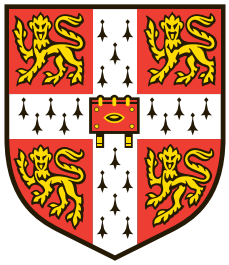
Immunity Against Malaria: an Atlas of the Mosquito Cellular Immune System at Single-Cell Resolution



Gianmarco Raddi

Wellcome Sanger Institute

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Clare Hall College September 2019

Acta est fabula, plaudite

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Gianmarco Raddi

September 2019

Gianmarco Raddi – Immunity Against Malaria: an Atlas of the Mosquito Cellular Immune System at Single-Cell Resolution

Abstract

Malaria is a deadly, worldwide disease, yearly responsible for 219 million cases and over four hundred thousand deaths[1]. The *Anopheles gambiae* species complex is the main African vector for the most virulent malaria parasite: *Plasmodium falciparum*[2]. Mosquitos are not mere bystanders however, and rely on both humoral and cellular innate immune divisions to defeat invading pathogens[2, 3]. These efforts are coordinated by hemocytes, the insect equivalent to vertebrate's white blood cells, circulating in the hemolymph within the insects' body cavity. Yet, hemocyte biology is largely unknown, mainly due to the low number and fragility of mosquito immune cells[4]. Here we dissect the *Anopheles* immune system under baseline and challenged conditions with single-cell RNA sequencing to identify previously unknown cell types, their gene signatures, and spatial-temporal localization in the mosquito. We profiled 5,292 individual *Anopheles* hemocytes 1,3 and 7 days after sugar-feeding, blood-feeding, or infection with *Plasmodium berghei,* as well as 3123 *A. aegypti* hemocytes. We identified 9 cell sub-types, including novel effector, phagocytic, and anti-microbial cell subtypes, in addition to dividing progenitor cells, validating the main cell types via correlative microscopy and morphology. And we described four lineages of hemocytes, showing them to be divided into two branches: oenocytoids and prohemocyte-granulocyte. We also found both blood-feeding and malaria infection to dramatically shift the composition of a mosquito's immune system, activating effector and proliferating cells at days 1 and 3 before returning to baseline by day 7. Conversely, human *P. falciparum* appears to inactivate an important local effector subtype. Our work is the first comprehensive transcriptomic study of a whole insect immune system. It demonstrates hemocytes are a dynamic, diverse class of insect cells which complexity far exceeds what is currently described in the literature. Our methods and results will hopefully serve as a resource for many entomologists, and could prove useful in developing novel vector control strategies. Our website will ease data access and provide an intuitive way to visualise hemocyte information: https://hemocytes.cellgeni.sanger.ac.uk/

Acknowledgements

I would like to thank Dr. Carolina Barillas-Mury and Dr. Oliver Billker for the opportunity to pursuit this research project. I am grateful for the camaraderie of my whole NIH and Wellcome Sanger Institute laboratories. I also want to thank Dr. Sarah Teichmann, whose group kindly adopted me when no more mosquito men and women walked the halls of Sanger. And special thanks to Dr. Carl Anderson, as well as Clare Hall college, for their unfailing support.

I am indebted to all my collaborators, without whom this work could have never been done. Especially so to Dr. Andrew Goldsborough, not simply the most supportive mentor a pesky PhD student could ever desire, but also a friend, unfailingly excited about any science I would bring forward, always ready and willing to discuss new exciting experiments and protocols. And to Mirjana, the bioinformatic sage who taught me all I know about scRNA-seq. You both saved my PhD, and by extension (academic) life. I also want to thank Dr. Levine, Dr. Liu, Dr. Geschwind, Rhea, Gene Roddenberry, Captain Kathryn Janeway, and all the crew of the Mobile Clinic Project at UCLA. You unknowingly all inspired me, and daily cemented my resolve to pursue science, as well as medicine.

And of course, how could I not mention Liam for his imperturbability in handling my very Italian way to order reagents – and apparently treacherous dislike of beer. Tomas, Ricardo and Mariana for being my adopted Portuguese family. Tom for his unfailing patience with my inability to swat down runaway mossies, to say nothing of his abysmal musical taste. Hannah for a cornucopia of clandestine London revelries. Etienne for teaching me a lesson in humility each time I set foot on a tennis court. Alexis for genuinely being the kindest human being I have ever met. Probably too much for her own good. And of course, my new medical family: the Clare gang for taking me in. Marc for the toughest foosball competition Cambridge has to offer. And Sofia for the endless poetry, politics, and global health conversations.

Hey hey hey. You thought you could get away from all this? Yes, I'm talking to you Sam, Eric, Anne. My American crew. Sam: the most deliciously bombastic New Yorker to have ever

walked the immortal city (and you are far from being alone). The word panache has been created to describe you. Eric! Oooooh where to start. Where to end? Should I bring into the picture Tomas? Or the honorary Italian? Or the master chef? Or the closeted Republican? Joking, joking. The only true closeted Republican can be nobody else but the funniest m*********r 'merica has ever created: Anne. Keep being boss darling dear.

Luke, Anna Wagner, grazie for the daily artistic inspiration. The evil sister scene is forever seared in my mind (I shudder as I type). Luke, Aline should have clearly chosen Daly, potion or no potion. And I CAN'T BELIEVE 'Engaged to So and So' would refuse you. How dare her! To be fair, my thanks go to all the people I have sung and danced with at Clare Hall, CUSC, the Opera Society, and the G&S society. Especially so to the sweetest latest surprise, French Marie! To share music with you all has been a joy such as one I never experienced. Onwards to even more loving get-outs and 6am after-show debauchery.

I am especially thankful to all my Sanger friends, the better part of my adventure in Cambridge. One could easily say, there would be have been no PhD without you. In no particular order (any identifiable patterns are but the product of a beholder's own oddities): C-Dawg, mah BB, da Russian, Velisludka, Little Lil, Ferdinand. Sanger shalt never know the same level of sass. Oh how the breakfast table rang of inappropriateness, I look forward to seeing you all in jail, or hell, whatever comes first. Same goes for all my volleyball friends, the infamous Signorelli gang, regularly terrorizing the Grafton Centre each and every Sunday morning, cornetto in one hand, espresso or cioccolata calda in the other. Not to mention our glorious weekly Franco Man… S&Cs! Again in no ORDAH: Lukas (and his supremely better half Katrin), Marta, Antonella, Max, Nico, Stefi, Jerome, Gia, and many more. Ciao a tutti! Auf wiedersehen!

And Melinka, my fair Baccarid, you have been part of it all. The most talented and hard-working scientist I know, and a role model in many respects. I am lucky to have known you, and will forever treasure your frenzy on the dancefloor, your calming presence, and your enchanting arias. Merci de m'avoir inspiré à explorer le chant, et l'opéra. Ciao ciao!

But most importantly of all, grazie mille a tutti coloro che hanno sempre, immancabilmente creduto in me, nonostante le tendenze Cubistiche che regolarmente venivano dimostrate dallo scrittore. Whisky, Socri, Zia Cri, Zio Franco, Marcellina, Max, Freccia, Stefi & Eta, Daddy & Mutti, vi amo e senza di Voi, senza il Vostro amore totale, insensato, infito per questo ragazzino del Villaggio, non sarei mai riuscito a fare nulla. Senza il Vostro Affetto, nessuna empatia sarebbe sorta; senza l' Educazione, nessun compasso morale; senza la Cucina, nessun buon gusto. Questo dottorato e' soprattutto Vostro.

"Looking up at the stars, I know quite well
That, for all they care, I can go to hell,
But on earth indifference is the least
We have to dread from man or beast.

How should we like it were stars to burn
With a passion for us we could not return?
If equal affection cannot be,
Let the more loving one be me.

Admirer as I think I am
Of stars that do not give a damn,
I must not, now I see that, say
I missed one terribly all day.

Were all stars to disappear or die,
I should learn to look at an empty sky
And feel its total dark sublime,
Though this might take me a little time."

W.H. Auden

Table of contents

List of figures

List of tables

Chapter I

Introduction

“By heaven, I'll make a ghost of him that lets me”
— William Shakespeare, Hamlet

Malaria is a deadly global disease. Every year the *Plasmodium* parasites are responsible for 219 million cases of malaria and over four hundred thousand deaths, mostly vulnerable young African children under 5 years old and pregnant mothers [1]. After decades of steady progress malaria incidence is no longer declining. According to the 2018 World Malaria Report, the number of malaria cases increased from 217 million cases in 2016 to 219 million cases in 2017, resulting in 435,000 deaths [1]. Furthermore, in 2017, 3.5 million more malaria cases were recorded in the 10 African countries with the highest disease burden. In the Americas, some countries are also seeing large increases in prevalence, namely Brazil, Nicaragua, and Venezuela [1].

Most malaria-related deaths (93%) are concentrated in Africa; particularly so (76%) in the 17 sub-Saharan countries with the highest incidence of disease. The stark death toll is due to the high prevalence of *Plasmodium falciparum* malaria – the most virulent form of the disease – in the African continent. Here, *P. falciparum* causes 99.7% of malaria cases, while outside of Africa *Plasmodium vivax* is the most common infection. Other *Plasmodium* species can also cause human disease, such as *P. malariae*, *P. knowlesi* and *P. ovale*, but their prevalence is lower [5]. In all cases, *Plasmodium* parasites are transmitted to humans through the bite of an infected mosquito of the *Anopheles* genus. *Anopheles gambiae* and *coluzzii* (former *A. gambiae* molecular “M form”[6]) are the main vectors of *P. falciparum* malaria in Africa. However, over 30 *Anopheles* species are major disease vectors in other geographic regions.

If left untreated, malaria is a chronic and often deadly infection, as the human immune system is unable to achieve sterile immunity. Only after years of exposure and repeated bouts of infection is the immune system able to contain the parasite, resulting in chronic

asymptomatic infections. During the first 5 years of life, malaria infections often cause symptoms such as fever and anemia, and a proportion of children (1-2%) develop life-threatening severe disease and can die[7]. Mild or asymptomatic disease is not observed until early adolescence, but febrile episodes can still occur. Only as they reach adulthood, do people transition to a chronic state where disease symptoms rarely occur [2, 8]. Because sterilizing immunity never develops, adults maintain asymptomatic blood-stage malaria infections throughout their lives. They also become asymptomatic gametocyte carriers, infecting mosquitoes in their communities and maintaining the transmission cycle [9]. Malaria immunity in humans is complex, involving early development of protection from severe disease, followed by asymptomatic uncomplicated disease, but rarely, if ever, involving complete resistance to infection. Similarly, complex innate immune responses to *Plasmodium* occur in *Anopheles* mosquitoes, the outcome of which determines disease transmission.

Recent calls for malaria eradication have led to considerable strides in controlling this deadly disease[10], but we are far from defeating it. Importantly, we might not even yet have the right tools for such a goal, as the first approved vaccine (RTS,S/AS01) only provides partial (32-41%) protection [11], and the current arsenal of anti-malarial drugs is becoming less effective as *Plasmodium* resistance spreads[12]. Two vector-control strategies, insecticide-treated nets (ITN) and indoor residual spraying (IRS), have been key for the successful reduction of the burden of malaria in the last ten years, but these gains are in peril as mosquitoes develop insecticide resistance[13].

The reproductive rate (R_0) is defined as the number of new infection one case can generate, on average, over the course of its infectious period. It has been recently calculated that the R_0 for malaria ranges from 1 to over 3,000, depending on location, parasite species, populations and vectors [14]. To put the number into perspective, the flu has an R_0 of 10. By definition, for malaria to be eradicated, R_0 has to drop below 1. From a public health perspective, that means no “one-size-fit-all” approach can work, while from a researcher stand-point, it indicates that vaccines alone might be insufficient to eradicate malaria in areas of high

transmission. New control strategies that reduce the rate of re-infection, such as transmission-blocking vaccines[15], or the use of *Metarhizium* fungi that rapidly kill insecticide-resistant mosquitoes[16], will be required to achieve eradication. Crucial to all such eradication efforts is a better understanding of the determinants of malaria transmission by mosquitoes. There is ample evidence that mosquitoes have the potential to mount effective anti-plasmodial immune responses[2]. The mosquito relies on epithelial, humoral and cellular innate immune responses, coordinated by the hemocytes, the equivalent of the human white blood cells in insects [17]. My thesis project involves the development of an atlas of the mosquito cellular immune system at single-cell resolution as it responds to *Plasmodium* infection.

Introduction

1 The malaria parasite

"There's more beauty in truth, even if it is dreadful beauty"
— John Steinbeck, East of Eden

The malaria parasite and mankind are old foes. A recent study has shown the parasite co-evolved with gorilla parasites and went through an evolutionary bottleneck when it gained the ability to infect human hosts between 40,000 and 60,000 years ago. Human-adapted *P. falciparum* thrived when the human population exploded 5,000 years ago thanks to advances in farming [18]. As such *P. falciparum* infection has exerted a strong selective pressure on human populations worldwide, perhaps more so than any other pathogen [19]. No example is more widely studied in medical schools all around the world than sickle cell anemia. Recent medical advances have transformed the life of sickle-cell disease patients, whose life expectancy keeps rising and was estimated at 57 years in the US (2014) and 60 years in the UK (2016) [20, 21]. However, the homozygous hemoglobin S (HbS) variant was historically uniformly lethal in children, and yet was still maintained in the population at a frequency of around 15%, thanks to partial protection against severe malaria in heterozygotes [22]. But the parasite also had to adapt to the mosquito vector. For example, the *Plasmodium* surface protein Pfs47 allows the parasite to evade the mosquito innate immune system. It is thought that for a parasite to be transmitted, it requires a Pfs47 haplotype compatible with the Pfs47 receptor of the mosquito. As such, Pfs47 functions as a molecular "key" that turns off mosquito immunity through interaction with a receptor ("the lock") specific for each vector species. Only the right "lock and key" combination allows parasites to survive in the mosquito and propagate [23]. The parasite's life cycle is exceedingly complex, in both its human and mosquito hosts. Hence, many more such host-parasite immune interactions surely remain to be discovered. One of the most widely used animal models to study host-parasite interactions, as well as *Plasmodium* life cycle and development in the mosquitoes is *P. berghei*, a malaria parasite that infects mice. I used this experimental model system extensively in my PhD thesis to investigate the transcriptional response of the mosquito hemocytes to *Plasmodium* infection.

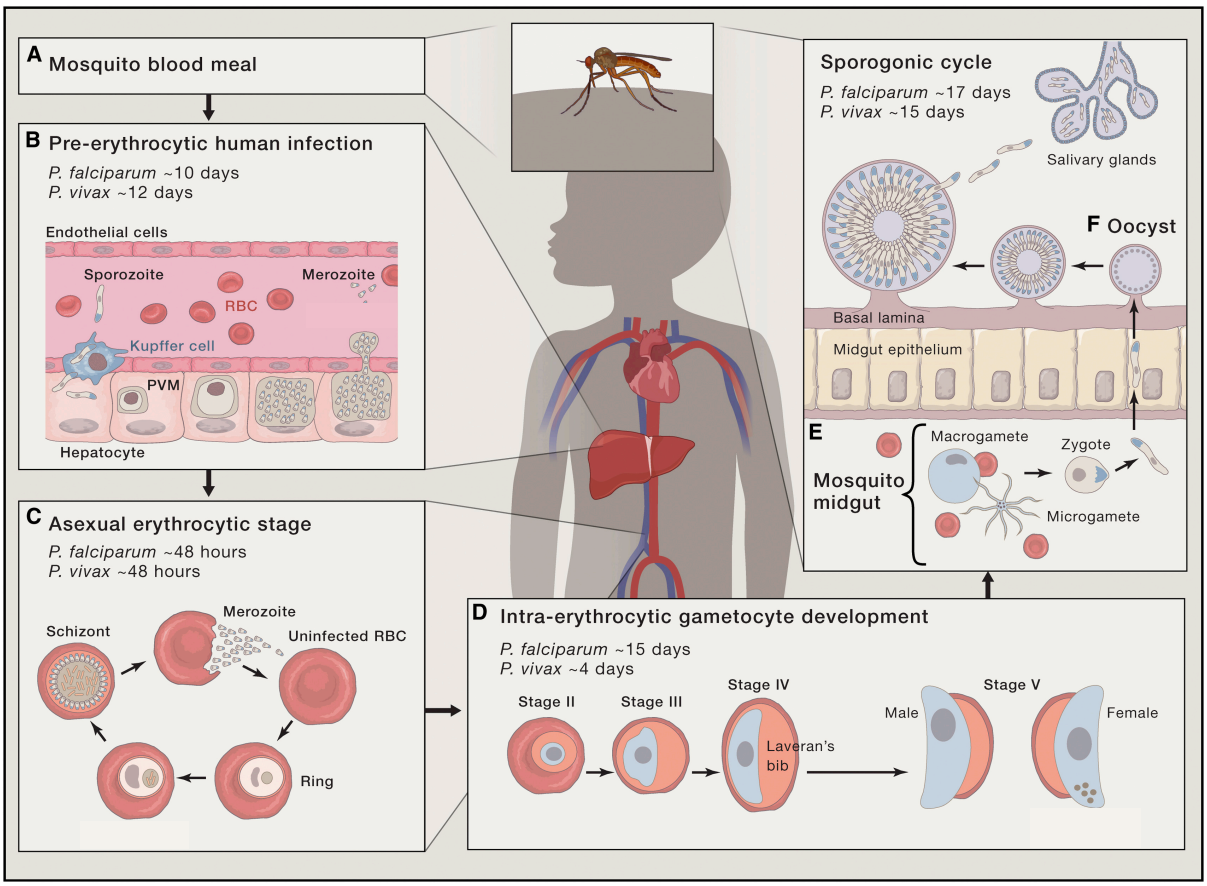


Figure I.1: ***Plasmodium falciparum*** **life cycle.** (A) Human malaria infection starts after an infected mosquito feeds and releases sporozoites from the salivary gland. (B) Some sporozoites escape the dermis, reach a blood vessel, travel to the liver through sinusoids or Kupffer cells and infect hepatocytes where they form a parasitophorous vacuole membrane (PVM) and undergo schizogony to released thousands of merozoites. (C) These merozoites travel in the blood and infect red blood cells, where multiple cycle of asexual reproduction (schizony) will occur before another burst and the repeat of the cycle. (D) Some merozoites are activated to differentiate to sexual gametocytes. (E) After bone marrow sequestration and maturation mosquitoes ingest gametocytes. In the midgut of mosquitoes, male and female gametocytes mate and form a zygote. In 24 hours the resulting motile ookinete penetrates the mosquito midgut epithelium and encysts. (F) In the oocyst asexual sporozoites replicate, are released in hemocoel, and colonize the salivary gland. Figure adapted from Cowman *et al*. [24]

Introduction

a. Malaria life cycle in humans

Before exploring in detail the mosquito phase of *Plasmodium* life cycle, it will be useful to review the human stages of parasite development. After a mosquito bite, sporozoites that had already colonized the salivary gland are injected into the human dermis. Some are able to survive local immune responses and move into blood vessels, travelling to the liver. Here they cross the hepatic sinusoidal barrier (fenestrated endothelial cells and Kupffer cells - resident macrophages), thanks to the action of SPECT, SPECT2, CelTOS, PL, and GEST proteins [24–28]. Sporozoites activate by binding higher sulfated forms of heparin sulfate proteoglycans (HSPGs), tetraspanin CD81, and scavenger receptor B1 (SR-B1) on hepatocytes [29, 30]. As circumsporozoite proteins (CSP) bind HSPGs, hepatocyte invasion commences [31]. Over the next 2-10 days sporozoites will morph into liver-stage (LS) schizonts, an exo-erythrocytic form (EEF) stage in which the parasite multiplies, eventually releasing over 40,000 merozoites per infected hepatocyte into the circulatory system [32].

Once released, merozoites infect circulating red blood cells in a three-step process [33]. The first and least understood is pre-invasion, in which low affinity binding between merozoites and erythrocytes orient the apical end of parasites towards red blood cells (RBCs) [34]. Then, specific binding mediated by erythrocyte binding-like proteins (EBA) and reticulocyte-binding protein homologs (PfRh) leads to actomyosin-driven host cell deformation and erythrocyte invasion [33, 35]. A PfRh5-PfRipr (Rh5-interacting protein) – CyRPA (cysteine-rich protective antigen) – basigin complex mediates the close interaction between erythrocyte and merozoite membranes, leading to microneme secretion and Ca^{2+} influx inside the red blood cells [36–39]. Merozoites are then irreversibly linked to erythrocytes through AMA1-RON tight junction complexes [40]. These are moving junctions, propelling the merozoites inside red blood cells just as rhoptry contents are released, which form the parasitophorous vacuole membrane (PVM) around the merozoites [41]. As the PVM seals, cytosolic water losses within host cells cause echinocytosis. Over the next 48 hours the

parasites take advantage of the established nutrient-rich cellular milieu to rapidly divide and produce 16 to 32 merozoites each, which then egress as they destroy the RBCs [24].

During these rounds of cellular replication (schizogony), a small proportion of *Plasmodium* parasites will differentiate into sexual forms, a required step for successful transmission to mosquito vectors. Male and female gametocyte differentiation is not fully understood, but is regulated by the master switch AP2-G [42] following sensing of environmental signals such as high parasitemia or presence of chloroquine in the blood stream. Gametocyte development lasts 11 days, during which time committed but not yet mature gametocytes hide sequestered in the bone marrow to avoid splenic clearance. Following development, mature stage V gametocytes are taken up by feeding mosquitoes to commence the mosquito life cycle [43] [Fig. I.2].

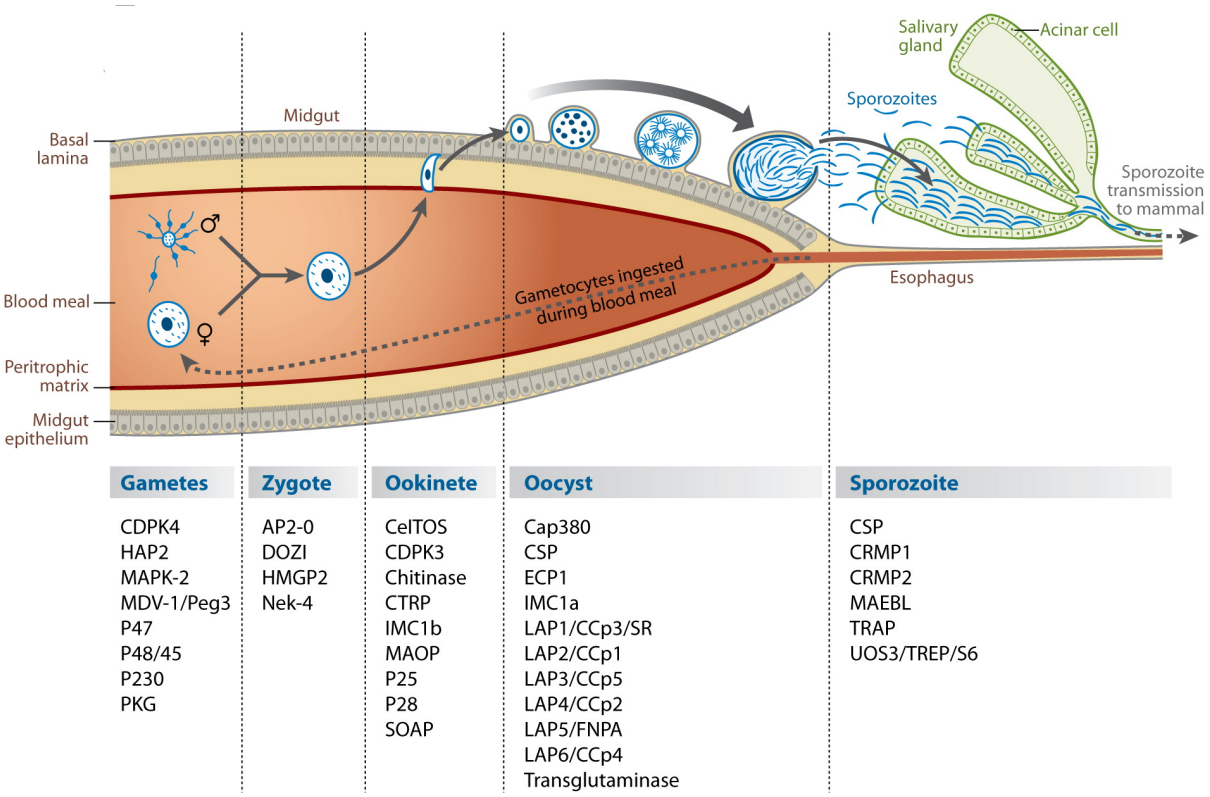


Figure I.2: Detailed *Plasmodium* life cycle in the mosquito and key parasite proteins. (A) *Plasmodium* gametocytes are first ingested when mosquitoes take an infected human blood meal. Gametes, both female (macrogametes) and male (microgametes) mature from gametocytes to form a zygote (B). After meiosis the zygote morphs into the motile, infective ookinete (C) which is able to penetrate through the mosquito midgut. After egressing from the midgut *Plasmodium* ookinetes encyst on the basal end, becoming sessile (D). After 10-14 days of growth and mitotic divisions, thousands of motile sporozoites are released into the mosquito circulation, travelling in the hemolymph until some reach the mosquito salivary glands (E). Here sporozoites attach to the basal side of salivary gland acinar cells, travel through them, and enter the ducts, where they await the next mosquito bite to continue the life cycle. Key proteins at each step are listed above. Figure adapted from Aly *et al.* [44]

b. *Plasmodium* life cycle in mosquitoes

Parasites ingested during a blood meal quickly undergo sexual reproduction in the mosquito midgut [44]. Gametogenesis starts the moment gametocytes are ingested by feeding mosquitos, leading to the formation of mature male and female gametes [45]. Gametogenesis is mediated by essential environmental signals such as a 5°C drop in temperature, the rise of extracellular pH (from 7.2 to 8), and xanthurenic acid (XA) sensing [46–50]. XA – a byproduct of mosquito metabolism – activates guanylyl cyclase, leading to increased second messenger cGMP production and protein kinase G (PKG) activation[51, 52]. In addition, XA increases inositol-(1,4,5)-trisphosphate (IP_3) production by activating phospholipase C, causing the opening of Ca^{2+} channels [53–55]. Heightened intracellular Ca^{2+} releases translational repression in male and female gametocytes by activating Ca^{2+}-dependent protein kinase 1 (CDPK1) [56]. Gametocyte activation is rapid, and within 15 minutes gametocytes egress from red blood cells by rupturing first the PVM and then the erythrocytic membrane (EM), steps respectively associated with osmiophilic bodies and egress vesicles [50]. The former is mediated by Pg377, MDV-1/Peg3 and GEST [57–59] while egress vesicles release perforin, which breaks the EM to release fertile gametes [60]. Activated microgametocytes undergo three rounds of replication, becoming octaploid, and producing eight flagellar mature microgametes by mitosis (exflagellation)[61]. Exflagellating microgametes adhere to nearby red blood cells, hiding within rosettes before detaching from the residual body, searching for macrogametes [45]. When a partner is found Pfs47, Pfs48/45, and Pfs230 proteins form complexes responsible for the binding of microgametes and macrogametes, commencing fertilization [62–64]. First, the plasma membranes of the two gametes fuse. The axoneme and male nucleus then enter the female cytoplasm, mediated by HAP2. Finally, nucleus fusion ensues, followed by meiosis and the production of a tetraploid zygote (as mediated by NIMA-related kinases Nek-2 and Nek-4)[65–67].

Next, the zygote morphs into a motile ookinete able to colonize mosquitoes. After fertilization ninety-one proteins were found to be specifically expressed, with silencing of

paternal alleles in both zygotes and ookinetes. These changes are orchestrated by the transcription factor AP-2O, the master regulator of ookinete development and motility. Furthermore, AP-2O also plays a key role in penetrating the mosquito midgut epithelium and protecting ookinetes from immune defenses [56, 68–70]. Among the upregulated genes are secreted proteins such as perforins (PPLP3, 4 and 5), PSOP1, 2, 6, 7 and 12, and SOAP, as well as WARP, POS1-10, and P25, P28 – all potential or actual targets of transmission-blocking vaccines [56, 70, 71]. Ookinete maturation completes between 19 and 36 hours after an infectious blood-meal, after which ookinetes exit the gut [72–74]. Ookinete motility is regulated by PKG and CDPK3 activity, as well as cGMP and Ca^{2+} levels [75–78]. In order to successfully infect a mosquito ookinetes must first penetrate the peritrophic membrane (PM) – a chitin structure that functions to protect mosquitoes from bacteria and gross food [79]. To break through the PM *Plasmodium* ookinetes produce a chitinase that is able to hydrolyse the chitin [80–82]. Traversal of midgut epithelial cells is mediated by CTRP, a protein secreted by the ookinete to form a bridge between the midgut epithelium and the actin/myosin motor of the ookinete [83–85]. Three perforins (PPLP3-5), SOAP, WARP, MAOP and CelTOS are all microneme proteins required to breach the epithelial membrane [70, 85–91]. Once ookinetes have crossed the midgut epithelium they are surrounded by the laminin and collagen of the basal lamina. The interaction of *Plasmodium* with laminin turns ookinetes sessile, which encysts on the basal side of the midgut epithelium, triggered by the proteins P25 and P28, with help from CTRP and SOAP [71, 92]. Cell transversal is a bottleneck, and only a few ookinetes are successful in invading the mosquito midgut [88, 93, 94].

Oocyst development and maturation lasts between 10 and 12 days, and is the only stage of the life cycle where the parasite is extracellular for an extended period of time. And yet, little is known of host-oocyst interactions. Nutrients flow through the oocyst capsule, formed by an outer layer of thick mosquito laminin, parasite transglutaminase, Cap380 (oocyst capsule protein)[95], and P25/P28, and an inner oocyst plasma membrane containing Cap93[96] and circumsporozoite protein (CSP), a GPI-anchored protein[97]. Oocysts grow to 50-60 μm in diameter, forming thousands of sporozoites after multiple rounds of mitotic divisions, mediated

by LAPs (LCCL/lectin adhesive-like proteins) expression [98]. CSP is essential in building syncytial lobes - called sporoblasts, coordinating the localization of microtubule organizing centers (MTOC) underneath sporoblast membranes to make mature sporozoites[99, 100].

Once sporozoites reach maturity, they egress from oocysts, in a process that involves digestion of the capsule mediated by a cysteine protease, ECP1 (egress cysteine protease 1)[101, 102]. CSPs and a hypothetical oocyst protein also have important roles in sporozoite release[103, 104]. Sporozoites in the hemocoel are then carried to all tissues in the mosquito body by circulatory flow. Some are deposited to the basal lamina of the salivary gland, where CSP again plays a role in attachment [105, 106]. Thrombospondin-related anonymous protein (TRAP) is essential for attachment and invasion[102, 107], binding to saglin, a mosquito receptor in the distal lobes of the salivary gland [108, 109]. Gliding motility and the actin-myosin motor are also involved in invasion, mediated by TRAP and TREP (TRAP-related protein), as well as cysteine repeat modular proteins (CRMP1 and 2) and MAEBL [110–112]. Cellular invasion mechanisms are largely conserved between human and mosquito life stages of the parasite. However, while MAEBL-deficient sporozoites can still invade human host cells, TRAP-deficient parasites cannot. In fact, acinar cells in the salivary gland are invaded by a slightly different mechanism than the parasitophorous vacuole (PV) involved in liver and blood-stage invasion. Rather, invasion happens through a specific vacuolar membrane produced by the host cell [113]. How the sporozoite is able to induce vacuole formation in the salivary gland epithelial cells is unknown.

2 Mosquito immune responses

"The world is, of course, nothing but our conception of it"
— Anton Chekhov

Throughout the life of *Plasmodium*, the mosquito is far from being a passive vector. There are active interactions between parasites and the immune system at each step of the life cycle, especially when *Plasmodium* is extra-cellular. In fact, insects are constantly exposed to a wide variety of micro-organisms and pathogens seeking to exploit the host mosquito for their own reproductive goals. Viruses, fungi, bacteria, protozoans, and nematodes all invade and infect mosquitoes [114–118]. Some pathogens are able to penetrate through the external exoskeleton of mosquitoes, formed by hydrophobic chitin, which also lines the foregut, hindgut, and tracheas. They accomplish invasion by degrading the cuticle [119, 120]. Other pathogens enter mosquitoes through the digestive tract, overcoming physical barriers such as pharyngeal armatures and the chitinous peritrophic matrix, as well as digestive enzymes, local microbiota, and a hostile pH. Some pathogens evolved mechanisms to penetrate through these defensive mechanisms to reach the hemocoel (blood) of the mosquitoes and replicate, while others remain within the gut itself. Mosquitoes have however developed sophisticated immune mechanisms to fight off and control these pathogens [117, 119, 121–123]. Mosquito hemocytes, the equivalent of human white blood cells, coordinate both cellular and humoral immune responses. Humoral immune responses are mediated by molecules that are secreted into the circulating mosquito hemolymph (equivalent to serum in vertebrates) by hemocytes, fat body cells or epithelial cells lining the haemocoel, such as midgut and salivary gland cells). For example, pattern recognition receptors (PPRs), phenoloxidase cascade components, antimicrobial peptides, and elements of the complement-like system are all key mediators of the mosquito humoral response [3, 124–127]. The different components of the immune system are all interconnected, crafting an exceedingly complex and well-coordinated immunological network able to kill pathogens by a variety of effector mechanisms [Fig. I.3].

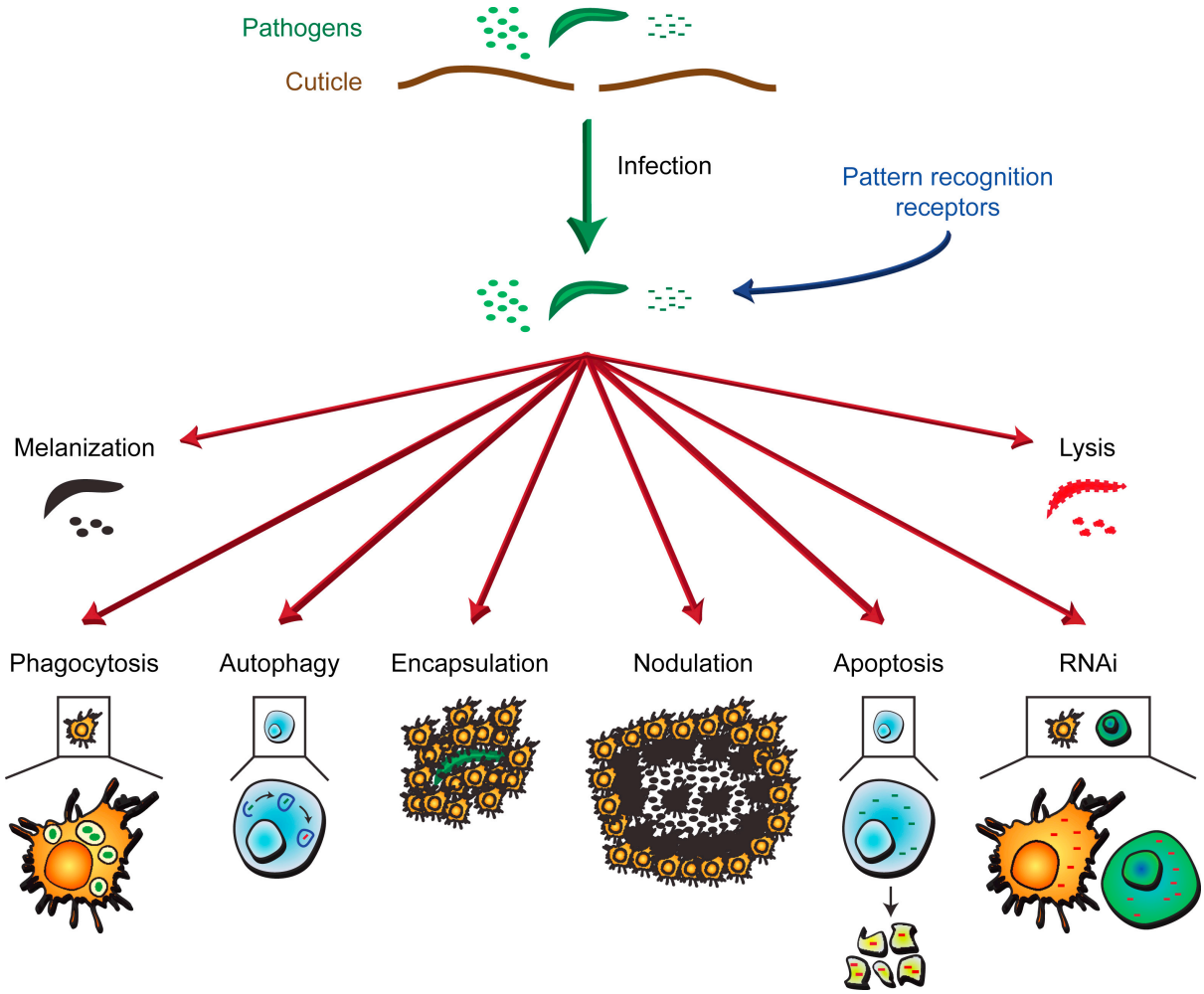


Figure I.3: Mechanisms of immune killing. Mosquitoes kill pathogens by melanisation, lysis, phagocytosis, autophagy, encapsulation, nodulation, apoptosis, and RNA interference. Adapted from Hillyer *et al.* [117]

a. Humoral immunity

Mosquitoes lack antibodies, but can activate highly effective humoral mechanisms to control infection. For instance, antimicrobial peptides (AMPs) – such as defensins, cecropins, gambicin, attacin and holotricin – are small charged molecules that are secreted into the hemolymph, with strong anti-bacterial or anti-fungal effects [128]. The composition and spatial expression of the "cocktail" of antimicrobial peptides secreted in response to an immune

challenge can differ widely among different mosquito species. For example, defensins can reach a concentration of 45 μM in *Aedes* [129], while in *Anopheles* they reach a maximum of 1-5 μM [130]. Furthermore, in *An. gambiae* antimicrobial peptides are produced at higher concentrations in the anterior midgut, and indeed *Plasmodium* oocysts colonise the posterior midgut[131]. Conversely, heightened AMP production can reduce parasite load. For instance, cecropin A (CecA) was shown to lower *P. berghei* oocyst counts in transgenic *An. gambiae* overexpressing CecA under the control of the *Aedes* carboxypeptidase promoter [132].

Melanisation is another powerful humoral mechanism to control infection. It is a biochemically conserved pathway to produce eumelanin from tyrosine and 5,6-dihydroxyindole (DHI) catalyzed by a cascade of prophenoloxidases (PPOs). The PPO enzymatic cascade leads to killing both by starving the invading pathogen of nutrients – walling it off from the rest of the body – as well as through the direct toxic effects of chemical byproducts. Melanisation is also involved in cuticle hardening and wound healing. Furthermore, it causes hemocytes to aggregate – an immune response akin to human granulomas, called in mosquitoes nodulation or encapsulation [Fig. I.4][128, 133–136]. The melanisation pathway begins with PRR sensing (C-type lectins, Gram-negative bacteria-binding proteins and beta-1,3 glucan recognition proteins), followed by a serine protease cascade leading to the activation of prophenoloxidases (mostly expressed by oenocytoids, a hemocyte subtype [136]). PPOs in turns activate melanin production by phenoloxidase. Melanisation is tightly regulated by serpins and C-type lectins. Similarly to human clotting, excessive activation would be deadly to the mosquito as widespread melanisation would damage the mosquito organs [133, 137, 138].

The complement-like pathway is one other crucial humoral effector mechanism, resulting in deposition of thioester-containing protein 1 (TEP1), a C3-like opsonin, on the surface of the microbe [139, 140]. Another four important proteins of the pathway are: two leucine-rich repeat proteins (APL1C and LRIM1), which stabilize TEP1 in circulation, and two clip domain serine protease homologs, (SPCLIP1 and CLIPA2), that modulate TEP1 activation. CLIPA2 is a negative regulator of TEP1, while SPCLIP1 promotes TEP1 activation.

It has been proposed that, following binding of TEP1 and SPCLIP1 recruitment and activation onto the pathogen surface, an endogenous TEP1 convertase is also deposited that further propagates local activation and binding of TEP1 [137, 141, 141–144].

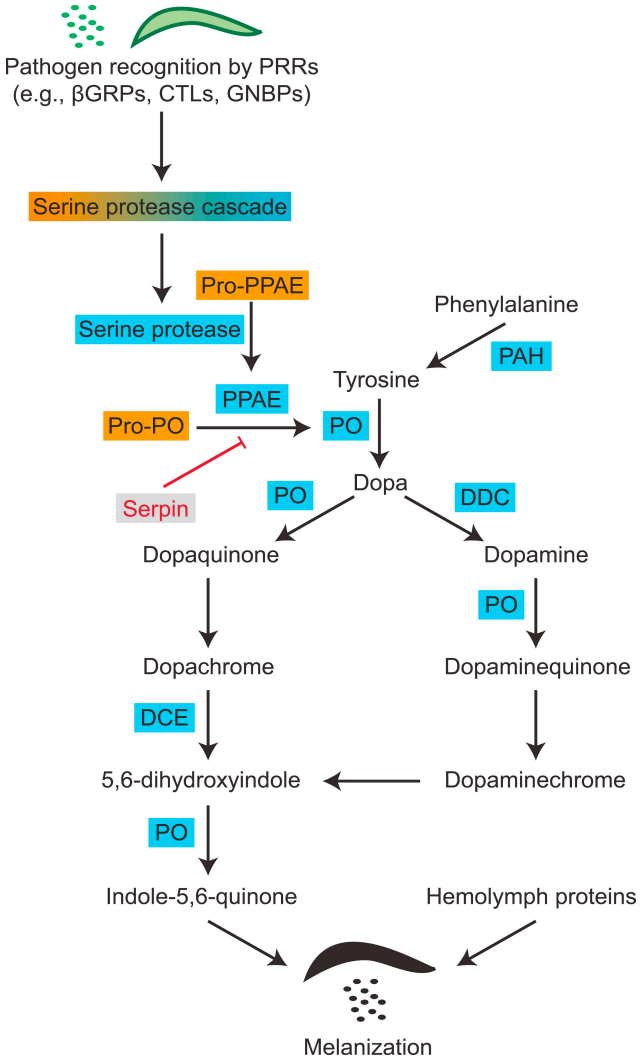


Figure I.4: Melanisation pathway. Abbreviations: PRR, pattern recognition receptor; βGRP, β-1,3 glucan recognition protein; CTL, C-type lectin; GNBP, Gram(−) binding protein; PPAE, phenoloxidase activating enzyme; PAH, phenylalanine hydroxylase; PO, phenoloxidase; DDC, dopa decarboxylase; DCE, dopachrome conversion enzyme. Adapted from Hillyer et al. [117].

b. Cellular immunity

Hemocytes are the primary immune cells circulating in the mosquito hemolymph. Mosquitoes have no antibodies or canonical adaptive immunity and entirely rely on innate immunity. Cellular responses are quick, with direct interactions between pathogens and the immune cells that include phagocytosis, encapsulation, and nodulation [117, 118, 145]. Experiments returned wildly discordant estimates of the total number of hemocytes in a mosquito, mostly due to technical differences in the collection methodology employed. The scientific community agrees that between 2,000 and 10,000 hemocytes patrol a mosquito, although only a fraction is motile (~ 500 - 2,000) and numbers vary considerably with blood-feeding and infection [4, 118, 146]. After morphological, enzymatic, and some functional characterization the consensus is that three main hemocyte subtypes exist: granulocytes, oenocytoids, and prohemocytes [Fig. I.5]. Of these, the vast majority are small prohemocytes (60-70%), followed by oenocytoids (20-30%) and granulocytes (1-10%), although estimates again vary considerably [4, 17, 17, 118]. Because the classification is largely morphological, subjective differences in interpretation and methodology are inevitable. Granulocytes are the main effector phagocytic cells in the mosquito, expressing AMPs, complement-pathway components, and low-level PO. Oenocytoids contain PPOs and POs at much higher levels. Prohemocytes are still a mystery. Originally thought to be progenitor cells, they have recently been shown to possess phagocytic capabilities, and are hypothesized to arise from asymmetric cell division of granulocytes [146].

