sequencing on flies from phenotypic extremes and observed that the response is widely similar between the most resistant and susceptible lines. Indeed, it is impossible to segregate resistant lines from susceptible lines in both control and infected conditions (**Figure 2-3a**) on the basis of gene expression patterns. A possible explanation would be that we are not capable of differentiating between lines at the level of the transcriptome. However, this was not the case because independent biological replicas clustered together. Thus we concluded that systemic differences in gene expression driven by genomic variation impact each line specifically but that only a few of these differences significantly impact the resistance of the fly to infection. Moreover, because phenotypic differences are not the results of large variation in the transcriptome, we could conclude that resistance to infection is a trait with a complex architecture, with many loci with small effects adding up to shape resistance.

We then performed an eQTL analysis to understand how genomic variation could impact gene expression and lead to variation in resistance to infection. We tested for association between SNPs situated in a 10 kb window around genes and gene expression variation. By definition, a *cis*-eQTL affects the expression in an allele-specific way whereas a *trans*-eQTL affect both alleles (Rockman & Kruglyak, 2006). However, because DGRP lines are entirely homozygous, it is impossible for us to determine if a variant acts in *cis*- or *trans*-, therefore we only detected *local*-eQTL. We found out that most of the *local*-eQTLs are condition-specific, with about a third of them present in both conditions. Because there is a large number of *local*-eQTLs that are infection-specific, we can conclude that cryptic variation is an important contributor to the resistance phenotype. This also reflects the fact that immune response is largely repressed in normal state since we would see more difference between the control condition if it would not be the case.

These results highlights the importance of perturbation to understand how the system reacts to it. Indeed, it is not possible to infer resistance based on gene expression in the control state only. Furthermore, a large number of eQTLs are infection-dependent, meaning that a large

part of the variation in gene expression is hidden in the non-infected state and, thus, that cryptic variants are the main drivers behind variation in resistance to infection. Because our study used *P.e.* as a the sole pathogen, we cannot generalize our findings to other pathogens such as Gram-positive bacteria, fungus or yeast, but we could hypothesize that the response to infection with a different pathogen would also be mediated by cryptic variations.

4.2 A majority of *local*-eQTLs acts in *trans*-

To further understand how *local*-eQTLs affect gene expression, we used allele-specific expression of F1 hybrids to categorize *local*-eQTLs into *cis*- and *trans*-eQTLs and found that only 17% of *local*-eQTLs acts as *cis*-eQTLs. We can think of several factors affecting this percentage. First, we study the effect of polymorphic sites detected in homogeneous background in various different genetic background, thus we cannot exclude that the new genetic background has an effect on the SNP. For example, epistasis may play an important role in determining if a the *cis*- effect of a *local*-eQTL can be masked by other *trans*- acting eQTLs affecting the same gene from a different genetic background or if a *local*-eQTL only has its effect when surrounded by certain variants. Indeed, when several eQTLs were predicted to affect one gene, we were not able to disentangle their effects. Furthermore, our statistical test measures the overall effect of the SNP on allele specific expression, thus it is possible that the variant is acting in *cis*- in a certain background and in *trans*- in another background, preventing us from categorizing it as *cis*-acting. It is also possible that a combination of multiple SNPs are necessary to observe an effect on gene expression (Y.-T. Huang et al., 2014) or that a single polymorphism drives differential expression and the other eQTLs are only in linkage disequilibrium (LD) with main driver of allele-specific expression. To be able to explain these behavior, we would need more crosses to increase the number of heterozygotes sites observable. Interestingly, we found a rather large discrepancy between the significance of *local*-eQTLs and their characterization as *cis*-eQTL. We did not see a link between a strong association in population *local*-eQTLs and their characterization as *cis*-acting eQTL, meaning that a strong *local*-eQTL is not poised to be *cis*-acting.

This result shows that many eQTLs experiments that use proximity between SNP and gene as a proxy for *cis*-effect may overestimate the number of variants acting in *cis*-. It is important to differentiate between the two modes of action as a *trans*-acting eQTL will have a stronger effect on overall gene expression compared to a *cis*-acting eQTL. Thus, a *trans*-acting variant is more likely to strongly affect a given phenotype than a *cis*-acting eQTL. Furthermore, we

found that many genes were linked with several variants. Thus understanding the interplay between these variants may be helpful to understand the hierarchy that leads to variation in gene expression. Because noncoding regulatory variants plays crucial roles in phenotypic difference, differentiating between *cis*- and *trans*-acting variants may help identify major determinants of gene expression variation. Furthermore, allele-specific experiments could allow researcher to finely map epistasis interaction mediated variation in gene expression. Finally, if we take into account the physical organization of the chromosome in the nucleus, were chromosomes occupy certain territory (Meaburn & Misteli, 2007), it is surprising that such a low amount of eQTLs act in *cis*-. Indeed, if chromosome are separated, it is surprising that a variant situated in one chromosome can affect the expression of a gene situated in another chromosome in particular if the variant is in a non-coding region.

4.3 *Nutcracker* is the only differentially-expressed gene between resistant and susceptible flies

One of the most striking results of our study is that we found only one gene differentially-expressed between the resistant and susceptible flies, *ntc*. *Ntc* is a part of an SCF ubiquitin ligase complex (E3) that recruit an E2 ubiquitin-conjugated enzyme and mediate the transfer of the ubiquitin from the E2 complex to the target protein. Ubiquitination is an important mechanism in cell biology that regulates various process such as protease degradation, cell cycle control (Teixeira & Reed, 2013) and other signaling pathway. Specifically, in the Drosophila immune response, it was shown that ubiquitination of DREDD by DIAP2 is required for Imd signaling (Meinander et al., 2012). However, it was also shown that Imd and DIAP2 are ubiquitinated after infection (L. Chen et al., 2017; Paquette et al., 2010). Moreover, there are several reports of Imd regulation via proteasome targeted degradation (M. Kim, Lee, Lee, Kim, & Chung, 2006). Furthermore, the IKK complex that phosphorylates Relish is also dependent on ubiquitination (R. Zhou et al., 2005). Although ntc was not shown to be directly involved in any of these interactions, it is possible that it could mediate ubiquitination of these proteins. We also show that ntc expression is dependent on Rel (**Supplementary 2-4f**) since *ntc* expression is strongly reduced in Rel^{E20} mutants. Moreover, our population data indicates that an increase in *ntc* expression is beneficial for survival, therefore, it seems that *ntc* effect on the Imd pathway would be positive, maybe helping to ubiquitinate various members of the pathway. Another possible mechanism would be that *ntc* help in the translocation of the cleaved N-terminal domain of Rel into the nucleus. In conclusion, further mechanistic studies are needed to understand the role of *ntc* in the enteric immune response.

4.4 A variant in a non-coding region affect *nutracker* expression

In our eQTLs analysis, we identified two *local*-eQTLs associated with *ntc*, the only differentially-expressed gene. Further analysis allowed us to establish that a variant located in the Broad binding site was significantly affecting the binding affinity between the transcription factor and the DNA. Because the binding affinity of Broad to its site is negatively affected by the mutation, this indicates that the TF has more difficulty affecting the expression of the gene. We showed that the expression of *ntc* is lower in susceptible flies and in flies with the wild-type allele for Broad TFBS. Therefore, a lower binding affinity should decrease the control that Broad exert on *ntc* and since *ntc* is more expressed in the lines harboring the mutation, this indicates that Broad acts as a repressor. Further analysis of the Allele-specific expression in the F1 showed altered ratio of both alleles after infection. Indeed, the allele that is in *cis*- with the alternate TFBS is more expressed than the allele in *cis*- with the reference allele. This prove that the mutation is the causal variant of the differential allele expression because the allele specific pattern is conserved through several different genetic background. We should note however that the effects induced by the mutation in the TFBS on the overall resistance of the fly are less severe than the phenotypes of the *ntc* mutants. Indeed, the expression of *ntc* in flies harboring the altered TFBS is still higher than the expression of the loss-of-function mutant.

Overall, this result is a particularly interesting example of how a non-coding variant in the TFBS leads to variation in gene expression and phenotypic change. Furthermore, because Broad was shown to be both a repressor and an activator depending on the isoforms expressed, it would be interesting to further study how alternative splicing of *Broad* affect the expression of *ntc.*

4.5 The proteome of the gut after infection show strong time- and pathogen-dependent variation

The main goal of our study on the gut proteome after infection was to characterize the main changes happening at the protein level after infection by a Gram-negative bacteria. The variation in gene expression in enteric infection in the adult fly has been characterized at length using micro-array (S Chakrabarti et al., 2012; W. J. Lee, 2009) and the variation in gene

expression mediating gut immuncompetence was studied at the population level in the DGRP (Bou Sleiman et al., 2015). However, because numerous studies have shown that gene expression and protein abundance do not correlate strongly, we wanted to investigate the proteome. When compared with the transcriptome, we found that the proteome is less variable than gene expression. Indeed, the percentage of differentially-expressed genes is higher than the number of proteins upregulated or downregulated at each timepoint and in response to each pathogen. Interestingly, we observed a larger number of proteins differentially-expressed at 16h than at 4h. Several causes may explain this observation. First, there is a delay between transcription and translation (Gedeon & Bokes, 2012). Second, the mechanisms needed to reestablish homeostasis during and after the infection may involve more proteins than the immune response (Nicolas Buchon et al., 2009; S Chakrabarti et al., 2012). We then noticed a strong time- and pathogen-dependent response indicating that the general immune response is very variable between the conditions. However, we could also detect several pathways and proteins that were affected in all conditions, indicating that a core of proteins, especially enriched for immune, biogenesis and metabolism-regulatory processes, mediate the resistance to infection.