While hemocytes remain in the hemocoel, and do not come in direct contact with the microbiome in the midgut lumen, transient bacteremia following blood feeding is thought to activate hemocyte replication after a blood meal [147–149]. A bacterium, yeast, fungus, or malaria parasite in the body cavity of a mosquito is usually quickly tagged, identified, and ingested by phagocytic hemocytes. Furthermore, hemocytes have been shown to aggregate around bacteria and form nodules [150]. Worms, fungi, or parasites become surrounded by melanocytic capsules [151–153]. In *Aedes aegypti* mosquitoes, hemocytes are also thought to play an important role in the systemic dissemination of arboviruses such as Sindbis or dengue virus [154, 155]. Hemocytes release microvesicles at sites of *Plasmodium* ookinete midgut

invasion that reduce parasite survival by promoting local activation of the complement-like system [156], thus coordinating epithelial and humoral antiplasmodial immune mechanisms to achieve an integrated and effective response.

Very little was known of hemocyte development in mosquitoes, except that blood-feeding increases their numbers[147]. In Drosophila, hematopoiesis is thought to occur in three waves: embryonic, larval, and lymph gland[157]. The first two are thought to be responsible for routine phagocytic and immunological functions of mosquitoes, whereas lymph gland hemocytes arise from synchronous differentiation of progenitors hemocytes within the gland following immune and environmental challenges. Crucially, after hemocytes differentiate these lymph glands disintegrate before adulthood[157]. Lymph glands have not been observed in mosquitoes.

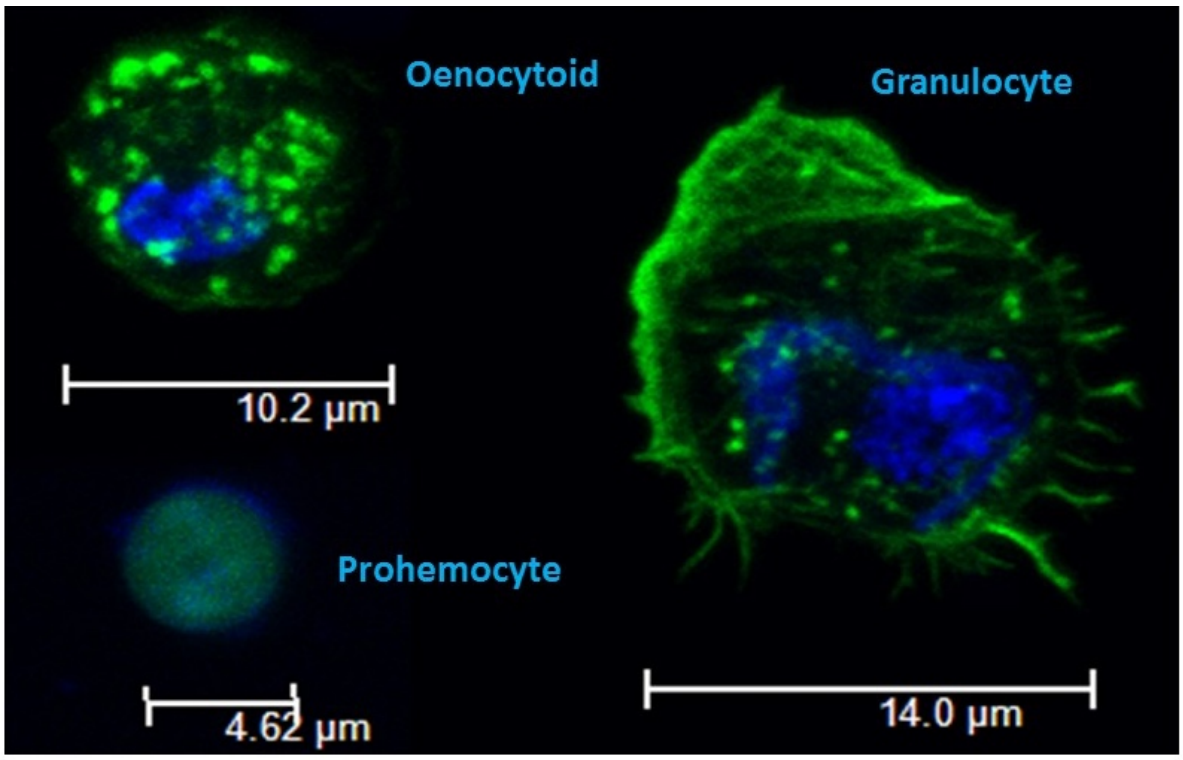


Figure I.5 Hemocyte subtypes. Average diameter and representative images of the three morphological subtypes of hemocytes in *A. gambiae*. Personal communication from Jose Luis Ramirez.

3 Specific *Anopheles* immune responses to *Plasmodium*

a. Midgut epithelial defenses

Mosquitoes ingest a very large, protein-rich meal during blood feeding which has to be digested to meet the nutritional needs of developing oocytes. At the same time, mosquitoes build a protective peritrophic matrix (PM), an acellular/semi-permeable layer of chitin polymers, that surrounds the blood meal and prevents direct contact between the microbiota in the lumen and the gut epithelium. The PM is also an important barrier against potential pathogenic organisms. At the end of the digestive process the PM matrix sack – containing remnants of the digested blood meal – is excreted from the mosquito. A new matrix will be formed when the mosquito ingests the next blood meal [158]. Mosquitoes also secrete mucin in the ectoperitrophic space between midgut epithelium and the PM, and actively modulate the permeability of this mucous layer through the activity of an immune-modulatory peroxidase (IMPer)/dual oxidase (Duox) system that catalyzes dityrosine cross-linking [2, 159, 160]. IMPer is secreted into the ectoperitrophic space, but is only active when immune elicitors activate Duox, which generates hydrogen peroxide a substrate required for IMPer to catalyze the formation of the dityrosine network. The end result is a dynamic modulation of the interaction between the immune elicitors released by gut flora and the gut epithelium following a blood meal, that allows the bacterial flora to survive by preventing constant activation of antibacterial immunity. This system also benefits *Plasmodium* parasites, because it allows them to develop within the midgut lumen without activating nitric oxide synthase (NOS) expression in epithelial cells. If this barrier is disrupted by silencing IMPer, mosquitoes mount a much stronger epithelial nitration response that eliminates *Plasmodium* [159, 161, 162].

b. Reactive oxygen/nitrogen species and complement-like defenses

Plasmodium ookinetes must traverse the midgut epithelium to complete their development in the mosquito. In doing so, they breach the peritrophic matrix, allowing the microbiota to come in direct contact with epithelial cells and cause irreversible damage as they invade midgut cells. Invaded mosquito epithelial cells express high levels of NOS, a response which is necessary,

but not sufficient to mount an effective response against *Plasmodium*. Specifically, increased NOS leads to nitric oxide production [161], which is unstable and is thought to rapidly convert to nitrite, a more stable molecule that accumulates in the cell [127, 161]. Similarly to what is observed in vertebrate macrophages, NOS activation is followed by activation of a peroxidase-mediated nitration reaction that uses nitrite as a substrate [161]. This is a highly efficient nitration reaction catalyzed by HPX2 (Heme peroxidase 2), that requires high local levels of hydrogen peroxide – provided by NOX5 (NADPH Oxidase 5) – and nitrite as substrates [159, 161, 162].

This NOX5/HPX2-mediated nitration is part of an apoptotic response in invaded/damaged midgut cell via JNK signaling that activates caspases expression, and is essential for mosquitoes to activate an effective immune response to *Plasmodium* by the mosquito complement-like system [161, 163, 164]. TEP1(C3-like factor), a key effector of the complement-like system, is produced by the hemocytes and circulates in the hemolymph as a stable complex with two proteins of leucine-rich (LRR) family, LRIM1 and APL1[140, 162, 165–167]. TEP1, APL1 and LRIM1 are form a MW complex responsible for TEP1 deposition on the surface of pathogens that promotes phagocytosis or leads to the formation of a complex that will lyse ookinetes [144, 166, 167]. The precise mechanism of killing and complement-activation is not completely understood, however work from our laboratory revealed that nitration of epithelial cells and the midgut basal lamina triggers the release of hemocyte-derived microvesicles (HdMv) into the basal lamina labyrinth, that is critical for activation of complement-mediated *Plasmodium* lysis [156].

c. Vector susceptibility and *Plasmodium* immune evasion

There are broad differences in compatibility, the extent to which the mosquito immune system limits infection, between different *Plasmodium*/mosquito combinations[168]. Intriguingly, while all ookinetes must come into contact with TEP1 in the mosquito hemolymph, only some are lysed [165]. That begs larger questions: how does *Plasmodium* evade the immune system

of mosquito vectors? And why are some *Anopheles* mosquitoes more susceptible to infection than others? Mosquito susceptibility to *Plasmodium* infection has a strong genetic component. For example, the mosquito *A. gambiae L3-5* strain was genetically selected to be highly resistant to *P. cynomolgi* (monkey malaria infection), but is also highly refractory to infection with *P. berhgei*, by expressing a TEP1 allele, with heightened anti-parasitic effects [169]. In addition, R strains have been shown to be in a state of chronic oxidative stress with increased basal levels of ROS and overactivation of JNK signaling, that is exacerbated by blood feeding [170].

While higher systemic ROS levels in the R strain result in loss of longevity and fecundity, the immune response to *Plasmodium* invasion is well localized both in time and space [171–173]. Invaded midgut epithelial cells activate ROS and nitration pathways, but these responses are localized and do not "spread" to healthy adjacent cells [162]. Furthermore, detox enzymes such as MnSOD (manganese-dependent superoxide dismutase), Gpx (hydrogen peroxide detox), and catalase are highly upregulated in healthy midgut cells and throughout the mosquito body (e.g. fat tissues), thus controlling any potential spillover and preventing damage. At the same time, catalase levels are downregulated in infected midgut cells to allow accumulation of ROS [162, 171, 172].

Parasite genetics are just as important in understanding *Plasmodium* transmission and infection. Our laboratory showed that epithelial nitration and microvesicle release are key for *P. berghei* destruction via TEP1-mediated lysis [156, 162]. Intriguingly, some reports had shown little [174, 175] to no [176] effects in disrupting the complement-like system when *A. gambiae* is infected with *P. falciparum*. Later studies demonstrated that susceptibility of *P. falciparum* killing by TEP1 is a *Plasmodium*-strain and mosquito-species specific response [162, 177]. For example, the *A. gambiae L3-5* refractory strain activates the complement-like system and kills the *P. falciparum* 7G8 strain from Brazil, while the African GB4 strain is able to evade the mosquito immune system and survive [178].

The *P. falciparum Pfs47* gene – a member of the 6-cystein protein family expressed on the surface of female gametocytes and ookinetes – allows the parasite to evade the immune responses mediated by TEP1: parasite killing with subsequent melanisation the *A. gambiae L3-5* refractory strain, as well as lysis without melanisation in the susceptible G3 strain [177]. *Pfs47* is polymorphic and exhibits a marked population structure and extreme fixation in non-African regions [179, 180]. The global populations structure of *Pfs47*, together with our laboratory experiments infecting anopheline mosquitoes vector species from different continents, provided strong evidence that distinct *P. falciparum* Pfs47 haplotypes were selected to be compatible with different mosquito vectors [177]. We then proposed the "lock-and-key theory", where *Pfs47* is a "key" that allows *P. falciparum* to evade the mosquito immune system by interacting with a mosquito receptor ("the lock"), different in each evolutionarily distant *anopheline* species [23, 181]. Only those parasites with a *Pfs47* haplotype compatible with a given mosquito species are able to evade the mosquito immune system, and this allows them to survive and become established in a given geographic area.

d. Signaling pathways of immune evasion / antiplasmodial immunity

Further work examined the mechanism through which *Pfs47* affects the response of the mosquito immune system to *Plasmodium* infection. JNK promotes TEP-1 lysis by inducing expression of HPX2 and NOX5 in midgut cells invaded by *P. berghei* ookinetes [182]. However, *Pfs47* disrupts JNK signaling, preventing caspases activation and downstream midgut nitration in response to *P. falciparum* invasion [163, 170]. A recent study has shown that in *P. berghei* Pbs47 is also required for ookinetes to avoid destruction by the complement-system [183]. *P. falciparum* ookinetes that do not express *Pfs47* activate JNK signaling, caspase activity and downstream epithelial nitration, triggering a strong activation of the mosquito complement system that is very effective killing the parasite [163, 184]. Other conserved immune-signaling cascades are important mediators of immune activation and killing of *Plasmodium*: *Toll***,** *Imd***,** and STAT. *Toll* and *Imd* activation promote TEP1-mediated lysis, but *Toll* appears to be more effective in limiting *P. berghei* (with silencing of repressor

protein *cactus*), while *Imd* is more effective against *P. falciparum*. These three pathways (*Toll*, *Imd*, JNK) all converge to TEP1 as the key effector of ookinete lysis [2].

If parasites evade and survive the early complement-mediated response, a different "late-phase" immune response is thought to further decreases parasite numbers by attacking the oocyst stage of *Plasmodium* [125, 185]. *Plasmodium* infection triggers a multi-pronged defense strategy by the mosquitoes, where an initial complement-mediated response that targets ookinetes is followed by activation of the STAT and LPS-induced TNFα transcription factor (LITAF)-like 3 (LL3) pathways that limits oocyst survival [185]. Interestingly, both STAT and LL3 seem to act independently. The STAT pathway is composed of STAT-B and STAT-A, with STAT-B regulating basal levels of STAT-A, which in turns regulated NOS, SOCS, and hemocytes differentiation. While STAT-dependent NOS expression reduces oocyst survival [125, 127, 185, 186], LL3-dependent midgut NOS induction has the opposite effect, increasing oocyst survival [186]. Other unknown factors are most likely at play, including the possibility of multiple isoforms of NOS [185].

In addition, while a double knock-down of SOCS (a suppressor of STAT) and NOS leads to higher oocyst survival than single SOCS silencing, single NOS knock-down unexpectedly leads to almost complete loss of infectivity due impaired epithelial cell invasion [125]. It appears that high levels of NOS are deleterious to oocysts, while a minimum level of NOS is required for ookinete midgut invasion to occur [125]. Finally, we still do not know what are the exact signals that lead to STAT and LL3 activation, although eicosanoids (see next section) or the wound-healing response might be implicated, as the AP-1/Fos-TGase2 axis has also been linked to increase TEP1 dependent *P. falciparum* killing [187].

e. Hemocytes are key coordinators of immunity in *Anopheles* and mediate mosquito immune memory

Our current understanding of both early and late cellular immune responses to *Plasmodium* is still limited, and much work is required to elucidate the precise molecular details of their immune effector functions. Furthermore, despite the ability of hemocytes to coordinate immune responses and respond to a variety of insults, including wound healing, and viral, bacterial, fungal, and parasitic infection, their exact molecular role in anti-*Plasmodium* immunity remains largely unknown [3, 185, 187–190]. We briefly discussed how three morphologically distinct subpopulations of hemocytes are believed to exist in *Anopheles*: the prohemocytes (putative undifferentiated precursors), granulocytes (phagocytic hemocytes), and oenocytoids (characterized by phenoloxidase activity) [4]. We have also discussed how hemocytes participate in the immune response against *Plasmodium* through cellular and humoral effector mechanisms. But what is the role of specific cell types in *Anopheles* defense mechanisms? And do only three cell types really exist? Already some recent studies suggest hemocytes could harbor greater complexity than originally thought, with three phagocytic subtypes found within $PPO6^{low}$ populations (equivalent to morphologic granulocytes)[191, 192].

Besides their conventional role as effectors of mosquito innate immunity, hemocytes have also been shown to mediate immunological memory. This phenomenon is called ‘immune priming’, and is defined as the ability of mosquitoes that have been infected with *Plasmodium* to develop a life-long, systemic state of enhanced immune surveillance, with an increased proportion of circulating granulocytes – the phagocytic cells that are more similar to vertebrate macrophages – which enhances their immune response to subsequent infections [193]. In addition, there are changes in the morphology and binding properties of granulocytes, with larger and more granular cytoplasm, pseudopodial extensions, and increased lectin-binding capabilities [193]. Interestingly, NK cells in vertebrates have recently been shown to also possess similar mechanisms [2]. The priming response in *A. gambiae* is activated when

Introduction

Plasmodium ookinetes breach the gut barrier and come into contact with the epithelial midgut cells, and damage to *Plasmodium* is due to a bystander effect [193]. Primed mosquitoes mount a stronger antiplasmodial response by greatly increasing the release of hemocyte-derived microvesicles [193].

Indeed, our laboratory has shown immune priming to be a 'two-step approach.' First, ookinete invasion induces expression of HXP7 and HPX8, two heme-peroxidases that catalyze prostaglandin E_2 (PGE_2) synthesis. Epithelial cells release PGE2 into the hemolymph and this attracts hemocytes to the basal surface of the gut. The chemotactic response is then followed by enhanced patrolling activity of the midgut basal lamina. If hemocytes detect a nitrated surface, they undergo apoptosis and release microvesicles into the basal labyrinth space, in close proximity to parasites that have traversed the midgut. The exact contents of these vesicles remains to be elucidated, but their release is essential for effective activation of TEP1-mediated anti-*Plasmodium* immunity [156, 194].

Following immune activation, primed mosquitoes constitutively release a hemocyte differentiation factor (HDF), and this factor persists in the hemolymph for the entire life of the mosquito [195]. HDF consists of a lipoxin/lipocalin complex [194]. Lipocalins are a family of proteins involved in lipid transport, while prostaglandins and lipoxins are all part of the eicosanoid lipid family [196]. Eicosanoids possess important signaling roles in homeostasis, inflammation and immunity not only in mammals, but also microbes and invertebrates like *Anopheles* [197]. Interestingly, suppression of host eicosanoid synthesis has been shown to be a mechanism of immune evasion by bacteria [197, 198].

Our laboratory has shown that immune priming involves an increase in lipoxin production (especially lipoxinA_4) from arachidonic acid, as well as increased expression of evokin, a lipid carrier protein of the lipocalin family. In addition, it appears as if LL3 is also necessary for HDF production, as silencing it stops HDF release. Priming can also be abolished by interfering with the function or movement of hemocytes by injecting water, PBS or Sephadex beads into the hemolymph [193]. Importantly, hemocyte differentiation factor

(HDF) is sufficient for effective priming, as transfer of both hemolymph, cell-free hemolymph, and HDF alone leads to hemocyte differentiation in the mosquitos and transference of enhanced antiplasmodial immune capabilities [2, 193]. Interestingly, although priming is elicited when ookinete invasion allows direct contact of the gut microbiota with midgut epithelial cells, *Plasmodium* species differ in their ability to establish a priming response, depending on their compatibility with the mosquito vector. For example, *A. gambiae* G3 mosquitoes mount a stronger immune response to *P. yoelii* ookinetes than to *P. berghei*, while *P. falciparum* NF54 fails to elicit an effective immune response. *P. yoelii*, the parasites that triggers the strongest immune response, leads to the strongest priming, while infection with the highly compatible *P. falciparum* NF54 strain results in weaker priming than *P. berghei* infection [163]. It is not clear whether strong epithelial nitration in midgut epithelial cells or the release of microvesicles also enhances the long-lasting priming response of hemocytes. Much remains to be discovered regarding the role of eicosanoids and hemocytes in insect immunity and immune memory.

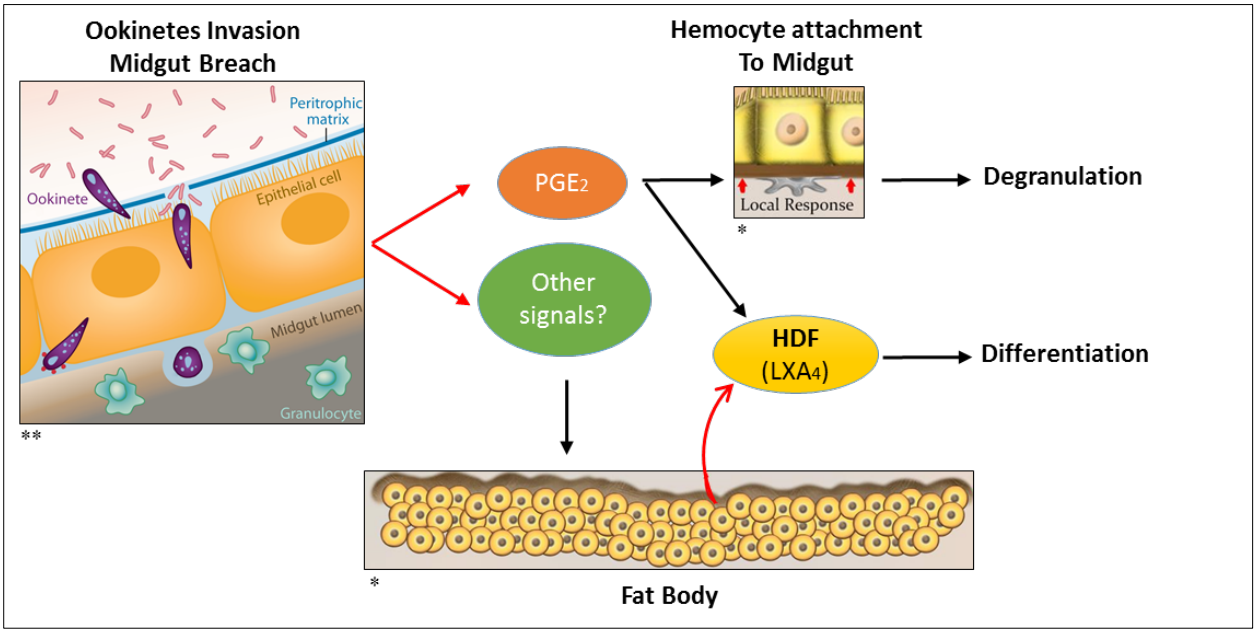


Figure I.6 Model of hemocytes activation and priming. Model follows *Plasmodium* ookinetes midgut invasion in *A. gambiae*. PGE_2 is released 24 hours post-invasion by the midgut, leading to chemotaxis and attachment of hemocytes to the basement membrane. Next, hemocytes release micro-vesicles in the basal surface of the midgut epithelial cells. PGE_2 and likely other signals lead to HDF release 48 hours post-invasion, which activates hemocytes for long-term priming and differentiation into granulocytes and oenocytoids. Priming can be abolished by interfering with function or movement of hemocytes[2, 193]. In addition, *LL3* is necessary for HDF production, as silencing it reduces HDF release [Dr. Barillas-Mury, personal communication]. None of the other immune pathways so far implicated in *Plasmodium* defense (*Toll*, *Imd*, *STAT*, *JNK*) are required for HDF release, although *Toll*, *STAT*, and *JNK* are all necessary for hemocyte differentiation in response to HDF[17]. Even for those pathways, we do not know the effector mechanisms, which receptors activate signaling cascades, or the sequence of kinases and transcription factor activation. Adapted from: * Crompton *et al.*[2] and ** Moreno-Garcia *et al.*[199]

4 Single-cell transcriptomics

"From out of all the many particulars comes oneness"
— Heraclitus

Past microarray studies in mosquitoes have uncovered *Plasmodium* and bacteria-mediated expression changes in genes regulating immunity [200–203]. However, few transcriptomic studies have been conducted to explore in depth how hemocytes respond to insults such as *Plasmodium* [192, 204]. And although terms such as "activation", "priming" and "innate memory" are used to describe immune phenomena in mosquitoes, their precise cellular basis is poorly understood [201]. As we have seen, many innate immune pathways are encoded in mosquito genomes, and have been linked to distinct immune responses by bulk transcriptomics and dissected through reverse genetics [17, 200, 202, 203, 205]. It is therefore highly likely that the number of relevant functional states in hemocytes is larger than currently known molecular markers suggest. In fact, two recent studies by the Levashina and Smith groups have started to explore the cellular heterogeneity of the mosquito immune system, but were limited by their chosen technology [191, 192]. They were largely unable to differentiate between hemocyte populations, since bulk approaches only look at the average expression level, and single-cell approaches conversely need large number of cells to make meaningful conclusions.

On the other hand, well designed single cell approaches such as single cell RNA-sequencing (scRNA-seq) enable researchers to thoroughly map whole immune systems, creating atlases of all immune cell type and states, describing their evolution in time and with infection. Critical biological questions can be explored, such as what transcript isoforms are variably expressed between different cell types [206–209], how cell types differentiate into one another[210, 211], and what is the precise lineage and cell cycle state of individual cells [207, 211]. In scRNA-seq we sample the transcriptome of each individual cell independently from one another, and the technique is quickly becoming the new state-of-the-art in cell biology.

Introduction

Seemingly homogeneous cell populations actually feature great transcriptional heterogeneity, both due to external factors such as individual microenvironments, but also internal stochastic processes [212]. Bulk approaches are unable to disentangle these differences, especially since the vast majority of transcripts in each cell is present in few copies, and most are not even messenger RNAs. The apparent randomness of transcriptional expression, even when cells are exposed to similar microenvironments – what we call transcriptional noise – is now recognised as crucial in determining cell fate decisions [213]. Single cell techniques are new, and both technical methodologies and analysis algorithms need to mature further, but a plethora of technologies have already emerged to make scRNA-seq the most sensitive, unbiased, and high-throughput technology to precisely capture these unique cell types, states, and transitions [214, 215].

The field has come a long way since its origins, having first been developed by Tang *et al*. in 2009 for hand-picked mouse blastomeres, which – thanks to their high RNA content (over 1 ng/cell) – could be more easily processed [216]. Now, tens of thousands of cells with only a few picograms of RNA can be successfully sequenced with highly automated pipelines [Fig. I.7]. However, all protocols share an initial reverse transcription to produce cDNA from RNA, which then needs to be amplified either by polymerase chain reaction (PCR) or in vitro transcription (IVT). As such, some of the original constraints of the technology remain [215, 217]. For example, it is still challenging to separate technical noise from biological variability [209, 210, 218]. In addition, any method only captures poly-adenylated RNA, and is severely limited by the suboptimal mRNA capture rate and reverse transcriptase efficiency [209]. The latter is the limiting step of scRNA-seq: it is estimated only 10-20% of all transcripts are reverse transcribed [219]. Direct RNA sequencing would represent a major step forward but it is still under active development [220].

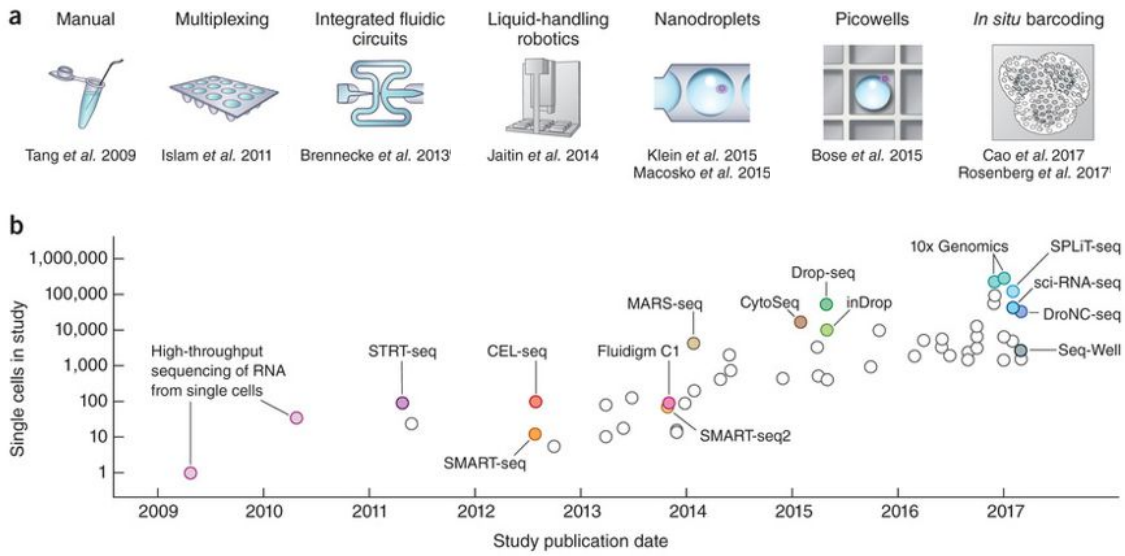


Figure I.7 Evolution of scRNA-seq technologies (A) Technical breakthroughs have increased the number of cells processed per run by orders of magnitude. Sample multiplexing was the first major innovation [221], followed robotics and fluidics[222, 223], which allowed researchers to study for the first-time thousands of cells in parallel. More recently, nanodroplets and picowells [224–226], and now *in-situ* barcoding, are pushing the field even further to its current scale [227, 228], as shown in panel (B). Key technologies are discussed below and summarized in Table I.1. Figure adapted from Svensson *et al*. [217].

a. Single-cell isolation and suspension

The first hurdle in a successful scRNA-seq experiment is creating a clean, pure, high-quality single-cell suspension of well-dissociated cells from the tissue of interest. Far from trivial, this initial step is crucial to the quality of downstream scRNA-seq data. The original Tang method – and one that is still in use when dealing with exceedingly fragile or rare cells – is low-throughput micromanipulation. As cells of interest are manually selected, the technique is precise, but it is also exceedingly labor-intensive. Alternatively, laser capture microdissection can likewise be used to isolate cells from solid samples. Fluorescence Activated Cell Sorting (FACS) on the other hand is able to quickly isolate of tens of thousands of cells. In addition, surface markers tagged with fluorescently-labelled antibodies can be used to purify cells of interest with high fidelity, and most scRNA-seq protocols are compatible with FACS. Nevertheless, FACS requires large amounts of starting material, can be rough on delicate cells,

and requires *a priori* knowledge of the system of interest. Alternatively, cells in suspension can be isolated and processed with microfluidics chips, which automate many of the required cell separation, selection, and collection steps, upstream of fully-automated scRNA-seq protocols. However, cellular stress can be high, capture-rate (number of cells sequenced per cells loaded) is low, and cell selection is highly dependent on chip-architecture. Recently, developments in microdroplet and microwell technologies have achieved significant reductions in hands-on time and reaction volumes (leading to lower costs), while increasing cellular throughput [209, 215, 229–231].

However, for tissues and cells rich in RNAses such as the pancreas or granular immune cells (e.g. neutrophils and macrophages), maintaining optimal cell and RNA integrity during sample preparation is an unsolved technical hurdle [232, 233]. Cells from tissues need to be dissociated and resuspended using enzymes such as collagenase and trypsin, which takes hours and inevitably affects both cell viability and transcriptome, further confounding biological differences. As a result, cells are stressed and their information altered by experimental manipulation, while RNA is lost due to the action of endogenous RNAses[234]. New protocols such as methanol and Lomant's fixation partially solved these issues, particularly for cells in suspension, however solid tissue dissociation remains a challenge [235, 236].

b. scRNA-seq technologies

Protocols can be roughly divided into two separate categories: full-length versus tag-based. The original Tang protocol was a full-length method, while the popular commercial 10X technology is a tag-counting protocol. Each has its strengths. Full-length scRNA-seq methods typically provide more genes per cell and allow a researcher to delve into the data deeper by exploring transcript isoform expression, allelic expression, and RNA editing thanks to the strand-specific information along the full length of each transcript [237]. Tag-based methods, on the other hand, feature higher throughput and lower costs, thus providing the necessary power to discover new and rare cellular subtypes or transcriptional states [231]. The most utilized methods are described and compared below (see also Table I.1).

Full-length protocols

Quartz-seq

An improvement over the original Tang method, with simplified workflow and improved performance, although quickly rendered obsolete by Smart-seq2 [238–240].

Smart-seq2

The classic full-length protocol –considered state-of-the-art in terms of genes per cell captured – begins with an RT reaction using the Moloney Murine Leukemia Virus (MMLV) RT enzyme and oligo-dT primers with template switching oligonucleotides (TSO) to synthesize cDNA. The cDNA is then amplified before library preparation. Though the protocol is time consuming, robotic handling can simplify the workflow [237, 239, 240]. Importantly, strand information is lost with standard Illumina sequencing, and technical errors due to unequal PCR amplification are not corrected by unique molecular identifiers (UMIs: unique molecular identifiers) as in tag-counting protocols, so that PCR amplification bias remains a concern [219]. When using UMIs, every transcript captured gets labeled with its unique barcode (e.g. 10-12 bp long with Chromium 10X) in addition to a cellular barcode. This allows to distinguish sequencing reads originating from unique mRNAs vis-à-vis PCR duplicates.

Tag-based protocols

CEL-seq, CEL-seq2, and MARS-seq

This tag-based protocol employs IVT rather than PCR amplification. CEL-seq (Cell Expression by Linear Amplification and Sequencing) starts with an RT reaction, before second strand cDNA synthesis, pooling, and IVT. Exonic reads are highly strand-specific (over 98% from sense strand), barcoding highly efficient, and no gene-length normalization is required. However, there is a strong 3' bias and spliced isoforms cannot be detected. CEL-seq shows poor sensitivity for lowly-expressed transcripts [241]. MARS-seq (Massively Parallel RNA single-cell sequencing) is a fully automated CEL-seq with UMIs, enabling the counting of individual RNAs [223, 242]. CEL-seq2 improved upon the original protocol by decreasing costs and hands-on time, while increasing sensitivity and implementing UMIs [243].

Introduction

STRT-seq

STRT-seq (Single-cell Tagged Reverse Transcription Sequencing), is a tag-based method that employs anchored oligo-dT primers and a MMLV-based enzyme as Smart-seq2, before PCR amplification. Later iterations of the protocol included UMIs and have been automated to allow multi-plexing and strand-specificity. Disadvantages are the same as for all tag-based protocols, including the inability to detect SNPs or splice variants [244].

DROP-seq, InDROP and Chromium 10X

All these three relatively newer technologies work in similar ways to increase throughput and lower reaction cost by carrying out all reactions in nanoliter emulsion droplets. These droplets contain the lysis buffer, RT, and barcoded microspheres with oligonucleotides to uniquely tag both the individual cells, as well as each transcript within those cells. inDrop and 10X are characterized by higher cell capture rate, 10X has the highest sensitivity and lowest technical noise. Drop-Seq on the other hand is the most cost-effective [224, 225, 245–247]. A detailed discussion of 10X follows in the materials and methods.

SeqWell

The latest addition to the arsenal and one of the most promising recent developments in scRNA-seq, SeqWell sports the same advantages of emulsion droplet methodologies, but employs microarrays and picowells to increase throughput even higher. Seq-Well utilizes PDMS arrays that each contain ~88,000 subnanoliter wells with uniquely barcoded poly(dT) mRNA beads. The uniquely barcoded mRNA capture beads and cells are both secluded in the wells, which are then sealed with semipermeable membranes, leading to a more efficient cell lysis and mRNA capture. Beads can then be pooled, thanks to double barcoding for cells and transcripts (UMIs). Seq-Well only requires a PDMS array, a polycarbonate membrane, a pipette, a clamp, an oven/heat source, and a tube rotator to produce stable cDNA product, making it functional in nearly every clinic and laboratory context. The protocol can also be adapted to use harsher lysis conditions, useful when dealing with fixed or otherwise challenging material [248, 249].

Combinatorial indexing

Recently, single-cell combinatorial indexing has emerged in different groups as another powerful high-throughput scRNA-seq methodology involving the split-pool barcoding of either cells or fixed nuclei. For RNAseq, the methods are similar and are alternatively called SPLiT-seq, sciRNA-seq, or sci-RNA-seq3 [227, 228, 250] . However, single-molecule combinatorial indexing can be used for many other omic techniques to explore chromatin accessibility (called sci-ATAC-seq)[251], genome sequence (sci-DNA-seq)[252], genome-wide chromosome conformation (sci-Hi-C)[253], and DNA methylation (sci-MET)[254].

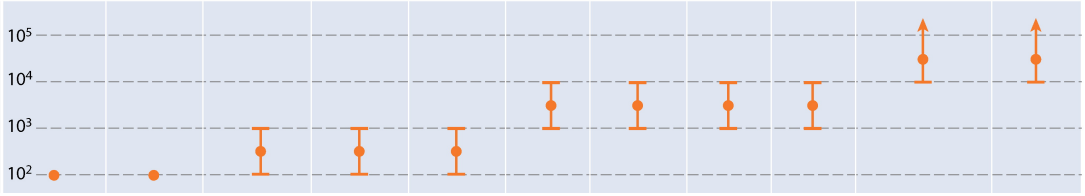
	SMART-seq2	CEL-seq2	STRT-seq	Quartz-seq2	MARS-seq	Drop-seq	inDrop	Chromium	Seq-Well	sci-RNA-seq	SPLiT-seq
Single-cell isolation	FACS, microfluidics	FACS, microfluidics	FACS, microfluidics, nanowells	FACS	FACS	Droplet	Droplet	Droplet	Nanowells	Not needed	Not needed
Second strand synthesis	TSO	RNase H and DNA pol I	TSO	PolyA tailing and primer ligation	RNase H and DNA pol I	TSO	RNase H and DNA pol I	TSO	TSO	RNase H and DNA pol I	TSO
Full-length cDNA synthesis?	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes
Barcode addition	Library PCR with barcoded primers	Barcoded RT primers	Barcoded TSOs	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers	Barcoded RT primers and library PCR with barcoded primers	Ligation of barcoded RT primers
Pooling before library?	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Library amplification	PCR	In vitro transcription	PCR	PCR	In vitro transcription	PCR	In vitro transcription	PCR	PCR	PCR	PCR
Gene coverage	Full-length	3'	5'	3'	3'	3'	3'	3'	3'	3'	3'
Number of cells per assay											

Table I.1 Comparison of scRNA-seq methodologies. Abbreviations: cDNA, complementary DNA; DNA pol I, polymerase; RNase H, ribonuclease H. Adapted from Chen et al. [214]

c. scRNA-seq data analysis

After making a quality single-cell suspension, successfully loading the cells onto the scRNA-seq platform of choice, making a library of appropriate complexity, and then sequencing it to the desired depth (note that 50k reads per cell is thought sufficient to successfully cluster cells into subpopulations, including rare cell types) [255], data must be quality controlled before downstream biological analyses. Multiple packages have been created to integrate QC methodologies and analyses and simplify data exploration and interpretation. Seurat [256], Scanpy [257], Scater [258], Monocle [211] and Cellranger [246] are the most popular.

Read Alignment, Expression Quantification, and Quality Control

The reads to reference transcriptome mapping ratio is an early indicator of scRNA-seq data quality. Samples with low mapping percentages likely contain a high amount of damaged or degraded RNA and must be removed. Since sequencing output is the same as for bulk RNAseq the same software can be used for the first data analysis and QC steps. Alternatively, Kallisto and Salmon can both accurately estimate transcript abundance without relying on alignment to an existing transcriptome [259, 260]. Most users however use standard splice-aware alignment programs using reference assemblies. The most popular tools are TopHat2 [261], STAR [262], and HISAT2 [263], although 10X has implemented their version of STAR into a proprietary software suite called Cell Ranger [246]. Studies have compared these aligners highlighting trade-offs between speed, memory requirements, and alignment efficiency in all [264–266]. Which expression quantification method to use varies according to the scRNA-seq technique used. For whole-transcript protocols such as Smart-seq2 traditional bulk-RNAseq methods suffice. Tag methods such as Chromium 10X will either use the Cell Ranger pipeline or specifically-tailored algorithms such as SAVER (Single-cell analysis via expression recovery) to take advantage of UMIs and reduce technical noise [231, 267]. Data is then cleaned up to exclude reads originating from multiplets, broken cells, or dead cells (unless cell were fixed). Even the highest quality, healthiest cells will suffer from low mRNA capture efficiency, bias in transcript coverage, and dropout events (lack of transcripts that are known to be expressed

in the cells). Nevertheless, poor quality samples and cells will skew biological interpretation and must be removed. Some protocols use extrinsic spike-ins (e.g. ERCC) to estimate technical noise and cellular quality [268], as cells with high proportions of ERCC spike-ins likely feature broken, porous cellular membranes. Furthermore, while cytoplasmic RNA is usually lost when a cell ruptures, mitochondria remain within the cell. Thus, a high percentage of mitochondrial RNAs to total RNAs can indicate poor quality. Finally, low total gene counts or transcripts abundance within cells can also be an indication of low quality, although this can sometimes be due to technical limitations or the low total RNA content of the cell of interest [210, 269].

Normalization of scRNA-seq data, and removal of batch effects

Initial QC must be followed by careful data normalization in order to disentangle the biological signal of interest from the variability in capture efficiency, sequencing depth, dropouts, and all other technical effects in each individual sample. This intra-sample normalization is important, but as scRNA-seq datasets become larger, batch normalization is also becoming crucial. The latter takes into consideration all of the above, but also harmonizes samples often run on different days, platforms and laboratories. Normalization is an issue also in bulk-RNAseq, however it is far more complex in scRNA-seq. Bulk RNA-seq investigators standardize libraries by calculating quantities such as transcripts per million (TPM), fragments per kilobase of exon per million fragments mapped (FPKM, which takes into consideration both transcript length and library size), or size factors [210, 231, 270, 271].

That is not sufficient for single-cell RNA-seq, which features unique analytical challenges requiring specifically-tailored normalization algorithms. For example, scRNA-seq data matrices are characterized by abundant zeroes, but ‘zero inflation’ is due to both technical reasons (dropouts due to the low reverse transcription efficiency previously mentioned) as well as meaningful biological differences (e.g. quiescent or stem cells). Moreover, scRNA-seq is characterized by higher technical noise even for non-lowly expressed genes, further augmented by true biological heterogeneity. Any overcorrection by normalization algorithms will reduce such biological differences. Conversely, under-correction will lead to spurious biological conclusions. Traditionally, scRNA-seq normalization methods have employed off-the-shelf or

adapted bulk-RNAseq methods. Specifically, median normalization methods are used to identify genes whose expression does not change across cells. Non-differentially expressed (DE) genes are then used to calculate global scaling factors that are unique for each cell, but common for all transcripts within that cell. These methods all assume total RNA in each cell is more or less the same and that all variation is technical. However, that is not the case when dealing with single cells, especially in heterogenous samples. And further, RNA content in different libraries is influenced by both the cell-cycle and the dynamics of transcription (including transcriptional bursts) for each individual gene [14, 270–272]. A first attempt to circumvent this limitation was the use external spiked-in ERCC artificial RNAs at a set concentration in each individual cell library, thus adjusting for technical variation and improving the accuracy of global scaling factors [273]. However, issues with spike-ins can lead to inconsistent detection and few studies have used this technique [210]. Rather, researchers have been using UMIs to successfully remove or reduce cell-specific effects due to amplification and gene length, although UMIs can only be used with tagging scRNA-seq protocols [219, 224, 242, 270]. As the field matures, more and more sophisticated normalization methods are being specifically tailored for scRNA-seq datasets, such as SCnorm, SAMstrt, and SCTransorm (as integrated in Seurat V3.0) [272, 274, 275].

As larger scale scRNA-seq experiments become the norm, ‘batch effects’ – the aggregated technical variation of different cell dissociation methods, library preparation techniques, sequencing platforms, environments, handling, operating equipment, institutes and laboratories – are becoming ever more important confounding factors. These confounders are especially problematic for large consortium-scale projects such as the Human Cell Atlas, and must be minimized. The field is thronged with new computational methods that have begun addressing the issue, from the linear regression models of ComBat[276] to the nonlinear canonical correlation analysis (CCA) of Seurat [277] or the projection of mutual nearest neighbors method (MNNs)[278]. Separately, MAST, DESeq and limma can include batch effects as covariates in their DE testing model [279–281]. Principal component analysis (PCA) and visualization in low dimensional usually follows. Recently, a dedicated method (kBET)

has been developed to further explore batch effects (and their correction), in detail [282]. All these normalization, batch correction, and visualization methodologies are summarized below in Fig. I.8.

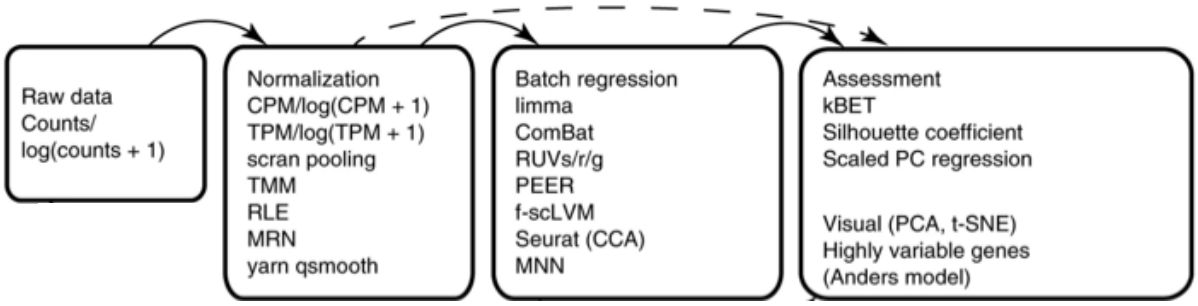


Figure I.8 Summary of normalisation, batch regression, and assessment techniques. Detailed information on each method and full figure in Büttner *et al*. [282]

Dimensionality reduction, feature selection, clustering, and differential expression analysis

Data matrices downstream of all QC and normalization processing feature many thousands of dimensions, with thousands of genes and tens of thousands of cells. Data must be simplified - dimensions reduced – to aid computations and interpretation while keeping intact key biological differences between cells and conditions. PCA is a linear dimensional reduction algorithm assuming normal distribution of data. It identifies new variables, called principal components (PCs), that are linear combinations of the variables from a dataset. Data is standardised so that each gene's mean expression across cells is zero and the PCs are then normalised eigenvectors of the genes' covariance matrix. Importantly, the PCS are ordered by how much dataset variation they describe. T-distributed stochastic neighbor embedding (t-SNE) is a non-linear dimensionality reduction technique used for example by Seurat to visualize the scRNA-seq data [283]. Both are limited. PCA is unable to fully display data complexity, while t-SNE plots are inconsistent and do not preserve global information. Newer algorithms such as uniform manifold approximation and projection (UMAP) [284], and scvis [285] were designed specifically for scRNA-seq. UMAP is fast, reproducible, and cluster organization and display reflects inherent cellular similarity, unlike for t-SNE.

Feature selection is also used to reduce data dimensionality and free-up computational resources for downstream analysis such as clustering. Unsupervised algorithms for feature selection are divided into three main types. Highly variable genes (HVG) methods as in Seurat assume HVG are those that vary because of biology [222, 256]. Spike-in approaches (e.g. scLVM and BASiCS) identify genes that have higher variance than spike-ins with similar expression levels [270, 273]. Finally, dropout methods such as M3Drop use the dropout distribution characteristic of scRNA-seq data to efficiently select all important features in a dataset [286].

After dimensionality reduction we can finally interrogate our data and answer key biological questions. For example, what populations and subpopulations of cells do exist in our dataset? And what are their cell states? Cell clustering can be done either using known markers, or more commonly with unsupervised clustering methods. These are mainly divided into k-means, hierarchical clustering, density-based clustering, and graph-based clustering methodologies. K-means requires setting the number of clusters *a priori*, and assigns cells to nearest cluster center, while all others methods work in unsupervised fashion to establish the optimal number of clusters. Some methods such as single-cell consensus clustering (SC3) use a combination of methodologies, and the popular Seurat clusters mainly with a shared nearest neighbor algorithm (SNN) [287]. Differentially expressed genes (marker genes) are then found with differential expression analysis (DE) or analysis of variance (ANOVA). DE analysis is an active area of software development. Often, clustering algorithms return not only cell subpopulations, but also variable cell states for each of these populations. Software packages must be able to differentiate between the two, while dealing with the high noise of scRNA-seq data and the large sample size. While bulk RNAseq DE techniques are still used, in recent years specific tools such as MAST (linear model fitting and likelihood ratio testing), SCDE (Bayesian approach with low-magnitude Poisson), DEsingle (Zero-Inflated Negative Binomial), have been developed. Seurat uses the non-parametric Wilcoxon rank sum test as a default, but other methodologies such as MAST and DEseq2 can also be employed. Soneson et al tested over 36 methods in their recent review [288]. See Table I.2 below for a summary.

Clustering	References	DE analysis	Type	References
Seurat	Satija et al., 2015[256]	MAST	Single-cell	Finak et al., 2015[279]
SC3	Kiselev et al., 2017[289]	ROTS	Single-cell	Seyednasrollah, 2016[290]
Destiny	Angerer et al., 2016[291]	BCseq	Single-cell	Chen et al., 2018[292]
SNN-Cliq	Xu and Su, 2015[293]	SCDE	Single-cell	Kharchenko et al., 2014[294]
RaceID	Grun et al., 2015[295]	DEsingle	Single-cell	Miao et al., 2018[296]
SCUBA	Marco et al., 2014[297]	Cencus	Single-cell	Qiu et al., 2017[298]
BackSPIN	Zeisel et al., 2015[299]	D3E	Single-cell	Delmans et al., 2016[300]
PAGODA	Fan et al., 2016[301]	BPSC	Single-cell	Vu et al., 2016[302]
CIDR	Lin et al., 2017[303]	DESeq2	Bulk	Love et al., 2014[280]
pcaReduce	Zurauskiene, 2016[304]	edgeR	Bulk	Robinson et al., 2010[305]
TSCAN	Ji et al., 2016[306]	Limma	Bulk	Ritchie et al., 2015[281]
ZIFA	Pierson et al., 2015[307]	Ballgown	Bulk	Frazee et al., 2015[308]

Table I.2 Summary of clustering and DE analysis software packages. For more information please consult these excellent reviews: Andrews and Hemberg (clustering) [287] and Soneson *et al* (DE analysis) [288].

Cell lineage, pseudotime, alternative splicing and gene regulatory networks analysis

After probing the cellular complexity of tissues and cell populations, data can be used to explore the dynamics of cellular development and identify cell types lineages, for example by building a pseudotime ordering of cells which can showcase cellular differentiation. Pseudotime techniques order cells along a continuous trajectory, aligning cells based on transcriptional similarities rather than clustering them. These approaches not only allow investigators to probe the initial, transitional, and final cell states of a population, but also the genes that are involved in such transitions. Popular tools are Monocle (based on minimum spanning tree) [211], Monocle2 (reversed graph embedding) [298], Slingshot (cluster-based approach) [309], TSCAN [306], PAGA, and Cellrouter [310]. Saelens *et al.* recently evaluated most pseudotime and lineage approaches and found Monocle2, Slingshot, and PAGA to be superior, depending on the individual data structure of the dataset (e.g. linear, bifurcating, complex separate trees) [311].

Introduction

Gene regulatory network inference is a common feature of bulk RNAseq analyses, normally employing weighted gene co-expression network analysis (WGCNA), which assumes all genes that are highly correlated in expression to be co-regulated. By combining cells together to build a pseudo bulk-RNAseq dataset we can evaluate gene regulatory networks in the same way. However, the analysis needs to be run separately for each subpopulation. SCENIC is one such scRNA-seq method that can build gene regulatory networks from single cell data and predict transcription factors - target genes interactions [312]. PDIC is an alternative software suite to answer the same questions [313].

Finally, when data is generated with scRNA-seq protocols producing full-length transcripts (such as Smart-seq2) investigators can also analyze alternative splicing. Over 90% of human genes undergo alternative splicing, which plays important roles both in tissue homeostasis and disease [314]. Data on isoform usage could be crucial in understanding the expression dynamics of specific pathogenic isoforms for example, or to further characterize the importance of cellular subsets in immune process. However, bulk RNAseq methodologies are again unsuitable to the task. Recently new methods have emerged such as SingleSplice, Census, BRIE, and Expedition [298, 315–317].

5 Aims and outline of the thesis

This dissertation first focuses on dissecting the complexity of the *A. gambiae* M-form *(A. coluzzi)* immune system under baseline conditions. That knowledge is then leveraged to obtain an in-depth understanding of how mosquitoes responds to both blood-feeding and *Plasmodium* infection. In analogy with vertebrates, I posit the existence of different hemocyte subpopulations and states, each characterized by distinct gene expression profiles. I will further argue that hemocytes transition between distinct states along a range of predetermined routes, through which the diversity of functions associated with cellular immunity in invertebrates is generated. In addition, I will show that single-cell approaches, coupled with complementary bulk techniques and imaging validation, are an effective method to study the cellular arm of the immune system of mosquitoes.

In Chapter II, we explored different strategies to isolate hemocytes and create a clean, pure single cell suspension for downstream scRNA-seq. We evaluated different methods, enzymes, and fixatives to adapt single cell protocols to the unique challenges of mosquito immune cells, while maintaining high quality RNA and cellular integrity. As part of this work we developed a protocol to fix and sequence hemocytes at single cell resolution making use of the droplet-based Chromium 10X technology. We then validated our scRNA-seq results by adapting the commercial RNAscope RNA-FISH technology to mosquitoes.

In Chapter III, we used these methods to characterize the functional classes of *A. gambiae* mosquito hemocytes and build a comprehensive atlas of the cellular arm of the mosquito immune system to discover new hemocyte cellular subtypes. We then defined marker genes for each cell type, and identified surface markers for future functional studies. We uncovered different cell states within each hemocyte type, successfully building a lineage tree to explain how hemocytes differentiate into each cell type and cell state. Finally, we validated these scRNA-seq results with a combination of bulk-RNAseq and RNA FISH techniques and visualize each cell type and their spatial-temporal localization in the mosquito. Importantly, we not only recapitulated what previous knowledge existed, but also discovered novel effector

cellular subtypes, including a cellular subtype potentially akin to lamellocytes in *Drosophila*, likely involved in the 'late' cellular immunity against *Plasmodium*, as well as hemocytes responsible for secreting anti-microbial peptides in circulation, revealing a previously unknown complexity of this biological system.

In Chapter IV, we challenged *A. gambiae* mosquitoes first with blood-feeding and then with *P. berghei* and *P. falciparum* infection. We evaluated how mosquito hemocytes, guts, and carcasses respond to these challenges to explore how hemocytes differentiate into their distinct cellular states. We identified a trajectory of immune activation following the mosquito on a time-course after infection, reaching a peak of transcriptomic activity against the parasite at days 2-3 after infection, before returning to baseline at day 7. Finally, we describe how hemocytes dynamically respond to infection, going into circulation to respond to injury and replenish the immune cell pool. We identified rapidly dividing precursor cells, as well as the transcriptomic signatures of the response of hemocytes and fat body to *Plasmodium*, including what pathways are differentially activated in various cellular subtypes. Then, we explored how the upregulation of the Toll pathway affects hemocytes and their ability to mount an effective immune response to suggest how different hemocyte subtypes are the control of specific and distinct immune pathways.

In chapter V, I conclude by providing a summary of our findings and discussing what significance they hold in view of the emerging importance of vector-borne diseases for human health and disease, not only in the developing world, but increasingly also in the West.

6 List of publications

1. "Mosquito cellular immunity at single-cell resolution"
   Raddi G*, Barletta-Ferreira A*, Efremova M, Luis Ramirez J, Cantera R, Teichmann S, Barillas-Mury C*, Billker O* (*in review*)
   * These authors contributed equally to the work
2. "Seq-Well: A Sample-Efficient, Portable Picowell Platform for Massively Parallel Single-Cell RNA Sequencing"
   Aicher TP, Carroll S, **Raddi G**, Gierahn T, Wadsworth MH 2nd, Hughes TK, Love C, Shalek AK. *Methods Mol Biol*. (2019) 1979: 111-132
3. "Single-Cell RNA Sequencing with Drop-Seq"
   Bageritz J, **Raddi G**. *Methods Mol Biol*. (2019) 1979: 73-85

Chapter II

Establishing an experimental system to explore the mosquito immune system

An experimental system to analyze mosquito immune cells

1 Introduction

"If we knew what it was we were doing, it would not be called research, would it?"
— Albert Einstein

No comprehensive scRNA-seq study had been done on mosquito or *Drosophila* immune cells, requiring technology development. The envisaged work-flow involved several steps for which no established protocols were available (Fig II.1). In particular, techniques for cell collection were not designed with subsequent scRNA-seq in mind, and new procedures were needed to keep handling of cells to a minimum [209]. When the project first started, no bulk RNAseq data of mosquito hemocytes existed either, and available protocols needed to also be adapted. For scRNA-seq in particular different sequencing technologies had to be evaluated and sequencing library preparation optimised for the specific requirements of mosquito immune cells.

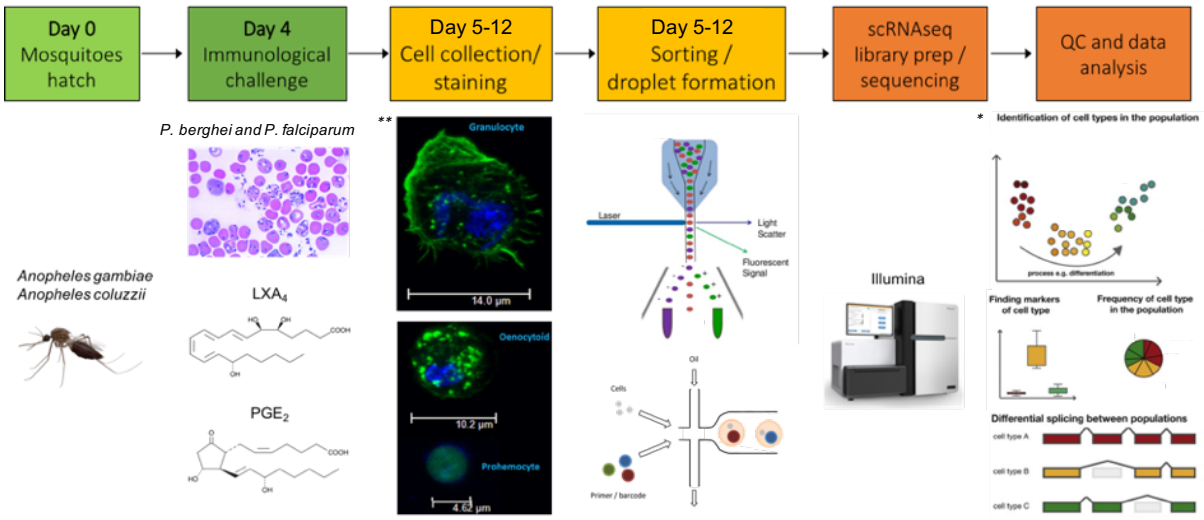


Fig. II.1 Experimental work-flow. At day 0 *A. gambiae* M-form (*A. coluzzi*) adult female mosquitoes hatch. After 3-4 days mosquitoes are challenged with *P. berghei* or *P. falciparum* infection or eicosanoid (lipoxin A_4 / prostaglandin E_2) injection. Samples are collected 1,2,3,7 days post-challenge. Following collection, cells are either stained (Hoechst 33342 / calcein) and sorted or fixed in vivoPHIX before sorting or direct scRNA-seq library preparation (Smart-seq2 and 10X chromium). Libraries sequenced on Illumina platforms before data analysis. Figure partially adapted from * Kolodziejczyk *et al* [209] and ** Jose Luis Ramirez (personal communication)

The main limitations were:

a) the low number of immune cells available for collection in each mosquito (only a few thousand hemocytes per mosquito can be bled with an injection-recovery method), many of which with a low RNA content (e.g. prohemocytes). This is a challenge for both bulk RNAseq and scRNA-seq methods, and pushed current protocols to the limit.
b) the heterogenous sizes and shapes of hemocytes, ranging from 3μm in diameter, round prohemocytes all the way to 20 μm in diameter, elongated granulocytes, with pseudopodia. The heterogeneity of our samples precluded use of the popular Fluidigm C1 instrument [318]. In addition, the different amounts of RNA content in each cell type presented a technical challenge, potentially requiring different numbers of PCR amplification steps for each.
c) Some subtypes of hemocytes such as granulocytes are very delicate and prone to cell death, and are filled with granules containing digesting enzymes detrimental to RNA quality. These cells can easily burst, especially when activated. Furthermore, many hemocytes attach to surfaces, including Eppendorf tubes and pipette tips, requiring investigators to reduce as much as possible hands-on time, centrifugation steps, and general handling, while coating all surfaces with silicone [193]
d) Finally, the final single cell suspension must be created rapidly, while also being void of contaminants. Mosquito guts must not be punctured to avoid introducing gut contents into the hemolymph. The contamination would lead to both cell loss due to immune activation and bursting of hemocytes, as well as to poor scRNA-seq library quality due to the debris, RNAses and other enzymes introduced in the reaction mix. In addition, mosquito handling must be vigorous enough to release sessile hemocytes into the circulation, but gentle enough to avoid excessive fat body and muscle cells contamination. That is a challenge on its own, as mosquito micromanipulation and hemocyte collection has to be rapidly completed to collect enough cells from as many mosquitoes as possible within a limited timeframe to overcome aforementioned issues.

An experimental system to analyze mosquito immune cells

1.1 Aims

1. To develop a pure, high quality cell hemocyte suspension for scRNA-seq and bulk RNAseq.
2. To implement an efficient sorting strategy for downstream scRNA-seq.
3. To select which scRNA-seq method produces the best quantity (number of cells after QC) and quality (as measured by genes per cell and mitochondrial transcripts ratio) single cell hemocyte data.

1.2 Colleagues

The data presented in a result of my own work unless stated otherwise.

2 Methods

2.1 *Anopheles gambiae* and *Aedes aegypti* mosquito rearing and *P. berghei* infection

A. gambiae (G3 NIH strain) and *A. gambiae* M-form (*A. coluzzi*) were reared at 28 °C, 80% humidity, 12-hour light/dark cycle with standard laboratory procedures. For infections we utilized GFP-CON transgenic *P. berghei* (259cl2 strain), maintained with serial passage in female 4-8 weeks old BALC/c mice[319]. Parasitemia was assessed by light microscopy following methanol-fixed blood-smears stained with 10% Giemsa and air-dried. Mosquitoes were fed on infected mice at a parasitemia of 3-5%, with 1-2 exflagellations per field. Infected mosquitoes were kept at 21 °C to allow for infection and midgut invasion. To confirm infection intensity at least 10 mosquito midguts were dissected 5 days post blood-feeding and oocysts counted by fluorescence. *A. aegypti* (Liverpool strain) mosquitoes were also reared with standard insectary conditions at our Laboratory of Malaria and Vector Research (NIH) at 28 °C, 80% humidity, 12-hour light/dark cycle. *Aedes* mosquitoes were maintained with 10% Karo syrup solution by Mr. Andre Laughinghouse.

2.2 *Anopheles* mosquito micro-injection with CM-DiL and eicosanoids

To stain hemocytes with the lipophilic dye chloromethylbenzamido-1,1′-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine-perchlorate (CM-DiL), two-day old mosquitos were cold-anesthetized and injected in the thorax with a Drummond Nanoject II in 'fast' filling and release mode with 69 nL of 140 μM CM-Dil in DPBS, and then left to recover 2-3 days. To challenge the immune system with eicosanoids, single-use aliquots of eicosanoids were thawed and 1.43 μl of 0.1 μg/μL lipoxin A_4 (LXA_4), or 1.43 μL of 1 μg/μL of prostaglandin E_2 (PGE_2), were dried with a mild stream of nitrogen gas in amber ampullas to protect the compounds from oxidation and light. The compounds were then resuspended with 50 μL of transfer buffer composed of 95% Schneider's Insect media, 5% citrate buffer (98 mM NaOH, 186 mM NaCl, 1.7 mM EDTA, and 41 mM citric acid; buffer pH to 4.5), final pH buffered to 7.0-7.4, and sterilized through a 0.22 μm syringe filter. Injection needles were prepared with a Narishige

PC-10 needle puller, using Drummond microinjection capillaries of borosilicate glass 3.5 inch in length. The needle puller was set in "Heater N.2" mode, with heater level 55.00 and the tip of the needle cut open with fine tweezers approximately 1cm from the fine end, leaving an even, clean bore for the injection. Cold-anesthetized four or five-day old mosquitos were then injected in the thorax with 138 μL of the 2 ng/μL dilutions (50nM) and let recover for 6 hours - 5 days before perfusions.

2.3 *Aedes* mosquito micro-injection with Lacz

Dr. Ana Beatriz Ferreira cold anesthetized and injected two to three-day old female *A. aegypti* mosquitoes with 69 nl of 3 μg/μl dsRNA solution specific for LacZ, a bacterial gene not related to the genome of *Aedes* mosquitoes. DsRNA of LacZ is used as control during dsRNA-injection gene knockdown. It was produced as previously described in Molina-Cruz et al. [178]: a 218-bp fragment was amplified from LacZ gene cloned into pCRII-TOPO vector using M13 primers. Then, 2 days before serum feeding mosquitoes were injected.

2.4 *Aedes* bacterial feeding

Dr. Ana Beatriz Ferreira performed the bacterial feeding experiments at the NIH with *Aedes*. First the mixture of bacteria to be used for feeding was collected by cleaning sugar-fed mosquitoes with 70% ethanol and sterile PBS, and then by dissecting the same mosquitoes in sterile PBS. Groups of 5 midguts were homogenised in sterile PBS before LB media incubation for at least 16 hours at 28 °C and 250 rpm in a shaker. This stock solution was then frozen and kept at -80 °C. The pre-inoculum was set-up before each experiment by scraping frozen stock into LB media and incubating for at least 16 hours at 28 °C and 250 rpm in a shaker. The day of the experiment pre-inoculum was diluted again in LB and allowed to grow for two hours, as above, after which bacteria were pelleted and washed with sterile PBS to remove any produced toxins. Concentration was estimated with optical density (OD) measurements at 600 nm, with 1 OD considered equivalent to 10^9 bacteria/mL. Three to four-day old mosquitoes were then fed 10% sterile sucrose solutions with 100 U/mL penicillin and 100 μg/mL streptomycin for 2

days before feeding the bacterial mixture. Control mosquitoes were instead fed with sterile 10% bovine serum albumin (BSA) solution in HBSS with no calcium nor magnesium. Bacteria-fed mosquitoes were fed the same solution but with $4x10^9$ bacteria/mL in each feeder. Four days post-feeding mosquitoes were perfused as of below in section 2.5. Twenty-five mosquitoes were perfused per replicate, with 3 replicates per condition in each of the control serum fed, bacterial fed and dsRNA injected serum fed conditions. Hemolymph was placed in 0.5 ml of vivoPHIX for RNA isolation and single cell transcriptome analysis.

2.5 Hemocyte collection, fixation, cell counting

Hemocytes were collected by gradually injecting in the thorax of cold-anesthetized mosquitoes 10 μL of anti-coagulant media (2 μL at a time) composed of 60% Schneider's insect media, 30% citrate buffer, 10% heat-inactivated fetal bovine serum, final pH 7.0-7.4, sterilized through a 0.22 μm syringe filter. Fire-polished and thin-wall single barrel TW150-6 borosilicate glass capillaries 152 mm long with 1.5 / 1.12 OD / ID in mm were prepared with a Narishige PC-10 needle puller. Needle puller was set in "Heater N.2" mode, with heater level 24.8, and the tip of the needle cut open with fine tweezers a few millimeters from the fine end, leaving an even, clean bore for the injection. Hemolymph was then collected from the lower abdomen where an incision was made with sterile micro-forceps [193]. A total volume of 10 μL was collected per mosquito and collected with a sterile non-stick pipette tip into non-stick Eppendorf tubes coated with silicone to prevent cell attachment. For manual cell counting, 8-12 mosquito samples per condition per experiment were individually placed in sterile single-use disposable hemocytometer slides (Neubauer Improved, iNCYTO C-Chip DHC-N01), and the number of cells counted manually with a light microscope and a 40X objective. Hemocytes were subdivided morphologically into three subtypes (granulocytes, oenocytoids, and prohemocytes), as previously described [4].

For oil-free anti-coagulant buffer injections a custom Tritech Research microINJECTOR system was assembled, featuring a microinjector All-Digital Multi-pressure system (MINJ-D) controller, a precision N2 cylinder pressure regulator for gas pressure control

(TREG-N2), fitted with BS341 cylinder fittings for use in the United Kingdom (TREG-BR580), and a brass straight-arm needle holder (MINJ-4), originally with Swage / Luer fitting. The fitting was then sewn off with an abrasive disk making sure not to dent the thin brass tube. Once access to the outside diameter of the MINJ-4's main brass body tube was gained, the tubing could be pushed into a dual quick-connect fitting with CT-1 tubing compatible with the MINJ-D. When building the system, brass tubing had to be gently pushed a few millimeters, and then when resistance was felt the brass tube was firmly pushed all the way in. Finally, holding the quick-connect fitting outer ring outward, pulling the brass tube out from the quick connect locked it in position. Regulator was set at 20 psi. Hemolymph was collected as above with a steady pressure of 1 psi during injection until 10 μL were collected per mosquito.

Cells were treated with the biomolecule stabiliser and cell fixative vivoPHIX (RNAssist Ltd, Cambridge, UK) which protects RNA, DNA and proteins from degradation within fixed cell, as well as inactivating viruses and killing bacteria. vivoPHIX, developed from a deep eutectic solvent, is non-cross-linking, dissolves fat droplets, and has very low volatility, so that fixed cells can be stored for weeks at room-temperature and months at 4C prior to analysis by scRNA-seq. When fixing hemocytes with VIVOphix cells were collected as above and then plunged into 500 μL of fixative at room temperature. After processing four mosquitoes the cell-fixative mix was pipetted up and down 5 times with a P1000 and well mixed. The procedure was repeated after adding four more samples, or reaching required amounts (8-12 mosquitoes per condition). Hemocytes were then fixed for 2 hours at room temperature, before being transferred to 4C storage. On the day of processing, fixed hemocytes were mixed with one volume of pure, molecular grade ethanol before centrifugation for 30 minutes at 3k RCF at room temperature. Supernatant was discarded and the pellet resuspended in pure molecular grade water before 10X Chromium scRNA-seq library processing. Alternatively, after primary VIVOphix fixation and 60 μm filtering three volumes of glacial acetic acid were added to one volume of fixed hemocyte and well-mixed. After 10 minutes incubation samples were transferred to ice. Then, one volume of pure molecular grade ethanol was added to the mixture and mixed well before centrifugation for 20 minutes at 3k RCF at room temperature.

Supernatant was discarded and pellet resuspended in pure molecular grade water with 0.1% BSA, freshly-prepared, before staining and sorting as below and scRNA-seq processing.

2.6 Hemocyte staining, flow cytometry, and sorting

Hemocytes collected for sorting were stained with 1:10,000 dilutions of 20 mM Hoechst 33342 and 1 mM LIVE/DEAD calcein AM for 15 minutes [150]. Cells were then loaded on BD Fortessa analysers for flow cytometry or BD INFLUX Index Cell Sorter / Mo-Flo XDP Cell Sorter or a Sony SH800 Sorter for sorting into silicone-coated Eppendorf tubes or 96 / 384 well plates with lysis buffer (0.8% Triton-X). Cell populations were determined through physical parameters such as forward scatter (FSC) and side scatter (SSC), as well as fluorescence intensity. BD INFLUX Index Cell Sorter and Mo-Flo XDP Cell Sorter were operated by the Wellcome Sanger Institute Cytometry Facility staff. At first, hemocytes for 10X Chromium experiments were sorted at a concentration of 450 cells per μL into siliconized 1.5 mL Eppendorf tubes. Hemocytes were also stained with Hcs Lipidtox Green Neutral Lipid Stain for quality control by incubating them for 10 minutes with 125 μL of 1:1000 dilution of stock Lipidtox.

Alternatively, Sony SH800 was used to sort VIVOphix hemocytes stained for 20 minutes with 1 drop per 500 μl of sample of NucBlue Live ReadyProbes Reagent (Hoechst 33342 formulation by ThermoFisher). Sony sorter was operated with 100 μm disposable chips. Cells were sorted on fluorescence intensity, with 405 nm laser excitation and Hoechst 33342 filter, gated to exclude negative events with non-stained control. Forward scatter (FSC) and side scatter (SSC) information was also used to exclude doublets and multiplets. Cells were sorted into chilled 1.5 mL Eppendorf tubes before scRNA-seq Chromium 10X library preparation.

2.7 Mouse embryonic stem cell culture

Wild-type E14 mouse ES cells (provided by Kedar Natarajan, Wellcome Sanger Institute) were cultured on gelatin-coated dishes with Knockout DMEM, 15% fetal calf serum, 1× penicillin–streptomycin–glutamine, 1× MEM NEAA, 2-mercaptoethanol, and 1,000 U leukemia inhibitory factor. mESCs tested free of mycoplasma contamination were passaged every 2 or 3 days by Kedar Natarajan. Cells were used to troubleshoot Smart-seq2 scRNA-seq hemocyte protocol, and were sorted as of above with the Sony SH800, 100 μm disposable sorting chip.

2.8 scRNAseq library preparations

2.8.1 Smart-seq2

Hemocytes collected and sorted into 96 and 384 wells plate were processed for Smart-seq2 single cell RNAseq with a modified protocol from Picelli *et al.* [239] Briefly, sorted cells were lysed in pre-made plates with 0.8% Triton-X100, 10 nM pre-mixed dNTP solution, 1-100 μM oligoDts (5′–AAGCAGTGGTATCAACGCAGAGTACT30VN-3′), with 25 μM identified as the best dilution, 2.5% v/v SuperRNAsin. and water, for a total of 4 μL per well. Alternatively, 2μL of RLT buffer, TCL buffer, or Norgen buffer were also used in lieu of Triton-X100. A ten second gentle sonication bath step after 3 minutes denaturation at 72C was used to aid with lysis. For reverse transcription 5.5 μL of reaction mix was dispensed into each well including: 0.29 μL of nuclease free water, 0.06 μL of 1mM $MgCl_2$, 2 μL of 5M betaine, 0.1 μL of 1μM bioTSO (/5Biosg/AGCAGTGGTATCAACGCAGAGTACATrGrG+G), 2μL of RT buffer (5X), 0.5μL of 1mM DTT, 0.25μL of RNAse enzyme (at 20units/μL), and 0.5μL of RT enzyme of choice. RT enzymes evaluated were SmartSCRIBE, Superscrip IV, and Maxima. Plates were placed in thermocycler to carry out the RT reaction: 42 °C for 90 min, then 10 cycles of 2 minutes at 42C followed by 2 minutes at 50C, and finally 70 °C for 15 min, 4 °C hold. This was followed by a pre-amplification reaction using for each well the KAPA HiFi HotStart Ready Mix (12.5 μL) and ISO SMART primer (0.25 μL of 100 μM), plus nuclease free water (2.25 μL). Plates were placed in a thermocycler, with the following PCR program: 98 °C for 3

minutes, then 25 cycles of (98 °C for 20 seconds, 67 °C for 20 seconds, 72 °C for 6 minutes), 4 °C hold. Following PCR, products were cleaned-up with Agencourt Ampure XP beads and RNA quality checked on Bioanalyzer with Agilent High Sensitivity DNA chips. Library preparation was then performed with NexteraXT library prep kit. First, PCR products were tagmented with 2.5 µL Illumina tagmentation buffer and 1.25 µL amplification tagment mix per well of an empty plate, onto which 2 µL of cDNA product for small hemocytes and 1.25 µL for large hemocytes and embryonic stem cells were added. Fifty cells controls were normalized by the dilution with 49 µL of pure water. Tagmentation reaction was carried out for 10 minutes at 55C and then stopped with NT (neutralize tagment) buffer. After adding 2.5 µL of pre-diluted Illumina indexes (S indexes on the column and N indexes on the rows, 10 µL of each per well, which is sufficient for 8 reactions) to each well, 3.75 µL of NPM (Nextera PCR master mix) buffer was also added well-wise, ahead of the Nextera XT PCR thermocycling program (72 °C for 3 minutes, 95 °C for 30 seconds, then 12 cycles of (95 °C for 10 seconds, 55 °C for 30 seconds, 72 °C for 30 seconds), 72 °C for 5 minutes, 4 °C hold. Libraries were then combined and again cleaned up with Ampure XP beads and quality quality-controlled on Bioanalyzer with Agilent High Sensitivity DNA chips before sequencing with paired-end 75 base-pairs read length MiSeq Illumina.

2.8.2 Chromium 10X

After having prepared an appropriate single cell suspension, 10X Genomics Chromium droplet single-cell RNAseq master mix was prepared (and all other steps of the protocol followed) per manufacturer's instructions (CG00052_SingleCell3_ReagentKitv2UserGuide_RevD). Briefly, the RT master mix was quickly made on ice in a sterile pre-PCR UV hood. 50 µL of RT reagent mix, 3.8 µL of RT primer, 2.4 µL of Additive A, and 10 µL of RT enzyme were added for a total of 66.2 µL per reaction. The master mix was dispensed into each well of an 8-tube strip on ice and then the appropriate volumes of nuclease-free water and single-cell suspension were added per manufacturer's recommendations. Then, 90 µL were transferred to row 1 of the 10X Chromium Single Cell Chip. After resuspending the gel beads by 30 seconds

of vigorous vortexing, 40 µL of beads were slowly loaded onto row 2 of the chip. Finally, 270 µL of partitioning oil were added onto the row 3 of the chip. Any unused channels were filled with 50% glycerol in water with the same amounts of above. After covering the loaded and primed chip with a disposable gasket the chip was inserted into the Chromium controller, and the Chromium Single Cell A program allowed to run for 6.5 minutes, generating the droplet emulsion containing encapsulated single cells with hydrogel beads and reagents (gel in emulsion beads, or GEMs). 100 µL of GEMs were then slowly recovered and transferred onto an emulsion-compatible 96-well plate, taking care not to disturb the fragile emulsion. Finally, the PCR plate was sealed with pierceable foil heat seal and loaded onto a thermocycler with the 10X RT program (Step1: 53°C for 45 minutes; Step 2: 85°C for 5 minutes, then 4 °C hold) before -20°C storage for maximum one week before post-GEM-RT clean-up.

In the next step of the Chromium 10X library preparation 125 µL of recovery agent was added to each well without any mixing. A biphasic mixture formed containing a recovery agent - oil pink phase and an aqueous clear phase containing the cDNA. After discarding 125 µL of the recovery agent - oil mix from the bottom of the well without disturbing the aqueous phase, cDNA purification was performed with 200 µL of magnetic Vortex DynaBeads MyOne Silane beads for each sample well. These were prepared as follows: a) 9 µL of nuclease-free water, b) 182 µL of buffer sample cleanup, c) 4 µL Dynabeads MyOne SILANE, and d) 5 µL of additive A. Following two ethanol washes on the magnetic strip, beads were resuspended with 35.5 µL of elution solution I (98 µL of Buffer EB, 1 µL of 10% Tween 20, 1 µL of Additive A). 35 µL of purified GEM-RT products were transferred to a new plate to prepare for cDNA amplification. 65 µL of cDNA amplification reaction mix were added to each well (8 µL nuclease free water, 50 µL amplification master mix, 5 µL cDNA additive, 2 µL cDNA primer mix). Plate was sealed and loaded onto a thermocycler with the cDNA amplification program (98 °C for 3 minutes, then N cycles of (98 °C for 15 seconds, 67 °C for 20 seconds, 72 °C for 1 minute), 72 °C for 1 minute, 4 °C hold). A custom amount of 14 PCR cycles were used, irrespective of the manufacturer's recommendations. PCR products were cleaned with 60 µL of SPRIselect reagent (0.6x) and washed with ethanol before being resuspended in 40.5 µL of

Buffer EB and quantification with Agilent Bioanalyzer High Sensitivity chip. Samples could be stored at this point at 4C for 72 hours or -20C for up to a week. Fragmentation buffer was prepared for each sample with 10 μL of fragmentation enzyme blend and 5 μL of fragmentation buffer. 15 μL were then added onto a new plate and 35 μL of purified cDNA added into each well before placing into a 4C pre-cooled thermal cycler with fragmentation program (Step1, Fragmentation: 32°C for 5 minutes; Step 2, End Repair and A-Tailing: 65°C for 30 minutes, then 4 °C hold). Products underwent a double-sided size selection first by the addition of 30 μL (0.6x) of SPRIselect reagent, then separation with magnetic beads, and finally transfer of 75 μL of supernatant into a new plate. 10 μL of SPRIselect reagent (0.8x) were then added to each sample and 80 μL of the supernatant removed before washing with ethanol and elution with 50.5 μL of buffer EB to isolate the desired products.

Library preparation followed, starting with adaptor ligation. For this, 50 μL of adaptor ligation mix was added to each 50 μL of sample (17.5 μL nuclease-free water, 20 μL ligation buffer, 10μL DNA ligase, 2.5μL adaptor mix) before incubation for 15 minutes at 20C in a thermocycler. Clean-up with 80 μL of SPRIselect reagent (0.8x) followed before ethanol wash and resuspension with 30.5 μL of Buffer EB. Sample index PCR was performed by adding to each well 60 μL of sample index PCR mix (8 μL nuclease-free water, 50 μL amplification master mix, 2 μL of sample index [SI] PCR primers) and 30 μL of post-ligation sample. 10 μL of individual Chromium i7 sample indexes were then also added to each separate well and the plate was placed in a thermocycler with sample index PCR program (98 °C for 45 seconds, then N cycles of (98 °C for 20 seconds, 54 °C for 30 seconds, 72 °C for 20 seconds), 72 °C for 1 minute, 4 °C hold). The optimal number of cycles must be determined by balancing the need to obtain enough material for sequencing and lowering PCR amplification biases. Manufacturer's instructions based on post-cDNA amplification quantification were followed. Post-sample index PCR double sided size selection was performed first by the addition of 60 μL (0.6x) of SPRIselect reagent, then separation with magnetic beads, and finally transfer of 150 μL of supernatant onto a new plate. 20 μL of SPRIselect reagent (0.8x) were then added

to each sample and 165 μL of the supernatant removed before washing with ethanol and lastly elution with 35.5 μL of buffer EB to isolate the desired products onto a new plate. Samples were then quantified with Agilent Bioanalyzer High Sensitivity chip and stored at 4C for 72 hours or -20C for long-term. Library preparation after RT reaction and cDNA production were performed either by the WTSI Research & Development Department or by Bespoke Low-Throughput Pipelines staff, also at the Wellcome Sanger Institute.

2.9 Sequencing

Sequencing and QC were performed by Sanger Bespoke Sequencing team. For Smart-seq2 samples, MiSeq 150PE (using kit version 2): libraries were run on the Illumina MiSeq instrument with standard protocols using a 300-cycle kit set to a 150pb paired-end configuration. Libraries supplied at 2.8 nM and loaded with a loading concentration of 8 pM. For Chromium 10X V1 and V2 kits, HS2500 rapid (using kit version 2): libraries were run on the Illumina HiSeq 2500 instrument set to Rapid Run Mode with standard protocols using a 200-cycle kit set. Libraries supplied at 2.8 nM and loaded with a loading concentration of 8 pM. For quality control, lanes passed QC if tags were decoded appropriately, reference matches were as expected either *A. gambiae* or *A. aegypti*, quality metrics met in-house expectations, other run metrics such as error rates were as expected, and yield expectation was met (given the number of cycles run and/or platform expectations). The data was then fit to the sequencing requested and any significant deviation from expected explained and appropriately annotated. For assessment two main pieces of software were used. Sequencing analysis viewer (SAV) was used to assess the instruments' performance. The summary tab gave statistics for the whole run in question whereas the 'analysis and imaging' tabs allowed QC to delve deeper and assess if the lanes have performed as expected across all the cycles of the run. NPG pages was used both for staff analysis and annotation, and user's visualisation of data. NPG is an in-house bespoke analysis/software package to include tag analysis, reference matching/mapping details and contamination which is the final point where lanes or tags in the run either passed or failed QC.

3 Results and Discussion

3.1 Establishing an experimental system for scRNA-seq of hemocytes

The first hurdle was to prepare a pure single cell suspension of hemocytes compatible with scRNA-seq. First, Smart-seq2 was attempted as it produces full-length sequences with the highest reads per cell count and it allows index sorting to correlate expression data with the size and granularity of each cell. To perform Smart-seq2 we first needed to develop an ability to sort mosquito hemocytes, which at the time had not been done.

3.1.1 Hemocytes are activated by systemic LXA_4 and PGE_2 injection and *P. berghei* infection

To optimize sorting, I first established whether flow cytometry patterns could be used to investigate changes in hemocyte populations in response to different treatments. To distinguish hemocytes from other cells which may accidentally get dislodge during hemolymph collection, mosquitoes were micro-injected 24 hours prior to immune stimulation with 69 nL of CM-DiL at 140 μM. CM-DiL is a lipophilic dye that exclusively stains hemocytes (especially granulocytes and oenocytoids) for reasons that are not completely understood [4, 146, 150]. Manual counting of *Anopheles* hemocytes was compared with flow cytometry profiles after mosquito injection of LXA_4 or PGE_2 [Fig. II.2 A-C], or mosquito feeding of an infectious blood meal containing *P. berghei* [Fig. II.2 D-F]. 24 hours after an infectious blood meal the proportion of granulocytes observed microscopically increased from ~1-2 % in control mosquitos to ~5-8 % in challenged mosquitoes, consistent with previous results [17, 193, 194] [Fig. II.2 A, D]. In parallel, flow cytometry detected an increase in large, dye-positive events [Fig. II.2 E-F]. Similar increases were found 24 hours after injection of eicosanoids [Fig. II.2 A-C]. These results confirm the ability of eicosanoids or *P. berghei* infection to change the proportion of mosquito hemocyte cell types [193, 194].

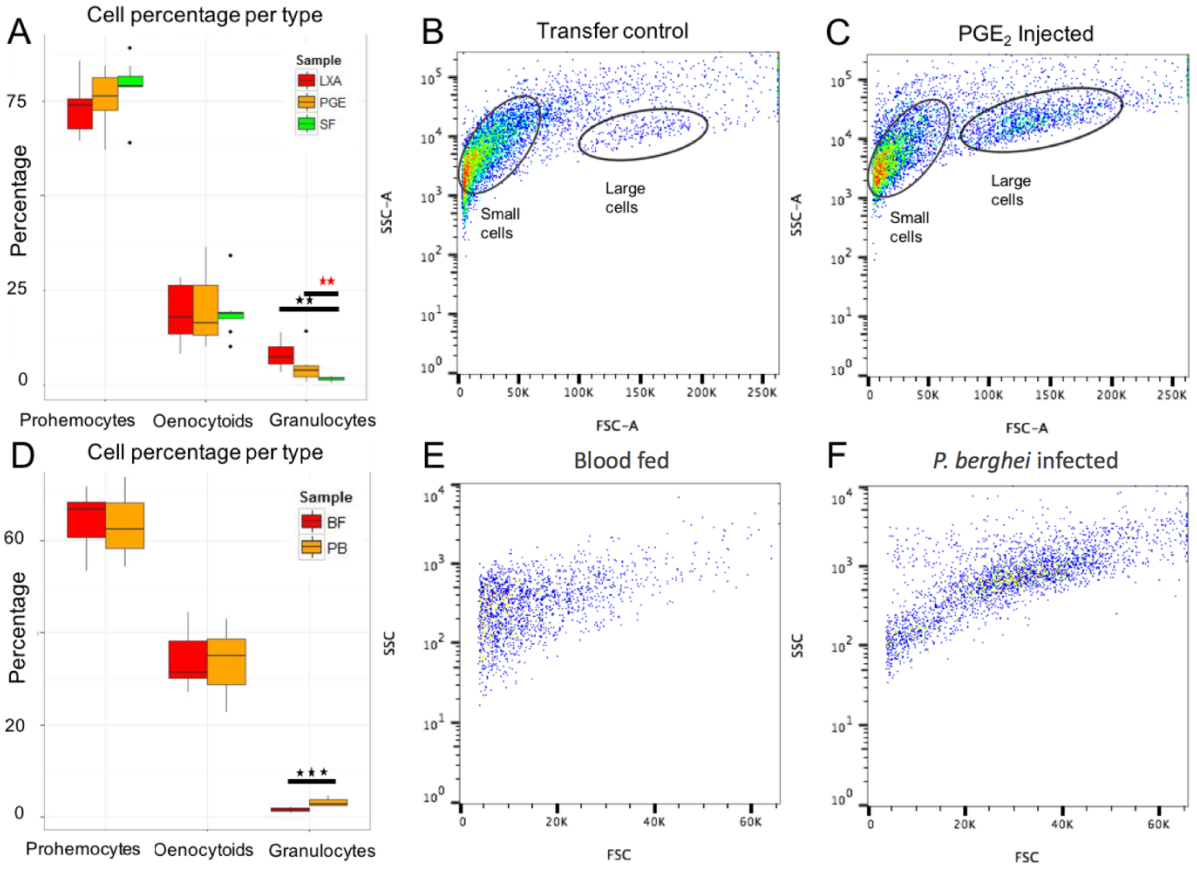


Fig. II.2 Effect of PGE_2, LXA_4 and *P. berghei* infection on hemocyte types. All events are CM-DiL positive. **(A,D)** Manual counting of cells by morphological differentiation into prohemocytes (small), oenocytoids (medium size and round), granulocytes (large and more complex) in PGE_2 and LXA_4 **(A)** and *P. berghei* **(D)** experiments. Two black asterisks: $p < 0.01$ with Student's T-test; three black asterisks: $p < 0.001$. Two red asterisks: $p < 0.01$ with Mann-Whitney test, (PGE_2 data not in normal distribution). **(B-C, E-F)** Pseudo-coloring shows event density: from light blue to green and red with higher events density. **(B,C)** Flow cytometry forward scatter (size) vs. side scatter (granularity) of **(B)** control and **(C)** hemocytes after PGE_2 injection. **(E,F)** Flow cytometry forward scatter (size) vs side scatter (granularity) of **(E)** control and **(F)** hemocytes from mosquitoes exposed to *P. berghei*. Gate labelled 'large cells' represents putative granulocytes. Experiments include four biological replicates.