These results highlight the relevance of proteomics to understanding the immune response in the gut. Indeed, proteins occupy the majority of the space in a cell and are the final effectors of the cell response to stimulus. However, there are several drawbacks to the technology. For example, it is easier to detect highly abundant proteins. These proteins are not always the most differentially-expressed proteins as seen in our experiments. But because they are highly abundant, these proteins may mask variation in other proteins' abundance. Furthermore, the detection of small proteins and post-transcriptionally modified proteins require different extraction methods. It was thus not possible for us to characterize the variation in small protein abundance, such as AMPs, nor the post-transcriptional modification of some important pathway members, for example DREDD ubiquitination or Relish phosphorylation.

4.6 Poor correlation between gene and protein variation

We also observed that the correlation between mRNA and protein levels was poor. Many molecular mechanisms can explain these observations. For example, miRNA translational blockage regulates the translation of a specific mRNA into proteins by binding to its 3' untranslated region (UTR) or 5' UTR and blocking the translation. We can also observe difference in ribosome occupancy, meaning that the number of ribosome on a strand of mRNA

is not fixed and thus that the amount of proteins produced depend on the number of ribosome on the mRNA. Another interesting mechanism is the control of translation via changes in tRNA abundance (Wilusz, 2015). These molecular mechanisms are a complement to the gene expression regulation but other methods are needed to measure the extent of their effect on protein abundance. Indeed, it is possible to measure expression of miRNA using modified RNA-seq protocols and to measure ribosome occupancy using ribosome sequencing (Ribo-seq) (Ingolia, 2014). Measurement of tRNA abundance can be assessed using fluorescent labelling as well (Y. Zhou, Goodenbour, Godley, Wickrema, & Pan, 2009).

However, our experimental design must be considered as well. Because there is a time delay between transcription and translation, our snapshot of the transcriptome and proteome are shifted as well. Thus a better design would be more timepoint with less time spent between them. However, such a design would multiply the cost and the complexity of experiments and it is not certain that the advantage obtained would outweigh the cost. Furthermore, the translation dynamics may not be the same for different RNA. Indeed, one could imagine that an mRNA is expressed and translated in burst, with a low half-life for the mRNA. This would lead to the protein being detected but not the mRNA. It is also possible that a protein has a very low half-life but its corresponding mRNA is very stably expressed. This would result in sharp upregulation of protein followed by rapid degradation and, depending on the time of the sample collection, to a protein appearing up- or down-regulated respectively and an mRNA level displaying no change.

4.7 Loss-of-function screen identify several proteins affecting resistance to infection

Our screen showed that several proteins that had so far been missed in other transcriptomics-based approach given their stable mRNA affect the fly resistance to infection. These proteins were selected based on variation in their protein abundance and lack or anti-correlated mRNA level change.

We found out that a high percentage of mutated proteins displayed reduction in resistance. Indeed, one could expect high percentages because of the intrinsic nature of the molecule assessed. Because, most of the functions of the cells are mediated via proteins, a direct measurement of the molecules that are present in virtually all molecular processes in the cells makes it more likely to identify a crucial cog in such a processes. Conversely, variation in gene expression may not indicate whether the associated protein will be present, thus transcriptomics-based target identification may overestimate the importance of a protein.

Furthermore, the lack of correlation between gene expression and variation in protein abundance implies that a gene that may have been overlooked in transcriptomics-based studies because of its observed downregulation may instead play an important role. Indeed, a proteome analysis will likely find key proteins because the costs of protein metabolism is higher than mRNA metabolism (Bier, 1999). Because the cost associated with the production, regulation and disposal of protein is higher than for mRNA, a cell with limited resources will likely be more vigilant of its protein synthesis as to not waste too much resources.

4.8 Conclusions

The immune response of *Drosophila melanogaster* has been a major field of study in science. Following the discovery and analysis of the various pathways of the Drosophila immune response, researchers investigated the immune response to oral infection in the gut. Then, a genome-wide association study on a *Drosophila* population identified lines highly resistant and susceptible to infection by the entomopathogenic bacteria *Pseudomonas entomophila* and identified key loci mediating this resistance. However, there was no knowledge of the genetic variants causing variation in gene expression after infection by this pathogen. Are the variants sufficient to explain phenotypic differences? How strong are the effects of the variants on gene expression? How different are the gene expression profiles between resistant classes?

To answer these questions, we first analyzed the transcriptome of phenotypically extreme lines. We did not find class-specific gene expression; instead, we found that it was not possible to separate gene expression based on resistance. However, we did find that a majority of *local*-eQTLs were condition-specific, showing that resistance to infection is mostly mediated by cryptic variation and with many loci with small effects adding up to affect resistance.

We then characterized the mode of action of *local*-eQTLs using allele-specific expression in F1 offspring. We found that the majority of *local*-eQTLs do not act in *cis*-. This result highlights the fact that most eQTLs analyses overestimate the amount of *cis*-acting variants. Indeed, a common shortcut for these analyses is to consider the proximity between a variant and the gene linked as a proxy for *cis*- or *trans*-effect where variants located close to the gene are thought to be *cis*-acting whereas variants situated further away are thought to be acting in *trans*-. This highlights the usefulness of allele-specific analysis to understand how a variant affects gene regulation.

Finally, we identified *nutcracker* as the sole differentially-expressed gene between resistance classes. We showed that *ntc* loss-of-function mutants were more susceptible to infection and discovered that a variant linked with *ntc* expression change and located in a TFBS affected the binding affinity of the TF Broad. We demonstrated that this variant affect *ntc* gene expression through differential binding and was responsible for an increase in expression. This is a prime example of a non-coding variant affecting a phenotype through alteration of a gene expression. However, if the effect of *ntc* on the resistance to infection was shown, we still do not know how this gene is affecting the resistance. We gave some ideas of potential functions, but further mechanistic studies are needed to understand the molecular mechanism by which *ntc* affect the resistance to infection.

Because most of the analyses of the fly gut enteric immune response was based on gene expression analysis, we alleviated this by performing an in-depth characterization of the gut proteome after infection using mass-spectrometry. We showed that the proteome variation is strongly time- and pathogen-dependent although we also detected a core of processes conserved in all time and pathogen conditions. Finally, we performed a small screen on proteins that may have been overlooked in previous RNA-based analysis of the gut. These results demonstrate that a mass-spectrometry analysis of the proteome can bring novel insights into an already established topic. Generally, it is the most thorough characterization of the Drosophila gut proteome upon infection by Gram-negative bacteria, thus opening the way for future studies on protein abundance variation in a population making it a great resource for the *Drosophila* community as a whole.

References

Picture of Charles Woodworth: https://en.wikipedia.org/wiki/Charles_W._Woodworth#/media/File:Charles_W._Woodworth_reading.jpg

Picture of Thomas Morgan: http://embryo.asu.edu/handle/10776/2188

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., … Venter, J. C. (2000). The genome sequence of Drosophila melanogaster. *Science*, *287*(March), 2185–2195. https://doi.org/10.1126/science.287.5461.2185

Agaisse, H., & Perrimon, N. (2004). The roles of JAK/STAT signaling in. *Immunological Reviews*, *198*, 72–82.

Albert, A. Y. K., Sawaya, S., Vines, T. H., Knecht, A. K., Miller, C. T., Summers, B. R., … Schluter, D. (2008). The genetics of adaptive shape shift in stickleback: Pleiotropy and effect size. *Evolution*, *62*(1), 76–85. https://doi.org/10.1111/j.1558-5646.2007.00259.x

Albert, F. W., & Kruglyak, L. (2015). The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, *16*(4), 197–212. https://doi.org/10.1038/nrg3891

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter., P. (2002). *Molecular Biology of the Cell, 4th edition* (New York:). New York: New York: Garland Science.

Alberts, R., Chen, H., Pommerenke, C., Smit, A. B., Spijker, S., Williams, R. W., … Schughart, K. (2011). Expression QTL mapping in regulatory and helper T cells from the BXD family of strains reveals novel cell-specific genes, gene-gene interactions and candidate genes for auto-immune disease. *BMC Genomics*, *12*(1). https://doi.org/10.1186/1471-2164-12-610

Alpern, D., Gardeux, V., Russeil, J., & Deplancke, B. (2018). Time- and cost-efficient high-throughput transcriptomics enabled by Bulk RNA Barcoding and sequencing. *Bioarxiv*.

Arya, G. H., Magwire, M. M., Huang, W., Serrano-Negron, Y. L., Mackay, T. F. C., & Anholt, R. R. H. (2015). The genetic basis for variation in olfactory behavior in drosophila melanogaster. *Chemical Senses*, *40*(4), 233–243. https://doi.org/10.1093/chemse/bjv001

Asara, J. M., Christofk, H. R., Freimark, L. M., & Cantley, L. C. (2008). A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. *Proteomics*, *8*(5), 994–999. https://doi.org/10.1002/pmic.200700426

Atilano, M. L., Glittenberg, M., Monteiro, A., Copley, R. R., & Ligoxygakis, P. (2017). Micrornas that contribute to coordinating the immune response in drosophila melanogaster. *Genetics*, *207*(1), 163–178. https://doi.org/10.1534/genetics.116.196584

Bader, M., Arama, E., & Steller, H. (2010). A novel F-box protein is required for caspase

activation during cellular remodeling in Drosophila. *Development (Cambridge, England)*, *137*(10), 1679–1688. https://doi.org/10.1242/dev.050088

Bader, M., Benjamin, S., Wapinski, O. L., Smith, D. M., Goldberg, A. L., & Steller, H. (2011). A conserved F box regulatory complex controls proteasome activity in Drosophila. *Cell*, *145*(3), 371–382. https://doi.org/10.1016/j.cell.2011.03.021

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., … Noble, W. S. (2009). MEME Suite: Tools for motif discovery and searching. *Nucleic Acids Research*, *37*(SUPPL. 2), 202–208. https://doi.org/10.1093/nar/gkp335

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., & Kuster, B. (2007). Quantitative mass spectrometry in proteomics: A critical review. *Analytical and Bioanalytical Chemistry*, *389*(4), 1017–1031. https://doi.org/10.1007/s00216-007-1486-6

Barrett, L. W., Fletcher, S., & Wilton, S. D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements, 3613–3634. https://doi.org/10.1007/s00018-012-0990-9

Basset, A., Khush, R. S., Braun, A., Gardan, L., Boccard, F., Hoffmann, J. A., & Lemaitre, B. (2000). The phytopathogenic bacteria Erwinia carotovora infects Drosophila and activates an immune response. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(7), 3376–3381. https://doi.org/10.1073/pnas.070357597

Bayer, C. A., Holley, B., & Fristrom, J. W. (1996). A switch in Broad-Complex zinc-finger isoform expression is regulated posttranscriptionally during the metamorphosis of Drosophila imaginal discs. *Developmental Biology*, *177*(1), 1–14. https://doi.org/10.1006/dbio.1996.0140

Bellen, H. J., Tong, C., & Tsuda, H. (2010). 100 years of Drosophila research and its impact on vertebrate neuroscience: a history lesson for the future. *Nature Reviews Neuroscience*, *11*(juLy), 1–10. https://doi.org/10.1038/nrn2839

Betel, D., Wilson, M., Gabow, A., Marks, D. S., & Sander, C. (2008). The microRNA.org resource: Targets and expression. *Nucleic Acids Research*, *36*(SUPPL. 1), 149–153. https://doi.org/10.1093/nar/gkm995

Bier, D. M. (1999). The Energy Costs of Protein Metabolism: Lean and Mean. In *Institute of Medicine (US) Committee on Military Nutrition Research. The Role of Protein and Amino Acids in Sustaining and Enhancing Performance.* (Academies). Washington (DC). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK224633/

Bonaldi, T., Straub, T., Cox, J., Kumar, C., Becker, P. B., & Mann, M. (2008). Combined Use of RNAi and Quantitative Proteomics to Study Gene Function in Drosophila. *Molecular Cell*, *31*(5), 762–772. https://doi.org/10.1016/j.molcel.2008.07.018

Bou Aoun, R., Hetru, C., Troxler, L., Doucet, D., Ferrandon, D., & Matt, N. (2011). Analysis of thioester-containing proteins during the innate immune response of Drosophila

melanogaster. *Journal of Innate Immunity*, *3*(1), 52–64. https://doi.org/10.1159/000321554

Bou Sleiman, M. S., Osman, D., Massouras, A., Hoffmann, A. A., Lemaitre, B., & Deplancke, B. (2015). Genetic, molecular and physiological basis of variation in Drosophila gut immunocompetence. *Nature Communications*, *6*, 7829. https://doi.org/10.1038/ncomms8829

Boutros, M., Agaisse, H., & Perrimon, N. (2002). Sequential Activation of Signaling Pathways during Innate Immune Responses in Drosophila. *Developmental Cell*, *3*, 711–722. Retrieved from papers2://publication/uuid/22DF8B07-F78A-40B8-8600-2072C57029F2

Bridges, S. M., Bryce, G. B., Wang, N., Williams, W. P., Burgess, S. C., & Nanduri, B. (2007). ProtQuant: A tool for the label-free quantification of MudPIT proteomics data. *BMC Bioinformatics*, *8*(SUPPL. 7), 1–9. https://doi.org/10.1186/1471-2105-8-S7-S24

Buchon, N, Broderick, N. A., & Lemaitre, B. (2013). Gut homeostasis in a microbial world: insights from Drosophila melanogaster. *Nat Rev Microbiol*, *11*(9), 615–626. https://doi.org/10.1038/nrmicro3074

Buchon, Nicolas, Broderick, N. A., Poidevin, M., Pradervand, S., & Lemaitre, B. (2009). Drosophila Intestinal Response to Bacterial Infection: Activation of Host Defense and Stem Cell Proliferation. *Cell Host and Microbe*, *5*(2), 200–211. https://doi.org/10.1016/j.chom.2009.01.003

Buchon, Nicolas, & Osman, D. (2015). All for one and one for all: Regionalization of the Drosophila intestine. *Insect Biochemistry and Molecular Biology*, *67*, 2–8. https://doi.org/10.1016/j.ibmb.2015.05.015

Buchon, Nicolas, Osman, D., David, F. P. A., Yu Fang, H., Boquete, J. P., Deplancke, B., & Lemaitre, B. (2013). Morphological and Molecular Characterization of Adult Midgut Compartmentalization in Drosophila. *Cell Reports*, *3*(5), 1725–1738. https://doi.org/10.1016/j.celrep.2013.04.001

Burke, M., & Rose, M. (2009). Experimental evolution with Drosophila. *American Journal of Physiology*, *296*, 1847–1854. https://doi.org/10.1152/ajpregu.90551.2008.

Butter, F., Bucerius, F., Michel, M., Cicova, Z., Mann, M., & Janzen, C. J. (2013). Comparative Proteomics of Two Life Cycle Stages of Stable Isotope-labeled Trypanosoma brucei Reveals Novel Components of the Parasite's Host Adaptation Machinery. *Molecular & Cellular Proteomics*, *12*(1), 172–179. https://doi.org/10.1074/mcp.m112.019224

Cannavò, E., Koelling, N., Harnett, D., Garfield, D., Casale, F. P., Ciglar, L., … Furlong, E. E. M. (2016). Genetic variants regulating expression levels and isoform diversity during embryogenesis. *Nature*, *541*(7637), 402–406. https://doi.org/10.1038/nature20802

Casas-Vila, N., Bluhm, A., Sayols, S., Dinges, N., Dejung, M., Altenhein, T., … Butter, F. (2017). The developmental proteome of *Drosophila melanogaster*. *Genome Research*,

gr.213694.116. https://doi.org/10.1101/gr.213694.116

Catalanotto, C., Cogoni, C., & Zardo, G. (2016). MicroRNA in control of gene expression: An overview of nuclear functions. *International Journal of Molecular Sciences*, *17*(10). https://doi.org/10.3390/ijms17101712

Chakrabarti, S, Liehl, P., Buchon, N., & Lemaitre, B. (2012). Infection-induced host translational blockage inhibits immune responses and epithelial renewal in the Drosophila gut. *Cell Host Microbe*, *12*(1), 60–70. https://doi.org/10.1016/j.chom.2012.06.001

Chakrabarti, Sveta, Poidevin, M., & Lemaitre, B. (2014). The Drosophila MAPK p38c Regulates Oxidative Stress and Lipid Homeostasis in the Intestine. *PLoS Genetics*, *10*(9). https://doi.org/10.1371/journal.pgen.1004659

Chatterjee, M., & Ip, Y. T. (2009). Pathogenic stimulation of intestinal stem cell response in Drosophila. *Journal of Cellular Physiology*, *220*(3), 664–671. https://doi.org/10.1002/jcp.21808

Chen, J., Rozowsky, J., Galeev, T. R., Harmanci, A., Kitchen, R., Bedford, J., … Gerstein, M. (2016). A uniform survey of allele-specific binding and expression over 1000-Genomes-Project individuals. *Nature Communications*, *48*(7), 829–834. https://doi.org/10.1038/ncomms11101

Chen, L., Paquette, N., Mamoor, S., Rus, F., Nandy, A., Leszyk, J., … Silverman, N. (2017). Innate immune signaling in Drosophila is regulated by transforming growth factor β (TGFβ)-activated kinase (Tak1)-triggered ubiquitin editing. *Journal of Biological Chemistry*, *292*(21), 8738–8749. https://doi.org/10.1074/jbc.M117.788158

Chintapalli, V. R., Wang, J., & Dow, J. a T. (2007). Using FlyAtlas to identify better Drosophila melanogaster models of human disease. *Nature Genetics*, *39*(6), 715–720. https://doi.org/10.1038/ng2049

Cox, J., & Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology*, *26*(12), 1367–1372. https://doi.org/10.1038/nbt.1511

Coyne, J. A., & Orr, H. A. (1997). Patterns of Speciation in Drosophila. *Evolution*, *51*(1), 295. https://doi.org/10.2307/2410984

Davenport, E. R., Cusanovich, D. A., Michelini, K., Barreiro, L. B., Ober, C., & Gilad, Y. (2015). Genome-Wide Association Studies of the Human Gut Microbiota. *Plos One*, *10*(11), e0140301. https://doi.org/10.1371/journal.pone.0140301

Dawkins, R. (1978). Replicator Selection and the Extended Phenotype 3. *Ethology*, *47*(1), 61–76.