However, when analyzing mosquito hemolymph content by flow cytometry, large quantities of non-cellular material were apparent as a sigmoid-shape collection of events [Fig. II.3A]. Non-cellular objects consist in part of droplets of mineral oil, required by the injection apparatus used to flush hemolymph out of mosquitoes. Silicone coating of Eppendorf tubes to prevent adherence of activated hemocytes also contributed to the background noise. Since CM-DiL also stained small debris and oil particles, hiding small hemocyte populations, I set-out to develop an improved sorting scheme to separate hemocytes from background. Additionally, injecting mosquitoes has the potential to pre-activate the immune and wound response systems, altering baseline mosquito conditions, and should ideally be avoided.

3.1.2 Hemocytes can be isolated via FACS with Hoechst 33342 and calcein AM dyes

The live sorting protocol was optimized by using Hoechst 33342 (Hoechst) and acetoxymethyl (AM) ester of calcein [Fig. II.3A-C]. Hoechst is a cell-permeant nuclear dye part of the bis-binzimide family, used to stain DNA. Calcein is used as a LIVE / DEAD discrimination agent as it can first permeate cells in a non-active form, but is then cleaved by intracellular esterases, resulting in a charged compound that cannot easily cross plasma membranes and is strongly fluorescent. If a cell is dead, it will not convert calcein into its active form, or the chemical will flow out of the cell's damaged cell membrane. Most importantly, cells can be directly stained after collection for 15 minutes, which avoids the need for dye injection. Calcein alone or a combination of Hoechst and calcein successfully distinguishes between cells and debris [Fig. II.3C]. Interestingly, while most calcein positive cells also stained positive for Hoechst, there were calcein negative non-autofluorescent events that were highly Hoechst + [Fig. II.3B-C]. The strong nuclear staining featured intensities that are multiples of each other, suggesting the tantalizing possibility these are small, replicating, polyploidy small hemocytes such as prohemocytes. Interestingly, polyploid populations were seen by another laboratory, albeit with different experimental set-up, collection methods, and nuclear staining [147, 149]. However, this possibility was not followed up further as I decided to continue optimizing hemocyte collection protocols. This new sorting protocol did solve some of the past issues with

CM-DiL. First, there was no need to inject a dye and hence run the risk of prematurely activating immune cells, and second, calcein AM appeared more specific and did not stain background debris. However, the issue remained of low RNA stability during FACS and 10X preparation for these highly active live immune cells.

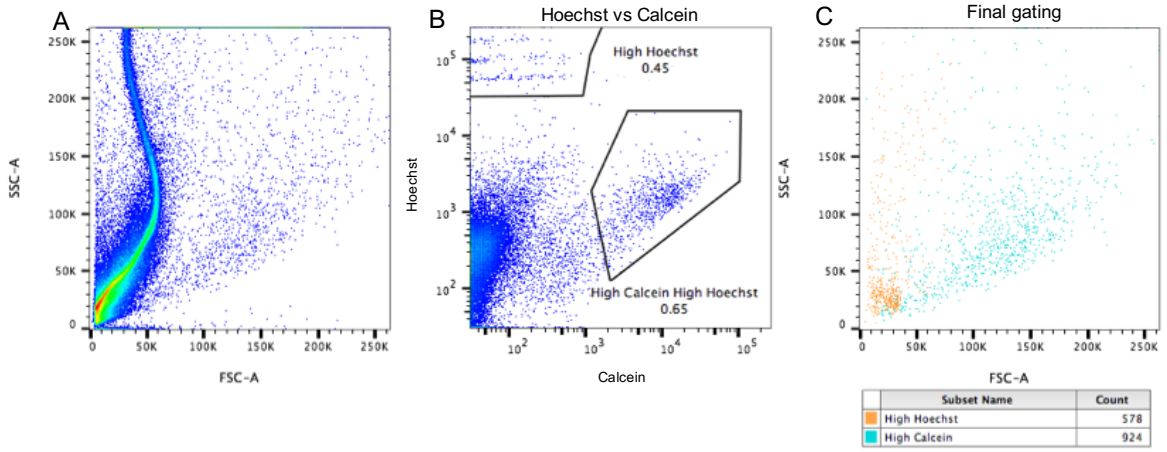


Fig. II.3 Sorting and flow cytometry analysis of hemocyte with Hoechst and calcein AM. **(A-B)** Pseudo-colouring refers to event density, going from light blue to green and finally red with higher events density. **(A)** Events on forward scatter (FSC, size) vs. side scatter (SSC, granularity). **(B)** Separation of Hoechst 33342 vs. calcein AM reveals calcein+/Hoechst+ and Hoechst+ only cell populations **(C)** Final sorting scheme. Granularity (SSC) vs size (FSC) of live calcein + cells (red) and Hoechst 33342 + cells.

3.1.3 With hemocyte fixation and pneumatic collection sorting becomes redundant

Typical hemocyte collection is laborious, time-consuming, and untidy. It involves filling with mineral oil pre-pulled needles secured onto a micromanipulator, as well as the tubing connecting these needles with a manual dispenser. The needle is prone to breaking, and inserting the needle into the tubing can be challenging. Furthermore, injection media can mix with the oil even after only one or two needle refills, meaning the whole set-up has to be replaced. Particularly during time-sensitive experiments involving multiple conditions and batches of mosquitoes this is not feasible. A custom oil-free injector was then developed to

displace the hemolymph and collect hemocytes in a cleaner, faster, more efficient manner. Details are in the methods section (Chapter 2.3), but briefly, this new methodology avoided mixing between oil and anti-coagulant media, providing investigators with the certainty that all cells or flow events observed after collection were endogenous to the mosquito (Fig. II.4.B). Furthermore, the injection needle did not need replacing as often as with manual collection, and refilling was rapid thanks to the negative pressure vacuum function of the custom injector. However, the FACS fat droplets issues partially persisted (Fig. II.4.A-B), meaning exogenous oil droplets and silicone particles were not the sole issue, and endogenous fat droplets also played a role.

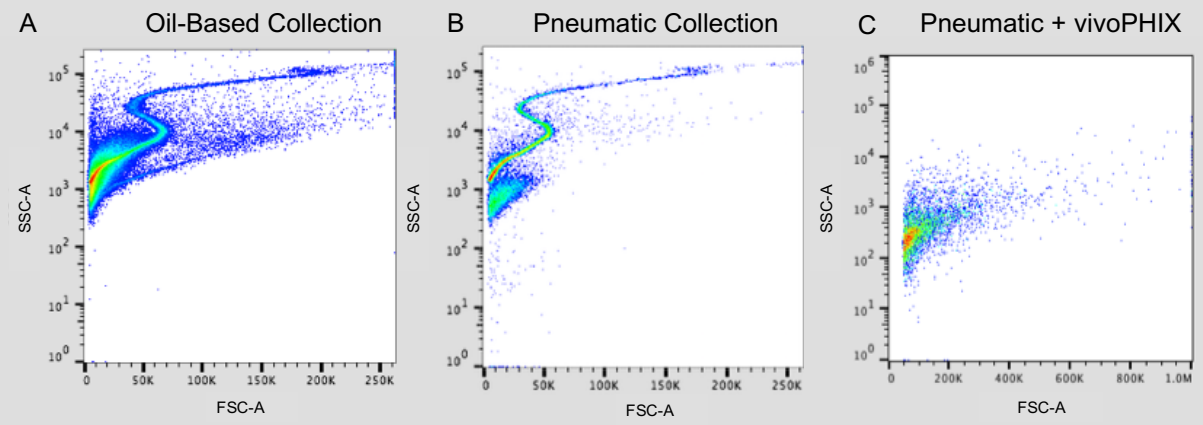


Fig. II.4 Hemocyte isolation optimisation with FACS. (A-C) Pseudo-colouring refers to event density, going from light blue to green and finally red with higher events **(A)** Standard oil-based collection **(B)** Custom oil-free pneumatic injector system and **(C)** Pneumatic injector plus cells fixed and processed with vivoPHIX.

To solve this issue, we used a novel, non-crosslinking, non-chaotropic agent called vivoPHIX, which fixes cells and preserves RNA while maintaining cellular morphology. Preliminary experiments showed vivoPHIX-treated samples were purer when compared to oil-based or pneumatic hemocyte collection systems. The new protocol made FACS redundant, as little background or debris are present in the cell suspension mix. Hemocytes were fixed with vivoPHIX, and resuspended in pure molecular grade water after ethanol mixing and density centrifugation (3k RCF, 20 minutes, room temperature). We found they had a clean FACS profile devoid of the sigmoid fat droplet curve of previous samples [Fig. II.4C] on a Sony

SH800 with 100µm disposable chip. The vivoPHIX sample showed an almost identical profile (4889 events) to hemocytes collected pneumatically and sorted based on calcein expression (6615 events), further demonstrating how vivoPHIX fixation and resuspension is optimal to both collect as many cells as possible and also decrease manipulation-induced stress responses in the transcriptome [Fig. II.5A-B].

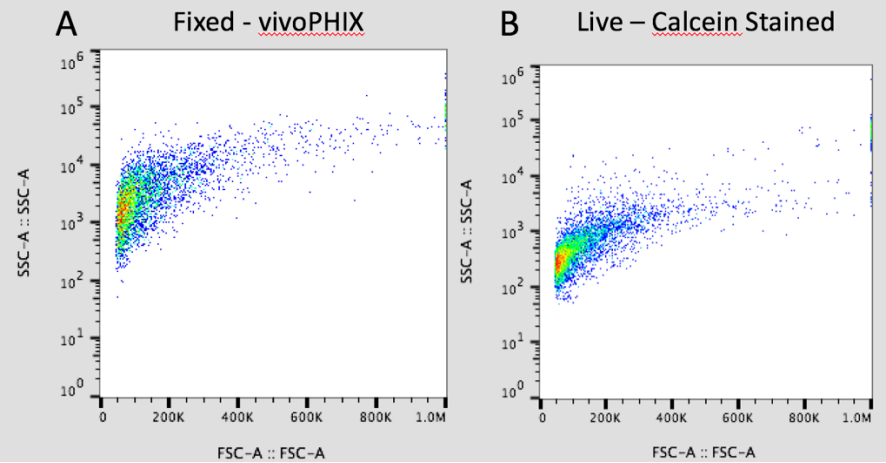


Fig. II.5 FACS of vivoPHIX vs. live calcein-stained hemocytes. (A-B) Pseudo-colouring refers to event density, going from light blue to green and finally red with higher density. Both samples were prepared with the pneumatic injector **(A)** Hemocytes fixed with vivoPHIX as of methods protocol, total of 4889 events **(B)** Live hemocytes stained for 15 minutes with calcein and sorted, for a total of 6615 events

In addition, centrifugation of vivoPHIX fixed cells with 100% ethanol before water resuspension effectively removed most adipocyte contamination, as demonstrated by a direct comparison with 70% ethanol spinning (Fig. II.6A-B). Here, hemocytes were stained for 10 minutes with a 1:1000 dilution of stock LipidTox, followed by sorting on Sony SH800 with 100 µm chip. Only the 70% ethanol sample showed the presence of highly LipidTox+ cells (fat). Due to the efficiency of the sort, the similar FACS profile compared to live calcein+ cells, and the added benefit of immediate fixation and RNA preservation, vivoPHIX fixation was used for all following experiments in chapters III and IV.

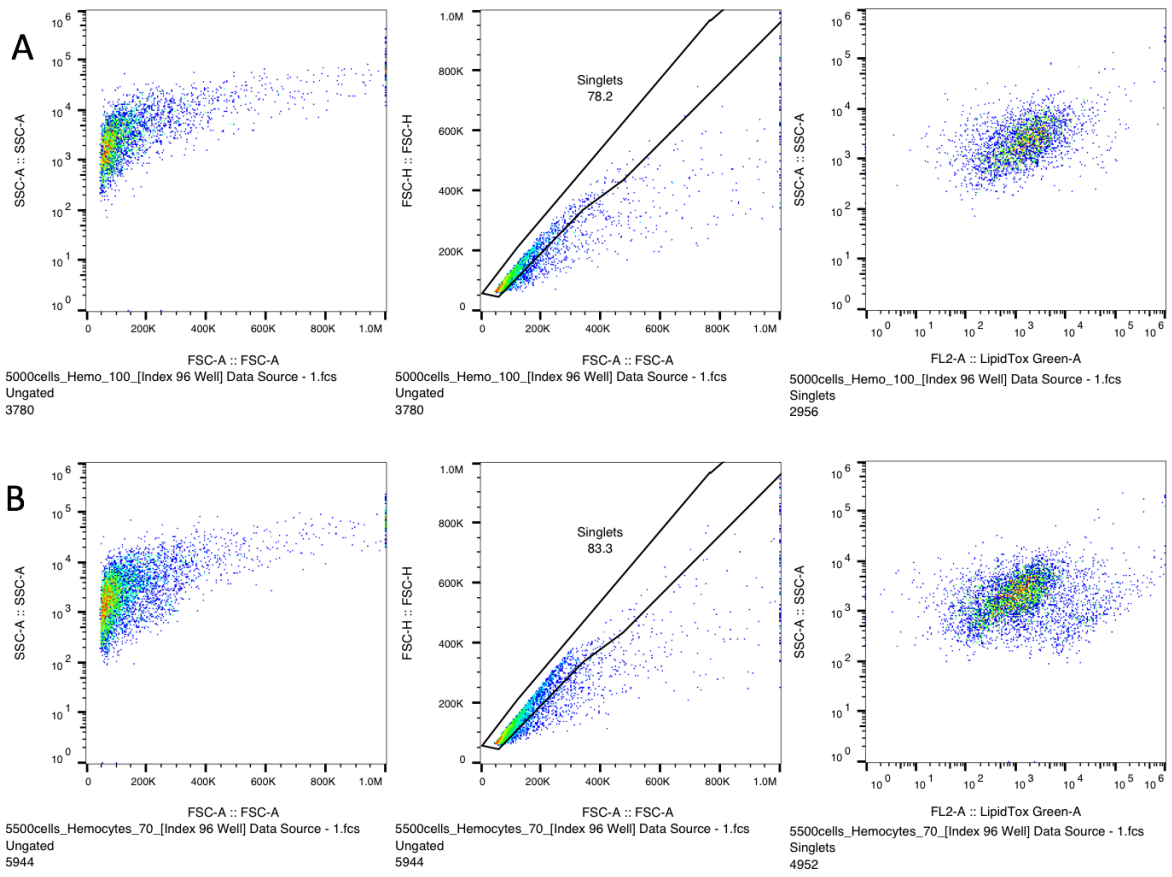


Fig. II.6 FACS analyses of vivoPHIX fixed cells stained with LipidTox show few adipocytes with 100% ethanol spin-down. To the left, side scatter area (granularity) vs. forward-scatter area (granularity). In the middle, forward scatter height (size) vs. side-scatter area (granularity), where the straight-line indicates singlets. To the right, LipidTox fluorescence **(A)** 100% ethanol spin down samples stained with LipidTox show few + events **(B)** Conversely, 70% ethanol spin down samples stained with LipidTox demonstrate a lower purity with increased LipidTox+ events.

3.1.4 Secondary fixation with vivoPHIX-SC

Since inhibition of RNAses by vivoPHIX is reversible, I worked with its inventor, Dr. Andrew Goldsborough (University of Bordeaux), to test new formulations to permanently deactivate the enzymes through the addition of a secondary fixation step where a strong acid (glacial acetic acid) was combined with standard vivoPHIX (vivoPHIX-SC). We hypothesized permanent deactivation could lead to higher transcripts per cell counts by preventing endogenous RNAses from degrading RNA when cells were resuspended in water. In addition, a Hoechst 33342 stain and nuclear sorting step was added to the protocol to precisely quantify the number of hemocytes loaded onto the 10X platform. After fixing, staining and sorting as of the modified protocol indicated in section 2.4 cells were resuspended in pure molecular grade water plus 0.1% BSA, stained with Hoechst 33342, and sorted on Sony SH800 before scRNA-seq processing. Importantly, cellular morphology was well-maintained, and sorting efficient [Fig. II.7A-B]. After combining three biological repeats for each condition a total of 6160 cells from bacteria-infected *Aedes* hemocytes, 5460 cells from serum-fed LacZ dsRNA-injected *Aedes* hemocytes, and 8462 cells from serum-fed control *Aedes* hemocytes were sorted. Cells were sorted on Hoechst 33342 + cells after gating auto-fluorescence on a non-stained control sample containing a mixture of all three conditions. Following sorting all cells were loaded onto the Chromium 10X chip for scRNA-seq library preparation (see section 3.3 below for scRNA-seq results).

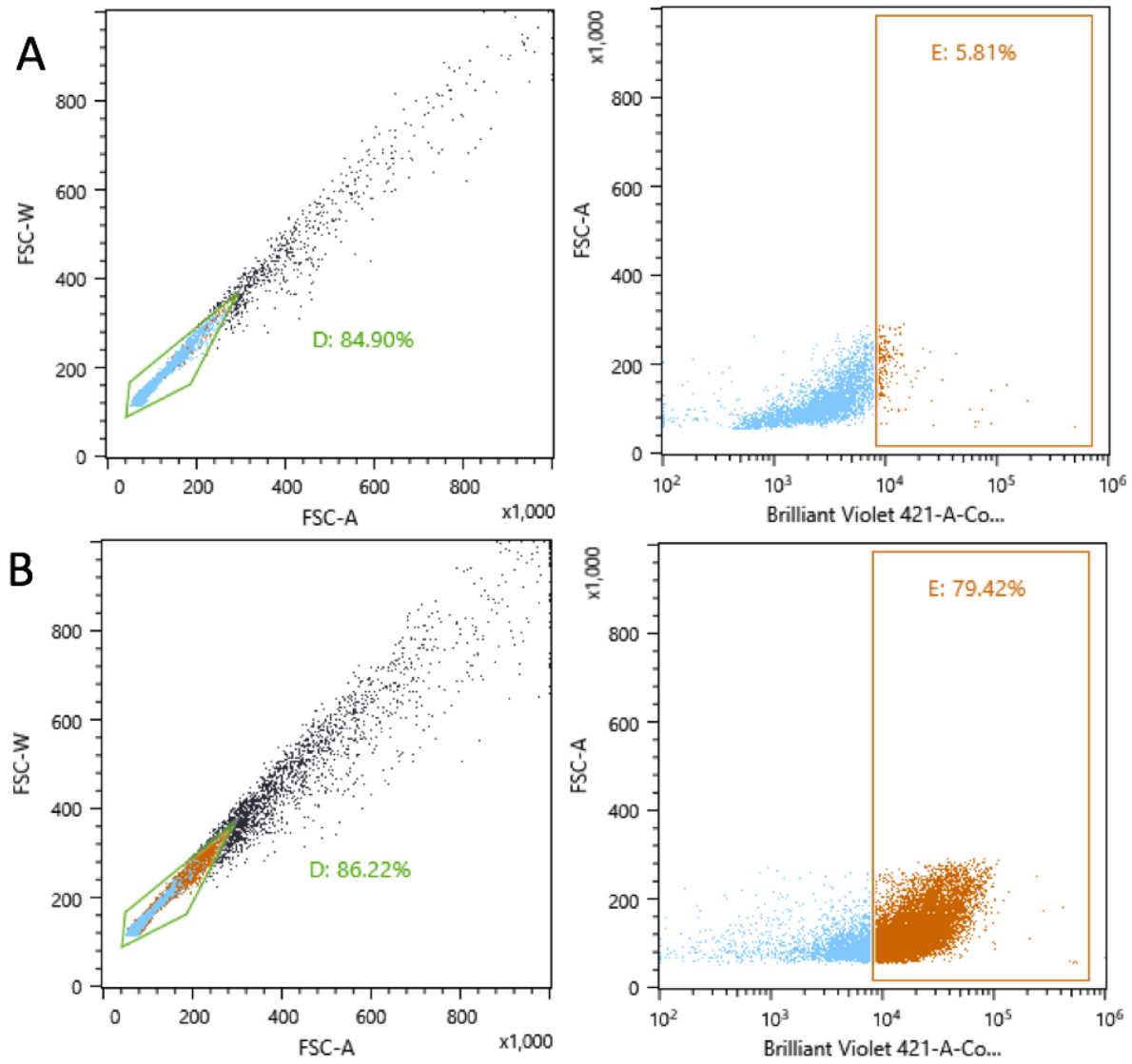


Fig. II.7 FACS of vivoPHIX double-fixed and Hoechst 33342 stained hemocytes. Cells fixed with vivoPHIX followed by secondary fixation with acetic acid and stained in water for 15 minutes with NucBlue Reagent as of methods section **(A)** Non-stained control. To the left, forward scatter (size) vs side-scatter (granularity), where the straight-line indicates singlets. To the right, gating to exclude auto-fluorescence **(B)** Representative stained sample (hemocytes from serum-fed *Aedes* mosquitoes). To the left, forward scatter (size) vs side-scatter (granularity), where the straight-line indicates singlets. To the right, gating to include only DAPI+ cells that we were certain to be positive singly-nucleated cells.

3.2 Smart-seq2 scRNA-seq in mosquito hemocytes

The next step after making a single-cell suspension with flow cytometry or other methods is scRNA-seq library preparation. The first technique we attempted was a modified Smart-seq2 protocol obtained by Dr. Hayley Bennett at the Wellcome Sanger Institute (see methods section for details). Hemocytes from non-challenged, sugar-fed mosquitos injected with CM-DiL were index sorted into single wells of 96-wells plates with Smart-seq2 lysis buffer. Success rate was then determined as the percentage of wells with quantifiable cDNA on Bioanalyser traces. One plate contained large cells ("large cells" in Fig. II.4-B) and one small cells/events ("small cells" in Fig.II4-B). The first experiments were disappointing, with only a 14% overall success rate in our 96-well plates. Of this, large hemocytes had a 28% success rate while small hemocytes were a complete failure (0% wells with quantifiable cDNA). A duplicate experiment had similar results, with only 19/40 single cells successfully sequenced for the large cells plate, and only 2/90 small cells. A Nextera XT scRNA-seq library was nevertheless successfully prepared from these 21 cells, showing that when sorting, lysis, and RT did work high-quality data could be obtained. Of the original 21 cells 18 yielded >50% reads mapped to *A. gambiae*. Another 3 cells were eliminated by manual QC (minimum of 130,000 reads per cell, at least 1000 genes per cell, not more than 30% of total reads mapping to mitochondrial genes) or automated QC (scater package default settings)[320]. I then used scater's normalisation strategy, cell-wise relative log expression (RLE, or size factor). Following QC, 15/21 cells were retained for further analysis, overall expressing 5621 genes of *A. gambiae*, with a median of 2100 genes per cell. We identified two main clusters of cells with similar transcriptomes and a few outliers using SC3 (Single-Cell Consensus Clustering scRNAseq analysis package) [Fig. II.6]. Most genes expressed were characteristic of hemocytes or typically involved in immunity and wound responses, indicating the correct cells had been isolated [Fig. II.7].

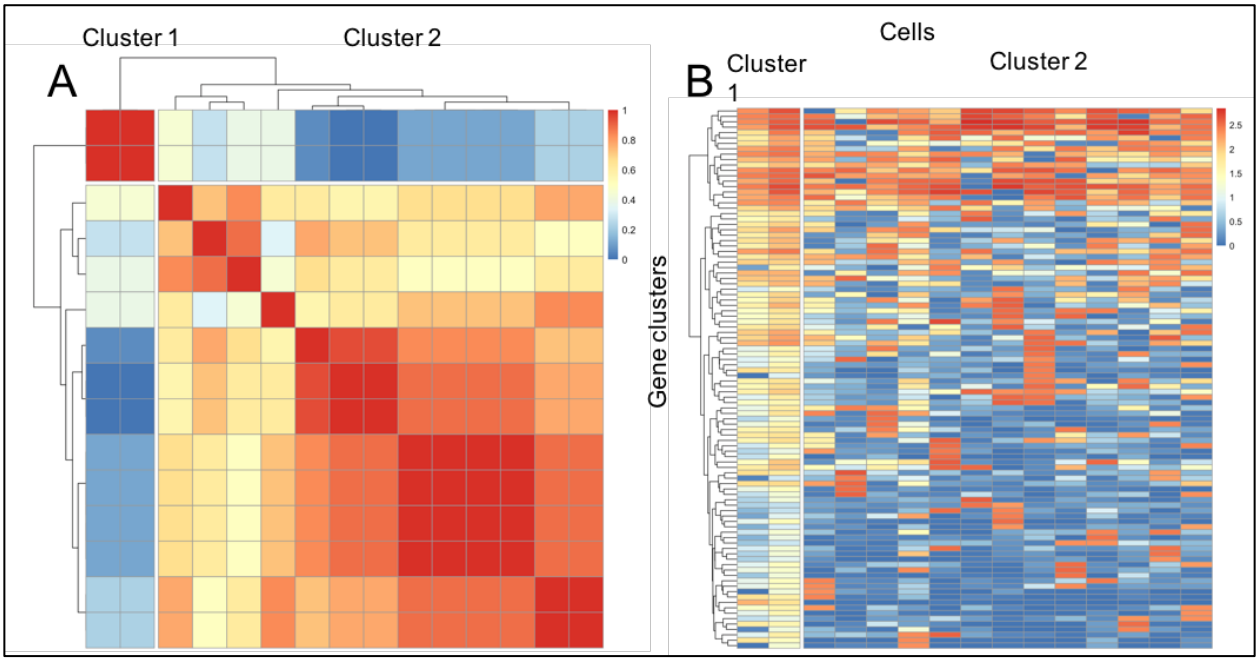


Fig. II.8 scRNA-seq with Smart-seq2: hemocytes cluster into two main groups. All data shown after QC and normalization. **(A)** Cell clustering with Single-Cell Consensus Clustering (SC3) [289]. Matrix shows percentage of times cells were assigned to the same cluster by different parameter combinations, with dark red (1) indicating assignment to same cluster every time and dark blue (0) indicating cells never assigned to the same cluster. White lines are visual guides separating clusters. SC3 outputs most likely clustering with k = 2 clusters. Normalized and QCed expression matrix with cells at columns and genes in rows is taken as input. Genes are filtered to remove ubiquitous or extremely rare genes and reduce matrix dimensions. Distance between the cells is calculated using Euclidean, Pearson and Spearman metrics to build distance matrices, which are then transformed by principal component analysis (PCA) or eigenvectors calculations. k-means clustering is calculated on the first x eigenvectors with the R function k-means with Hartigan-Wong algorithm [321] . Red is similarity among cells. **(B)** Cell expression matrix with SC3. Figure represents input expression matrix with clusters of genes in rows and cells in columns, after gene filtering as above. Genes clustered with SC3 package by *k*-means with $k = 100$ (as seen by dendrogram on the left). After log2-scaling, heatmap shows expression levels of gene cluster centers.

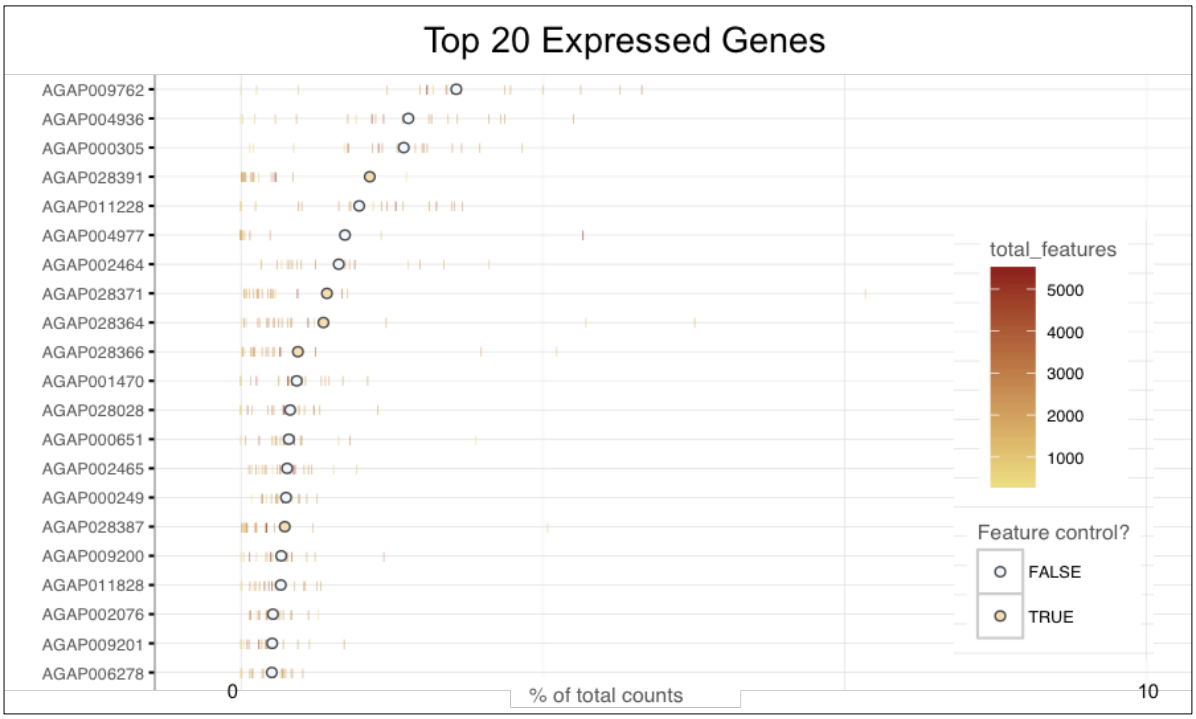


Fig. II.9 scRNA-seq from Smart-seq2: top expressed genes. Note the relatively flat distribution which typically indicates good coverage. Yellow (feature control) are mitochondrial genes. Proportion of reads mapping to mitochondrial genes may are useful for identifying low-quality cells as in broken cells cytoplasmic RNA leaks out, while mitochondrial RNAs are preserved [269]. Round circles represent the average expression across all cells while each colored bar is a single cell (painted by total genes). Top 50 genes accounted for ~38% of the counts. Plot and calculations performed in R with the scater package. The top 10 genes were hemocyte or immunity related. Apart from mitochondrial genes, they were AGAP009762 (Nimrod), AGAP004936 (hemocyte-specific) [204], AGAP000305 (hemocyte-specific) [205], AGAP011228 (fibrinogen), AGAP004977 (phenoloxidase) [204, 322], and AGAP002464 and AGAP002465 (ferritin) [204], AGAP001470 (hemocyte-specific) [202], AGAP000651 (actin 5C) [323], AGAP028028 (leucine-rich immune protein) [202, 324]. Annotations from VectorBase.

Light and fluorescence microscopy confirmed that sorting on CM-DiL did not distinguish small cells from debris or oil well enough, explaining why few cells were sequenced by Smart-seq2. most small events were not cells but debris or oil, explaining the small number of successful cells (not shown). Incomplete lysis may have contributed to the low yield since by light microscopy lysis was only 60-70% efficient after 5 minutes in Smart-seq2 lysis buffer at 4°C. Smart-seq2 on mouse embryonic stem cells was also performed. These cells are rich in RNA, and thus perfect as positive controls. Live sorted mouse E14 WT embryonic stem cells showed 100% cDNA amplification efficiency, with 38/38 positive single cells and 1/1 positive 50 cells control (both 100%), confirming our Smart-seq2 protocol worked. The main challenge lied in the inherent characteristics of mosquito immune cells.

To increase overall cDNA amplification efficiency a 10 seconds sonication step was added to aid cellular lysis. In addition, we set up a dilution series from 1/2 to 1/100 of the original oligo(dT) concentration, as too high a concentration can inhibit the RT [Table II.1]. In duplicate experiments, cDNA amplification efficiency increased from 17% to 45%. We thus showed the optimal oligo(dT) concentration to be between 5 µM and 10 µM. The RNA of larger hemocytes was still marginally easier to reverse transcribe and amplify, but we were also able to capture small hemocytes (56% vs 44% of cells sorted).

OligoDTs	100 µM	50 µM	10 µM	5 µM	1 µM
Success rate to cDNA	3/18 (17%)	7/19 (37%)	9/20 (45%)	24/52 (46%)	3/18 (15%)

Table II.1 Optimisation of oligo(dT) concentrations, Smart-Seq2. Percentage of wells with quantifiable cDNA after Smart-seq2 library preparation. Numbers indicate successfully amplified wells with single cells over the total wells sorted.

Furthermore, after a preliminary comparison of lysis buffers (0.8% Triton X-100, RLT buffer, TCL buffer, Norgen buffer) and RT enzymes (SmartSCRIBE, Superscript IV, Maxima), the 0.8% Triton X-100 / SmartSCRIBE combination was confirmed as the most efficient, with a cDNA amplification success of just under 50% as in the experiments above.

In parallel, we directly compared vivoPHIX fixed hemocytes and CMDiL live sorted hemocytes. Fixed cells showed comparable (if slightly higher) cDNA amounts [Fig. II.10A-B]. Smart-seq2 controls (50 single-cells sorted, lysed, and reverse transcribed) from fixed cells featured three times higher cDNA levels than live cells [Fig. II.10C]. The difference was likely due to increased sorting of cells, rather than debris or vesicles.

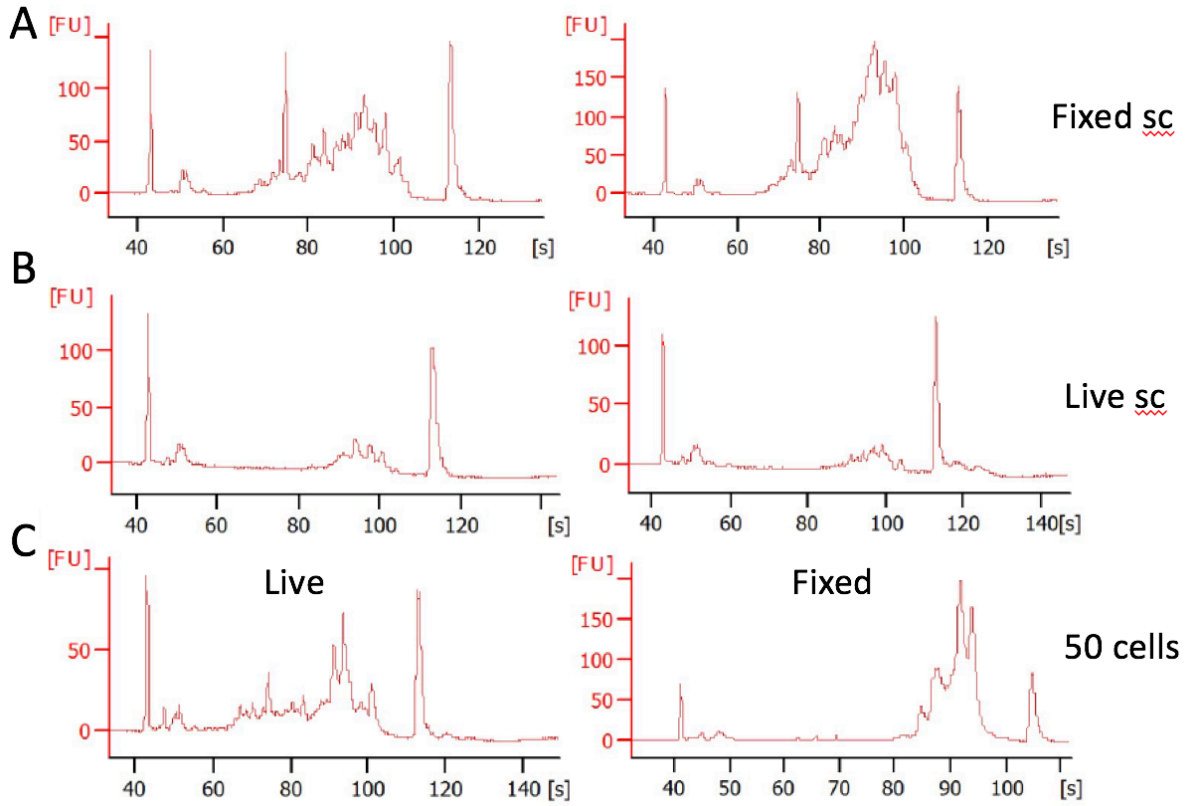


Fig. II.10 Bioanalyser traces from fixed and live hemocytes after RT. Heightened [FU] readings indicate higher amounts of cDNA **(A)** To the left and right representative examples of cDNA traces from vivoPHIX-fixed single cells **(B)** To the left and right representative examples of cDNA traces of live-sorted hemocytes **(C)** To the left, cDNA traces after RT of 50 live-sorted hemocytes. To the right, cDNA traces after RT of 50 vivoPHIX fixed hemocytes. Abbreviation: sc (single cell).

A 50% cell capture efficiency is not optimal, but it could have been sufficient for low-throughput scRNA-seq of mosquito hemocytes. Indeed, I collected the cDNA from the 69 single cells and positive controls for which RT and cDNA amplification worked and prepared

and sequenced a library with rapid-run Illumina Hiseq2500. Hence, in total we gathered information on 90 single cells and positive controls through Smart-seq2. However, parallel Chromium 10X scRNA-seq technology optimisation was successful, and hundreds of cells per run could be analysed, albeit with a lower genes-per-cell count. We thus focused on Chromium 10X. Nevertheless, all cells successfully prepared with Smart-seq2 were analysed together, after filtering out cells with did not have a majority of reads matching the transcriptome of *A. gambiae*. Reads for positive cells were then aligned with STAR, using the AgamP4.9 annotation. Thirty-nine cells from the latest library were successfully sequenced, in addition to the 22 cells from the previous library, for a total of 61 cells. After processing and QC (filtering cells with > 100 features and < 30% mitochondrial reads) 48 cells were left, with a mean gene count of 1194 genes per cell and mean mitochondrial gene content of 5.7%.

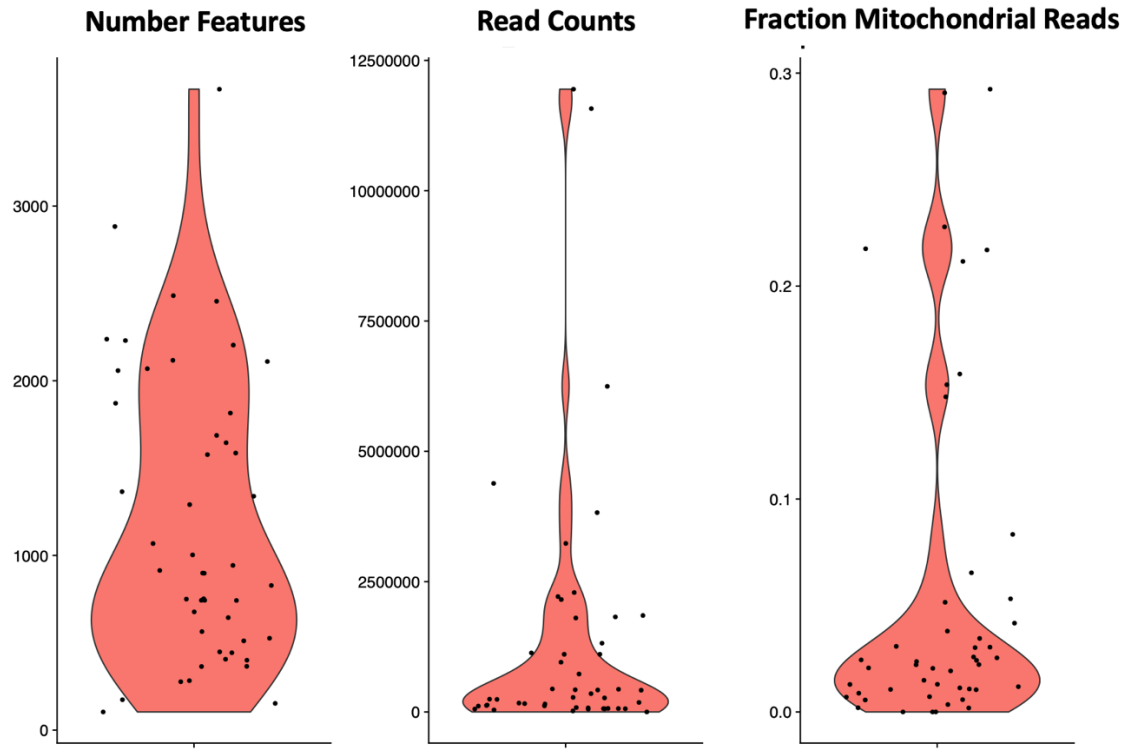


Fig. II.11 scRNAseq QC (Smart-seq2) with Seurat. Violin plots showing QC metrics for both Smart-seq2 libraries combined. To the left total number of features per cell. In the middle total number of reads per cell and to the right the ratio of total reads in mitochondrial genes.

3.3 Chromium 10X scRNA-seq in mosquito hemocytes

In preliminary experiments live hemocyte were either loaded onto the Chromium 10X chip after CM-DiL staining and collection with a manual oil injector, or after calcein staining and collection with the pneumatic oil-free injector. In addition, we also tested vivoPHIX-fixed hemocytes. More recently, we tested vivoPHIX-fixed hemocytes with acetic acid double fixation and Hoechst 33342 sorting (see methods). The latter protocol produced a higher cell yield, and more genes per cell were detected [Table II.2].

	Live CM-DiL sorted hemocytes	Live calcein-sorted hemocytes	vivoPHIX-fixed hemocytes	Double-fixation hemocytes
Number of cells	291	573	459	936
Genes per cell	61	677	780	947
Total genes detected	n/a	7320	7186	11650

Table II.2 Summary of Chromium 10X scRNA-seq metrics. See methods chapter for details. Double-fixation hemocytes includes standard vivoPHIX fixation, post-fixation with acetic acid, and Hoechst 33342 staining and sorting ahead of Chromium 10X chip loading. Metrics refer to Cell Ranger pipeline outputs, before Seurat QC.

FACS sorting of hemocytes stained with CM-Di resulted in a suspension of 450 cells / μL. Loading ~ 1,000 hemocytes onto the Chromium 10X chip produced a low recovery in the first pilot experiment, returning 113 to 291 cells per technical repeat after QC, with 29-96 median genes per cell (Cell Ranger). Manual analysis with Seurat confirmed the low number of genes and UMIs. Multiple factors could have been at play: a) improper alignment or other software errors, b) poor cell quality / high cell death due to sorting scheme and wait times, c) low transcript capture rate, with selected amplification of just a few transcripts. The difficulties experienced with Smart-seq2 suggested hemocytes are exceedingly difficult to lyse and easily damaged during sample preparation and sorting. In addition, a more in-depth analysis of Cell Ranger output suggested even less cells than hypothesised had been detected. Total cells were likely ~50, with a genes per cell count of ~200.

Learning from our experience with hemocyte isolation, sorting, and Smart-seq2 processing, Chromium 10X sample preparation was improved by using the oil-free pneumatic hemocyte collection system and by fixing hemocytes in vivoPHIX or sorting calcein+ cells. Two higher quality libraries were prepared. Results were comparable between the two conditions [Fig. II.12]. Granulocytes are the largest hemocytes and the most fragile. The results may indicate an improved ability of vivoPHIX fixation to preserve larger, RNA-rich cells. By avoiding the use of silicone coating, reducing preparation time with live cells (with consequent cell damage and RNA degradation), and collecting cells directly into vivoPHIX without losing material by sorting we have developed a quick, efficient, and scalable method to explore the cellular heterogeneity of the immune system of a mosquito at single cell resolution. This was the protocol used to process all of our *Anopheles* samples.