De Gregorio, E., Spellman, P. T., Rubin, G. M., & Lemaitre, B. (2002). Genome-wide analysis of the Drosophila immune response by using oligonucleotide microarrays. *Proceedings of the National Academy of Sciences*, *98*(22), 12590–12595.

https://doi.org/10.1073/pnas.221458698

Degner, J. F., Marioni, J. C., Pai, A. A., Pickrell, J. K., Nkadori, E., Gilad, Y., & Pritchard, J. K. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, *25*(24), 3207–3212. https://doi.org/10.1093/bioinformatics/btp579

Dembeck, L. M., Huang, W., Magwire, M. M., Lawrence, F., Lyman, R. F., & Mackay, T. F. C. (2015). Genetic Architecture of Abdominal Pigmentation in Drosophila melanogaster. *PLoS Genetics*, *11*(5), 1–22. https://doi.org/10.1371/journal.pgen.1005163

Deplancke, B., Alpern, D., & Gardeux, V. (2016). The Genetics of Transcription Factor DNA Binding Variation. *Cell*, *166*(3), 538–554. https://doi.org/10.1016/j.cell.2016.07.012

Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–501. https://doi.org/10.1038/ng.806

Di Narzo, A. F., Peters, L. A., Argmann, C., Stojmirovic, A., Perrigoue, J., Li, K., … Hao, K. (2016). Blood and Intestine eQTLs from an Anti-TNF-Resistant Crohn's Disease Cohort Inform IBD Genetic Association Loci. *Clinical and Translational Gastroenterology*, *7*(6), e177. https://doi.org/10.1038/ctg.2016.34

Dionne, M. S. (2014). Immune-metabolic interaction in Drosophila. *Fly*, *8*(2), 37–41. https://doi.org/10.4161/fly.28113

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Dostálová, A., Rommelaere, S., Poidevin, M., & Lemaitre, B. (2017). Thioester-containing proteins regulate the Toll pathway and play a role in Drosophila defence against microbial pathogens and parasitoid wasps. *BMC Biology*, *15*(1), 1–16. https://doi.org/10.1186/s12915-017-0408-0

Durham, M. F., Magwire, M. M., Stone, E. A., & Leips, J. (2014). Genome-wide analysis in Drosophila reveals age-specific effects of SNPs on fitness traits. *Nature Communications*, *5*, 1–8. https://doi.org/10.1038/ncomms5338

Dutta, D., Dobson, A. J., Houtz, P. L., Glasser, C., Revah, J., Korzelius, J., … Buchon, N. (2015). Regional Cell-Specific Transcriptome Mapping Reveals Regulatory Complexity in the Adult Drosophila Midgut. *Cell Rep*, *12*(2), 346–358. https://doi.org/10.1016/j.celrep.2015.06.009

Early, A. M., Shanmugarajah, N., Buchon, N., & Clark, A. G. (2017). Drosophila genotype influences commensal bacterial levels. *PLoS ONE*, *12*(1), 1–15. https://doi.org/10.1371/journal.pone.0170332

Enomoto, M., & Siow, C. (2018). Drosophila Models for Human Diseases. In *Drosophila*

Models for Human Diseases (Vol. 1076). https://doi.org/10.1007/978-981-13-0529-0

Fabian, M. R., Sonenberg, N., & Filipowicz, W. (n.d.). Regulation of mRNA Translation and Stability by microRNAs. https://doi.org/10.1146/annurev-biochem-060308-103103

Fairfax, B. P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., … Knight, J. C. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science (New York, N.Y.)*, *343*(6175), 1246949. https://doi.org/10.1126/science.1246949

Fairfax, B. P., Makino, S., Radhakrishnan, J., Plant, K., Leslie, S., Dilthey, A., … Knight, J. C. (2012). Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genetics*, *44*(5), 502–510. https://doi.org/10.1038/ng.2205

Falcon, S., & Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. *Bioinformatics*, *23*(2), 257–258. https://doi.org/10.1093/bioinformatics/btl567

Feany, M. B., & Bender, W. W. (2000). A Drosophila model of Parkinson's disease. *Nature*, *404*(6776), 394–398. https://doi.org/10.1038/35006074

Fearon DT, & Locksley RM. (1996). The instructive role of innate immunity in the acquired immune response. - PubMed - NCBI. *Science*, *272*(5258), 50–53. https://doi.org/10.1126/science.272.5258.50

Ferguson-Smith, M. A., Aitken, D. A., Turleau, C., & de Grouchy, J. (1976). Localisation of the human ABO: Np-1: AK-1 linkage group by regional assignment of AK-1 to 9q34. *Human Genetics*, *34*(1), 35–43.

Galenza, A., & Foley, E. (2019). Immunometabolism: Insights from the Drosophila model. *Developmental and Comparative Immunology*, *94*(January), 22–34. https://doi.org/10.1016/j.dci.2019.01.011

Gan, K. A., Pro, S. C., Sewell, J. A., & Fuxman Bass, J. I. (2018). Identification of single nucleotide non-coding driver mutations in cancer. *Frontiers in Genetics*, *9*(FEB), 1–10. https://doi.org/10.3389/fgene.2018.00016

Garlapow, M. E., Huang, W., Yarboro, M. T., Peterson, K. R., & Mackay, T. F. C. (2015). Quantitative genetics of food intake in Drosophila melanogaster. *PLoS ONE*, *10*(9), 1–25. https://doi.org/10.1371/journal.pone.0138129

Gedeon, T., & Bokes, P. (2012). Delayed protein synthesis reduces the correlation between mRNA and protein fluctuations. *Biophysical Journal*, *103*(3), 377–385. https://doi.org/10.1016/j.bpj.2012.06.025

Gibson, G., & Dworkin, I. (2004). Uncovering cryptic genetic variation. *Nature Reviews Genetics*, *5*(9), 681–690. https://doi.org/10.1038/nrg1426

Gibson, G., Powell, J. E., & Marigorta, U. M. (2015). Expression quantitative trait locus analysis for translational medicine. *Genome Medicine*, *7*(1), 1–14. https://doi.org/10.1186/s13073-

015-0186-7

Gilad, Y., Rifkin, S. A., & Pritchard, J. K. (2008). Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics*, *24*(8), 408–415. https://doi.org/10.1016/j.tig.2008.06.001

Gilchrist, J. J., Mills, T. C., Naranbhai, V., Chapman, S. J., Fairfax, B. P., Knight, J. C., … Hill, A. V. (2015). Genetic variants associated with non-typhoidal Salmonella bacteraemia in African children. *The Lancet*, *385*, S13. https://doi.org/10.1016/s0140-6736(15)60328-2

Gordon, M. D., Dionne, M. S., Schneider, D. S., & Nusse, R. (2005). WntD is a feedback inhibitor of Dorsal/NF-κB in Drosophila development and immunity. *Nature*, *437*(7059), 746–749. https://doi.org/10.1038/nature04073

Griffin, T. J., Gygi, S. P., Ideker, T., Rist, B., Eng, J., Hood, L., & Aebersold, R. (2002). Complementary Profiling of Gene Expression at the Transcriptome and Proteome Levels in Saccharomyces cerevisiae. *Molecular & Cellular Proteomics*, *1*(4), 323–333. https://doi.org/10.1074/mcp.m200001-mcp200

Grün, D., Kirchner, M., Thierfelder, N., Stoeckius, M., Selbach, M., & Rajewsky, N. (2014). Conservation of mRNA and protein expression during development of C.elegans. *Cell Reports*, *6*(3), 565–577. https://doi.org/10.1016/j.celrep.2014.01.001

Gundogdu, A., & Nalbantoglu, U. (2017). Human genome-microbiome interaction: metagenomics frontiers for the aetiopathology of autoimmune diseases. *Microbial Genomics*, *3*(4), 1–5. https://doi.org/10.1099/mgen.0.000112

Gundry, R. L., White, M. Y., Murray, C. I., Kane, L. A., Fu, Q., Stanley, B. A., & Van Eyk, J. E. (2009). Preparation of proteins and peptides for mass spectrometry analysis in a bottom-up proteomics workflow. *Current Protocols in Molecular Biology*, (SUPPL. 88), 1–23. https://doi.org/10.1002/0471142727.mb1025s88

Gupta, V., Stewart, C. O., Rund, S. S. C., Monteith, K., & Vale, P. F. (2017). Costs and benefits of sublethal Drosophila C virus infection. *Journal of Evolutionary Biology*, *30*(7), 1325–1335. https://doi.org/10.1111/jeb.13096

Ha, E.-M., Lee, K.-A., Seo, Y. Y., Kim, S.-H., Lim, J.-H., Oh, B.-H., … Lee, W.-J. (2009). Coordination of multiple dual oxidase-regulatory pathways in responses to commensal and infectious microbes in drosophila gut. *Nature Immunology*, *10*(9), 949–957. https://doi.org/10.1038/ni.1765

Ha, E. M., Oh, C. T., Ryu, J. H., Bae, Y. S., Kang, S. W., Jang, I. hwan, … Lee, W. J. (2005). An antioxidant system required for host protection against gut infection in Drosophila. *Developmental Cell*, *8*(1), 125–132. https://doi.org/10.1016/j.devcel.2004.11.007

Haldane, A., Manhart, M., & Morozov, A. V. (2014). Biophysical Fitness Landscapes for Transcription Factor Binding Sites. *PLoS Computational Biology*, *10*(7), 36–38. https://doi.org/10.1371/journal.pcbi.1003683

Hang, S., Purdy, A. E., Robins, W. P., Wang, Z., Mandal, M., Chang, S., … Watnick, P. I. (2014). The acetate switch of an intestinal pathogen disrupts host insulin signaling and lipid metabolism. *Cell Host and Microbe*, *16*(5), 592–604. https://doi.org/10.1016/j.chom.2014.10.006

Hansen, T. F. (2006). The Evolution of Genetic Architecture. *Annual Review of Ecology, Evolution, and Systematics*, *37*(1), 123–157. https://doi.org/10.1146/annurev.ecolsys.37.091305.110224

Harrison, D. a, & Perrimon, N. (1993). Simple and efficient generation of marked clones in Drosophila. *Current Biology : CB*, *3*(7), 424–433. https://doi.org/10.1016/0960-9822(93)90349-S

Hasin-Brumshtein, Y., Hormozdiari, F., Martin, L., van Nas, A., Eskin, E., Lusis, A. J., & Drake, T. A. (2014). Allele-specific expression and eQTL analysis in mouse adipose tissue. *BMC Genomics*, *15*(1), 1–13. https://doi.org/10.1186/1471-2164-15-471

Hindorff, L. A., Sethupathy, P., Junkins, H. A., Ramos, E. M., Mehta, J. P., Collins, F. S., & Manolio, T. A. (2009). Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences*, *106*(23), 9362–9367. https://doi.org/10.1073/pnas.0903103106

Hoskins, R. A., Carlson, J. W., Wan, K. H., Park, S., Mendez, I., Galle, S. E., … Celniker, S. E. (2015). The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Research*, *25*(3), 445–458. https://doi.org/10.1101/gr.185579.114

Hu, Y.-J., Sun, W., Tzeng, J.-Y., & Perou, C. (2015). Proper Use of Allele-Specific Expression Improves Statistical Power for cis-eQTL Mapping with RNA-Seq Data. *J Am Stat Assoc*, *110*(511), 962–974. https://doi.org/10.1080/01621459.2015.1038449.Proper

Huang, W., Carbone, M. A., Magwire, M. M., Peiffer, J. A., Lyman, R. F., Stone, E. A., … Mackay, T. F. C. (2015). Genetic basis of transcriptome diversity in Drosophila melanogaster. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), E6010-9. https://doi.org/10.1073/pnas.1519159112

Huang, W., Massouras, A., & Inoue, Y. (2014). Natural variation in genome architecture among 205 Drosophila melanogaster Genetic Reference Panel lines Natural Variation in Genome Architecture Among 205 Drosophila melanogaster Genetic Reference Panel Lines. *Genome Research*, 1193–1208. https://doi.org/10.1101/gr.171546.113

Huang, Y.-T., VanderWeele, T. J., & Lin, X. (2014). Joint Analysis of Snp and Gene Expression Data in Genetic Association Studies of Complex Diseases. *Annals of Applied Statistics*, *8*(1), 1–24. https://doi.org/10.1214/13-AOAS690.JOINT

Hulur, I., Gamazon, E. R., Skol, A. D., Xicola, R. M., Llor, X., Onel, K., … Kupfer, S. S. (2015). Enrichment of inflammatory bowel disease and colorectal cancer risk variants in colon expression quantitative trait loci. *BMC Genomics*, *16*(1), 1–15.

https://doi.org/10.1186/s12864-015-1292-z

Ingolia, N. T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. *Nature Reviews Genetics*, *15*(3), 205–213. https://doi.org/10.1038/nrg3645

Inoue, F., & Ahituv, N. (2015). Genomics Decoding enhancers using massively parallel reporter assays. *Genomics*, *106*(3), 159–164. https://doi.org/10.1016/j.ygeno.2015.06.005

Irving, P., Troxler, L., Heuer, T. S., Belvin, M., Kopczynski, C., Reichhart, J. M., … Hetru, C. (2001). A genome-wide analysis of immune responses in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(26), 15119–15124. https://doi.org/10.1073/pnas.261573998

Isakova, A., Groux, R., Imbeault, M., Rainer, P., Alpern, D., Dainese, R., … Deplancke, B. (2017). SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nature Methods*, *14*(3), 316–322. https://doi.org/10.1038/nmeth.4143

Izutsu, M., Toyoda, A., Fujiyama, A., Agata, K., & Fuse, N. (2016). Dynamics of Dark-Fly Genome Under Environmental Selections. *G3: Genes|Genomes|Genetics*, *6*(2), 365–376. https://doi.org/10.1534/g3.115.023549

Janeway, C. J., Travers, P., & Walport, M. (2001). *Immunobiology: The Immune System in Health and Disease* (New York:). New York.

Jha, A. R., Miles, C. M., Lippert, N. R., Brown, C. D., White, K. P., & Kreitman, M. (2015). Whole-genome resequencing of experimental populations reveals polygenic basis of egg-size variation in drosophila melanogaster. *Molecular Biology and Evolution*, *32*(10), 2616–2632. https://doi.org/10.1093/molbev/msv136

Jin, H.-J., Jung, S., DebRoy, A. R., & Davuluri, R. V. (2016). Identification and validation of regulatory SNPs that modulate transcription factor chromatin binding and gene expression in prostate cancer. *Oncotarget*, *7*(34), 54616–54626. https://doi.org/10.18632/oncotarget.10520

Kabakchiev, B., & Silverberg, M. S. (2013). Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology*, *144*(7), 1488-1496.e3. https://doi.org/10.1053/j.gastro.2013.03.001

Karim, F. D., Guild, G. M., & Thummel, C. S. (1993). The Drosophila Broad-Complex plays a key role in controlling ecdysone-regulated gene expression at the onset of metamorphosis. *Development (Cambridge, England)*, *118*(3), 977–988. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8076529

Kaufman, T. C. (2017). A Short History and Description of Drosophila Aberrations , Forward Genetic Screens , and the Nature of Mutations. *Genetics*, *206*(June), 665–689.

Kelley, K., & Preacher, K. J. (2012). On effect size. *Psychological Methods*, *17*(2), 137–152.

Khansefid, M., Pryce, J. E., Bolormaa, S., Chen, Y., Millen, C. A., Chamberlain, A. J., … Goddard, M. E. (2018). Comparing allele specific expression and local expression quantitative trait loci and the influence of gene expression on complex trait variation in cattle. *BMC Genomics*, *19*(1), 1–18. https://doi.org/10.1186/s12864-018-5181-0

Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, Y., Giangreco, N., … Hornung, V. (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00366-1

Kim, M., Lee, J. H., Lee, S. Y., Kim, E., & Chung, J. (2006). Caspar, a suppressor of antibacterial immunity in Drosophila. *Proceedings of the National Academy of Sciences*, *103*(44), 16358–16363. https://doi.org/10.1073/pnas.0603238103

Kim, S.-H., & Lee, W.-J. (2014). Role of DUOX in gut inflammation: lessons from Drosophila model of gut-microbiota interactions. *Frontiers in Cellular and Infection Microbiology*, *3*(January), 116. https://doi.org/10.3389/fcimb.2013.00116

Kinchen, J. M., & Ravichandran, K. S. (2008). Phagosome maturation: Going through the acid test. *Nature Reviews Molecular Cell Biology*, *9*(10), 781–795. https://doi.org/10.1038/nrm2515

King, E. G., Sanderson, B. J., McNeil, C. L., Long, A. D., & Macdonald, S. J. (2014). Genetic Dissection of the Drosophila melanogaster Female Head Transcriptome Reveals Widespread Allelic Heterogeneity. *PLoS Genetics*, *10*(5). https://doi.org/10.1371/journal.pgen.1004322

Kleino, A., Myllymäki, H., Kallio, J., Vanha-aho, L.-M., Oksanen, K., Ulvila, J., … Rämet, M. (2008). Pirk is a negative regulator of the Drosophila Imd pathway. *Journal of Immunology (Baltimore, Md. : 1950)*, *180*(8), 5413–5422. https://doi.org/10.4049/jimmunol.180.8.5413

Klepsatel, P., Gáliková, M., De Maio, N., Huber, C. D., Schlötterer, C., & Flatt, T. (2013). Variation in thermal performance and reaction norms among populations of drosophila melanogaster. *Evolution*, *67*(12), 3573–3587. https://doi.org/10.1111/evo.12221

Kounatidis, I., & Ligoxygakis, P. (2012). Drosophila as a model system to unravel the layers of innate immunity to infection. *Open Biology*, *2*(MAY). https://doi.org/10.1098/rsob.120075

Kukurba, K. R., Zhang, R., Li, X., Smith, K. S., Knowles, D. A., How Tan, M., … Montgomery, S. B. (2014). Allelic Expression of Deleterious Protein-Coding Variants across Human Tissues. *PLoS Genetics*, *10*(5). https://doi.org/10.1371/journal.pgen.1004304

Kumar, A., Wu, H., Collier-Hyams, L. S., Hansen, J. M., Li, T., Yamoah, K., … Neish, A. S. (2007). Commensal bacteria modulate cullin-dependent signaling via generation of reactive oxygen species. *EMBO Journal*, *26*(21), 4457–4466. https://doi.org/10.1038/sj.emboj.7601867

Lawrenson, K., Li, Q., Kar, S., Seo, J.-H., Tyrer, J., Spindler, T. J., … Freedman, M. L. (2015).