However, following up on the improvements in cell sorting and RNA preservation (using vivoPHIX and acetic acid secondary fixation), we tested whether the updated fixation protocol could improve 10X Chromium library preparation, further increasing cell counts and genes per cells counts. Dr. Ana Beatriz Ferreira prepared three *Aedes* hemocyte samples in vivoPHIX as described in the methods and shipped them to the Wellcome Sanger Institute. Here, I first combined the three repeats and then sorted and loaded onto Chromium 10X 6160 cells from bacteria-infected *Aedes*, 5460 cells from serum-fed LacZ dsRNA-injected *Aedes*, and 8462 cells from serum-fed control *Aedes*. Libraries were of high quality, with 1289 total cells (mean of 769 genes per cell) detected in the LacZ-injected sample, 872 cells (686 genes per cell) in the serum sample, and 965 (656 genes per cell) in the bacteria-infected sample after QC with Seurat. All metrics are improved compared to vivoPHIX alone and in future hemocyte work we would use this new protocol.

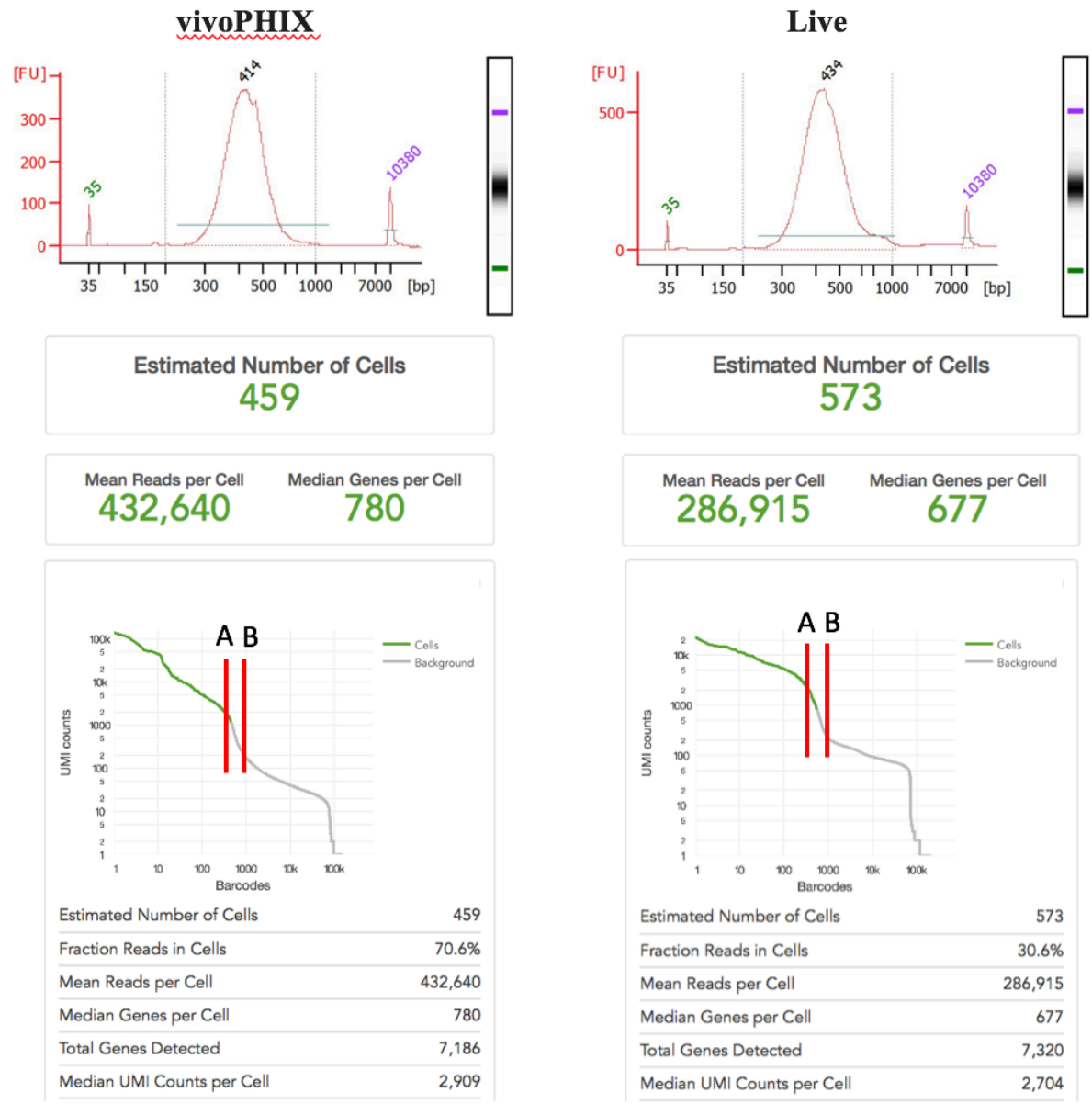


Fig. II.12 Library traces and Cell Ranger statistics: vivoPHIX vs live hemocytes. To the left vivoPHIX sample metrics, to the right the live calcein sorted cells. At the top Bioanalyser profiles for the libraries, and below Cell Ranger pipeline results. Marked as **(A)** are the cell Ranger default settings underestimating the number of cells detected and overestimating genes per cells. **(B)** A manual cut-off of 200 UMIs results in a comparable number of cells.

Chapter III

Characterizing the functional classes of mosquito hemocytes

1 ScRNA-seq: a new era of cell biology

"*Omnis cellula e cellula*"
– **Rudolf Virchow**

The invention of the microscope revolutionised biological investigations. This new technology allowed Robert Hooke to publish in 1665 *Micrographia*, a collection of his microscopic observations. Among these were the depictions of the microscopic units of cork, classically considered the first description of cells. Indeed, in Latin *cella* means a 'little room with a rigid wall.' And cellular biology was born [325].

It took time however to progress from this basic definition of a cell to modern cell biology. In 1896 E.B. Wilson finally defined the cell as "the basis of life of all organisms." [326] However, the foundations for this conclusion were laid even earlier, in 1861 by Max Schultze, who recognised the importance of a cell not for the rigid wall enclosing it, but rather for what it contained. He set out his vision poetically, defining the cell as a "naked speck of protoplasm with a nucleus" (where protoplasm is now called cytoplasm) [327]. Nuclei had nevertheless been observed before, first by abbot Fontana in 1781, and then by Robert Brown in 1831, who recognised the nucleus as an essential component of cells. Finally, in 1838-9 Jakob Schleiden and Theodor Schwann formulated modern 'cell theory' for the first time, declaring "the elementary parts of all tissues to be formed of cells." [328–331] However, it was only in the 1850s through the work of Remak, Virchow, and Kölliker that cells were shown to form through scission of pre-existing cells, finally disputing the theory of spontaneous generation. Virchow went even further, showing cells not only to be the basic unit of life, but also of human pathology [332, 333].

Finally, as the 19th century came to a close, further technological advances in microscopy led to the discovery of all the major organelles we now know comprise a cell, spearheaded by work of Camillo Golgi [334]. Golgi was also responsible for disproving the theory that nervous tissue formed a completely interconnected syncytium. The development of

the 'black reaction' and the work of Santiago Cajal completely dispelled the syncytium theory and confirmed the neurons as the basic cellular unit of the brain [335, 336].

Single-cell transcriptomic techniques are now becoming just as transformative in morphing our understanding of cells, their identities, origins, and functions. Since Hooke's first observations of a cell now almost four centuries ago, generations of scientists have toiled to catalogue and describe all the different cell types in humans, animals, and plants by looking at morphology and function. Before the advent of scRNA-seq it was thought 210 different cell types existed in the human body [337]. And yet, the diversity within all of these cell types is still bewildering. Even markers traditionally thought to define individual cell types in fact isolate multiple subtler subtypes of cells. Nowadays however we are able to measure the expression level of genes in each individual cell and thus define its circuitry through single cell transcriptomics. But then, what is a cell state, and what is a cell type? When does a transcriptional perturbation define the advent of a new cell? And when is that perturbation a transition point between different cell types, and when the consequence of stochastic processes with no long-term consequences on cellular function? These are still very much active areas of investigations, but at least we now do have for the first time the tools to look anew at the cellular landscape of organisms, with a fresh set of eyes, and yet the same thirst for discovery.

We applied these technologies to mosquitoes. Three hemocyte types have been described in *Anopheles* and *Aedes* based on their morphology[4]. Granulocytes are highly phagocytic cells of about 10-20 μm, while oenocytoids are relatively smaller (8-12 μm), round cells that produce melanin, an insoluble pigment involved in wound healing and pathogen containment by encapsulation. Finally, prohemocytes are small round cells (4-6 μm) with a high nuclear to cytoplasmic ratio, thought to be precursors of the other two cell types. Hemocytes can be circulating or sessile, and alternate between these two states[146, 150]. However, the full functional diversity of mosquito hemocytes and their developmental trajectories have not been established, and it is not clear to what extent morphologically similar hemocytes are functionally equivalent. Here, we use single cell RNA sequencing (scRNA-seq)

to analyse the transcriptional profiles of individual mosquito hemocytes in response to blood feeding or infection with *Plasmodium*. We reveal a previously unknown functional diversity of hemocytes, with different types of granulocytes expressing distinct and evolutionarily conserved subsets of effector genes. And we identify two basic lineages and differentiation pathways in prohemocytes and granulocytes, and we discover new hemocyte populations and markers of immune activation. Finally, a comparison of hemocyte types from *Anopheles* and *Aedes* show that some are shared, while others appear to be unique to each mosquito species.

1.1 Aims

1. To investigate the diversity of the adult *A. gambiae* M-form (*A. coluzzi*) hemocytes in response to *Plasmodium* infection by scRNA-seq.
2. To identify markers of cell types and states and generate RNA-FISH probes and antibodies for functional studies.
3. To learn about cell lineages of hemocyte subtypes and their differentiation to functional effector subtypes.
4. To validate bioinformatic results microscopically in *A. gambiae* M-form (*A. coluzzi*) and *A. gambiae* (G3 NIH strain), and characterise hemocyte types in sections, whole-mounts and isolated hemolymph of the mosquito through RNA-FISH
5. To compare *Anopheles* hemocytes with *Aedes* hemocytes

1.2 Colleagues

Dr. Ana Beatriz Ferreira and the NIH imaging core prepared the single hemocytes RNA-FISH / morphology correlative images, and prepared *Aedes* samples up to fixed cells. Tom Metcalf aided in some of the dissections for bulk RNAseq. Mirjana Efremova calculated correlation between *Aedes* and *Anopheles* hemocytes. All other data presented is a result of my own work unless stated otherwise.

2 Methods

2.1 *A. gambiae* mosquito rearing and *P. berghei* infection

A. gambiae (G3 NIH strain) and *A. gambiae* M-form (*A. coluzzi*) were reared at 28 °C, 80% humidity, 12-hour light/dark cycle with standard laboratory procedures. For infections we utilized GFP-CON transgenic *P. berghei* (259cl2 strain), maintained with serial passage in female 4-8 weeks old BALC/c mice [319]. Parasitemia was assessed by light microscopy following methanol-fixed blood-smears stained with 10% Giemsa and air-dried. Mosquitoes were blood-fed on infected mice at a parasitemia of 3-5%, with 1-2 exflagellations per field. Infected mosquitoes were kept at 21 °C to allow for infection and midgut invasion. To confirm infection 10 mosquito midguts were dissected 5 days post blood-feeding and oocysts counted by fluorescence. *Aedes* mosquitoes were reared and challenged as of Chapter II.2.1-2.4.

2.2 Hemocyte collection, fixation, cell counting

For details of collection apparatus and collection methodology see Chapter II.2.5. Hemocytes were collected by gradually injecting in the thorax of cold-anesthetized mosquitoes 10 μL of anti-coagulant media (2 μL at a time) composed of 60% Schneider's insect media, 30% citrate buffer, 10% heat-inactivated fetal bovine serum, final pH 7.0-7.4, sterilized by 0.22 μm filtration. A total volume of 10 μL was collected per mosquito (8-12 mosquitoes per condition) and transferred with a sterile non-stick pipette tip into 500 μL vivoPHIX at room temperature. Cells were fixed for 2 hours at RT and then stored at 4C until Chromium 10X processing.

2.3 RNA extraction and bulk RNAseq library preparation

For bulk RNAseq hemocytes were collected as described above from 8 mosquitoes, but transferred directly in 500 μL of TRIZOL reagent (Invitrogen). From the same mosquitoes, midguts and carcasses were transferred into separate 1.5 mL Eppendorf tubes containing 150 μL TRIZOL reagent by Tom Metcalf. The samples were well triturated with an electrical

homogenizer and disposable pestles before adding 350 μL more TRIZOL reagent and mixing. Samples were allowed to lyse for 15-30 minutes at room temperature to allow for full dissociation, then stored at 4C overnight and then at -20C until RNA extraction. Non-hemocyte samples were then spun for 12,000 RCF, 10 minutes at 4C to remove all insoluble material. The supernatant, as well as the homogenate of hemocyte samples were transferred to Phase Lock GelHeavy 2 mL tubes that had been pre-spun for 1500 RCF for 1 minute, and allowed to incubate for 5 minutes at room temperature. 100 μL of chloroform (200 μL per 1 mL TRIZOL) was added, the tubes capped, and then vigorously shaken for 15 seconds. Samples were then centrifuged for 12,000 RCF, 10 minutes, 4C. If the clear, aqueous phase was still mixed with TRIZOL matrix then 100 μL more of chloroform was added, and the samples again mixed vigorously and spun as before. The aqueous phase was then transferred to a fresh 1.5 mL Eppendorf tube and the RNA precipitated by adding 0.25 mL of isopropyl alcohol (500 mL per 1 mL TRIZOL reagent used). For midguts and hemocyte samples 20 μL of glycogen (5 mg / mL) were also added to aid in precipitation and pelleting. Samples were mixed by repeated inversion 10 times, incubated at 10 minutes at room temperature, and then spun at 12,000 RCF, 10 minutes, 4C. All the supernatant was removed, and the RNA pellets washed twice with 75% ethanol (minimum 1 mL of ethanol per 1 mL of TRIZOL used). Each time the samples were mixed by vortexing and centrifuged 7,500 RCF, 5 minutes, 4C. At the end, the supernatant was removed and samples air-dried until almost dry, but not completely (still translucent). RNA was resuspended with 20 μL of RNAse free water for hemocyte samples, 30 μL for midgut samples, and 70 μL for carcass samples, pipetting a few times to homogenize and then incubating at 55C for 10 minutes to completely resuspend. Samples were then stored at -20C until library preparation by Bespoke Low-Throughput Team at the Wellcome Sanger institute. Total RNA quantity was assessed on a Bioanalyser and ranged from 300 ng to 39 μg. mRNA was then isolated with the NEBNext Poly(A) mRNA magnetic isolation module. RNA-seq libraries were prepared from mRNA using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs) as by manufacturer instructions, except that a proprietary Sanger UDI (Unique Dual Indexes) adapters / primer system was used. Furthermore, Kapa Hifi polymerase rather than NEB Q5 was employed.

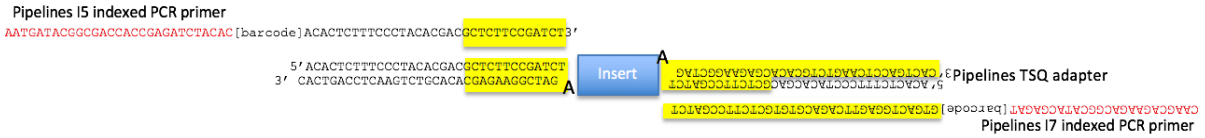


Fig. III.1 Bulk RNAseq proprietary Sanger UDI adapter / primer system. Used with NEBNext Ultra II Directional RNA Library Prep Kit.

2.4 scRNA-seq library preparation

2.4.1 Smart-seq2

See Chapter II.2.8.1 for details. 61 cells passed initial QC after Smart-seq2, as defined by wells containing a majority of sequenced reads mapping onto the *A. gambiae* genome. These cells were processed downstream as ChapterIII.2.8.2, and 48 cells passed stricter QC (>100 features per cell and <30% total reads in mitochondrial genes)

2.4.2 Chromium 10X

Fixed hemocytes were mixed with one volume of pure molecular grade ethanol before centrifugation for 30 minutes at 3k RCF at room temperature. Supernatant was discarded and pellet resuspended in pure molecular grade water before 10X Chromium scRNA-seq library processing. See Chapter II.2.8.2 for details.

2.5 Sequencing

For bulk RNAseq samples HS4000, (using kit version 1) 75PE (RNA): libraries were run on the Illumina HiSeq 4000 instrument with standard protocols using a 150-cycle kit set to a 75bp paired-end configuration. Libraries supplied at 2.8 nM and loaded with a loading concentration of 280 pM. For scRNA-seq Chromium 10X V2 and V3 kits, HS4000 (using kit version 1) 10X V2 and V3 read lengths: libraries were run on the Illumina HiSeq 4000 instrument with standard protocols using a 150-cycle kit set. As recommended by 10x Genomics an elongated reverse read was used during the sequencing run. For V2, the read lengths were as follows: Read 1: 26 bases, index 1: 8 bases, read 2: 98 bases. For V3, read lengths were as follows: Read

1: 28 bases, index 1: 8 bases, read 2: 91 bases. Libraries supplied at 2.8 nM and loaded with a loading concentration of 280 pM. For quality control, lanes passed QC if tags were decoded appropriately, reference matches were as expected either *A. gambiae* or *A. aegypti*, quality metrics met in-house expectations, other run metrics such as error rates were as expected, and yield expectation was met (given the number of cycles run and/or platform expectations). The data was then fit to the sequencing requested and any significant deviation from expected explained and appropriately annotated. For assessment two main pieces of software were used. Sequencing Analysis Viewer (SAV) was used to assess the instruments' performance. The Summary tab gave statistics for the whole run in question whereas the Analysis and Imaging tabs allowed QC to delve deeper and assess if the lanes have performed as expected across all the cycles of the run. NPG pages was used both for staff analysis and annotation, and user's visualisation of data. NPG is an in-house bespoke analysis/software package to include tag analysis, reference matching/mapping details and contamination which is the final point where lanes or tags in the run either passed or failed QC.

2.6 RNA-FISH

2.6.1 Whole mount

Mosquitoes were cold anesthetized, micro-injected with 69 nL of 16% fresh paraformaldehyde (PFA) as of Chapter II.2.2, and after 15 seconds immediately dissected while bathing in freshly prepared 4% PFA. Carcasses and midguts were separated by adding carcasses directly into an Eppendorf containing 4% PFA on ice, while midguts were quickly fixed for one minute in ice-cold fresh 4% PFA and then transferred to fresh 1X PBS where they were carefully opened along their longitudinal axis with two small gauge needles under the dissecting microscope to release the blood meal. Using the surface tension of PBS guts were gently raised up and down the PBS to release all blood from the gut until clean and then fixed in a 1.5 mL Eppendorf tube containing fresh 4% PFA. The samples were fixed overnight at 4C on a gentle rocker to guarantee good mixing and fixation. Non-stick tubes and pipette tips were used to prevent sample adhesion. In all next steps care was shown in removing solutions, as guts especially can stick onto or be sucked into pipette tips, or remain stuck on tube walls. Solutions were always

removed against a source of light to increase contrast and decrease likelihood to remove samples by error. Each wash was performed on a gentle rocker, as samples were fragile and could easily break apart.

The day after collection all PFA was carefully removed and guts and carcasses washed twice with 1mL of PBST (0.1% v/v Tween 20 in 1x PBS). Samples were then incubated for 5 minutes in a 40C rocking water-bath with 300-500 μL of RNAscope Protease Plus. After removing as much solution as possible without disturbing the samples, these were twice washed with 500 μL of probe diluent before following the RNAscope 4-plex Ancillary Kit for Multiplex Fluorescent Reagent Kit v2 technical note protocol. Briefly, the pre-mixed C1, C2, C3, and C4 probes were mixed and then 1 or 2 drops added into each sample tube and incubated for 2 hours at 40C. Samples were washed twice for 5 minutes at room temperature on a gentle rocker with pre-warmed RNAscope 1X Wash Buffer. Wash buffer had been pre-warmed to 40C for 10-20 minutes before being diluted from 50X to 1X with distilled water. Samples were then either stored overnight in 5X SSC buffer at room temperature or immediately prepared for hybridisation. 1-2 drops of RNAscope Multiplex FL v2 Amp1, Amp2, and Amp3 were added in series and incubated for 30 minutes (except Amp3 for 15 minutes) in a rocking 40C water bath. Between each reagent samples were washed twice with RNAscope 1X Wash Buffer for 5 minutes on a gentle rocker. Then Opal fluorophores were prepared at the appropriate dilutions (between 1:750 and 1:3000) and each incubated for 30 minutes in a gently rocking water bath at 40C in the dark. Before adding each Opal, samples were treated with the corresponding RNAscope Multiplex FL v2 HRP-C(1/2/3/4) for 15 minutes in a gently rocking water bath at 40C in the dark. Then, samples were treated with RNAscope Multiplex FL v2 HRP-Blocker for 15 minutes in a gently rocking water bath at 40C in the dark. Between all these steps samples were washed twice with RNAscope 1X Wash Buffer for 5 minutes on a gentle rocker in the dark. Finally, as much wash buffer was removed before adding 1-2 drops of DAPI for 30 seconds. DAPI was then in turn removed and samples added onto a slide with 1 drop of Prolong Gold antifade reagent. The samples were flattened in the Prolong Gold reagent (important: without DAPI or background fluorescence will be high) under a dissecting

microscope to prevent flaps and folding of the tissue. After adding coverslips corners were sealed with transparent nail polish and the samples let dry overnight at room temperature in the dark. The day after nail polish was added all around the slide to seal the samples. These were then stored at 4C in the dark until imaging.

Probes	Channel	Dilution	Amount	Annotation
General				
AGAP009623	C1	1:1500	Std	GAPDH - mosquito + control
AGAP008296	C2	1:3000	1/2	Trypsin - gut
AGAP004203	C2	1:3000	1/2	Vitellogenin - fat body
Hemocytes / Granulocytes T. I and II				
AGAP004017	C4	1:1000	1.5	LRR. All hemocytes' marker
AGAP011974	C4	1:1000	Std	SCRC1. General hemoc. marker
AGAP000790	C3	1:1000	Std	Prohem. / granulocyte marker
AGAP003057	C1	1:1000	Std	Gran. Type II
AGAP011871	C2	1:750	Std	Gran. Type I
Rapidly dividing				
AGAP005363	C3	1:750	n/a	
Fat Body - Baseline				
AGAP007033	C1	1:750	n/a	
AGAP028406	C1	1:750	n/a	APL11C
Oenocytoids				
AGAP004981	C2	1:1500	Std	PPO4
AGAP012851	C1	1:1500	Std	Aldo-keto-reductase
AGAP012000	C3	1:1500	Std	Fibrinogen/fibronectin
Effector				
AGAP007318	C3	1:1000	1.5	Transmembrane
Secretory				
AGAP011239	C1	1:1500	Std	Some also in oenocytoids

Table III.1 RNAscope probe channels and Opal dilution for whole-mounts and sections. See RNAscope 4-plex Ancillary Kit for Multiplex Fluorescent Reagent Kit v2 technical note protocol for details. 'Amount' column indicates the ratio of probes added to hybridization mix compared to standard protocol. 'Std' indicated standard, 0.5 is half of standard. 'n/a' indicates a probe was not successful even with the strongest Opal dilution (1:750) and highest probe amount. Note all dilutions were 1:750 for RNAscope of isolated hemocytes.

2.6.2 Isolated hemocytes

Wells of μ-Slide Angiogenesis Chambers (Cat# 81506 from IBIDI) were coated with 3.5 μg / cm^2 of Cell-Tak Cell and Tissue Adhesive (Corning, 734-1081) by first preparing a fresh 300 μL coating solution with 10 μL Cell-Tak, 285 μL Sodium Bicarbonate pH 8.0 and 5 μL 1N NaOH and immediately coating the glass slides. Wells were incubated at room temperature for least an hour, after which they were washed with sterile water, air-dried and stored at 4C for a maximum of one day.

Hemocytes were collected as of above but directly onto the wells. Eight mosquitoes were processed per sample. Hemocytes were then let to attach onto the coated wells for 15 minutes at 28C in an incubator, before removing all of the media, and fixing cells with 4% PFA for an hour at room temperature before proceeding to RNA-FISH protocol as of Chapter III.2.7.1. The process was made easier by not having to take care of aspirating tissue with the washes, however care was shown not to disperse liquid to strongly, but to always do it gently on the sides of the well to prevent cell detachment. Dr. Ana Beatriz Ferreira performed the isolated *P. berghei* experiments and the correlative experiments.

2.6.3 Sections

Mosquitoes were cold anesthetized, dipped in 100% ethanol to decrease surface tension, and then dipped and fixed in 10% formalin for 18-24 hours overnight at room temperature. Following that the Histology Core of the Sanger Institute processed the samples to make slides. The Sakura Tissue-Tek VIP Tissue processor on Rapid Biopsy programming was used (10 min VIP1 and 10 min VIP2 for each solution except: no VIP2 for 50% and 70% ethanol; first paraffin wax 20 min for both VIP1 and VIP2), with the following solutions in order: 50% ethanol, 70% ethanol, 90% ethanol, 3X 100% ethanol, 3X xylene, and 4x wax. For embedding, two orientations (longitudinal and transverse) were used for each condition (sugar-fed, blood-fed and *P. berghei* infection), before 5 μm sectioning. H&E sections were prepared for every other section, with the mirror section available for RNA-FISH (RNAscope) as of above.

2.7 Imaging

Mosquito sections and whole mounts were imaged with the 3DHISTECH MIDI II automatic digital slide scanner (3DHISTECH, Budapest, Hungary), with 20x and 40x objectives (numerical aperture 0.8 to 0.95), and a bespoke DAPI, Opal 520, 570, 620 and 690 filter sets and a 4.2MP 16-bit camera with wideband LED, or with a 20x bright-field camera for H&E mosquito sections and a 4.2MP 16-bit camera with RGB illumination. Sections and whole-mounts were imaged with extended focus, sequential acquisition, and variable z-steps, mosaic size and integration.

For whole-mount and hemocytes samples images were captured at the National Institute of Health using a Leica TCS SP8 DMI8 confocal microscope (Leica Microsystems, Wetzlar, Germany) with a 20x, 40x and 63x oil immersion objective (using zoom factor of 2, 3 or 4; numerical aperture, 1.25 to 1.4) equipped with a photomultiplier tube/hybrid detector. Samples were visualized with a white light laser and specific emission and excitation range were used depending on the fluorophore used. For these experiments we used the following spectra for excitation/ emission: 488/520, 550/ 570 594/620, and 670/690. DAPI was excited using a 405-nm diode laser. Images were taken using sequential acquisition, and variable z-steps, mosaic size and integration. Image processing was performed using proprietary Leica LAS X and Imaris 9.2.1 (Bitplane, Concord, MA, USA). At the Wellcome Sanger Institute images were captured using a Leica TCS SP8 DMI8 confocal microscope (Leica Microsystems) using a 40×, 63×, or 100× oil immersion objective (using zoom factor of 2, 3 or 4; numerical aperture, 1.25 to 1.4) and equipped with photomultiplier tube/hybrid detectors. Fluorochromes were excited using a 405nm DMOD laser for DAPI, 488-nm CSU laser for Opal 520, a 552-nm CSU laser for Opal 570 and Opal 620, 638-nm CSU laser for Opal 690. Images were taken using sequential acquisition, and variable z-steps, mosaic size and integration. Image processing was performed using proprietary Leica LAS X and Imaris 9.2.1 (Bitplane, Concord, MA, USA).

2.8 Bioinformatics

2.8.1 Bulk RNA-seq

Sequencing reads in CRAM format were fed into a bespoke BASH pipeline to first automatically convert cram files to fastq using biobam's bamtofastq program (Version 0.0.191). Then, forward and reverse fastq reads in paired mode were aligned to the *A. gambiae* AgamP4.3 reference genome using hisat2 (Version 2.0.4) and featureCounts (Version 1.5.1) with recommended settings. Combined counts matrix was then produced by a python script before downstream data processing and analysis within R version 3.5.3 (RStudio version 1.0.153). Downstream normalization, differential expression analysis and visualization were done with DESeq2 R package (Version 1.18.1) [280]. Base factor was defined as the sugar condition, and time 0 (non-infected). One outlier was removed (blood fed hemocyte sample at 48 hours, experiment GR88) after plotting residuals of internal batch correction and visually inspecting a PCA plot. Data was normalized by making a scaling factor for each sample. First the log(e) of all the expression values were taken, then all rows (genes) were averaged (geometric average). Genes with zero counts in one or more samples were filtered out and the average log value from log(counts) for all genes was subtracted. Finally, the median of the ratios calculated as above for each sample was computed and raised to the e to make the scaling factor. Original read counts were divided by the scaling factor for each sample to get normalized counts. Then, the dispersion for each gene was estimated, and a negative binomial generalized linear model fitted. P values for the differential expression analysis were adjusted for multiple testing using the Bonferroni correction. Genes were considered as differentially expressed if they had an adjusted P value < 0.001 (Wald T-test) and a $\log_2$ fold change > 2. All body parts, conditions and timepoints were considered together while running the following model for differential expression analysis focused on body part, with experimental repeats, time, and effects of treatment (*P. berghei*, blood feeding and sugar feeding) as covariates:

```
ddsMat <- DESeqDataSetFromMatrix(countData = countdata, colData = coldata,
                                 design = ~ 0 + experiment + time + treatment + part)
```

2.8.2 scRNA-seq

Droplet-based sequencing data were aligned and quantified using the Cell Ranger Single-Cell Software Suite [246] (version 2.0, 10x Genomics) against the *A. gambiae* PEST, AgamP4.9 reference genome provided by Vectorbase [338]. Cells with fewer than 100 and more than 2500 genes and for which total mitochondrial gene expression exceeded 20% (or 50%) were removed. Genes that were expressed in fewer than three cells were also removed.

Downstream analyses—such as normalization, shared nearest neighbor graph-based clustering, differential expression analysis and visualization—were performed using the R package Seurat (version 2.3.4 or 3.0.2) [256, 277, 339]. The two experimental batches were integrated using canonical correlation analysis, implemented in the Seurat alignment workflow. In the newer Seurat version, batches were integrated with a hybrid CCA / MNN strategy identifying 'anchors' of similar cells between conditions and CCs. Cells for which the expression profile could not be explained by low-dimensional canonical correlation analysis compared to low-dimensional principal component analysis were discarded. Clusters were identified using the community identification algorithm as implemented in the Seurat 'FindClusters' function. For Seurat V2 the shared nearest neighbour graph was constructed using 13 canonical correlation vectors as determined by the dataset variability. The resolution parameter to obtain the resulting number of clusters was fine-tuned so that it produced a number of clusters large enough to capture most of the biological variability. UMAP analysis was performed using the RunUMAP function with default parameters. Differential expression analysis was performed based on the Wilcoxon rank-sum test. The P values were adjusted for multiple testing using the Bonferroni correction. Clusters were annotated using canonical cell-type markers. We remove a blood-fed 24 hours post-feeding sample (experiment GR72) because it formed a technical outlier in the initial PCA-driven quality control and all cells clustered separately without mixing with other samples. Some clusters were further analyzed by partitioning the clusters separately and performing the analysis anew, with the same

alignment and clustering procedure. For example, all hemocytes were subdivided from other non-hemocyte cells and reanalyzed.

Diffusion pseudotime [340] implemented in the SCANPY package [257] was applied to find the major non-linear components of variation across cells, using the most highly variable genes. The first diffusion component correlated with oenocytoids identity as defined by know marker genes, whereas the second diffusion component second diffusion component correlated with immune activation and cell division. Genes which changed along the identified trajectories (diffusion components) were identified by performing a likelihood ratio test using the function differentialGeneTest in the monocle 2 package [341]. The Seurat implementation of velocyto [342] was then applied to estimate RNA velocity and infer in which direction cells were changing along the previously inferred trajectories or UMAP. scVelo was used as an additional RNA velocity analysis tool to confirm the results [343].

Lineage tree reconstruction was performed with partition-based graph abstraction (PAGA) as implemented in SCANPY package [344]. The graph abstraction algorithm combines clustering and trajectory inference to elucidate the variability of scRNA-seq through discrete and continuous variables. PAGA takes into consideration a partitioned graph of neighbourhood relations. It quantifies distances between nodes with a random-walk based measure and then it quantifies what connectivity partitions there is. The abstracted graph is anchored on nodes which are the clusters first identified with Seurat. The differentiation tree is a tree-like subgraph which best explains topology. Slingshot was another highly rated lineage tree reconstruction software that we used to validate PAGA results [309]. With a matrix input representing cells in a reduced-dimensional space (UMAP) and a vector of cluster labels the Slingshot algorithms then built a minimum spanning tree (MST) of the clusters to infer the lineage structure. Finally, smooth lineage curves were built and pseudotime inferred for all lineages. We then used the pseudotime values calculated by Slingshot to discover differentially expressed genes between the identified lineages with the tradeSeq package (TRAjectory Differential Expression analysis for SEQuencing data) [345]. TradeSeq uses pre-calculated UMAP coordinates and pseudotime values to fit generalized additive models (GAMs).

To compare the *A. gambiae* with the *Aedes* cell types, a logistic regression with L2-norm regularization and a multinomial learning approach (implemented by the scikit-learn function LogisticRegression) was trained on the anopheles gambiae clusters. The log-transformed normalized data was used. The model was used to predict the probabilities of each *Aedes* cell belonging to each one of the anopheles gambiae clusters (implemented by the predict_log_proba function).

3 Results

Hemocytes were obtained from mosquitoes at different states of immune activation in order to survey their diversity. In the first experiment we collected mosquitoes at both 24- and 27-hours post-infection to potentially gain information about the early hemocyte response to *P. berghei*. The 48- and 72-hours timepoints were chosen to explore hemocyte changes after infection. In the second experiment, the 27 hours timepoint was removed to make space (cost concerns) for a day 7 timepoint, which we hypothesised could give information on hemocyte deactivation. We chose sugar feeding as baseline control. However, we also used blood feeding as control for *P. berghei* infection due to the large changes blood feeding causes in the mosquito.

Experiment 1	**Day 0**	**Day 1 PF**		**Day 2 PF**	**Day 3 PF**
Condition		*24 h*	27 h	48 h	72 h
Cntrl (SF)	SF	X	X	X	Bleed
Cntrl (BF)	BF	Bleed	Bleed	Bleed	↓
P. berghei	BF	↓	↓	↓	↓

Experiment 2 and bulk	**Day 0**	**Day 1 PF**	**Day 2 PF**	**Day 3 PF**	**Day 7 PF**
Condition		*24 h*	48 h	72 h	7 days
Cntrl (SF)	SF	Bleed	Bleed	Bleed	Bleed
Cntrl (BF)	BF	↓	↓	↓	↓
P. berghei	BF	↓	↓	↓	↓

Table III.2 Experimental strategy: bulk and scRNAseq of *Anopheles*. PF = post-feeding; BF = blood-feeding. Experiment 1 refers to scRNA-seq repeat 1. Experiment 2 was the second scRNA-seq repeat and the same scheme was used for the bulk RNAseq samples.

Following hemocyte capture and 10X library preparation and sequencing we then normalized and performed QC on all cells from an experiment together, then batch corrected the experiments, clustered, and investigated differences between clusters, time points, and conditions as of below and method chapter.

3.1 scRNA-seq identifies at least six hemocyte subpopulations

3.1.1 QC of Chromium 10X single cell data

Processed scRNA-seq matrices from each individual sample were loaded onto the R-based Seurat (v2.4 or v3.0) analysis suite. First, cells were filtered based on QC metrics to remove poor quality cells. The total number of genes (or of UMIs) within a cell is traditionally considered a useful marker to distinguish low quality cells or empty droplets from healthy cells. In addition, an excessive gene count can indicate that the original droplet contained a doublet or multiplet and should also be excluded. Cells were thus filtered if they were found to have less than 100 or more than 2500 unique genes. Then, we identified which *Anopheles* genes map to the mitochondrial genome to calculate the percentage of reads mapping to mitochondrial genes. Typically (though not necessarily always) damaged, dying, and low-quality cells will show a high ratio of mitochondrial reads to total reads. In our data-set we initially excluded all cells that had more than 20% of total reads mapping onto the mitochondrial genome. We repeated this process for both our scRNA-seq experiments, plotting data both with violin plots and scatter plots to identify outlier cells. We discarded outlier samples: blood-fed 24 hours (experiment 1 and 2), sugar-fed 48 hours (experiment 2).

Filtering appeared successful in removing all outliers, with each parameter showing a compact distribution in both experiments [Fig. III.2]. The first experiment had a total of 7762 cells before QC, with means of 85 genes and 221 UMIs per cell. After QC we were left with 2081 cells (mean of 180 genes per cell, and 575 UMI per cell). In the second experiment before QC we had a total of 3883 cells, with a mean of 380 genes per cell and 1422 UMIs per cell. After QC 3162 cells remained, with a mean of 441 genes per cell and 1516 UMIs per cell. Statistics showed the first experiment had lower data quality than the second. Of note, samples from the first experiment had been stored for about a month at 4C while the second experiment was processed within a week of collection.

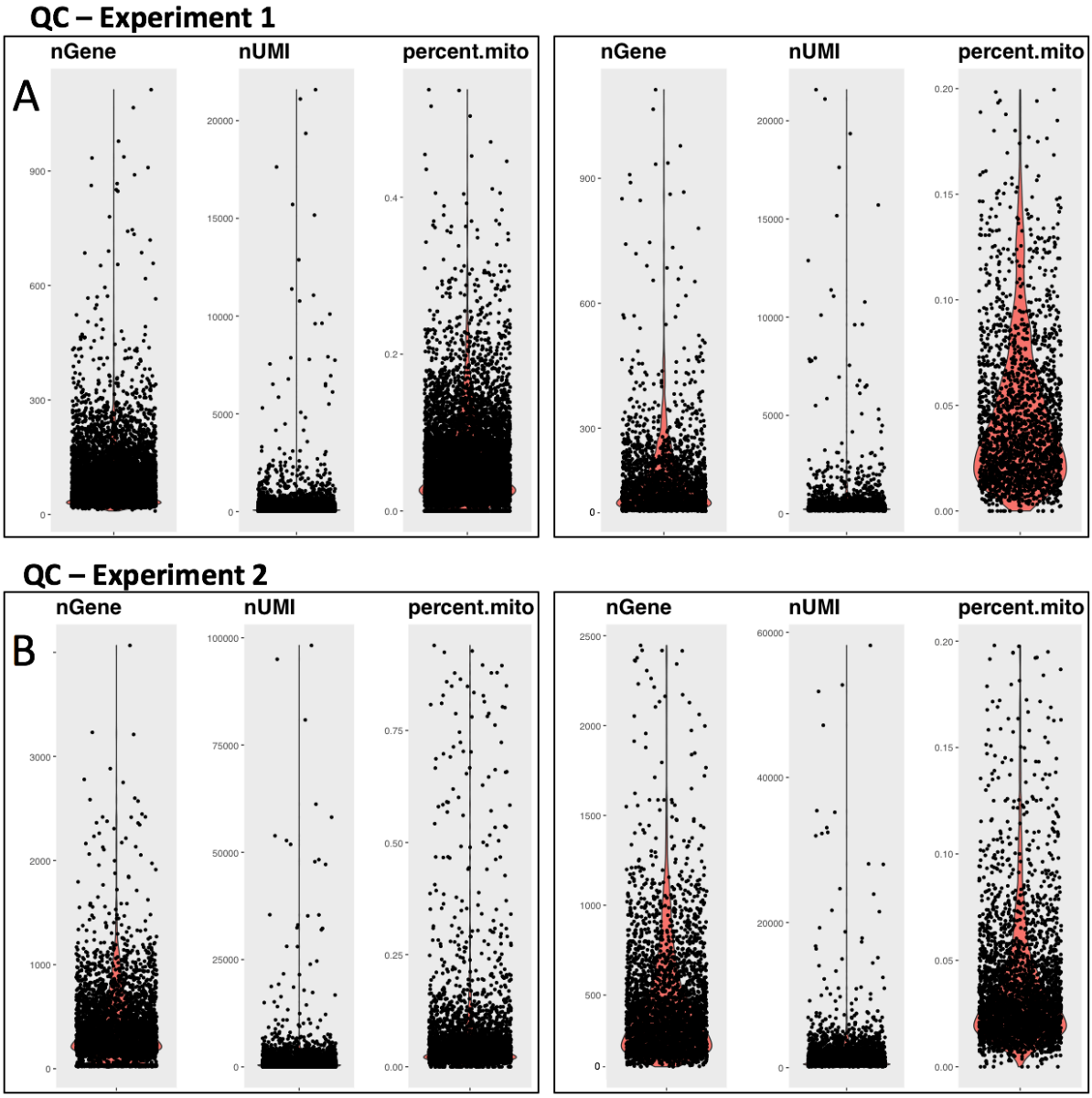


Fig. III.2 Seurat scRNAseq QC. (A) QC metrics for the first experiment. To the left metrics before QC, to the right after QC. **(B)** QC metrics for the second experiment. To the left metrics before QC, to the right after QC. nGene = total number of genes detected per cell. nUMI = total number of UMIs detected per cell. percent.mito = the proportion of total reads mapping to mitochondrial genes.

3.1.2 Normalisation, scaling, identification of variable genes, and PCA

Data was then normalized using the Seurat global-scaling normalization method, which normalizes gene expression data of our cells by total expression, multiplies it by a scale factor of 10,000, and then takes the natural logarithm of the resulting number. Highly variable genes (focus of downstream analyses) were calculated with a variance stabilizing transformation (VST) [277, 339]. We identified 2000 variable genes in each experiment. We then linearly transformed the data ('scaling') to pre-process data for dimensionality reduction techniques such as PCA, the first step of an integrated analysis. Scaling reduced the importance of highly expressed genes. This step shifted gene expression so that the mean across cells is zero, and scaled expression so that variance across cells is 1. Many of these highly variable genes were common among the two experiments. For instance, AGAP011294, AGAP01002, or AGAP011230 were identified as top variable genes in both [Fig. III.3].

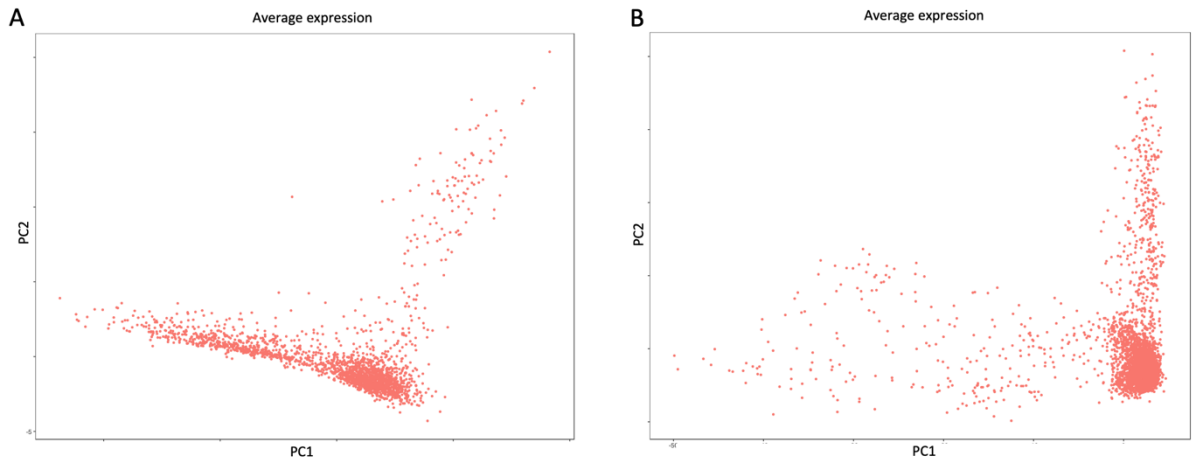


Fig. III.3 PCA profiles are similar between the two experiments (A) PCA showing the first two principal components for first experiment **(B)** and PCA of the two first principal components for the second experiment

3.1.3 Clustering reveals 9 separate cell types

In Seurat 3.0, dataset aggregation was drastically improved by using mutual nearest neighbours (MNN) – 'cell anchors' – in addition to canonical correlation. Different QC parameters returned the same results and so we lowered stringency of mitochondrial gene filtering to 50% (see discussion). After aggregating the two experiments we had a total cell count of 5383 hemocytes after QC, with a mean of 335 genes per cell, and 1142 UMI per cell. We classified *Anopheles* cell types in the hemolymph to identify nine major clusters. Most clusters could be further subdivided into smaller clusters by increasing the resolution of the clustering algorithm. However, increasing resolution typically identifies cell states rather than cell types and initial clustering therefore needs to be more conservative.