Cis-eQTL analysis and functional validation of candidate susceptibility genes for high-grade serous ovarian cancer. *Nature Communications*, *6*(May), 8234. https://doi.org/10.1038/ncomms9234

Lee, K.-A., Cho, K.-C., Kim, B., Jang, I.-H., Nam, K., Kwon, Y. E., … Lee, W.-J. (2018). Inflammation-Modulated Metabolic Reprogramming Is Required for DUOX-Dependent Gut Immunity in Drosophila. *Cell Host & Microbe*, 1–15. https://doi.org/10.1016/j.chom.2018.01.011

Lee, K. A., Kim, B., Bhin, J., Kim, D. H., You, H., Kim, E. K., … Lee, W. J. (2015). Bacterial uracil modulates drosophila DUOX-dependent Gut immunity via hedgehog-induced signaling endosomes. *Cell Host and Microbe*, *17*(2), 191–204. https://doi.org/10.1016/j.chom.2014.12.012

Lee, K. A., Kim, B., You, H., & Lee, W. J. (2015). Uracil-induced signaling pathways for DUOX-dependent gut immunity. *Fly*, *9*(3), 115–120. https://doi.org/10.1080/19336934.2015.1126011

Lee, M. N., Ye, C., Villani, A.-C., Raj, T., Li, W., Eisenhaure, T. M., … Hacohen, N. (2014). Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. *Science*, *343*(6175), 1246980–1246980. https://doi.org/10.1126/science.1246980

Lee, W. J. (2009). Bacterial-modulated host immunity and stem cell activation for gut homeostasis. *Genes and Development*, *23*(19), 2260–2265. https://doi.org/10.1101/gad.1858709

Lemaitre, B., & Hoffmann, J. (2007). The Host Defense of Drosophila melanogaster. *Annual Review of Immunology*, *25*, 697–743. https://doi.org/10.1146/annurev.immunol.25.022106.141615

Lemaitre, B., & Miguel-Aliaga, I. (2013). The digestive tract of Drosophila melanogaster. *Annual Review of Genetics*, *47*(August), 377–404. https://doi.org/10.1146/annurev-genet-111212-133343

Leulier, F., Rodriguez, a, Khush, R. S., Abrams, J. M., & Lemaitre, B. (2000). The Drosophila caspase Dredd is required to resist gram-negative bacterial infection. *EMBO Reports*, *1*(4), 353–358. https://doi.org/10.1093/embo-reports/kvd073

Li, J. J., Bickel, P. J., & Biggin, M. D. (2014). System wide analyses have underestimated protein abundances and the importance of transcription in mammals. *PeerJ*, *2*, e270. https://doi.org/10.7717/peerj.270

Liehl, P., Blight, M., Vodovar, N., Boccard, F., & Lemaitre, B. (2006). Prevalence of local immune response against oral infection in a Drosophila/Pseudomonas infection model. *PLoS Pathogens*, *2*(6), 0551–0561. https://doi.org/10.1371/journal.ppat.0020056

Lugowski, A., Nicholson, B., & Rissland, O. S. (2018). Determining mRNA half-lives on a transcriptome-wide scale. *Methods*, *137*, 90–98.

https://doi.org/10.1016/j.ymeth.2017.12.006

Ma, J., & Ptashne, M. (1987). A new class of yeast transcriptional activators. *Cell*, *51*(1), 113–119. https://doi.org/10.1016/0092-8674(87)90015-8

Mackay, T. F. C., Richards, S., Stone, E. a, Barbadilla, A., Ayroles, J. F., Zhu, D., … Gibbs, R. a. (2012). The Drosophila melanogaster Genetic Reference Panel. *Nature*, *482*(7384), 173–178. https://doi.org/10.1038/nature10811

Maerkl, S. J., & Quake, S. R. (2007). A Systems Approach to Measuring the Binding Energy Landscapes of Transcription Factors. *Science*, *315*(5809), 233–237. https://doi.org/10.1126/science.1131007

Maher, S., Jjunju, F. P. M., & Taylor, S. (2015). Colloquium: 100 years of mass spectrometry: Perspectives and future trends. *Reviews of Modern Physics*, *87*(1), 113–135. https://doi.org/10.1103/RevModPhys.87.113

Manolio, T. A. (2010). Genomewide Association Studies and Assessment of the Risk of Disease. *The New England Journal of Medicine*, (363), 166–176.

Marianes, A., & Spradling, A. C. (2013). Physiological and stem cell compartmentalization within the Drosophila midgut. *ELife*, *2013*(2), 1–19. https://doi.org/10.7554/eLife.00886

Maricque, B. B., Chaudhari, H. G., & Cohen, B. A. (2019). A massively parallel reporter assay dissects the influence of chromatin structure on cis-regulatory activity. *Nature Biotechnology*, *37*(1), 90–95. https://doi.org/10.1038/nbt.4285

Markow, T. A., Beall, S., & Matzkin, L. M. (2009). Egg size, embryonic development time and ovoviviparity in Drosophila species. *Journal of Evolutionary Biology*, *22*(2), 430–434. https://doi.org/10.1111/j.1420-9101.2008.01649.x

Markow, Therese Ann. (2011). Cost of virginity in wild drosophila melanogaster females. *Ecology and Evolution*, *1*(4), 596–600. https://doi.org/10.1002/ece3.54

Markow, Therese Ann. (2015). The secret lives of Drosophila flies. *ELife*, 1–9. https://doi.org/10.7554/eLife.06793

Massouras, A., Waszak, S. M., Albarca-Aguilera, M., Hens, K., Holcombe, W., Ayroles, J. F., … Deplancke, B. (2012). Genomic Variation and Its Impact on Gene Expression in Drosophila melanogaster. *PLoS Genetics*, *8*(11). https://doi.org/10.1371/journal.pgen.1003055

McLeay, R. C., & Bailey, T. L. (2010). Motif Enrichment Analysis: A unified framework and an evaluation on ChIP data. *BMC Bioinformatics*, *11*(165). https://doi.org/10.1186/1471-2105-11-165

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Research*, *20*(6), 816–825. https://doi.org/10.1101/gr.102491.109

McQuilton, P., St Pierre, S. E., & FlyBase Consortium. (2012). FlyBase 101 - The basics of

navigating FlyBase. *Nucleic Acids Research*, *40*(D1), 706–714. https://doi.org/10.1093/nar/gkr1030

Meaburn, K. J., & Misteli, T. (2007). Chromosome territories. *Nature*, *445*(25), 379–381. https://doi.org/10.1240/sav_gbm_2007_h_002046

Meinander, A., Runchel, C., Tenev, T., Chen, L., Kim, C.-H., Ribeiro, P. S., … Meier, P. (2012). Ubiquitylation of the initiator caspase DREDD is required for innate immune signalling. *The EMBO Journal*, *31*(12), 2770–2783. https://doi.org/10.1038/emboj.2012.121

Micchelli, C. A., & Perrimon, N. (2006). Evidence that stem cells reside in the adult Drosophila midgut epithelium. *Nature*, *439*(7075), 475–479. https://doi.org/10.1038/nature04371

Miguel-Aliaga, I., Jasper, H., & Lemaitre, B. (2018). Anatomy and physiology of the digestive tract of drosophila melanogaster. *Genetics*, *210*(2), 357–396. https://doi.org/10.1534/genetics.118.300224

Millburn, G. H., Crosby, M. A., Gramates, L. S., & Tweedie, S. (2016). FlyBase portals to human disease research using Drosophila models . *Disease Models & Mechanisms*, *9*(3), 245–252. https://doi.org/10.1242/dmm.023317

Mistry, R., Kounatidis, I., & Ligoxygakis, P. (2015). Exploring interactions between pathogens and the Drosophila gut. *Developmental and Comparative Immunology*. https://doi.org/10.1016/j.dci.2016.01.016

Mondotte, J. A., Frangeul, L., Lambrechts, L., Saleh, M.-C., Gausson, V., & Blanc, H. (2018). Immune priming and clearance of orally acquired RNA viruses in Drosophila. *Nature Microbiology*, *3*(12), 1394–1403. https://doi.org/10.1038/s41564-018-0265-9

Morgan, T. H. (1910). Sex Limited Inheritance in Drosophila. *Science*, *32*(812), 120–122.

Moyerbrailean, G. A., Richards, A. L., Kurtz, D., Kalita, C. A., Davis, G. O., Harvey, C. T., … Luca, F. (2016). High-throughput allele-specific expression across 250 environmental conditions. *Genome Research*, *26*(12), 1627–1638. https://doi.org/10.1101/gr.209759.116

Mugat, B., Brodu, V., Kejzlarova-Lepesant, J., Antoniewski, C., Bayer, C. A., Fristrom, J. W., & Lepesant, J. A. (2000). Dynamic expression of Broad-Complex isoforms mediates temporal control of an ecdysteroid target gene at the onset of Drosophila metamorphosis. *Developmental Biology*, *227*(1), 104–117. https://doi.org/10.1006/dbio.2000.9879

NCBI. (2015). Drosophila melanogaster genome at NCBI. Retrieved from https://www.ncbi.nlm.nih.gov/genome/47

Neyen, C., Bretscher, A. J., Binggeli, O., & Lemaitre, B. (2014). Methods to study Drosophila immunity. *Methods*, *68*(1), 116–128. https://doi.org/10.1016/j.ymeth.2014.02.023

Nica, A. C., & Dermitzakis, E. T. (2013). Expression quantitative trait loci: present and future. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *368*(1620). https://doi.org/10.1098/rstb.2012.0362

Nochi, T., & Kiyono., H. (2006). Innate Immunity in the Mucosal Immune System. *Current Pharmaceutical Design*, *12*(32). https://doi.org/10.2174/138161206778743457

Ohlstein, B., & Spradling, A. (2006). The adult Drosophila posterior midgut is maintained by pluripotent stem cells. *Nature*, *439*(7075), 470–474. https://doi.org/10.1038/nature04333

Ohlstein, B., & Spradling, A. (2007). Multipotent Drosophila Intestinal Stem Cells Specify Daughter Cell Fates by Differential Notch Signaling. *Science*, *315*(July), 372–377.

Okada, H., Ebhardt, H. A., Vonesch, S. C., Aebersold, R., & Hafen, E. (2016). Proteome-wide association studies identify biochemical modules associated with a wing-size phenotype in Drosophila melanogaster. *Nature Communications*, *7*, 12649. https://doi.org/10.1038/ncomms12649

Orozco, L. D., Bennett, B. J., Farber, C. R., Ghazalpour, A., Pan, C., Che, N., … Lusis, A. J. (2012). Unraveling inflammatory responses using systems genetics and gene-environment interactions in macrophages. *Cell*, *151*(3), 658–670. https://doi.org/10.1016/j.cell.2012.08.043

Paiva, C. N., & Bozza, M. T. (2014). Are Reactive Oxygen Species Always Detrimental to Pathogens? *Antioxidants & Redox Signaling*, *20*(6), 1000–1037. https://doi.org/10.1089/ars.2013.5447

Paquette, N., Broemer, M., Aggarwal, K., Chen, L., Husson, M., Ertürk-Hasdemir, D., … Silverman, N. (2010). Caspase-Mediated Cleavage, IAP Binding, and Ubiquitination: Linking Three Mechanisms Crucial for Drosophila NF-κB Signaling. *Molecular Cell*, *37*(2), 172–182. https://doi.org/10.1016/j.molcel.2009.12.036

Paredes, J. C., Welchman, D. P., Poidevin, M., & Lemaitre, B. (2011). Negative regulation by Amidase PGRPs shapes the drosophila antibacterial response and protects the Fly from innocuous infection. *Immunity*, *35*(5), 770–779. https://doi.org/10.1016/j.immuni.2011.09.018

Peters, J. E., Lyons, P. A., Lee, J. C., Richard, A. C., Fortune, M. D., Newcombe, P. J., … Smith, K. G. C. (2016). Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. *PLoS Genetics*, *12*(3). https://doi.org/10.1371/journal.pgen.1005908

Qu, W., Gurdziel, K., Pique-Regi, R., & Ruden, D. M. (2018). Lead Modulates trans-and cis-Expression Quantitative Trait Loci (eQTLs) in Drosophila melanogasterHeads. *Frontiers in Genetics*, *9*(SEP), 1–13. https://doi.org/10.3389/fgene.2018.00395

Raj, T., Rothamel, K., Mostafavi, S., Ye, C., Lee, M. N., Replogle, J. M., … Jager, P. L. De. (2014). Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes. *Science*, *344*(May), 519–523.

Rockman, M. V., & Kruglyak, L. (2006). Genetics of global gene expression. *Nature Reviews Genetics*, *7*(11), 862–872. https://doi.org/10.1038/nrg1964

Roff D.A., & Mousseau T.A. (1987). Quantitative genetics and fitness: lesson from Drosophila. *Heredity*, *58*(April 1986), 103–118.

Rojano, E., Seoane, P., Ranea, J. A. G., & Perkins, J. R. (2018). Regulatory variants: from detection to predicting impact. *Briefings in Bioinformatics*, (April), 1–16. https://doi.org/10.1093/bib/bby039

Rubin, G. M., & Lewis, E. B. (2000). A Brief History of Drosophila's Contributions to Genome Research. *Science*, *287*(March), 2196–2204.

Ruffieux, H., Carayol, J., Harper, M. E., Dent, R., Saris, W. H. M., Astrup, A., … Valsesia, A. (2019). A large-scale multivariate pQTL study sheds light on the genetic architecture of obesity. *Bioarxiv*, 1–25.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985). Enzymatic Amplification of beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science*, *230*(4732), 1350–1354.

Sano, Y., Akimaru, H., Okamura, T., Nagao, T., Okada, M., & Ishii, S. (2005). Drosophila Activating Transcription Factor-2 Is Involved in Stress Response via Activation by p38, but Not c-Jun NH2-Terminal Kinase. *Molecular Biology of the Cell*, *16*, 2934–2946. https://doi.org/10.1091/mbc.E04

Schmiedel, B. J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A. G., White, B. M., Zapardiel-Gonzalo, J., … Vijayanand, P. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell*, *175*(6), 1701-1715.e16. https://doi.org/10.1016/j.cell.2018.10.022

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., … Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature*, *473*(7347), 337–342. https://doi.org/nature10098 [pii]10.1038/nature10098

Shabalin, A. A. (2012). Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, *28*(10), 1353–1358. https://doi.org/10.1093/bioinformatics/bts163

Singh, T., Levine, A. P., Smith, P. J., Smith, A. M., Segal, A. W., & Barrett, J. C. (2015). Characterization of Expression Quantitative Trait Loci in the Human Colon. *Inflammatory Bowel Diseases*, *21*(2), 251–256. https://doi.org/10.1097/mib.0000000000000265

St Johnston, D. (2002). The art and design of genetic screens: Drosophila melanogaster. *Nature Reviews Genetics*, *3*(3), 176–188. https://doi.org/10.1038/nrg751

Stainier, D. Y. R. (2005). No Organ Left Behind: Tales of Gut Development and Evolution. *Science*, *307*(March), 1902–1905. https://doi.org/10.1126/science.1108709

Stanley, P. D., Ng'oma, E., O'Day, S., & King, E. G. (2017). Genetic dissection of nutrition-induced plasticity in insulin/insulin-like growth factor signaling and median life span in a Drosophila multiparent population. *Genetics*, *206*(2), 587–602. https://doi.org/10.1534/genetics.116.197780

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). Revigo summarizes and visualizes long lists of gene ontology terms. *PLoS ONE*, *6*(7). https://doi.org/10.1371/journal.pone.0021800

Suresh, R., & Mosser, D. M. (2013). Pattern recognition receptors in innate immunity, host defense, and immunopathology. *Advances in Physiology Education*, 284–291. https://doi.org/10.1152/advan.00058.2013

Tain, L. S., Sehlke, R., Jain, C., Chokkalingam, M., Nagaraj, N., Essers, P., … Partridge, L. (2017). A proteomic atlas of insulin signalling reveals tissue-specific mechanisms of longevity assurance. *Molecular Systems Biology*, *13*(9), 939. https://doi.org/10.15252/msb.20177663

Tan, F. H. P., & Azzam, G. (2017). Drosophila melanogaster: Deciphering Alzheimer's disease. *Malaysian Journal of Medical Sciences*, *24*(2), 6–20. https://doi.org/10.21315/mjms2016.24.2.2

Tanaka, K. (2003). The Origin of Macromolecule Ionization by Laser Irradiation (Nobel Lecture). *Angewandte Chemie International Edition*, *42*(33), 3860–3870. https://doi.org/10.1002/anie.200300585

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., & Yoshida, T. (1988). Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of-flight Mass Spectrometry. *Rapid Communications in Mass Spectrometry*, *2*(8), 151–153.

Tang, H. (2009). Regulation and function of the melanization reaction in Drosophila. *Fly*, *3*(1), 105–111. https://doi.org/10.4161/fly.3.1.7747

Teixeira, L. K., & Reed, S. I. (2013). Ubiquitin Ligases and Cell Cycle Control. *Annual Review of Biochemistry*, *82*(1), 387–414. https://doi.org/10.1146/annurev-biochem-060410-105307

Tewhey, R., Kotliar, D., Park, D. S., Liu, B., Winnicki, S., Reilly, S. K., … Sabeti, P. C. (2016). Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. *Cell*, *165*(6), 1519–1529. https://doi.org/10.1016/j.cell.2016.04.027

Thomson, J. J. (1913). *Rays Of Positive Electricity and Their Application to Chemical Analysis*. London: Longman's Green and Company.