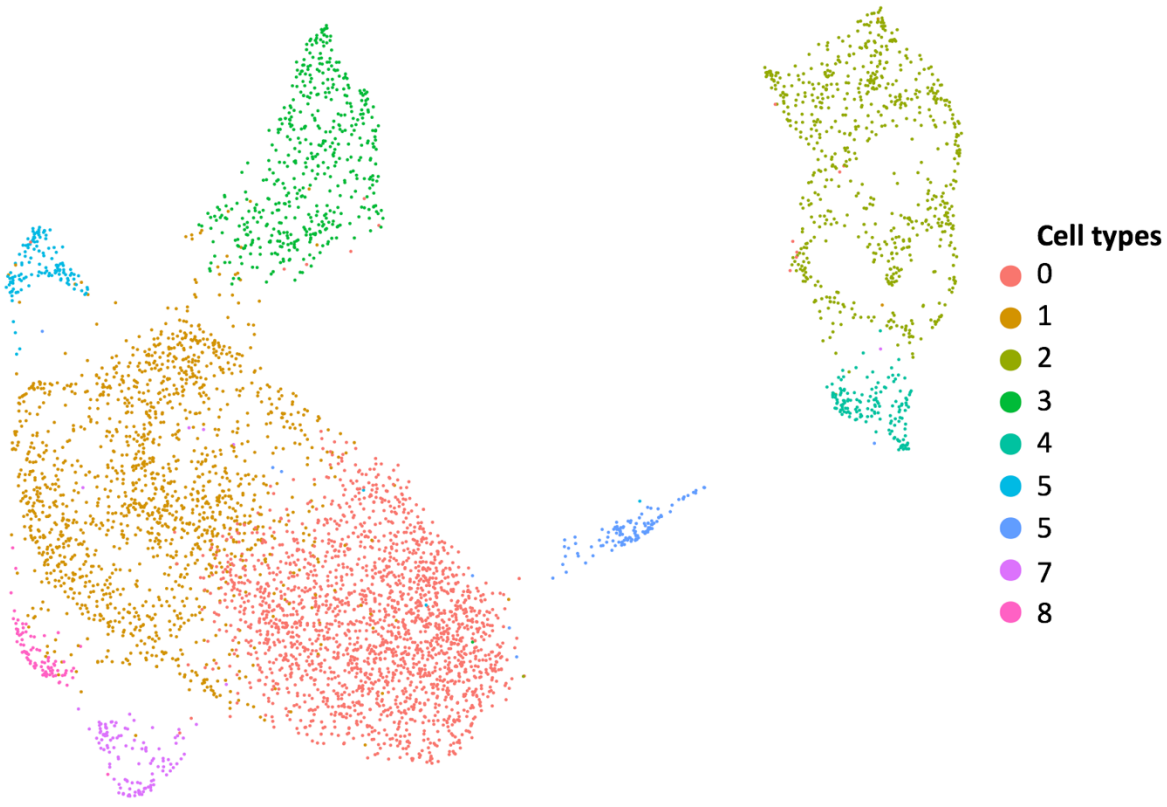


Fig. III.4 Clustering solution of *A. gambiae* hemocytes. UMAP dimensionality reduction separates clusters of cells by overall transcriptomic similarity. Each dot represents a cell, whereas different colors identify clusters of similar cells.

3.1.4 Varying QC parameters does not alter clustering solution

Compared to simple CCA integration of Seurat v2.4 the v3.0 clustering solution was well mixed with regards to both experimental batches as well as individual samples [Fig. III.5A-B].

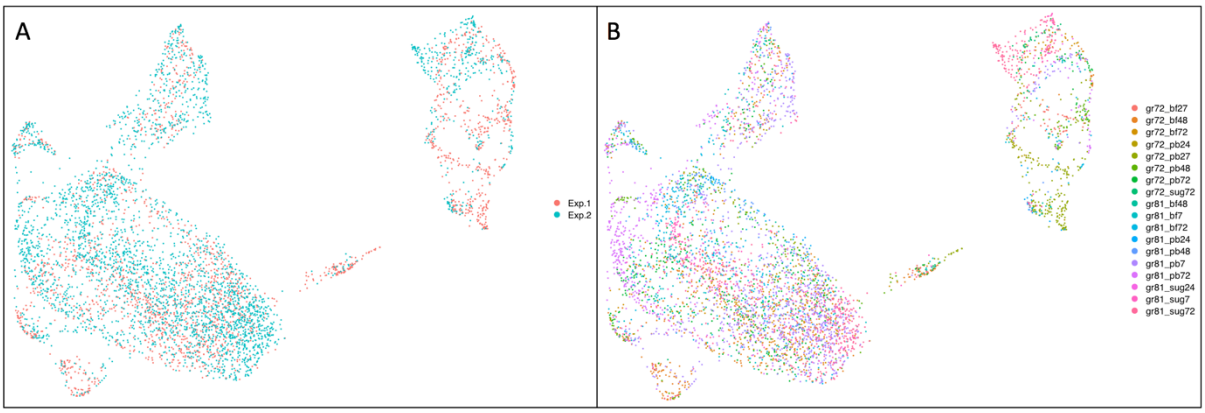


Fig. III.5 Samples and experiments are well-mixed. (A) Both between the two experiments, as well as **(B)** between samples (separate 10X lanes and chips)

The new clustering strategy is robust to a wide spectrum of parameters and is more unsupervised, lowering the risk of bias due to parameter selection. We nevertheless manually checked whether results were reasonable by raising the minimum number of genes per cell to 150 and then to 200, without changes to cluster numbers, structure or markers genes [Fig. III.6].

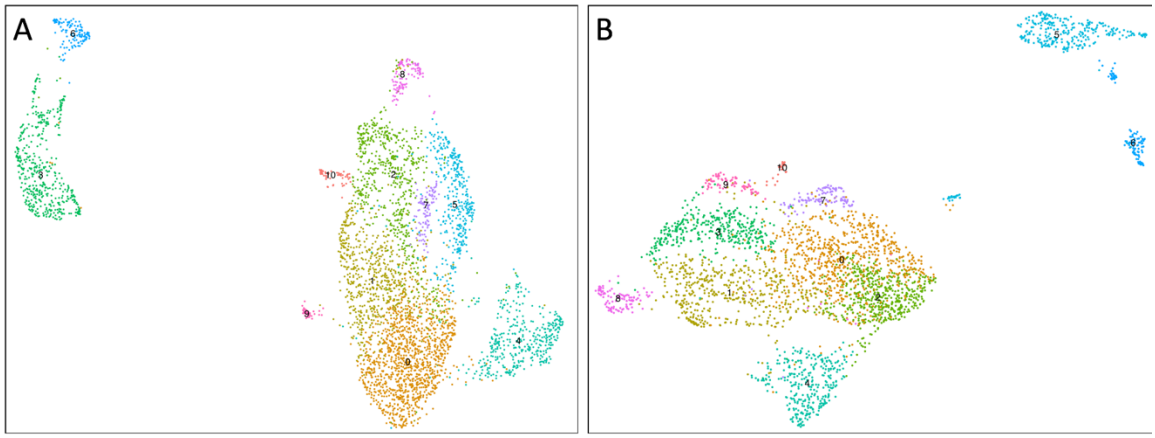


Fig. III.6 Clustering solutions are robust to gene thresholding. Manual QC iteration: increasing minimum gene per cell parameter stringency does not alter computer clusters. **(A)** Minimum 150 genes per cell **(B)** Minimum 200 genes per cell.

Functional classes of mosquito hemocytes

We then removed mitochondrial genes thresholding. Few cells were added and no changes in clustering were detected [Fig. III.7A]. Finally, we compared cells (droplets with more than 100 genes) and background (droplets with less than 50 genes) with principal component analysis. Without calculating a UMAP, already the first two principal components cells and debris clearly separate. Combined, the QC tests demonstrate our thresholds are reasonable for this dataset [Fig. III.7B].

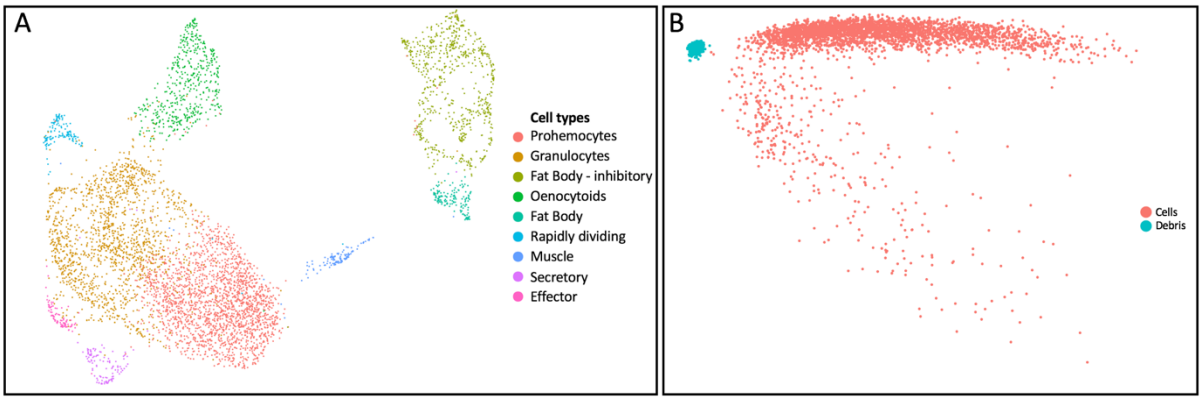


Fig. III.7 Clustering solutions is robust to more stringent mitochondrial filtering. Debris and cells are clearly identifiable. Clustering done as above, except threshold was set with **(A)** maximum 100% of reads mapping to mitochondrial genes, showing no changes **(B)** Principal component analysis of debris (blue, droplets with less than 50 genes per droplet) and cells (red, droplets with more than 100 genes per droplet) shows cells separate clearly from debris (PC1 vs PC2).

3.1.5 Differential expression analysis identifies conserved marker genes for each cell cluster, and suggest cellular identity

Though the *Anopheles* genome is poorly annotated we utilised gene ontology annotations from g:Profiler [346], as well as manual curation of *Anopheles* genes [347], to understand the identity of each cell cluster. The table below shows the top 10 genes for each cluster, annotated, while the full list can be found in the Appendix.

Cluster 0

Gene	Name	Pval adj	Avg logFC	Pct.1	Pct.2	Annotation
AGAP012100	RpS26	5.21E-87	0.325	0.97	0.98	40S ribosomal protein S26
AGAP002464	-	9.33E-75	0.471	0.95	0.90	secreted ferritin G subunit
AGAP011828	Cp1	1.00E-71	0.498	0.83	0.70	cathepsin L
AGAP010163	RpL38	2.29E-68	0.322	0.95	0.96	60S ribosomal protein L38
AGAP000305	-	6.01E-58	0.383	0.88	0.70	SPARC
AGAP004936	-	5.04E-50	0.428	0.79	0.62	None
AGAP007740	RpLP1	4.04E-45	0.258	0.96	0.97	60S ribosomal protein LP1
AGAP002422	CLIPD1	2.74E-41	0.656	0.61	0.54	CLIP-domain serine protease
AGAP011119	-	1.73E-40	0.421	0.74	0.62	None
AGAP002465	-	1.54E-36	0.421	0.82	0.77	ferritin heavy chain

Cluster 1

Gene	Name	Pval_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP011228	-	2.12E-189	0.746	0.99	0.75	None
AGAP007312	-	7.96E-162	0.799	0.77	0.35	None
AGAP004936	-	1.16E-142	0.596	0.92	0.59	None
AGAP006278	-	3.23E-137	0.666	0.86	0.53	None
AGAP000651	actin5c	2.72E-136	0.713	0.78	0.39	Actin-5C
AGAP004017	-	8.90E-129	0.590	0.82	0.41	None
AGAP004164	GSTD1	1.58E-125	0.704	0.44	0.13	glutathione S-transf del. c1
AGAP028028	lrim16a	1.70E-121	0.593	0.82	0.44	leucine-rich immune prot
AGAP004016	-	2.29E-119	0.557	0.69	0.29	None
AGAP006367	-	2.62E-118	0.869	0.33	0.08	None

Functional classes of mosquito hemocytes

Cluster 2

Gene	Name	Pval adj	Avg logFC	Pct.1	Pct.2	Annotation
AGAP010968	CLIPA9	0	2.460	0.48	0.04	CLIP-domain serine protease
AGAP013060	-	0	1.976	0.66	0.09	None
AGAP012571	-	0	1.943	0.78	0.17	None
AGAP008011	-	0	1.902	0.48	0.04	None
AGAP003473	-	2.70E-303	3.031	0.85	0.27	None
AGAP003474	-	1.54E-298	2.450	0.99	0.95	None
AGAP005888	-	1.20E-295	1.828	0.96	0.53	None
AGAP008004	-	7.26E-291	2.367	0.89	0.37	None
AGAP004674	-	1.01E-278	2.010	0.38	0.02	Phenoloxidase inhibitor prot
AGAP009527	-	2.92E-272	2.043	0.61	0.10	None

Cluster 3

Gene	Name	Pval adj	Avg logFC	Pct.1	Pct.2	Annotation
AGAP004978	PPO9	0	4.469	0.81	0.12	prophenoloxidase 9
AGAP011223	-	0	4.448	0.84	0.11	None
AGAP006258	PPO2	0	4.364	0.79	0.13	prophenoloxidase 2
AGAP004977	PPO6	0	4.055	0.98	0.34	prophenoloxidase 6
AGAP012616	PPO5	0	3.961	0.83	0.08	prophenoloxidase 5
AGAP012851	-	0	3.829	0.74	0.02	Aldo-keto reduct fam 1,C3
AGAP006570	-	0	3.669	0.73	0.11	myo-inositol-1(4)-monoph
AGAP006743	-	0	3.489	0.63	0.03	None
AGAP000162	-	0	3.471	0.80	0.06	Cystathionine beta-synth
AGAP000679	-	0	3.159	0.98	0.36	Aminoacylase

Cluster 4

Gene	Name	Pval_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP004203	Vg	2.94E-162	2.998	0.78	0.10	vitellogenin
AGAP007940	-	9.56E-127	2.767	0.72	0.11	Reticulon-like protein
AGAP006548	-	1.20E-126	2.565	0.91	0.21	glycine cleavage sys H
AGAP002593	-	6.61E-114	2.098	0.43	0.04	outer membr lipopr Blc
AGAP001065	-	8.30E-105	2.551	0.76	0.15	glycine hydromethyltran
AGAP004700	-	3.30E-100	2.239	0.38	0.03	None
AGAP010046	-	4.33E-88	2.512	0.29	0.02	None
AGAP009173	Fbp	7.86E-83	2.189	0.38	0.04	fructose-1,6-bisphosph I
AGAP001116	-	1.29E-81	1.946	0.44	0.05	D-amino-acid oxidase
AGAP002198	Gnmt	2.09E-76	2.051	0.46	0.06	glycine N-methyltransf

Cluster 5

Gene	Name	Pval adj	Avg logFC	Pct.1	Pct.2	Annotation
AGAP005363	-	0	1.729	0.45	0.003	None
AGAP004962	-	0	1.526	0.41	0.004	cyclin B
AGAP007855	-	4.72E-295	1.583	0.43	0.007	aurora kinase, other
AGAP013736	-	8.53E-285	1.075	0.31	0.002	None
AGAP005019	-	2.01E-274	2.028	0.56	0.018	None
AGAP003550	-	3.62E-271	1.302	0.32	0.003	None
AGAP006671	-	1.30E-267	1.117	0.30	0.002	None
AGAP006105	-	5.29E-230	1.018	0.28	0.003	None
AGAP004963	-	7.99E-223	0.989	0.25	0.002	cyclin B
AGAP004239	-	1.13E-212	1.284	0.28	0.003	polo-like kinase 1

Cluster 6

Gene	Name	Pval adj	Avg logFC	Pct.1	Pct.2	Annotation
AGAP009526	-	1.7E-104	2.864	0.74	0.12	None
AGAP006181	-	1.12E-97	2.621	0.58	0.07	troponin C
AGAP003939	-	5.44E-83	2.674	0.56	0.08	None
AGAP001622	-	2.17E-72	2.640	0.76	0.19	myosin light chain 5
AGAP003778	-	1.13E-70	2.417	0.50	0.07	None
AGAP001569	-	6.19E-66	2.279	0.48	0.07	myosin alkali light chain 1
AGAP004161	-	8.04E-64	2.322	0.74	0.20	myofilin variant C
AGAP002358	-	3.84E-58	2.334	0.45	0.07	ADP,ATP carrier protein 2
AGAP008311	-	2.87E-50	2.092	0.27	0.03	acylphosphatase
AGAP004790	-	5.28E-46	1.918	0.91	0.50	Up skl mscl growth 5 hom

Cluster 7

Gene	Name	Pval adj	Avg logFC	Pct.1	Pct.2	Annotation
AGAP007347	Lysc1	7.3E-217	4.377	0.91	0.08	C-type lysoz
AGAP005848	-	6.2E-105	2.455	0.39	0.03	Fic A
AGAP011294	DEF1	2.59E-69	1.857	0.28	0.02	defensin anti-micr
AGAP000694	CEC3	2.91E-63	2.455	0.27	0.02	cecropin anti-micr
AGAP000376	Tsf1	1.50E-51	2.139	0.76	0.24	-
AGAP011197	-	1.33E-40	1.779	0.78	0.29	-
AGAP005888	-	2.24E-37	2.573	0.93	0.58	-
AGAP000693	CEC1	1.49E-32	2.855	0.49	0.13	cecropin anti-microb
AGAP005612	-	8.23E-23	2.085	0.32	0.07	-
AGAP010816	TEP3	1.11E-17	1.344	0.34	0.09	thioester-contain prot 3

Cluster 8 Gene	Name	Pval adj	Avg logFC	Pct.1	Pct.2	Annotation
AGAP007318	-	0	3.648	0.79	0.02	None
AGAP009053	LL3	7.0E-212	3.014	0.54	0.02	LITAF-l3
AGAP028208	-	4.0E-195	2.728	0.34	0.01	cuticular prot CPLCP22
AGAP009051	LL1	1.6E-177	1.972	0.37	0.01	LITAF-l1
AGAP007320	-	4.3E-175	1.529	0.29	0.01	None
AGAP001002	-	2.3E-129	3.812	0.42	0.02	Toll
AGAP001652	-	9.6E-107	2.219	0.61	0.05	lipase
AGAP003319	-	6.01E-95	2.147	0.49	0.04	None
AGAP011226	-	1.25E-92	1.941	0.42	0.03	None
AGAP005209	-	1.06E-73	1.817	0.47	0.04	Uridine kinase

Table III.3 Marker genes for each cell cluster. P_val_adj = P value adjusted for multiple testing. Avg_logFC = average log fold change for the gene between cluster of interest and other clusters. Pct.1 = percentage of cells in cluster of interest where gene is detectable. Pct.2 = percentage of cells in other clusters where gene is detectable. Annotation = electronic annotation of gene.

We then assigned putative cell type names based on their gene markers. We molecularly confirmed known cell types such as granulocytes, expressing SPARC, collagens, laminins, scavenger receptors, LRIMs, Nimrod, LRR8 (leucine-rich-repeats), CLIPs [202, 348]. Putative oenocytoids also expressed well known markers such as PPOs (2, 4, 5, 6, 9), fibrinogens, and fibronectins. Potential prohemocytes shared many of the granulocyte markers, including collagens, LRR (leucine-rich-repeats), SPARC, CLIPD1, but also ferritin and ribosomal genes. Of note, expression of granulocyte markers in prohemocytes is not fully abrogated, but rather of lower intensity, suggesting granulocytes and prohemocytes might be different cell states, and not cell types.

We also characterised previously unknown hemocytes classes. For instance, 120 cells baptised ‘secreting hemocytes’ specifically expressed proteins with N-terminal signal peptides for secretion, such as e.g. LYSC1, TEP3, ficolins, cecropins, and defensins. A cluster of 131 ‘Rapidly dividing granulocytes’ was enriched in cell cycle and spliceosome markers such as aurora kinase, Cyclin Bs (G2/Mitotic specific), polo-kinase 1, inhibitor of apoptosis 5, Barrier-

to-autointegration factor B. Finally, 85 ‘effector hemocytes’ were characterised by high expression of LITAF (LPS-Induced TNF-alpha transcription factor) 3 and LITAF 1, AGAP007318 (an uncharacterised membrane protein upregulated in *P. berghei* infection [349]), Toll proteins, NFkappaB essential modulator, CLIPB8. Full table in Appendix.

Interestingly, fat body cells divided into two major cell states, correlated with activation. A baseline fat body state of 701 cells expressed many immune-related and regulatory genes such as CLIPs (CLIPA1, 7, 8, 9, 14), LRIMs (LRIM 1, 4A, 8A, 8B, 9, 17), lectins (CTL 4, MA2), APL1C, SRPN2, TEP1, and phenoloxidase inhibitor protein. Conversely, activated fat body cells (149 cells) highly expressed a canonical marker of fat body after feeding: vitellogenin. Finally, 121 cells have been classified as muscle cells due to the expression of markers such as troponin C, myosin light chain 5, myosin alkali light chain 1, myofilin variant C, and numerous transcripts related to energy production. A heatmap of the top 10 marker genes for each subtype follows below [Fig. III.9].

We also quantified each cell type cluster, looking at both number of cells and total UMI per cell in each cluster to reinforce our hypotheses regarding putative cellular identities. Putative cells types were then identified and quantified. Prohemocytes were the most common cell type with 2034 cells, followed by granulocytes (1553). Baseline fat body cells followed with 701, oenocytoids with 489, and fat body with 149. Rare cells included dividing granulocytes (131), muscle (121), secretory cells (120), and effector cells (85). We classified cell types by taking into consideration both the RNA content of cells - using the number of UMIs per cell as a proxy - as well as the analysis of the differentially expressed genes between each cell cluster. Putative prohemocytes were characterised by a low number of UMIs (yet distinct from background as shown by Fig. III. 8B), consistent with a high nuclear-cytoplasmic ratio and small overall size [Fig. III.8]. Conversely, granulocytes are transcriptionally active, have large diameters, and have high UMIs, similarly to oenocytoids.

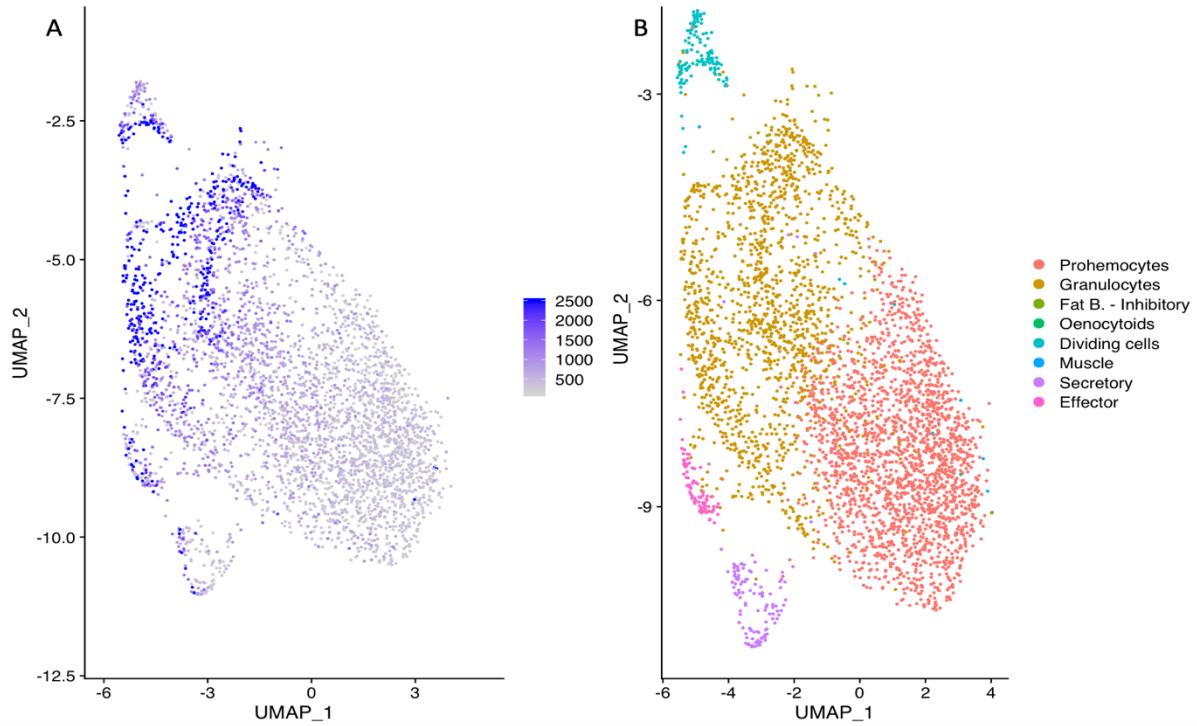


Fig. III.8 UMI count as proxy for size suggests prohemocyte-granulocyte split. Clustering done as above, data split to remove oenocytoids, fat body, and muscle cells **(A)** number of UMIs per cell plotted onto the UMAP visualisation of selected cells, capped at 2500 UMIs to aid visualisation **(B)** clustering solution mapped onto UMAP as above.

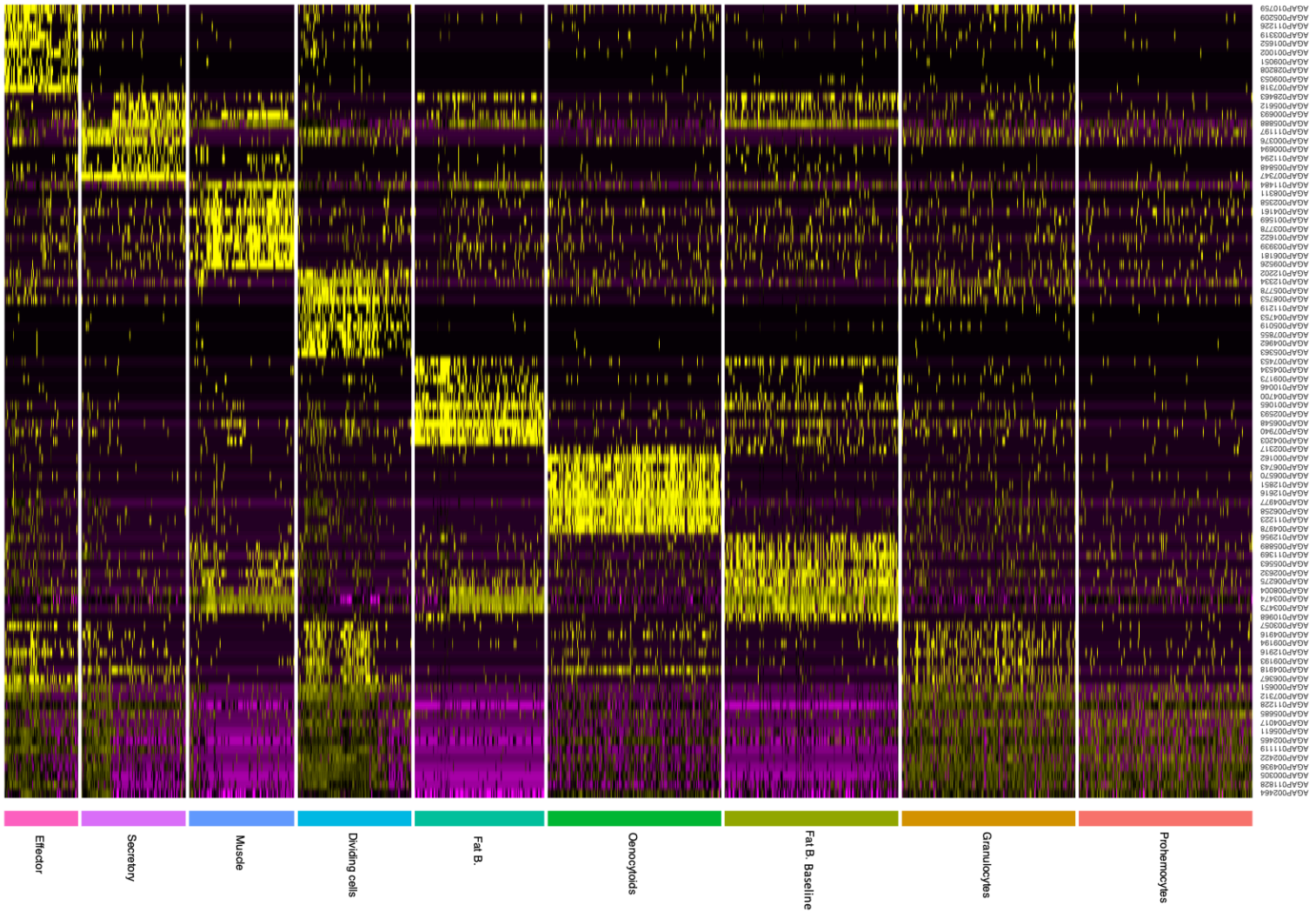


Fig. III.9 Heatmap of the top ten gene markers for each cell type identified. DE genes were identified with the Wilcoxon rank-sum test. P values were adjusted for multiple testing using the Bonferroni correction. All P-adjusted values < 0.001, ordered by average log fold change between cluster of interest and all other cells. Down-sampled to 300 cells per cluster for clarity.

3.1.6 Specific hemocyte markers for RNA-FISH validation identified by combining scRNA-seq and bulk RNA-seq results

We then set out to validate our cell types. The first step was to confirm the exclusive expression of cell type markers in hemocytes, excluding those also expressed in the mosquito midgut or the rest of the body (carcass). Bulk RNAseq of *Anopheles* hemocytes, guts, and carcasses was performed with the same time-points and conditions of the scRNAseq experiments: 1,3 and 7 days after sugar-feeding, blood-feeding, or mosquito infection with *P. berghei*. Between 8-12 mosquitoes per group were used for each condition, with three biological replicates to increase statistical power. After alignment, quantification, and normalisation (see methods) a PCA of the samples showed all biological replicates clustering together. Rather, samples correctly split by body part. Differences between carcass samples in red, gut samples in green, and hemocyte samples in green were the main drivers of sample diversity [Fig. III.10A]. Furthermore, sample-to-sample distances were plotted on a distance matrix to obtain a qualitative appreciation of similarities between samples. The correlation matrix once again demonstrates clear differences between three sample groups: guts, carcasses, and hemocytes.

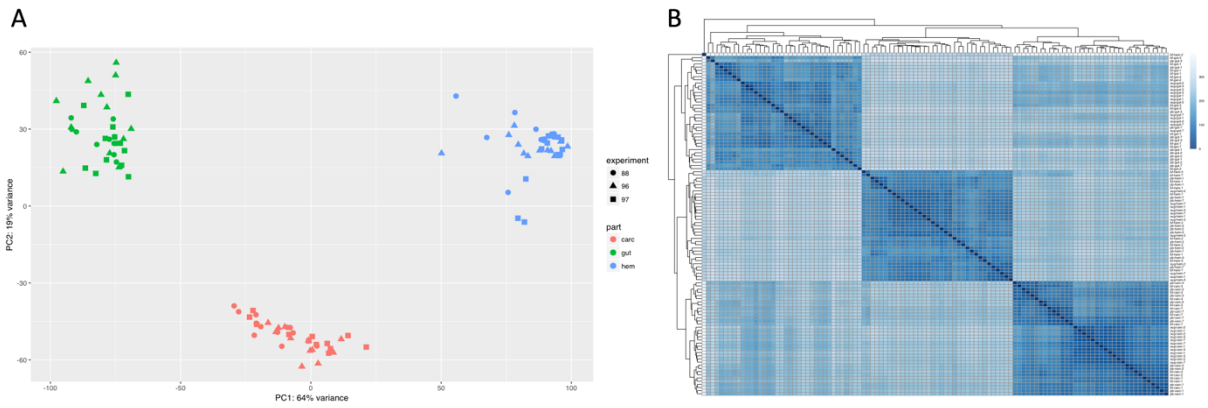


Fig. III.10 Bulk RNA-seq dataset QC. (A) PCA analysis and clustering of samples based on overall transcriptional similarity divides samples into three main groups: carcasses in red, guts in green, and hemocytes in blue **(B)** Distance matrix correlating the overall similarity and hierarchical clustering of each sample. Three large groups (gut top left block, hemocytes in the centre, and carcass at the bottom right)

After QC, normalisation, and fitting of a generalised linear model as of methods we performed a differential expression analysis with DESeq2 on hemocyte samples against the average expression of carcass and gut samples. We filtered for an adjusted p-value after Wald significance testing of P<0.001 and an absolute log2 fold change larger than 2 and identified 5126 differentially expressed genes, of which 1587 were upregulated in hemocytes and 3539 downregulated. Running separate DE analyses of hemocytes vs guts' samples and hemocytes vs carcasses returned similar results. Among the top upregulated genes in hemocytes we found well characterised genes associated either with hemocytes or with immune function, such as PPO2,3,5,6,9, fibrinogen and fibronectin, CLIPs, SPARC, laminins, collagens, scavenger receptors, toll proteins, LRIMs, TEP4, PPO activating factor, CD63, antimicrobial peptides, and REL1.

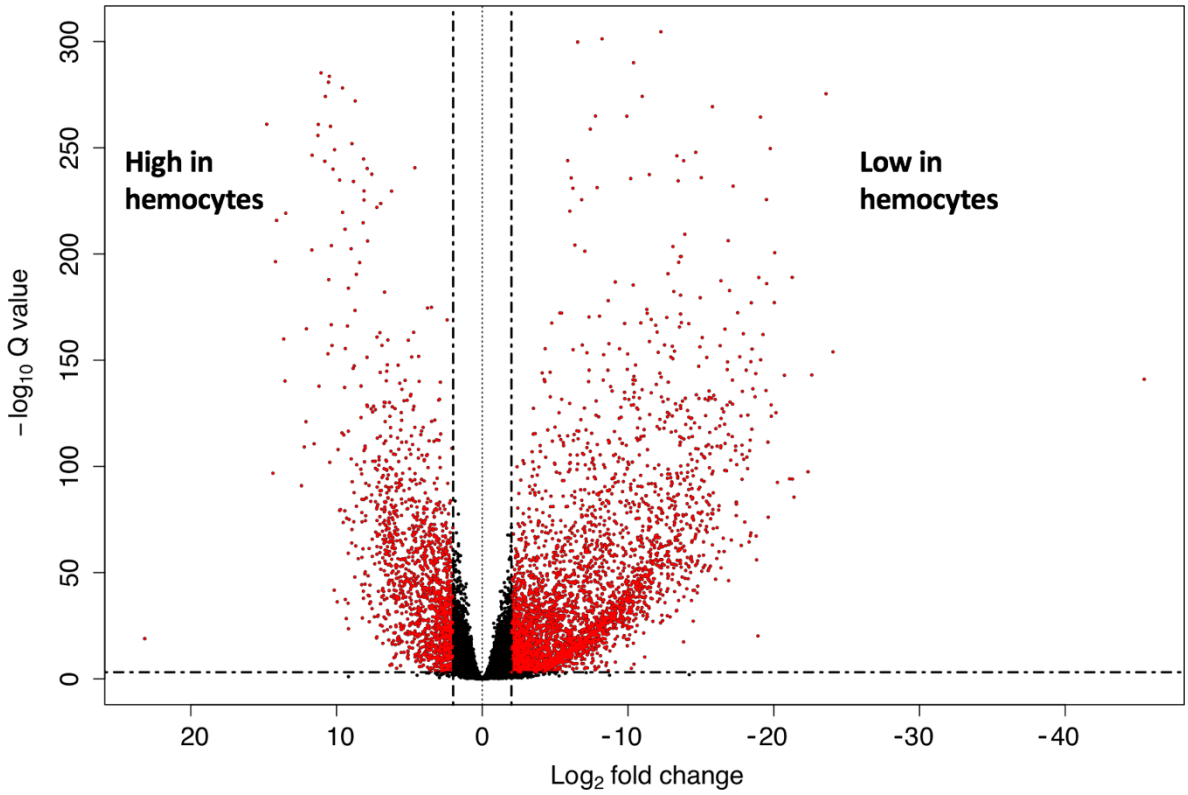


Fig. III.11 Differential expression analysis - hemocytes vs carcasses and guts. DEseq2 DE analysis of hemocytes vs averaged gut and carcass expression, filtered for $\log_2$ fold change >2 and Wald significance testing Q <0.001.

Functional classes of mosquito hemocytes

There was a strong correlation between markers identified by bulk RNAseq and biomarkers of scRNA-seq cell clusters. Especially so for common cells such as prohemocytes (91.2% of sc-RNAseq markers also present in the list of positively upregulated genes in bulk RNAseq hemocytes' samples) and granulocytes (71.3%). Less markers were identified for rare cell types such as secretory cells (only 28.1%) or muscle cells (25.9%), and intermediate levels for cell types such as dividing cells (44.3%) and effector cells (46.5%). Non-hemocyte contaminants such as fat body cells, are also well represented (86.6% and 50.0% for baseline fat body and activated fat body respectively). These cells are large and feature substantial amounts of RNA.

Cluster	Total markers - scRNAseq	Pos. in bulk RNAseq	Percentage
Prohemocytes	34	31	91.2
Granulocytes	178	127	71.3
Fat B. - Baseline	112	97	86.6
Oenocytoids	52	39	75.0
Fat Body	118	59	50.0
Dividing cells	221	98	44.3
Secretory	32	9	28.1
Muscle	58	15	25.9
Effector	99	46	46.5

Table III.4 Correlation of scRNA-seq markers with positively upregulated bulk RNAseq markers in hemocyte samples. First, scRNA-seq marker genes were filtered to select those with Wilcoxon test p adjusted value <0.05. The resulting table was then merged with DE markers in bulk RNAseq hemocyte samples as above, filtered for log2 fold change >2 and Wald significance testing of Q <0.001.

Once DE genes between hemocytes and mosquito midguts and carcasses were identified we cross-references the top ten marker genes for each cluster to the bulk RNAseq gene list to identify the best marker of each cellular subtype for RNA - FISH validation. Markers were selected according to the following criteria:

1) Highest and most specific expression of markers in each scRNA-seq cell type cluster
2) Highest and most specific expression of markers in bulk RNAseq data of hemocytes

Markers were selected using the clustering solution identified with Seurat v2.4. The following table summarises our findings. All markers previously identified and then validated via RNA-FISH were also found to be valid cellular markers in the new Seurat v3 analysis.

Markers	scRNA - specificity	scRNA - expression	Bulk vs gut - log2 fold	Bulk vs body - log2 fold	Description
General					
AGAP009623	n/a	n/a	n/a	n/a	GAPDH – pos. control
AGAP008296	n/a	n/a	-13.2	-7.6	Trypsin - gut
AGAP004203	+++	+++	4.1	-2.5	Vitellogenin - fat body
Hemocytes / Granulocytes					
AGAP004017	n/a	+++	7.3	4.8	LRR. All hemocytes
AGAP011974	n/a	++	5.6	4.2	SCRC1. General hemos
AGAP000790	n/a	+	6.6	4.7	Prohem. / granulocytes
AGAP003057	+	+	4.7	1.8	Active granulocytes
AGAP011871	-	+	2.6	1.2	Granulocytes
Rapidly dividing					
AGAP005363	+++	++	1.2	0.4	
Fat B. - Baseline					
AGAP007033	+	+	6.8	1.2	
AGAP028406	++	++	5.7	3.2	APL1C
Oenocytoids					
AGAP004981	++	++	10.4	4.8	PPO4
AGAP012851	+++	+++	6.9	4.7	Aldo-keto-reductase
AGAP012000	++	++	8.1	5.5	Fibrinogen/fibronectin
Effector					
AGAP007318	+++	++	5.3	2.8	TM7318
Secretory					
AGAP011239	++	++	4.0	2.9	Some also in oenos

Table III.5 RNA-FISH markers chosen by total expression and expression specificity in scRNA-seq and bulk RNAseq samples. scRNA-seq markers were cross-checked with gene tables of DE genes in bulk RNAseq (hemocytes vs guts and hemocytes vs bodies, separately). The most specific and highly expressed genes (qualitative assessment) were chosen.

3.1.7 RNA-FISH validation of putative cell types

We then validated our cell types via imaging. Dr. Ana Barletta Ferreira recovered hemocytes from mosquitoes that were sugar-fed, blood-fed or infected with *P. berghei*, spun the hemocytes onto slides coated with the adhesive Cel-Tek, then fixed them in paraformaldehyde. The cellular morphology was first captured by staining cells with actin and imaging them with confocal microscopy, and then RNAscope commercial RNA-FISH was then performed with the probes of Table III.5, and correlative fluorescent / FISH microscopy was performed by imaging the same area of the slide with confocal microscopy [Fig. III.12].

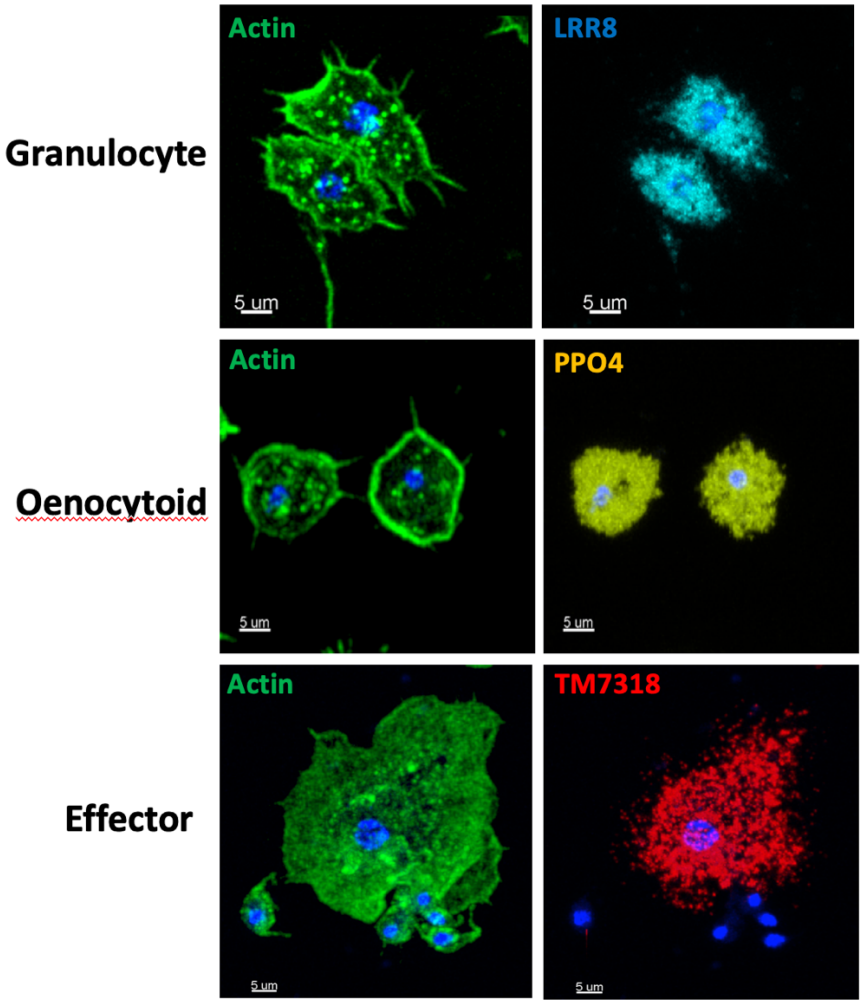


Fig. III.12 Correlation of hemocyte morphology with RNA-FISH markers. Main cell types were confirmed by matching to the left cellular morphology (actin), and to the right gene markers by RN-FISH. Blue is DAPI nuclear stain. Representative images from over 3200 cells.