Thurmond, J., Goodman, J. L., Strelets, V. B., Attrill, H., Gramates, L. S., Marygold, S. J., … Baker, P. (2018). FlyBase 2.0: the next generation. *Nucleic Acids Research*, *47*(October 2018), 759–765. https://doi.org/10.1093/nar/gky1003

Tolwinski, N. (2017). Introduction: Drosophila—A Model System for Developmental Biology. *Journal of Developmental Biology*, *5*(3), 9. https://doi.org/10.3390/jdb5030009

Turro, E., Su, S. Y., Gonçalves, Â., Coin, L. J. M., Richardson, S., & Lewin, A. (2011). Haplotype and isoform specific expression estimation using multi-mapping RNA-seq reads. *Genome Biology*, *12*(2), 1–15. https://doi.org/10.1186/gb-2011-12-2-r13

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., … Cox, J. (2016). The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, (June). https://doi.org/10.1038/nmeth.3901

Ugur, B., Chen, K., & Bellen, H. J. (2016). Drosophila tools and assays for the study of human diseases. *Disease Models & Mechanisms*, *9*(3), 235–244. https://doi.org/10.1242/dmm.023762

Välikangas, T., Suomi, T., & Elo, L. L. (2017). A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Briefings in Bioinformatics*, *19*(6), 1344–1355. https://doi.org/10.1093/bib/bbx054

Veenstra, J. A., Agricola, H., & Sellami, A. (2008). Regulatory peptides in fruit fly midgut. *Cell Tissue Research*, (334), 499–516. https://doi.org/10.1007/s00441-008-0708-3

Visscher, P. M., McEvoy, B., & Yang, J. (2010). From Galton to GWAS: quantitative genetics of human height. *Genetics Research*, *92*(5–6), 371–379. https://doi.org/10.1017/s0016672310000571

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Vlisidou, I., & Wood, W. (2015). Drosophila blood cells and their role in immune responses. *FEBS Journal*, *282*(8), 1368–1382. https://doi.org/10.1111/febs.13235

Vodovar, N., Vinals, M., Liehl, P., Basset, A., Degrouard, J., Spellman, P., … Lemaitre, B. (2005). Drosophila host defense after oral infection by an entomopathogenic Pseudomonas species. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(32), 11414–11419. https://doi.org/10.1073/pnas.0502240102

Walton, E. L. (2015). Lessons from the Drosophila microbiota. *Biomed J*, *38*(4), 271–274. https://doi.org/10.4103/2319-4170.162481

Wei, G., Sun, L., Li, R., Li, L., Xu, J., & Ma, F. (2018). Dynamic miRNA-mRNA regulations are essential for maintaining Drosophila immune homeostasis during Micrococcus luteus infection. *Developmental and Comparative Immunology*, *81*, 210–224. https://doi.org/10.1016/j.dci.2017.11.019

Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., & Ni, Y. (2018). Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Scientific Reports*, *8*(1), 663. https://doi.org/10.1038/s41598-017-19120-0

Wilusz, J. E. (2015). Controlling translation via modulation of tRNA levels. *Wiley Interdiscip Rev RNA*, *6*(4), 453–470. https://doi.org/10.1002/wrna.1287

Wright, S. (2006). Dobzhansky's Genetics of Natural Populations. *Evolution*, *36*(5), 1102. https://doi.org/10.2307/2408088

Wu, L., Candille, S. I., Choi, Y., Xie, D., Jiang, L., Li-Pook-Than, J., … Snyder, M. (2013).

Variation and genetic control of protein abundance in humans. *Nature*, *499*, 79–82.

Xiong, Y., & Yu, J. (2018). Modeling Parkinson's disease in Drosophila: What have we learned for dominant traits? *Frontiers in Neurology*, *9*(APR). https://doi.org/10.3389/fneur.2018.00228

Yamamoto, D., Jallon, J.-M., & Komatsu, A. (1997). Genetic Dissection of Sexual Behavior in Drosophila melanogaster. *Annual Review of Entomology*, *42*, 551–585. https://doi.org/10.1603/ice.2016.105582

Yamamoto, S., Jaiswal, M., Charng, W.-L., Gambin, T., Karaca, E., Mirzaa, G., … Bellen, H. J. (2014). A drosophila genetic resource of mutants to study mechanisms underlying human genetic diseases. *Cell*, *159*(1), 200–214. https://doi.org/10.1016/j.cell.2014.09.002

Zaidman-Rémy, A., Hervé, M., Poidevin, M., Pili-Floury, S., Kim, M. S., Blanot, D., … Lemaitre, B. (2006). The Drosophila Amidase PGRP-LB Modulates the Immune Response to Bacterial Infection. *Immunity*, *24*(4), 463–473. https://doi.org/10.1016/j.immuni.2006.02.012

Zasloff, M. (2002). Antimicrobial peptides of multicellular organisms. *Nature*, *415*(January), 389–395.

Zhang, G., Ueberheide, B. M., Waldemarson, S., Myung, S., Molloy, K., Eriksson, J., … Fenyö, D. (2010). Protein Quantitation Using Mass Spectrometry. *Methods in Molecular Biology*, *613*, 211–222. https://doi.org/10.1109/MCISE.1999.764211

Zhou, R., Silverman, N., Hong, M., Liao, D. S., Chung, Y., Chen, Z. J., & Maniatis, T. (2005). The role of ubiquitination in Drosophila innate immunity. *Journal of Biological Chemistry*, *280*(40), 34048–34055. https://doi.org/10.1074/jbc.M506655200

Zhou, Y., Goodenbour, J. M., Godley, L. A., Wickrema, A., & Pan, T. (2009). High levels of tRNA abundance and alteration of tRNA charging by bortezomib in multiple myeloma. *Biochemical and Biophysical Research Communications*, *385*(2), 160–164. https://doi.org/10.1038/mp.2011.182.doi

Zichner, T., Garfield, D. A., Rausch, T., Stütz, A. M., Cannavó, E., Braun, M., … Korbel, J. O. (2012). Impact of genomic structural variation in Drosophila melanogaster based on population-scale sequencing. *Genome Research*, *23*(3), 568–579. https://doi.org/10.1101/gr.142646.112

Zou, J., Hormozdiari, F., Jew, B., Ernst, J., Sul, J. H., & Eskin, E. (2018). Leveraging allele-specific expression to refine fine-mapping for eQTL studies. *BioRxiv*, 257279. https://doi.org/10.1101/257279

Curriculum vitae

MICHAEL FROCHAUX

Ch. De Cotsette 24, 1950 Sion · 0041797804019

michael.frochaux@gmail.com

My motivation comes from finding creative solutions in multidisciplinary projects in multicultural teams. Leader, analytical, flexible and independent. Strong interpersonal and communication skills, active in institutional organizations in representative positions.

PROFESSIONAL EXPERIENCE

01.03.2013 –
PHD STUDENTS, EPFL-SV-IBI-LSBG
PhD and many collaboration

EDUCATION

SEPTEMBER 2019
PHD, EPFL

OCTOBER 2013
MASTER, EPFL
Master in Life Science with a specialization in Molecular Medicine

SKILLS

- Programming (R, python)
- Excel
- Word

ACTIVITIES

Unihockey: I have been a player in the unihockey club of Sion and have been a member in its committee for more than 10 years, filling the position of vice-president, president, secretary, referee supervisor and junior team supervisor. I have also been a referee for more than 10 years.

X-wing miniature games: a highly tactical tabletop games I have been playing for more than 5 years. Several top results at big tournaments and captain of team Switzerland at the ETC.
When I have time, I like nothing more than reading a book, either in French or in English
French: native speaker
English: C1
German: B1

PUBLICATIONS

Mouchiroud, L., Sorrentino, V., Williams, E. G., Cornaglia, M., Frochaux, M. V., Lin, T., … Auwerx, J. (2016). The movement tracker: A flexible system for automated movement analysis in invertebrate model organisms. Current Protocols in Neuroscience, 2016(October), 8.37.1-8.37.21. http://doi.org/10.1002/cpns.17

Moullan, N., Mouchiroud, L., Wang, X., Ryu, D., Williams, E. G., Mottis, A., … Auwerx, J. (2015). Tetracyclines disturb mitochondrial function across eukaryotic models: A call for caution in biomedical research. Cell Reports, 10(10), 1681–1691. http://doi.org/10.1016/j.celrep.2015.02.034

Williams, E. G., Mouchiroud, L., Frochaux, M., Pandey, A., Andreux, P. A., Deplancke, B., & Auwerx, J. (2014). An Evolutionarily Conserved Role for the Aryl Hydrocarbon Receptor in the Regulation of Movement. PLoS Genetics, 10(9). http://doi.org/10.1371/journal.pgen.1004673

**Ce document a été imprimé au Centre d'impression EPFL,
imprimerie climatiquement neutre, certifiée myClimate.**