Granulocytes were identified because of their larger size (10-20 µm) as compared to oenocytoids (8-12 µm) and prohemocytes (4-6 µm). In addition, granulocytes featured an increased number of pseudopodia. Oenocytoids also had pseudopodia, but they were shorter, and less prominent, and cells were rounder. Furthermore, the nuclear size in granulocytes was larger than in oenocytoids [Fig. III.12]. LRR8 mostly identified granulocytes and prohemocytes, whereas PPO_4 identified for the most part oenocytoids. Some cells were double-positive, but typically $LRR8_{high}$ cells would be $PPO4_{neg}$ or $PPO4_{low}$, and conversely $PPO4_{high}$ cells would be $LRR8_{neg}$ or $LRR8_{low}$ [Fig. III.13]

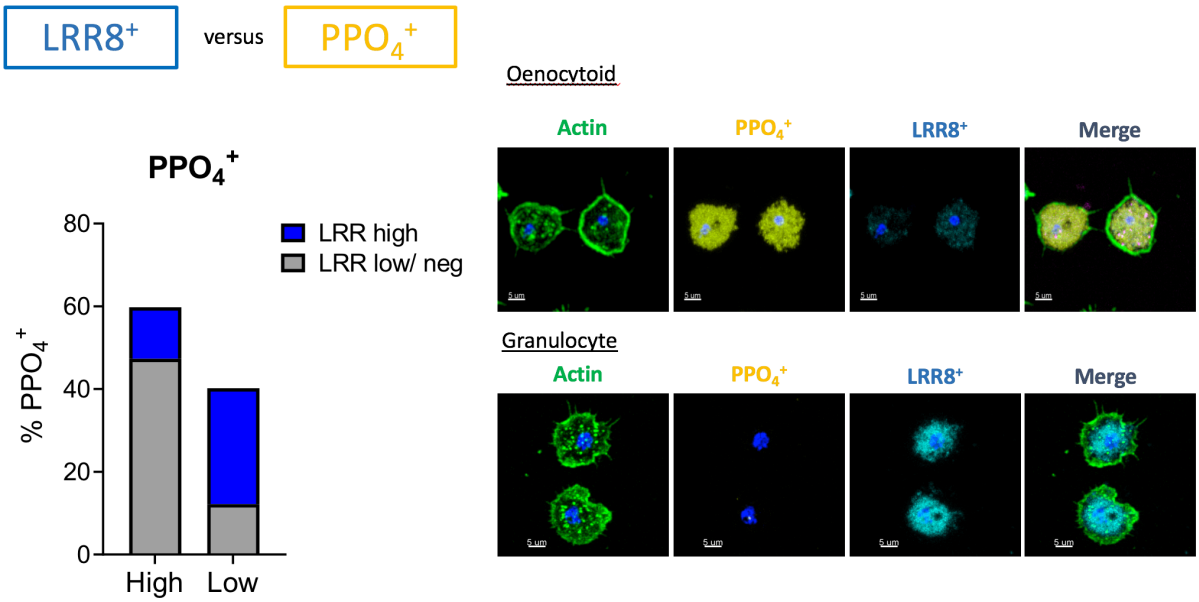


Fig. III.13 Granulocytes vs oenocytoids: morphology and RNA-FISH markers. LRR8+ cells could be split into LRR_8 high and low. PPO4+ cells (oenocytoids) were more likely to be LRR8 negative or low. The opposite for $PPO4_{low}$ cells. Representative images from 435 cells.

We then explored the spatial localisation of hemocytes in the *Anopheles* mosquitoes. Mosquitoes were then sugar-fed, blood-fed or infected with *P. berghei*, then fixed in paraformaldehyde, before paraffin embedding and sectioning. We performed RNA-FISH with the commercial technology RNAscope on the sections per RNAscope protocol and then imaged samples on an automated slide scanner or with confocal microscopy. We alternated one slide for haemotoxylin and eosin (H&E) staining and one slide for RNAscope. H&E staining was

useful to orient ourselves and identify the anatomical features of mosquitoes. In Fig. III.14 we can observe an H&E stain and mirrored RNA-FISH section of the mosquito. From the left to the right we can observe the compound eye, brain, thorax and wing muscles, abdomen and foregut, midgut, and fat body, as well as the ovaries.

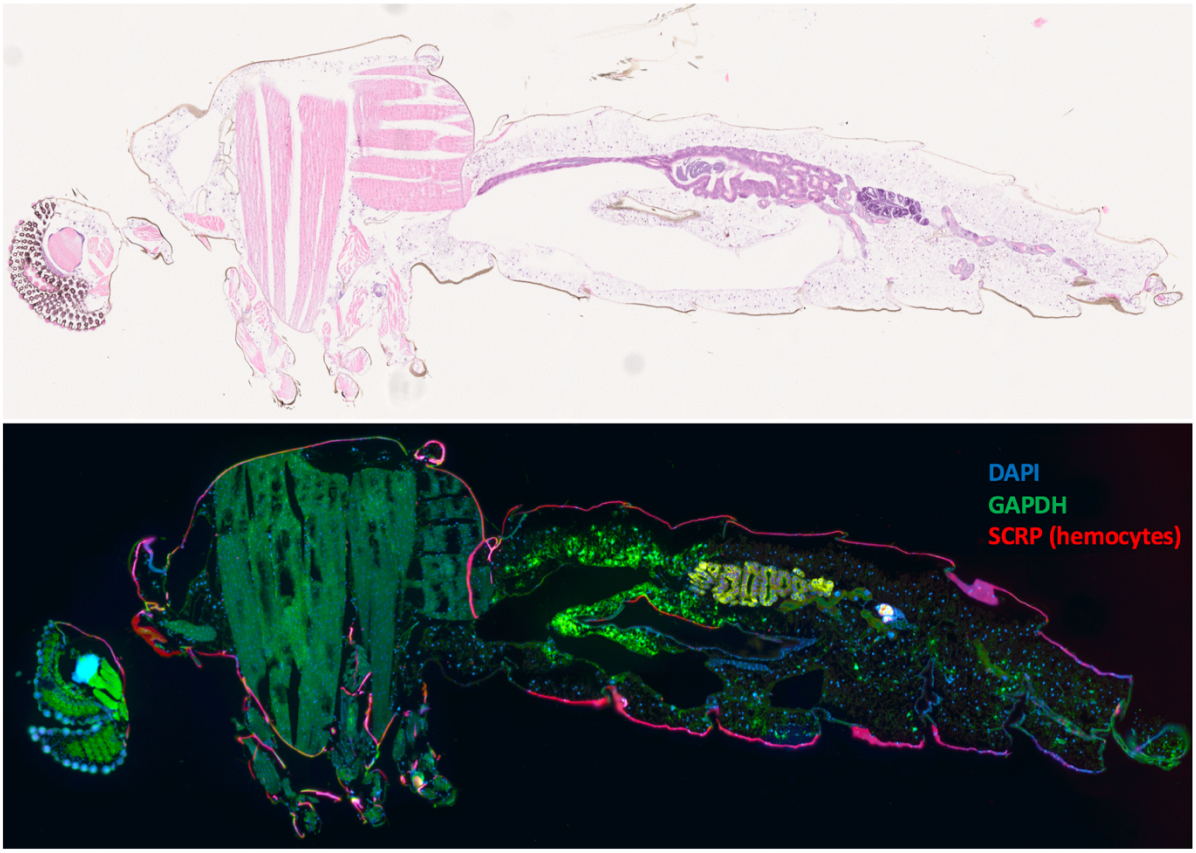


Fig. III.14 Overall view of the *A. gambiae* body with H&E and RNA-FISH. At the bottom, RNA-FISH of hemocytes (red, SCRC1 probe), cellular nuclei (blue DAPI counter-stain), and all mosquito cells (green, GAPDH positive control mosquito probe) on a longitudinal section of an *Anopheles* mosquito. At the top, mirrored H&E section. Both imaged with slide scanner.

Hemocytes can be seen patrolling all areas of the mosquito body, including the thorax - between flight muscles - and the abdomen, both in the fat body or attached to the gut. Hemocytes are found everywhere (except within the gut lumen or the central nervous system) but they particularly line areas of the body in potential contact with pathogens, such as the salivary glands, the proboscis, the gut lining, the rectal area, and the spermathecal vestibule of female mosquitoes. Hemocytes do not normally form clumps but appear as isolated cells,

although in these sections we mainly used the SCRC1 probe for our survey. SCRC1 is more specific for granulocytes and prohemocytes than oenocytoids, secretory, or effector cells.

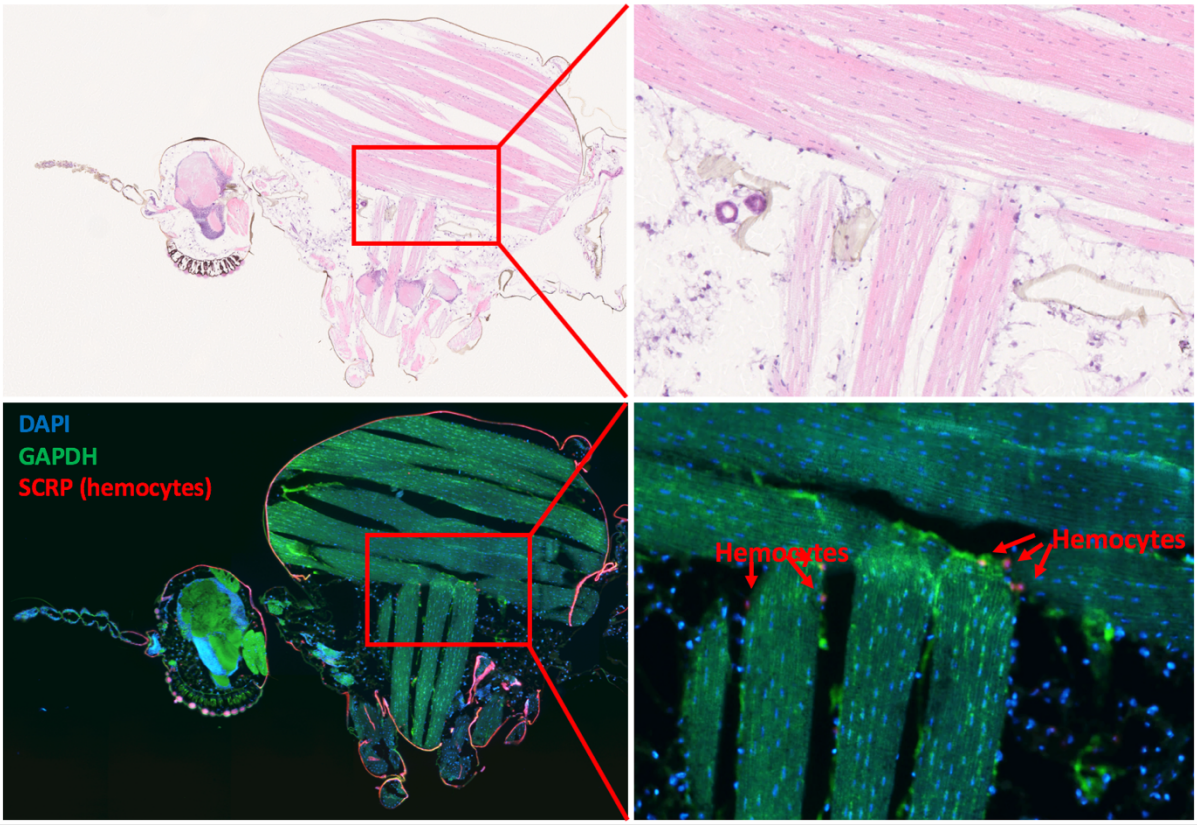


Fig. III.15 Hemocytes patrolling the thorax of *A. gambiae*. At the bottom, RNA-FISH of hemocytes (red, SCRC1 probe), cellular nuclei (blue DAPI counter-stain), and general mosquito cells (green, GAPDH positive mosquito control probe) on longitudinal section of *Anopheles* mosquito. At the top, mirrored H&E section. Both imaged with slide scanner.

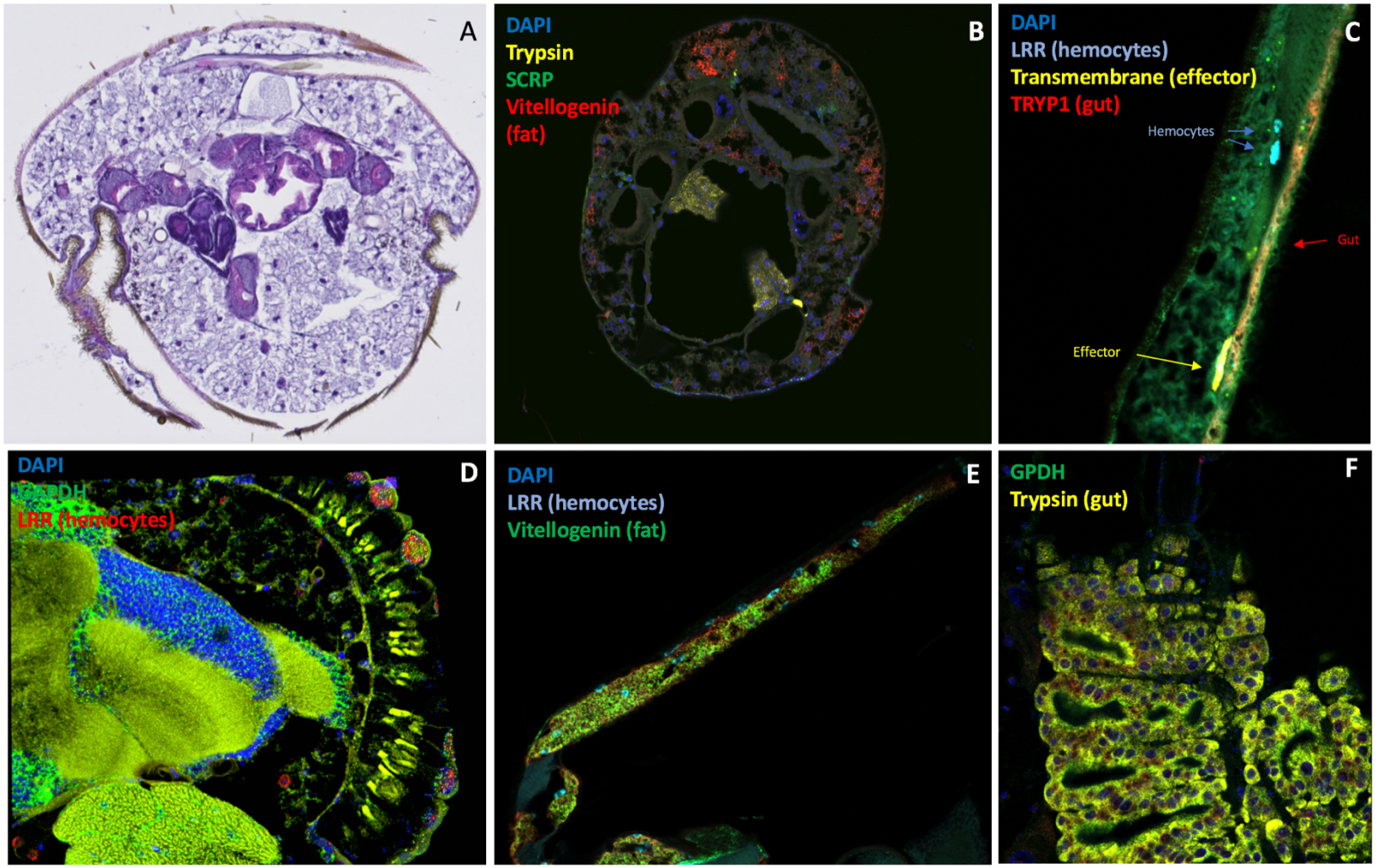


Fig. III.16 Hemocytes patrolling the *A. gambiae* body (A) Vertical H&E section of mosquito abdomen and **(B)** mirrored RNA-FiSH section. From **C to F** RNA-FISH of: gut lining in abdomen, CNS, proboscis, and gut. Imaged with slide scanner **(A-C, E-F)** and confocal microscopy **(D)**. RNA-FISH probes indicated in each separate panel.

Functional classes of mosquito hemocytes

Hemocytes can be both sessile and motile. Imaging requirements for each are different. To capture sessile hemocytes we injected paraformaldehyde inside the mosquito cavity before dissecting the mosquito midgut and the mosquito body wall (carcass). Then, whole-mount RNA-FISH of the whole organs were done with a modified RNAscope protocol (see methods). All hemocyte cell types for which we have probes were identified with the exception of the rapidly diving cellular subtype, for which we have yet to develop an appropriate probe. We observed the general hemocyte population, as well as specific oenocytoids, granulocytes, effector hemocytes, and secretory hemocytes. Body walls were especially rich in immune cells, with control blood fed body walls having 286 (±76 CI) hemocytes. Blood-fed control guts showed fewer numbers of cells, with a total of 23 (±6.6 CI) hemocytes. We also observed pericardial cells, staining positively with the AGAP007318 and AGAP011239 probes (effector and secretory probes). These cells could be recognised both by virtue of their characteristic arrangement along the dorsal wall as well as their larger size.

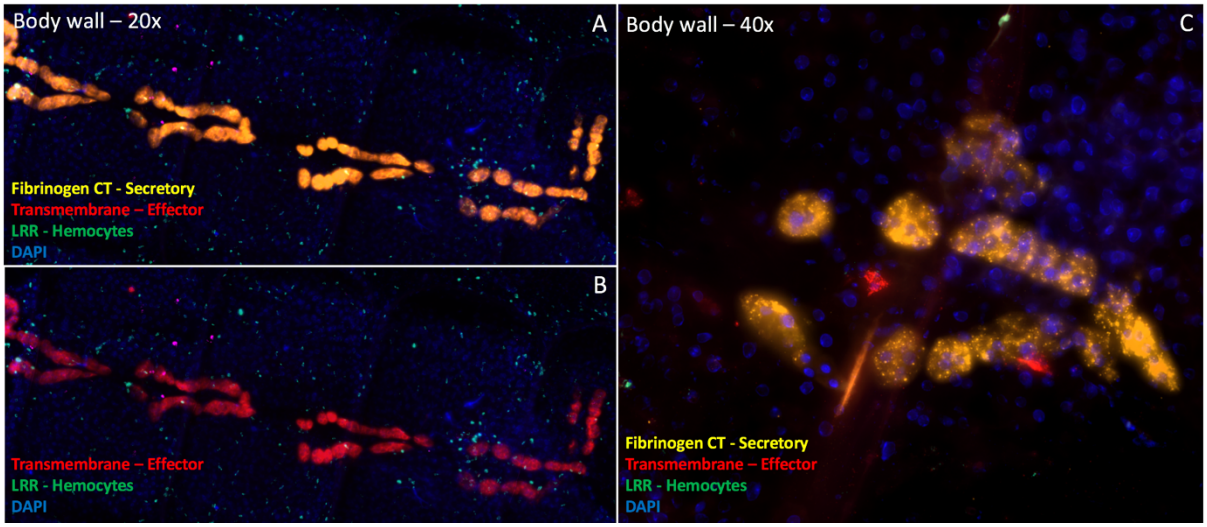


Fig. III.17 Pericardial cells along the *Anopheles* body wall (A) 20x whole-mount RNA-FISH shows AGAP007318 and AGAP011239 positive pericardial cells, in addition to immune cells **(B)** Same as above but without the Fibrinogen-CT probe to show positive staining for Transmembrane (Effector) probe **(C)** 40x whole-mount RNA-FISH of a separate mosquito wall. Two effector hemocytes can be seen in close proximity to the pericardial cells complex.

Functional classes of mosquito hemocytes

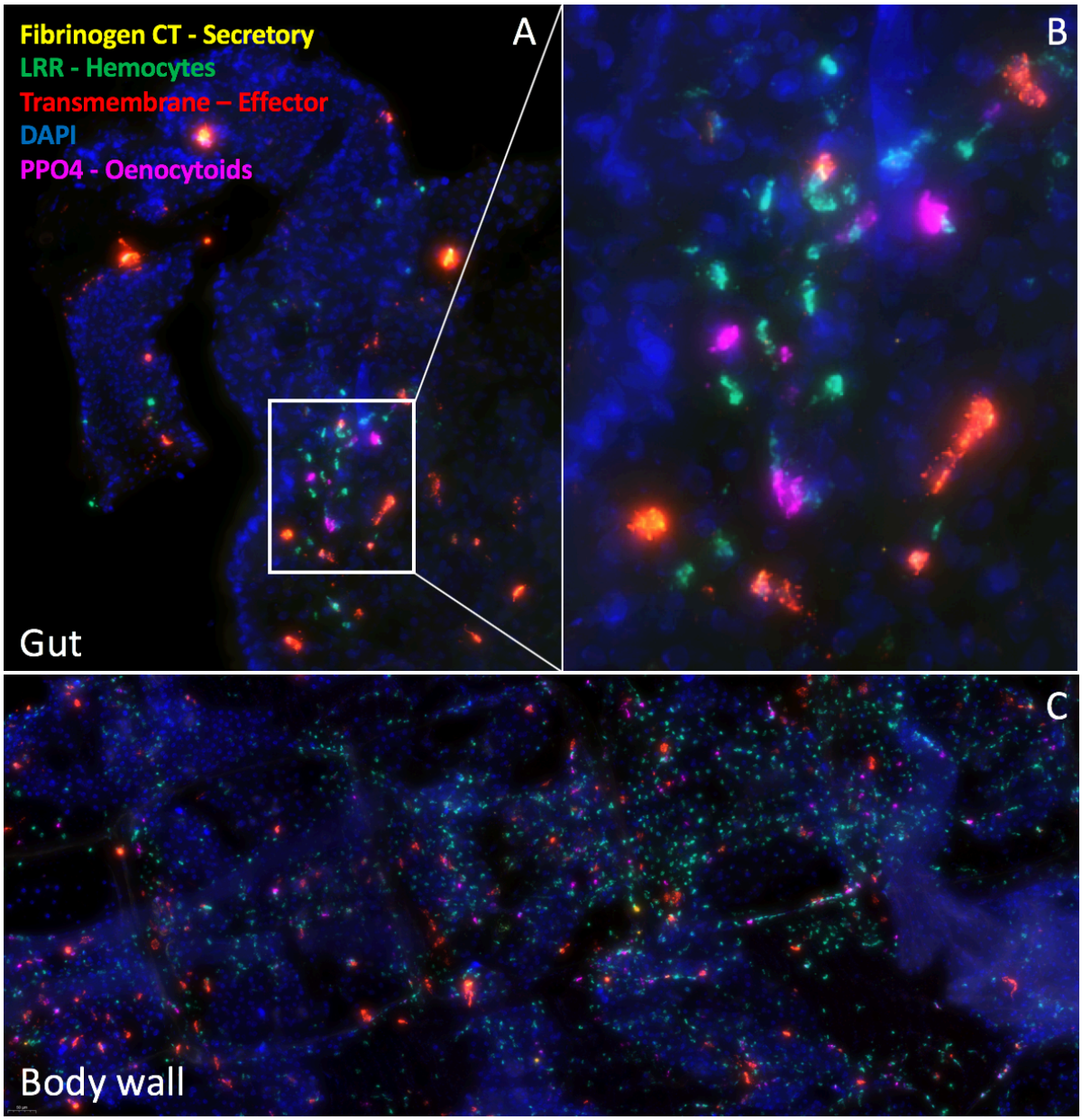


Fig. III.18 Mosquito midguts and bodies contain all subtypes of sessile hemocytes (A) A 20x view of the proximal part of a blood-fed control mosquito gut, with RNA-FISH of hemocytes (green, LRR probe), secretory cells (yellow, Fibrinogen C Terminal), effector cells (red, transmembrane), and nuclear counterstain (blue, DAPI) on whole mounts of *Anopheles* mosquito. **(A)** A 40x magnification of the gut. **(C)** A 20x whole mount view of a mosquito body wall with the same probe of above. All imaged with a slide scanner.

3.1.8 Distinct states within each cell type

While initially conservative in our clustering as to only capture true cell types rather than cell states, thresholding was then relaxed to identify subtler grouping of cells, which could theoretically split existing cell types into cell states, differentially responding to stimuli. There was hidden diversity within the original mapping, especially in the large granulocyte cluster. We observed a central disc of cells, surrounded by two separate hemi-discs. Importantly, the central group contained more cells from baseline conditions, whereas the two hemi-discs featured more active cells (blood fed and *P. berghei*-infected) [Fig. III.19A-B]. After iterating clustering until all clusters had at least more than 20 meaningful marker genes (adjusted p value <0.05) and were well-mixed among samples and conditions, we identified four additional cell states. Fat body cells divided into an additional cell state that sat between baseline cells and activated cell types based on the UMAP and the marker genes (see table III.6 below for top 10 genes, as well as figures III.19 and III.20). From the same figures and tables prohemocytes also split in two: a more active state defined by increased expression of hemocyte / granulocyte genes and a more inactive state with decreased gene expression. Granulocytes showed the largest transcriptional diversity, splitting into three different cell states: one putative baseline state, as well as two different types of more activated granulocytes [Fig. III.19C]. The baseline granulocyte cluster contained the highest number of inactivated cells (sugar-conditions), whereas activated cells came either from blood-fed or even more so from *P. berghei*-infected samples [Fig.19A, Fig. IV.1, Fig. IV.2]. A heatmap of the top 10 marker genes for each cell state more clearly showed how putative prohemocytes and granulocytes sat in a transcriptional programming continuum. Oenocytoids on the other hand still formed a distinct separate group on the UMAP, as well as on the marker genes heatmap. Furthermore, the heatmap also showed how within the prohemocyte-granulocyte group baseline granulocytes and prohemocytes were more similar to each other, whereas Type 1 and Type 2 active granulocytes show larger transcriptional differences.

Functional classes of mosquito hemocytes

Putative inactive prohemocytes

Gene	P_val_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP011828	4.98E-47	0.4644	0.843	0.725	cathepsin L
AGAP010163	2.85E-39	0.3039	0.943	0.961	60S ribosomal protein L38
AGAP007740	1.14E-36	0.2630	0.96	0.966	60S ribosomal protein LP1
AGAP012100	2.03E-36	0.2586	0.966	0.977	40S ribosomal protein S26
AGAP000305	5.64E-26	0.2972	0.877	0.739	SPARC
AGAP002464	2.68E-23	0.2907	0.95	0.909	secreted ferritin G subunit
AGAP029054	7.29E-17	0.3604	0.739	0.645	nimrod B2
AGAP002422	1.94E-15	0.5140	0.591	0.56	CLIP-domain serine prot
AGAP002465	5.35E-15	0.3314	0.804	0.78	ferritin heavy chain
AGAP013186	3.70E-07	0.2842	0.15	0.282	None

Putative active prohemocytes

Gene	P_val_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP004936	2.69E-63	0.58799	0.873	0.65	None
AGAP011119	7.93E-54	0.554097	0.843	0.638	None
AGAP011228	7.70E-47	0.445928	0.981	0.792	None
AGAP002464	1.29E-45	0.488801	0.974	0.908	secreted ferritin G subunit
AGAP005611	7.84E-37	0.50457	0.775	0.65	None
AGAP000305	1.15E-20	0.30061	0.899	0.745	SPARC
AGAP002465	2.32E-19	0.365025	0.854	0.773	ferritin heavy chain
AGAP011828	8.19E-19	0.297401	0.86	0.73	cathepsin L
AGAP002422	1.36E-18	0.475993	0.654	0.551	CLIP-domain serine prot
AGAP002878	6.99E-13	0.519473	0.509	0.408	Cystatin-like protein

Putative baseline granulocytes

Gene	P_val_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP011228	4.6E-101	0.74197	0.988	0.796	None
AGAP011119	7.82E-93	0.682409	0.946	0.628	None
AGAP004936	2.29E-76	0.630795	0.939	0.646	None
AGAP007312	1.03E-65	0.66471	0.781	0.428	None
AGAP006278	7.85E-62	0.583197	0.893	0.583	None
AGAP005611	2.25E-58	0.519989	0.915	0.632	None
AGAP002594	1.50E-57	0.602139	0.743	0.426	apolipoprotein D
AGAP000790	2.86E-56	0.799228	0.47	0.196	None
AGAP000305	4.56E-56	0.516235	0.961	0.74	SPARC
AGAP000964	4.50E-51	0.672595	0.66	0.353	None

Putative granulocytes T2

Gene	P_val_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP006367	6.0E-165	1.427458	0.547	0.104	None
AGAP004916	1.54E-89	1.210717	0.38	0.087	None
AGAP004164	8.04E-80	0.974785	0.543	0.181	glutathione S-transf delta cl. 1
AGAP003016	1.11E-79	0.930581	0.446	0.125	mesenceph. neurotroph hmlg
AGAP029139	7.64E-76	0.98333	0.604	0.238	None
AGAP007120	1.16E-72	0.720407	0.901	0.584	nucleoside-diphosphate kinase
AGAP004743	2.90E-70	0.838938	0.657	0.275	Transmembr. emp24 containing
AGAP009194	5.10E-67	1.183577	0.407	0.124	glutathione S-transf. epsilon 2
AGAP005861	1.00E-66	0.877063	0.428	0.131	Translocon-associated subun β
AGAP004918	1.90E-60	1.094499	0.596	0.282	fibrinogen

Putative granulocytes T1

Gene	P_val_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP011828	7.3E-109	0.943378	0.983	0.73	cathepsin L
AGAP009156	6.93E-97	1.016372	0.505	0.118	None
AGAP004993	4.51E-93	1.109549	0.84	0.427	laminin subunit alpha
AGAP009201	1.17E-92	1.130115	0.842	0.481	collagen type IV alpha
AGAP011974	7.29E-88	1.013233	0.732	0.291	Class C Scavenger Receptor
AGAP002599	9.63E-83	0.916165	0.818	0.387	polyubiquitin
AGAP002016	3.58E-82	0.988187	0.545	0.158	iron/zinc purple acid phosphata
AGAP002879	3.12E-73	0.8705	0.78	0.357	cathepsin F
AGAP028157	1.02E-70	0.824397	0.452	0.12	None
AGAP013509	1.26E-70	0.947402	0.72	0.322	carboxylesterase clade H, 1

Putative fat body baseline T.1

Gene	P_val_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP010968	0	2.657141	0.702	0.053	CLIPA9
AGAP008013	2.8E-303	1.993149	0.418	0.013	None
AGAP005563	2.5E-290	2.843697	0.731	0.084	Tret1
AGAP011792	5.1E-269	2.177422	0.541	0.039	CLIPA7
AGAP006275	7.7E-261	2.351344	0.86	0.156	None
AGAP008227	7.3E-258	2.216784	0.737	0.097	trehalose 6-phosphate synth
AGAP002588	4.2E-254	1.689412	0.38	0.014	None
AGAP013060	6.1E-250	1.889122	0.804	0.123	None
AGAP008688	1.0E-245	2.040984	0.392	0.017	None
AGAP006177	3.9E-245	1.748359	0.406	0.02	None

Functional classes of mosquito hemocytes

Putative fat body T1

Gene	P_val_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP004203	8.4E-162	3.006847	0.782	0.096	vitellogenin
AGAP007940	3.8E-126	2.764072	0.721	0.109	Reticulon-like protein
AGAP006548	2.6E-124	2.550861	0.912	0.214	glycine cleavage system H
AGAP002593	8.3E-116	2.110959	0.435	0.035	outer membrane lipoprot Blc
AGAP001065	1.9E-104	2.542809	0.769	0.15	glycine hydroxymethyltransf
AGAP004700	7.0E-102	2.252557	0.381	0.03	None
AGAP010046	1.58E-89	2.525451	0.293	0.019	None
AGAP009173	2.84E-84	2.202421	0.381	0.037	fructose-1,6-bisphosphatase I
AGAP001116	5.13E-80	1.918763	0.442	0.054	D-amino-acid oxidase
AGAP002198	7.81E-78	2.063797	0.463	0.062	glycine N-methyltransferase

Putative fat body T2

Gene	P_val_adj	Avg_logFC	Pct.1	Pct.2	Annotation
AGAP003473	3.3E-163	2.480769	0.865	0.305	None
AGAP003474	6.5E-160	2.15225	0.992	0.955	None
AGAP005888	1.2E-135	1.620302	0.945	0.563	None
AGAP002632	1.5E-105	2.280139	0.701	0.265	None
AGAP004203	9.03E-93	2.308033	0.437	0.091	vitellogenin
AGAP012571	3.67E-91	1.310944	0.673	0.222	None
AGAP008011	3.37E-85	1.437902	0.382	0.072	None
AGAP008004	7.54E-82	1.195067	0.813	0.409	None
AGAP028386	2.64E-81	1.502057	0.799	0.469	NADH dehydr subunit 6
AGAP028373	3.34E-77	1.474149	0.626	0.23	NADH dehydr subunit 3

Table III.6 Marker genes for each cell state cluster. P_val_adj = P value adjusted for multiple testing. Avg_logFC = average log fold change for the gene between cluster of interest and other clusters. Pct.1 = percentage of cells in cluster of interest where gene is detectable. Pct.2 = percentage of cells in other clusters where gene is detectable. Annotation = electronic annotation of gene.

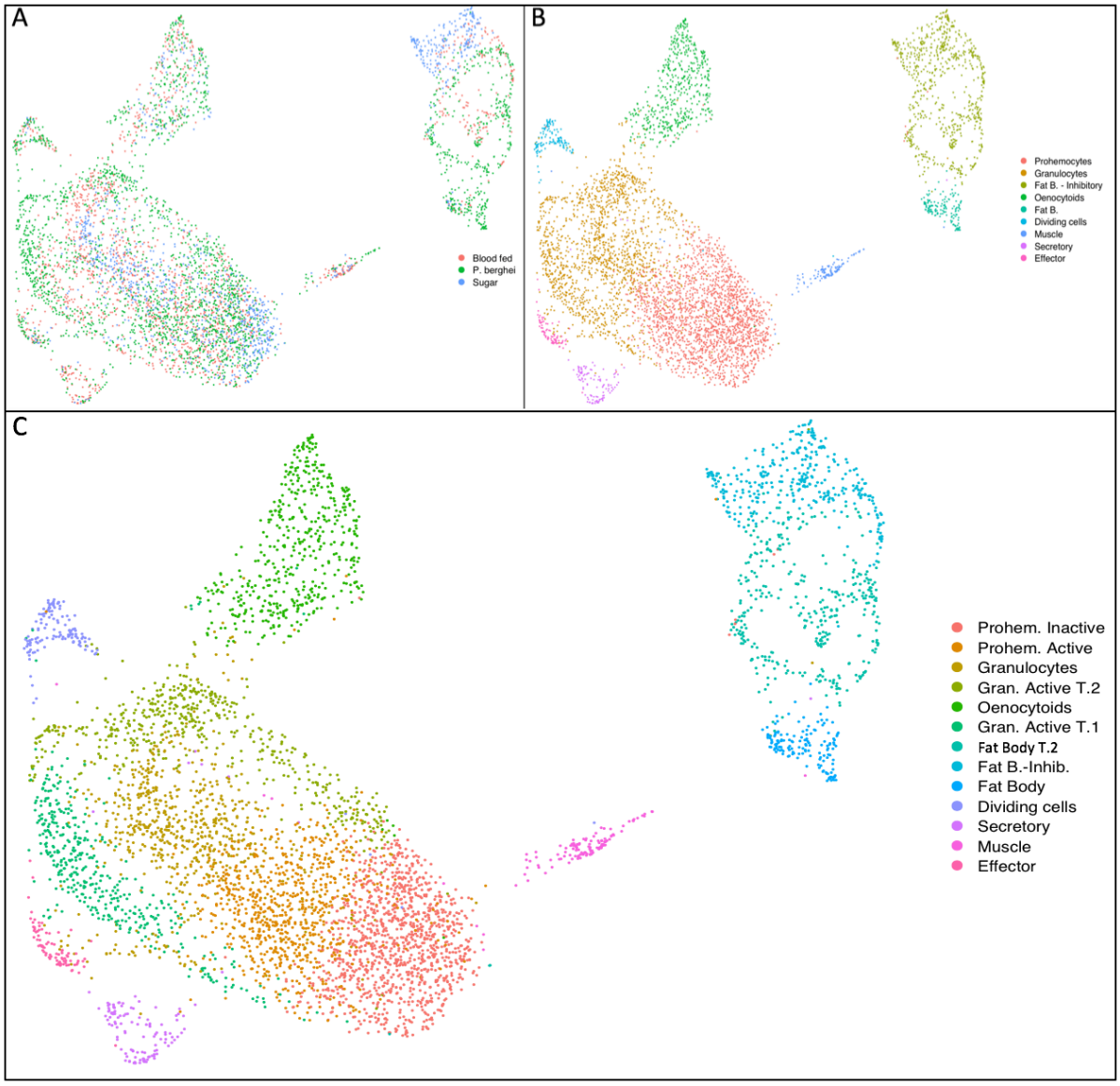


Fig. III.19 Diversity within cell types. (A) UMAP coloured by experimental condition. Within the putative granulocyte cluster, cells from sugar-fed (in blue) mosquitoes segregated from blood-fed mosquitoes (red), and more so *P. berghei* mosquitoes (green) **(B)** UMAP of cells clustered with 0.3 resolution (conservative subdivision identifying cell types) **(C)** UMAP of cells clustered with 0.7 resolution to identify cell states within the larger cell types.

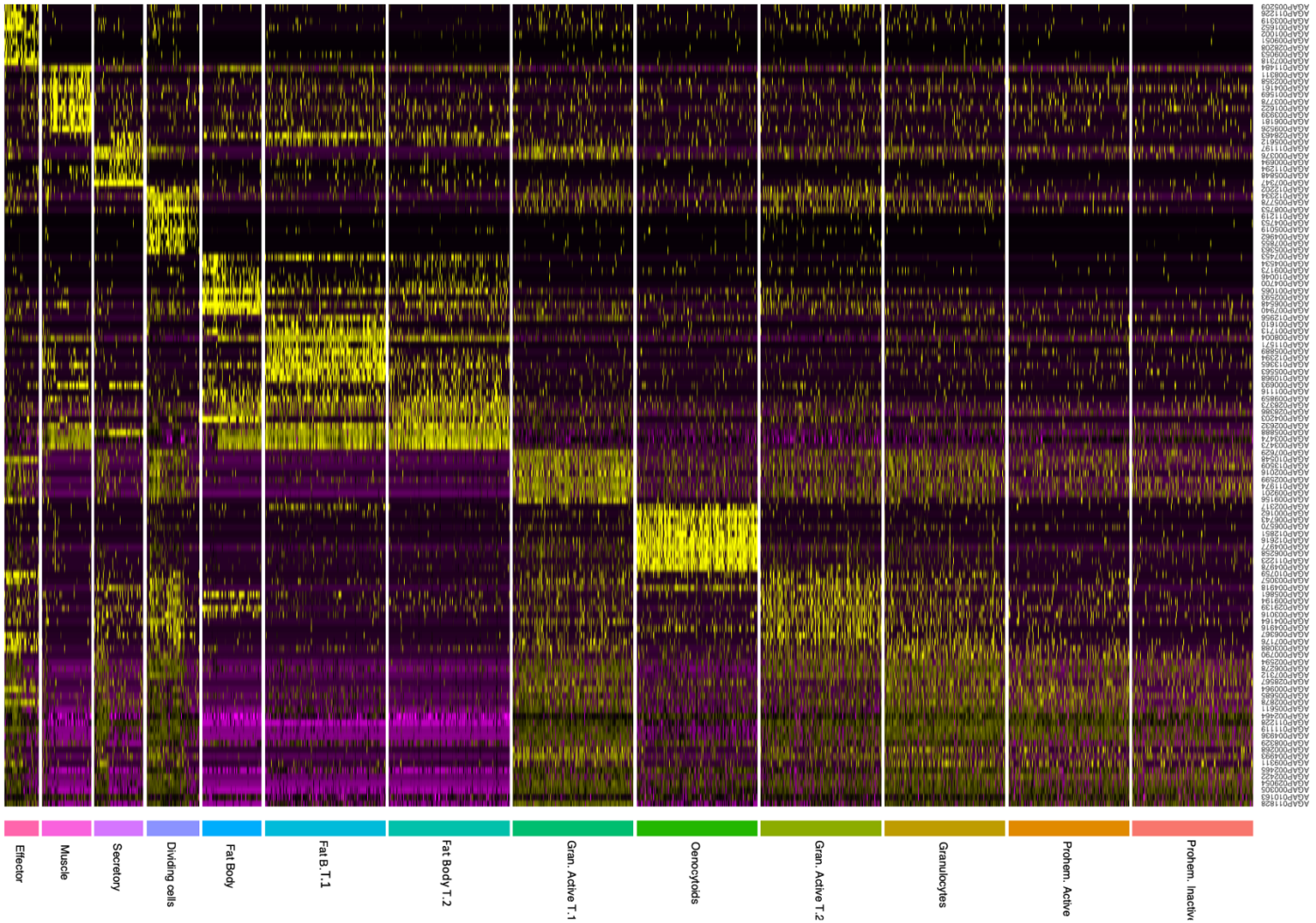


Fig. III.20 Heatmap of top ten gene biomarkers for each cell type or state. DE genes were identified with the Wilcoxon rank-sum test. The P values were adjusted for multiple testing using the Bonferroni correction. P-adjusted values < 0.001, ordered by average log fold change between cluster of interest vs all other cells. Down-sampled to 300 cells per cluster for clarity.

3.1.9 Distinct hemocyte lineages in *A. gambiae* mosquitoes

Hemocytes differentiation dynamics are unclear. To understand whether prohemocytes are true stem cells or a separate lineage we used cellular states subdivision to perform lineage tree reconstruction with the partition-based graph abstraction (PAGA) method. By combining clustering and pseudotemporal algorithms we were able to infer hemocyte trajectories and differentiation paths. We chose PAGA as it was recently shown to be the most accurate and robust lineage analysis software for complex datasets [311]. As a positive control, PAGA correctly identified fat body cells and muscle cells as separate clusters with no close connection to other cell types. Oenocytoids were also shown to be disconnected from other hemocyte subtypes, indicating a wholly separate lineage, while all other cell states were connected along a linear differentiation trajectory with inactive baseline prohemocytes at one end, moving towards active prohemocytes and granulocytes, before splitting into three different lineages. Secretory cells formed their own lineage from baseline granulocytes, while the two intermediate activated granulocyte cell subtypes split into either effector granulocyte subtypes or dividing granulocytes. Dividing cells reverted back into activated granulocytes type 2, replenishing the granulocyte cell pool after immune activation. We were thus able to identify a branching event centred on granulocytes thanks to an unsupervised network analysis. Nodes were identified with Seurat and connected by PAGA into a biologically meaningful network.

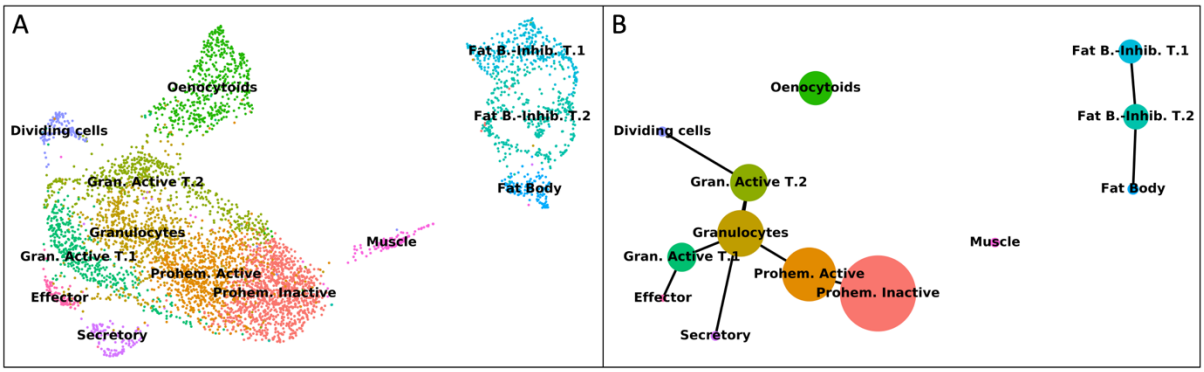


Fig. III.21 Cell lineages in adult *Anopheles*. (A) Graphical mapping of cell states with UMAP **(B)** Unsupervised PAGA network analysis of *Anopheles* hemolymph cells uncovers separate lineages and a branching event. Nodes correspond to clusters identified with Seurat while edges are putative cluster transitions.

Functional classes of mosquito hemocytes

We then confirmed the connections between clusters in the granulocyte lineage with a different method, diffusion maps. Like PCA, diffusion maps are another popular dimensionality reduction technique. However, diffusion mapping is a non-linear dimensionality reduction technique which aligns cells based on transcriptional similarities rather than clustering them. Hence, diffusion components (DCs) emphasize transcriptional transitions, which is particularly useful when analysing processes that are continuous, as for instance differentiation. Our data set showed DC1 to recapitulate the interconnectivity of prohemocytes, active prohemocytes, granulocytes, and active granulocytes type 1 and 2. These existed in a continuum of differentiation which includes dividing cells, whereas effector, secretory, and diving cells split along their independent trajectories [Fig. III.22A-B]. A DC1 vs DC3 plot showed that rapidly dividing cells and active granulocytes type 2 sat on a common differentiation trajectory, as expected from PAGA lineage tracing [Fig. III.22C]. DC1 vs DC3 also showed the opposite lineage (effector cells) emerging from active granulocytes type 1 [Fig. III22.D]. DC2 recapitulated hemocyte cell maturity: young, proliferating cells sat diametrically opposite to mature effector cells such as effector and secretory hemocytes [Fig. III.22E].

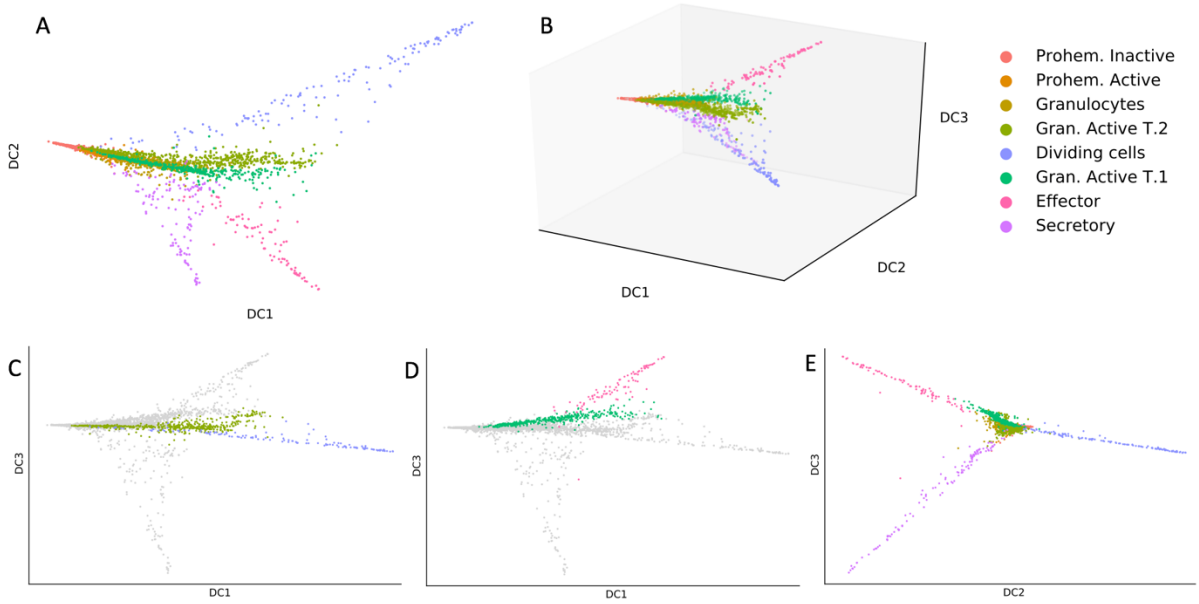


Fig. III.22 Diffusion maps confirm hemocyte lineages. (A) 2D diffusion map of granulocytes **(B)** 3D diffusion map of granulocytes **(C)** Diffusion Component 1 (DC1) vs DC3 plot highlights transition between dividing cells and granulocytes T.2 **(D)** DC1 vs DC3 plot highlights transition between effector cells and granulocytes T.1 **(E)** DC2 showcases hemocyte maturity, with proliferating cells on the right and differentiated states on the left.

Lastly, hemocyte lineages were also confirmed with the lineage analysis software package Slingshot, another highly rated lineage tracing software. It does not perform as well as PAGA when dealing with complex dataset containing multiple separate lineages, but it does work well in branching analyses [311]. As such, we subset our dataset to only include the three interconnected granulocyte-prohemocytes branches, and then run Slingshot. The results confirmed PAGA and diffusion maps findings. Slingshot identified three separate lineages originating in the inactive, baseline prohemocytes, moving into active prohemocytes and standard granulocytes, before branching alternatively into Type 2 active granulocytes and dividing granulocytes, or into Type 1 active granulocytes and then effector or secretor cells. Cells were ordered along a pseudotemporal dimension showing the differentiation of each hemocyte lineage. Pseudotime reconstruction was comparable between Slingshot and diffusion maps, with in blue baseline inactive prohemocytes, and in yellow the terminal effector states or proliferating cells. The central basal granulocyte cluster appeared once again to be the main branching point of the prohemocyte-granulocyte system [Fig. III.23].

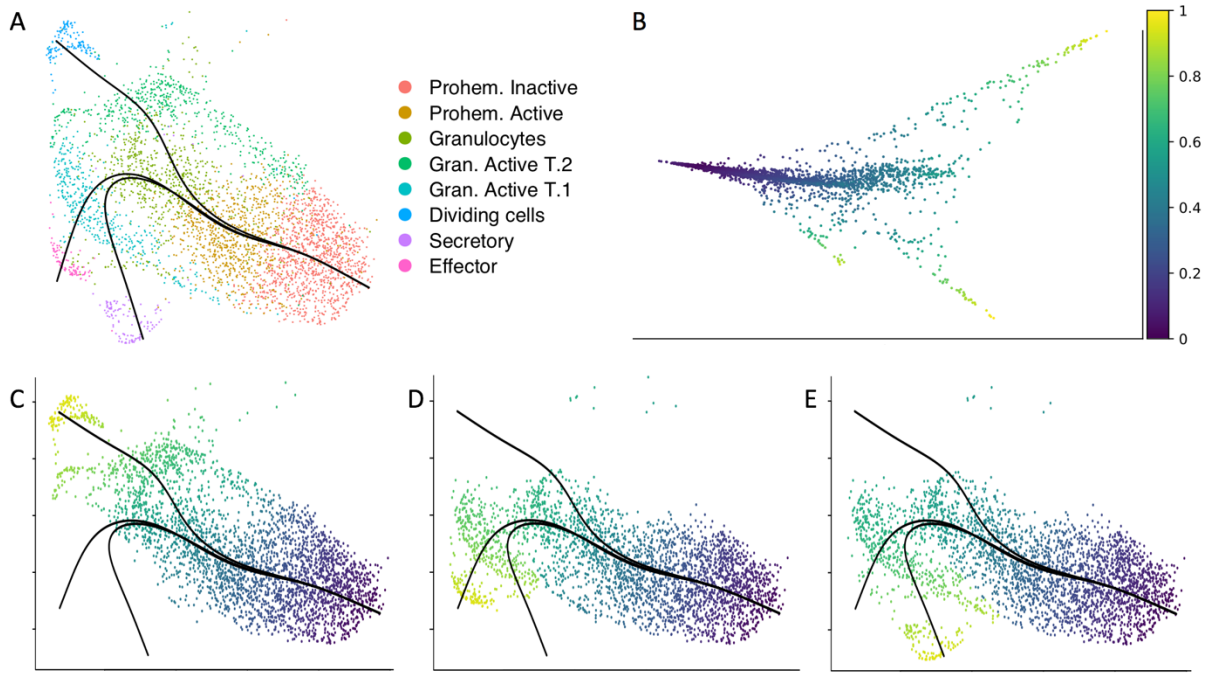


Fig. III.23 Slingshot lineage tracing and pseudotime reconstruction of granulocytes and prohemocytes (A) Slingshot analysis after subsetting non-hemocytes and oenocytoids. **(B)** Pseudotime reconstruction on DC1 vs DC2 **(C-E)** Pseudotime reconstruction with Slingshot for each separate lineages from prohemocytes to **(C)** Dividing **(D)** Effector **(E)** Secretory cells.

Functional classes of mosquito hemocytes

After trajectory identification, generalized additive models (GAMs) were fitted with the package tradeSeq, estimating one smoother per lineage with a negative binomial distribution. A total of 1018 highly expressed genes were filtered for the analyses. The TMM effective library size was internally used as offset by the model, which also allowed to fit zero inflated negative binomial to deal with zero inflation. After filtering for Wald test score >150 and a p-value <0.001 we identified 57 DE genes whose expression changed along lineage 1 (prohemocytes to granulocytes to rapidly dividing), 28 DE genes for lineage 2 (prohemocytes to granulocytes to secretory), and 40 for lineage 3 (prohemocytes to granulocytes to effector cells). Lineage 1 DE genes included PPO6, fibrinogen, cofilin, actin 5C, ARP2/3 complex, and many ribosomal transcripts. Lineage 2 DE genes featured cecropin, LYSC1, collagen Type IV alpha, laminin subunit alpha, cathepsin, LRIM16A, actin 5C, SPARC, class C scavenger receptor. DE genes for lineage 3 were largely similar to lineage 2, further demonstrating their similarity. LITAF3 (LL3), laminin gamma 1, LRIM16B were however specific for this lineage [Fig. III.24]. Overall many marker genes identified with Seurat were also independently found in this independent pseudotime-based analysis.

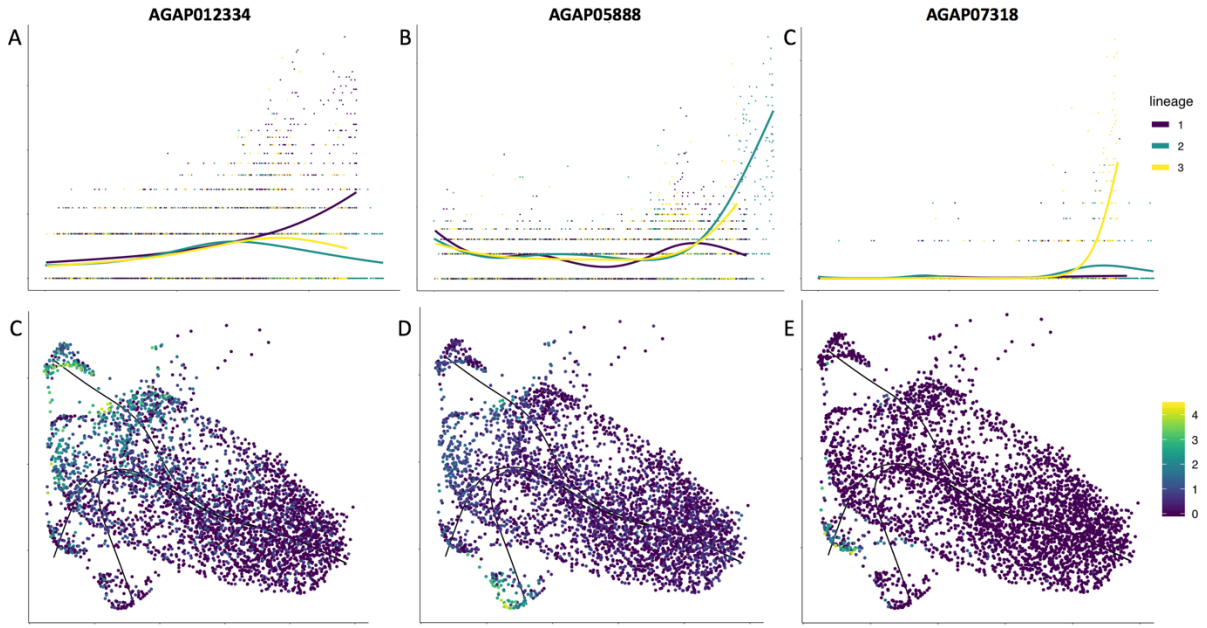


Fig. III.24 DE analysis of lineage-specific genes based on Slingshot pseudotime. (A-C) Smooth curves showing expression by pseudotime for the top three DE genes for each lineage **(D-E)** Corresponding expression of the top 3 DE genes on UMAP of prohemocyte-granulocyte lineage. Blue low transcript counts, yellow highest transcript counts.

Lastly, we analysed correlative microscopy images to help validate our lineage tracing hypotheses. Putative intermediate and early stages of both hemocytes and oenocytoids could be found, defined by a smaller cell size, smaller nuclei, lower expression of marker genes, and rounder morphology. Finally, less mature forms were likely to have less, or be void of, pseudopodia. The images are consistent with a cell development hypothesis that holds prohemocytes as the starting point, before branching differentiated cell types, both for LRR8+ hemocytes and oenocytoids [Fig. III.25 and Fig. III.26].

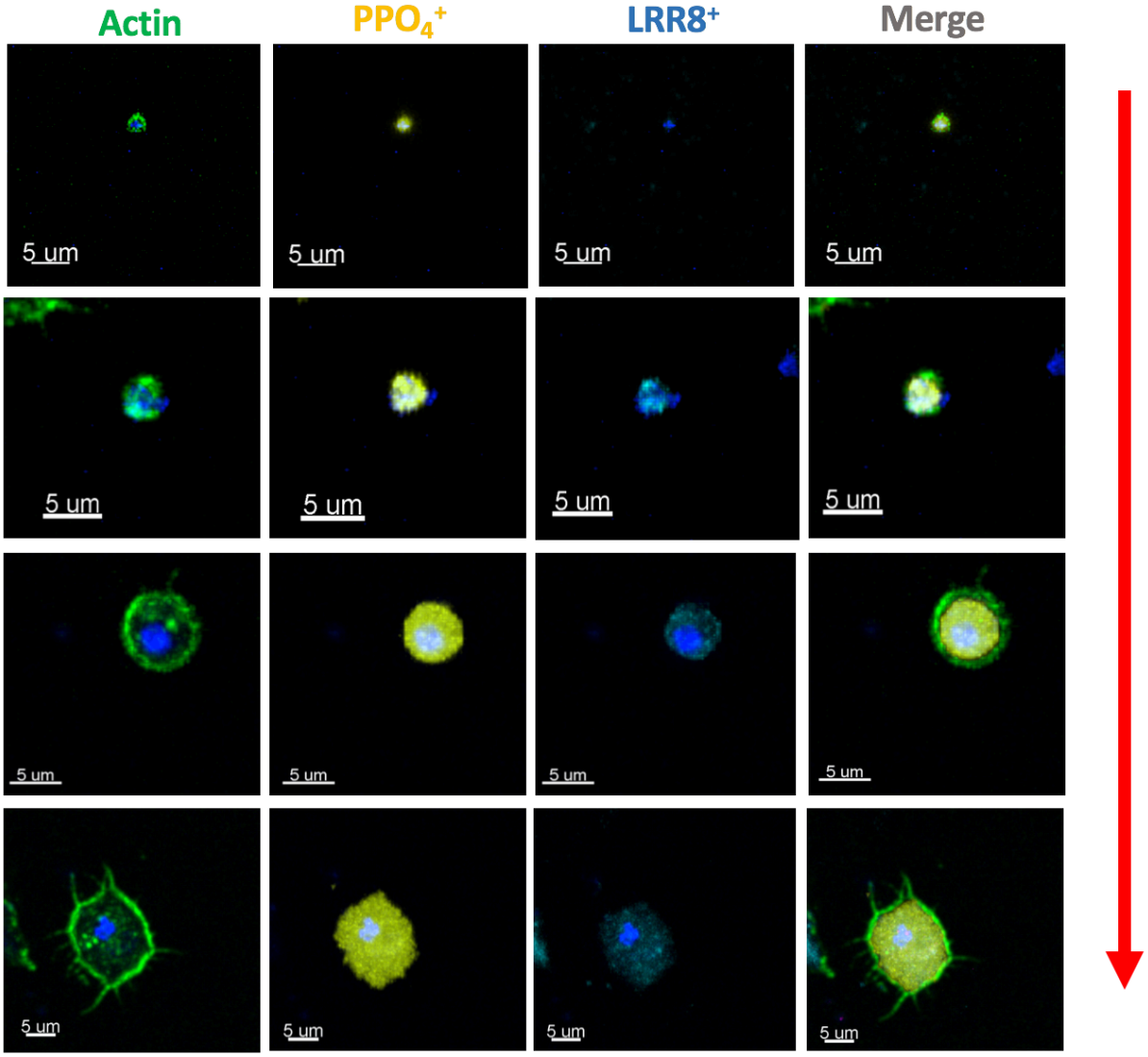


Fig. III.25 Oenocytoid lineage. Red arrow indicates trajectory of maturation. Correlative microscopy. 63x merged, RNA-FISH, and morphological (green, actin) view of circulating hemocytes (blue, LRR8 probe), and oenocytoids cells (yellow, PPO4).

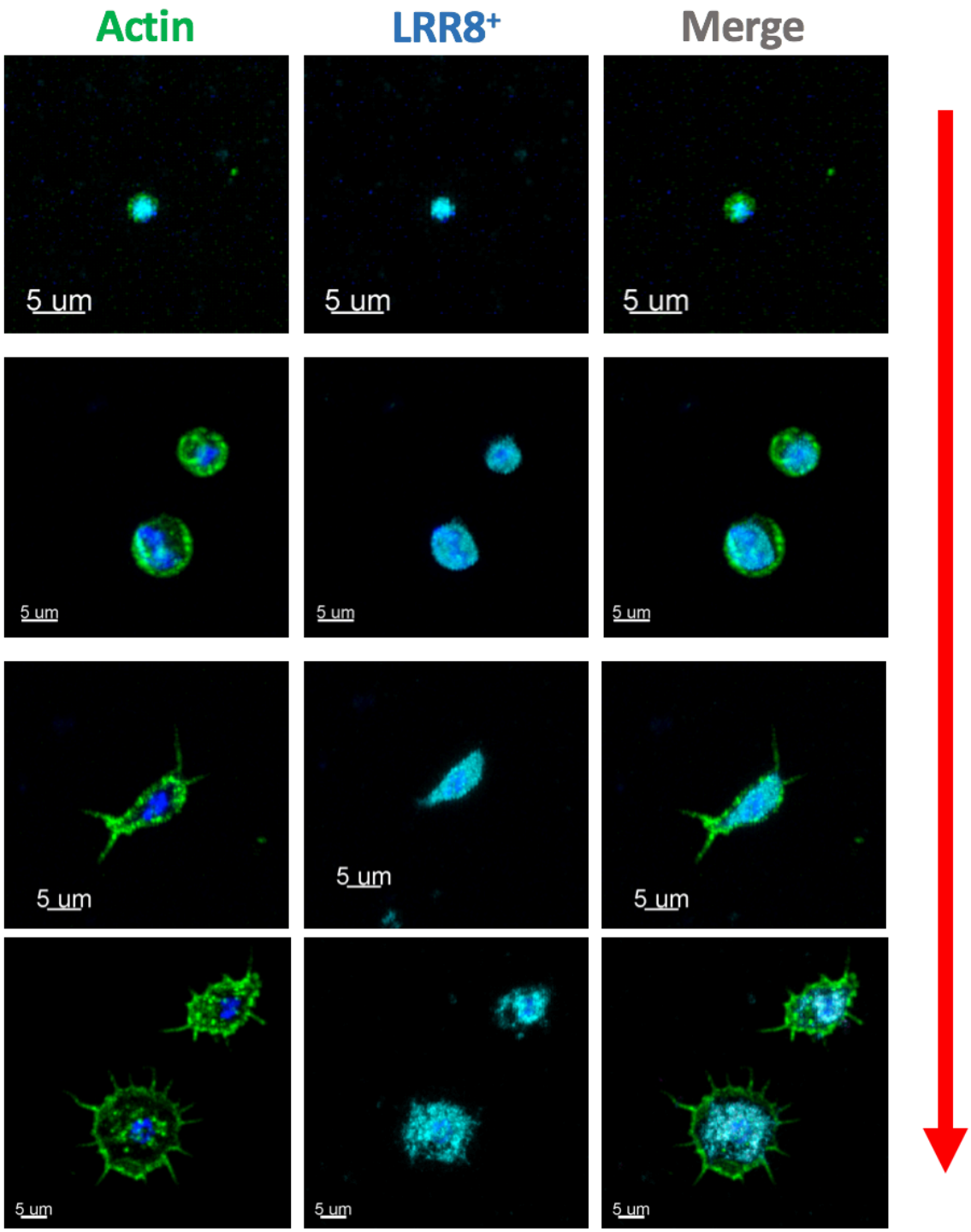


Fig. III.26 Granulocyte lineage. Red arrow indicates trajectory of maturation. Correlative microscop. 63x morphological (green, actin), RNA-FISH (blue, LRR8 probe), and merged view of circulating hemocytes.

3.1.10 Correlation of *Aedes* and *Anopheles* hemocytes

To assess which of the newly discovered putative cell types are shared between anopheline and culicine mosquitoes, we also analyzed the single-cell transcriptome of 3123 cells from *A. aegypti*, a vector for several viral diseases including yellow fever, dengue, chikungunya and Zika. As with *Anopheles*, a dimensional reduction plot shows both canonical hemocytes and other cell types with mostly fat body signatures [Fig. III.27-28]. We once again identified canonical oenocytoids (two subtypes, HC1 and HC2), granulocytes (HC4 and HC5), prohemocytes (HC3), dividing granulocytes (two subtypes, HC6 and HC7), secretory granulocytes (HC8). Fat body cells were characterised by a heightened complexity, with five different cell states recognised (FBC1-5).

A cross-species correlation after a logistic regression and multinomial learning approach further supported our cell type identification, and revealed similarities and differences with *Anopheles* hemocytes. Two clusters (AaHC1 and AaHC2) both have conserved transcriptome signatures for oenocytoids compared to *Anopheles* oenocytoids (AgHC1): 99% and 77% correlation respectively. We again detected different granulocyte subtypes, including antimicrobial peptide secreting cells (94% correlation with *Anopheles* secreting granulocytes), and dividing granulocytes (87% with *Anopheles* progenitor cells). Granulocytes and prohemocytes are again positioned on a continuum of transcriptomic similarity, with four different cell states, including a proliferating S-phase granulocyte cluster (AaHC6) without a clear *Anopheles* equivalent. Granulocytes once again express laminins, leucine-rich repeat proteins, scavenger receptors, Toll receptor 5, and the transcription factor Rel2 [Fig. III.28]. However, effector cells (AgHC5) lack an obvious counterpart in *Aedes*.

Functional classes of mosquito hemocytes

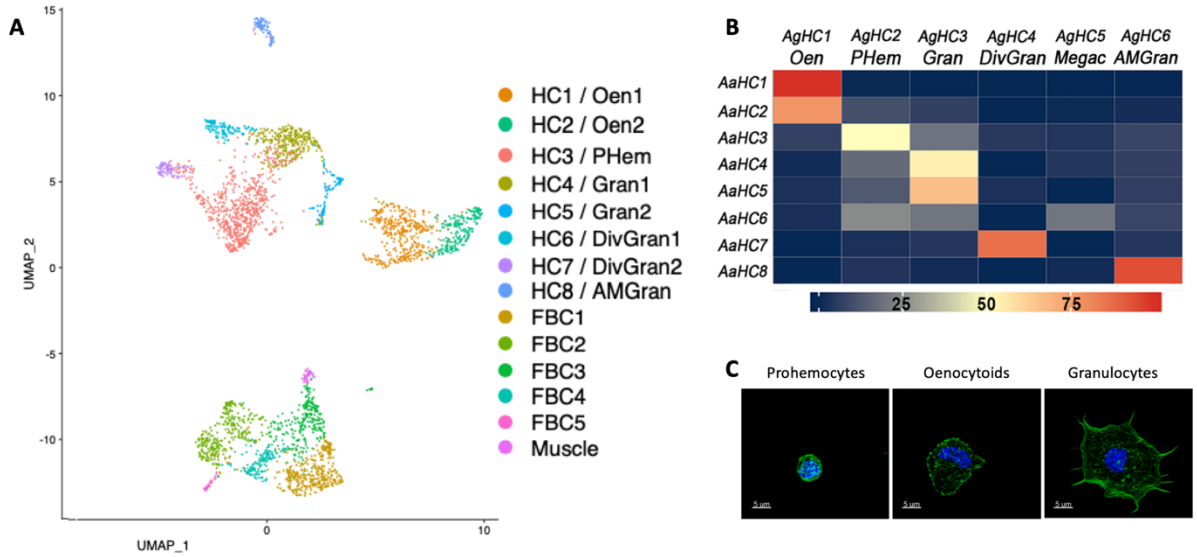


Fig. III.27 Characterisation *Aedes aegypti* hemocytes and correlation with *Anopheles*
(A) UMAP of 3123 *A. aegypti* hemocyte clusters colored by cluster identity with Seurat clustering. **(B)** Heatmap showing probability of each *A. aegypti* hemocyte cell in the cluster belonging to each one of the *Anopheles* cell types after logistic regression and multinomial learning approach. *Ag*, *Anopheles*; Aa, *Aedes*. Oen, oenocytoids; Div Gran, dividing granulocytes; Gran, granulocytes; Mega, megacytes (effector); AM Gran, secretory granulocytes; PHem, prohemocytes. **(E)** *Aedes* hemocyte morphology. Stained with phalloidin (actin) in green and Hoechst (nuclei) in blue. Scale bar: 5 µm.

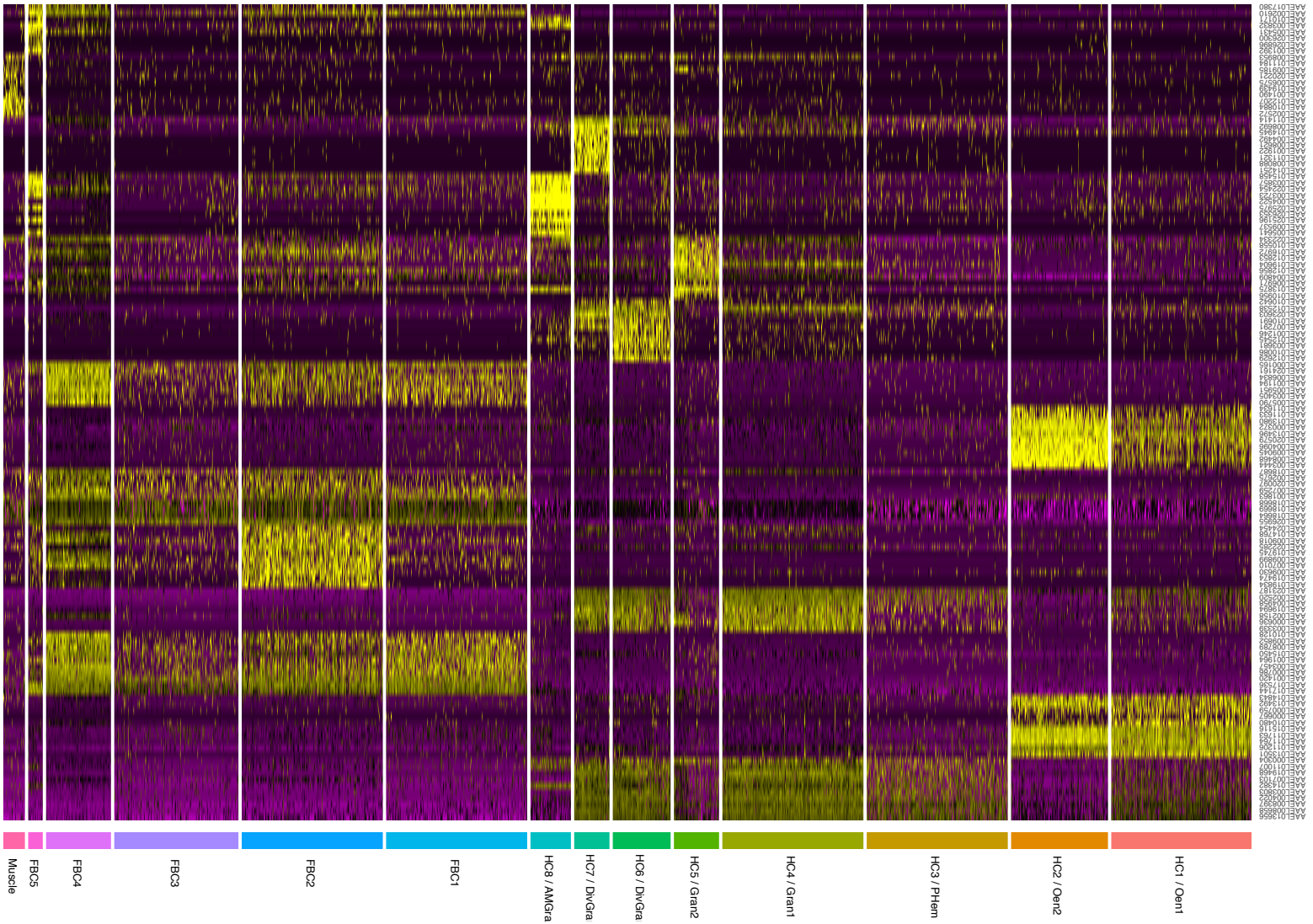


Fig. III.28 Heatmap of top ten gene biomarkers for each *Aedes* cell type or state. DE genes were identified with the Wilcoxon rank-sum test. The P values were adjusted for multiple testing using the Bonferroni correction. P-adjusted values < 0.001, ordered by average log fold change between cluster of interest vs all other cells. Down-sampled to 300 cells per cluster for clarity.

4 Discussion

Clustering analysis with Seurat, diffusion maps, lineage tracing with PAGA and Slingshot, and RNA-FISH validation make us posit that 6 hemocyte cell types exist in the hemolymph of mosquitoes. These include three main types already known: prohemocytes, granulocytes, and oenocytoids. In addition, we found novel cell types, namely dividing hemocytes, effector hemocytes, and secretory hemocytes. We classified cell types by taking into consideration both the RNA content of cells - using the number of UMIs per cell as a proxy - as well as the analysis of the differentially expressed genes between each cell cluster. Prohemocytes were characterised by a low number of UMIs (yet distinct from background), consistent with the high nuclear-cytoplasmic ratio and small overall size. Conversely, granulocytes were transcriptionally active, had large diameters, and exhibited high UMIs. Oenocytoids were intermediate in size, RNA content and number of UMIs.

Furthermore, when looking more in detail into cell expression, prohemocytes split into two main cell states within their larger group: inactive and active prohemocytes. Granulocytes showed the largest diversity, compatible with their effector functions. They subdivided into baseline, Type 1 and Type 2 active granulocytes. While baseline granulocytes were well represented in sugar-fed, blood-fed, and infected conditions, that was not the case for Type 1 and Type 2 active granulocytes, which were enriched in blood-fed and *P. berghei* infected conditions. Blood feeding has been shown to activate and induce granulocyte proliferation, in keeping with our results[147]. Thus, T.1 and T.2 granulocytes appear to be activated granulocyte states, and lineage tracing analysis indeed suggests they are alternative granulocyte activation trajectories. Whereas Type 1 active granulocytes appeared to give rise to dividing cells, the other differentiation branch split from baseline granulocytes into Type 2 granulocytes and then effector or secretory hemocytes. Indeed, effector hemocytes were characterised by high expression of LITAF (LPS-Induced TNF-alpha transcription factor), AGAP007318 (an uncharacterised membrane protein upregulated with *P. berghei* infection [349]), Toll proteins, and ficolins. LL3 had been previously shown to control oocysts numbers, but the cell population responsible for the phenotype was unknown [186]. We hypothesize these cells to

be the elusive immune effectors responsible for *Plasmodium* oocyst control. Secretory hemocytes on the other hand constitutively expressed proteins with N-terminal signal peptides for secretion either into circulation or lysosomes, such as LYSC1, TEP3, Ficolins, cecropins, and defensins. Granulocytes, oenocytoids, prohemocytes could be found both in circulation as well as in sessile form, and the same for effector and secretory hemocytes. We did not however find dividing cells in tissues. It is possible replicating granulocytes exist only briefly in this cell state, or alternatively that only circulating hemocytes replicate.

Genes of interest that should be followed up include AGAP009201, encoding for the collagen type IV, highly expressed in circulating hemocytes and the basal lamina and shown to be important to reduce oocyst load, to increase phagocytic capabilities of hemocytes, and to modulate LRIM1 [324]. In our study AGAP009201 was highly expressed in prohemocytes and all granulocytes, including dividing cells. LRR (AGAP004017 and AGAP004016) are leucine-rich repeat proteins highly expressed in circulating hemocytes (in our data in all hemocytes, including some oenocytoids). Of interest AGAP004016 was shown to be a *Plasmodium* agonist [324]. Both LL3 and LL1 are highly expressed in effector hemocytes and are part of the LITAF family (LPS-induced tumor necrosis factor alpha factor) and have important roles in *Plasmodium* control and immune modulation [185]. AGAP011223 was one of the top genes in oenocytoids and encodes fibrinogen-related FBN8 (FREP57), which was shown to promote phagocytosis and have a role in anti-*Plasmodium* defences [324]. Finally, among cell cycle genes and transcription factors we have NF-X1-type zinc finger protein NFXL1, orthologue to *Drosophila* 'nessun dorma', a top gene marker for dividing cells, but with an unknown role in hemocyte replication [350].

There likely exist four distinct hemocyte lineages in the mosquito. Two main lineages, the prohemocyte – granulocyte lineage, and the oenocytoids lineage, are distinct as shown by clustering, lineage tracing analyses, and correlative microscopy. Prohemocytes have long been thought to be the stem cells of the mosquito immune system. In this dataset there was no direct evidence for prohemocytes to be stem cell-like, but prohemocytes do appear to be a pool of

inactive, immature immune cells that the mosquito can draw upon when challenged, or when overloaded with nutrients such as after blood-feeding. Under these conditions, cell activate and replicate. We saw cellular activation shifts in all cell types, with prohemocytes becoming active prohemocytes and granulocytes. Baseline granulocytes morphed into two active subtypes, which also functioned as intermediate stages before terminal effector and secretory cells, and dividing cells. It appears thus more likely that with blood-feeding and infection granulocytes undergo a rapid activation and replication, and that prohemocytes are recruited at the same time to also become active granulocytes, some of which can then go on to replicate. Whether these replicating and active cells can return to an inactive prohemocyte state is yet unknown, and we did not find direct evidence for replicating stem cells in our *Anopheles* dataset. In the correlative experiment dataset however, we did find a large number of small cells (prohemocytes) expressing markers of cell maturity such as LRR (granulocytes) and PPO4 (oenocytoids), supporting microscopically the hypothesis that all hemocyte subtypes, including oenocytoids, derive from prohemocytes.

Recent studies have shown prohemocytes to have phagocytic capabilities and thus to partially resemble granulocytes [192]. Consistently we showed that prohemocytes and granulocytes exist on a continuum of activation and development. The prohemocyte-granulocyte combined lineage split into three subtypes: a) one lineage differentiated from prohemocytes into granulocytes, then active granulocytes type 2 and finally dividing granulocytes, replenishing the granulocyte cell pool after blood feeding, b) two other lineages instead branched off together into active granulocytes type 1 before splitting into effector cells and c) secretory cells. Oenocytoids on the other hand appear to be a completely separate lineage. We did not find evidence of transcriptomic transition between prohemocytes and oenocytoids, but we did find likely transitions between prohemocytes and oenocytoids with correlative microscopy. Prohemocytes are also the smallest of hemocytes, and few genes per prohemocyte could be captured. The transitions could have thus been missed. Importantly, all three lineage tracing algorithms (PAGA, diffusion maps, Slingshots), as well as Seurat agreed with one another, reinforcing our confidence in the hypothesised lineages. PAGA in particular

is well suited to identify connections between cell types in complex datasets. No clusters were removed in the PAGA analysis, and yet the algorithm still correctly identified a transcriptomic relationship between all fat body cells, whereas muscle cells formed a separate cluster of its own. Surprisingly, oenocytoids were also disconnected from all other cell types. Indeed, even when the PAGA threshold was lowered to capture less confident inter-cluster connections, oenocytoids still did not connect to any other clusters, even when fat body cells and hemocytes did. The lack of connection between fat body cells and hemocytes amounts to a positive control, and we thus conclude that oenocytoids and hemocytes either sit on different lineages that likely arose during the embryonic and larval stage, or that the depth of coverage of our dataset did not allow for the connection to be determined transcriptomically, as few transcription factors or lowly expressed genes could be found in prohemocytes. After subsetting the prohemocytes-granulocytes family we then run separate Slingshot and diffusion maps analyses to confirm the data found through PAGA. And indeed, when visualising diffusion component 1 vs component 3 we could observe direct transitions from active granulocytes type 2 to rapidly diving cells, as well as from type 1 granulocytes to effector hemocytes, indicating a differentiation process. Furthermore, DC1 vs DC2 and the 3D visualisation of the first three diffusion components also showed the secretory subtype emerging from granulocytes.

Slingshot – another top-rated lineage tracing software – further supported our hypothesis, recapitulating the differentiation process we had observed with PAGA. A pseudotime analysis of the three branches also showed some of the genes involved in the transitions. Keeping in mind that most cell cycle genes were not included in the lineage analysis due to the strict filtering requirements, many of the genes Seurat identified as markers for each cell type were also independently found in the pseudotime-based analysis. For example, lineage 1, which traces prohemocytes to dividing cells, featured PPO6 and fibrinogen. Of interest, in humans and mice extravascular fibrinogen has been shown to induce macrophage chemokine expression via Toll-like receptor 4, leading to increased immune surveillance at sites of increased inflammation [351]. In our dataset, granulocytes type 2 and many oenocytoids expressed fibrinogen and fibronectin-like transcripts. It may be that these cells are

immunogenic sensors leading to fibrinogen deposition and activation, followed by mitotic division of granulocytes (dividing cells). Lineage 2 genes featured cecropin, LYSC1, collagen Type IV alpha, laminin subunit alpha, once again transcripts that were gene markers of granulocytes type 1 and secretory cells with Seurat. Lineage 3 genes were very similar but LITAF3 (LL3), laminin gamma 1, and LRIM16B were specific for effector cells.

These conclusions were reinforced by the parallel results in our *Aedes* dataset. The cell types originally discovered or confirmed in *Anopheles* were largely conserved between the two species, and thus possibly of functional importance. Because of the increased number of genes per cell we were able to detect more granular details, including two different oenocytoid cell and dividing granulocytes cell states. Interestingly however, effector cells were not detected at all in the *Aedes* dataset. Furthermore, the gene marker (TM7318) defining them is only present in anophelines of the *Cellia* subgenus (malaria vectors in Africa and Asia). We speculate these cells may thus have specific functions in African and Asian Anopheles, potentially connected to immune priming and *Plasmodium* responses (see Chapter IV).

Fat body cells and muscle cells were captured in both species, either because they naturally slough off into the hemolymph, or because the shear stress of the anti-coagulant buffer injection, or the tearing of the abdomen, dislodges them. Fat body cells had two main transcriptomic states: baseline and active. The active fat body cell was highly metabolic, characterised by the expression of canonical markers such as vitellogenin. Conversely, baseline fat body cells expressed a plethora of immune genes, both pro and anti-inflammatory, although many of the top markers are known for dampening the immune system. Inactive fat body cells were characterised by high expression of CLIPs, lectins, LRIMs, APL1C, and SRPNs, in addition to regulatory genes of the PPO cascade, such as apolipophorins and phenoloxidase inhibitor protein. This cell type appears to specifically express *Plasmodium* infection agonists. For example, CLIPA9 expression increases oocystsis load [352], and both CLIPA7 and CTL4 stop parasite melanisation [353]. LRIM17 is downregulated after infection to activate an effective immune response [354], and LYSC1 and CLIPA14 knock-down mosquitoes exhibit

increased resistance to *P. berghei* and bacterial infections [355, 356]. SRPN2 also appears to aid malaria parasites [357]. Interestingly, with blood-feeding or infection there was a shift towards a metabolically active, and immunologically permissive fat body. The loss of immune inhibition by the fat body and the concurrent activation of immune cells in the hemolymph suggests the mosquito immune response is tightly integrated with its metabolic functions, with different organs interacting to provide an optimal immune response at each phase of the mosquito life.

Chapter IV

The immune response of *Anopheles* to malaria

1 The understated importance of the mosquito immune system in developing effective transmission-blocking strategies for malaria

"It's time to close the books on infectious diseases, and declare war against pestilence won"

When dealing with vector-transmitted infectious diseases, the importance of the vector's immune system has long been underappreciated. Only in the last few years has interest in so-called transmission blocking strategies (TBS) blossomed. These are different from traditional control measures in that they do not kill mosquitoes, and do not select them towards survival like insecticides [15, 358]. Three main strategies are being evaluated. The first entails killing gametocytes with drugs to stop mosquito midguts colonisation. The second pursues the same goal by vaccinating a population against the late human or early mosquito life stages of malaria. And the last seeks to make the mosquitoes refractory to infection and transmission. Especially for the latter, a thorough understanding of how the immune system of mosquitoes works both with blood-feeding and with immune challenge is required. Mosquitoes do not possess an adaptive immune response and as such rely solely on innate defence mechanisms. As such, the malaria parasite did not need to develop immune-evasion strategies quite as sophisticated in mosquitoes than in humans. And that represents an opportunity for intervention.

Three main strategies have been employed to increase the number of malaria-refractory mosquitoes: replacement of the native mosquito population, artificial gene drive, and use of other organisms for delivery. In all cases effective anti-malarial molecules are required. Phospholipase A2 (PLA2) has already been trialled as one such effector molecule [359]. While the mechanism was originally unknown, work in our laboratory (see main introduction) demonstrated the importance of eicosanoids (LXA_4 and PGE_2) in controlling malaria infection. As the upstream enzyme in the eicosanoid pathway, PLA2 at least partially decreases oocysts load by increasing the availability of eicosanoids in the mosquito. Another molecules used to control infection has been the gland and midgut peptide 1 (SM1), which blocks recognition sites of sporozoites and ookinetes [360].

The immune response of *Anopheles gambiae* to malaria

The most potent molecule is still useless without efficient delivery systems, or an effective way to release modified mosquitoes in the wild. Population replacement is the simplest approach, but requires initial native mosquito elimination campaigns to decrease the number of local susceptible mosquitoes. Furthermore, even the slightest fitness cost will result in the need for continuous releases of modified organisms. Not to mention the important ethical issues related to releasing biting *Anopheles* mosquitoes around human populations. An alternative is the use of gene drive systems, which can decrease the amounts of mosquito releases required, and thus spread malaria resistance faster and more effectively (or even eliminate malaria-transmitting species altogether) [361]. Finally, bacteria and fungi can also be harvested as expression systems for *Plasmodium*-killing molecules, as was done with a strain of *Escherichia coli* expressing a fusion antibody against Pbs2, thereby reducing oocyst load by 95% [362]. Another example is a fusion protein of SM1 and scorpine (antimicrobial toxin) in the fungus *Metarhizium anisopliae*. These fungal spores were able to decrease *P. falciparum* sporozoites by 98% in *A. gambiae*. Alternatively, the bacterium *Wolbachia* has shown direct transmission-blocking effects, but there are limitations in terms of natural *Wolbachia* density levels [363].

All in all, the number of effector molecules and delivery strategies that could induce a refractory state in mosquitoes remain limited, and more research is required into the way *Plasmodium* colonises *Anopheles,* and how the mosquito's immune system responds to infection. A thorough understanding of the immune system of mosquitoes is crucial to stop transmission of diseases such as malaria that are spread by arthropod vectors. The mosquito's immune system limits *Plasmodium* infection in several ways[364, 365], and hemocytes, the insect white blood cells, are key players in these defense responses[366, 367]. Ookinete invasion triggers a strong nitration response in invaded midgut epithelial cells and their basal lamina[368, 369]. Hemocytes that come in contact with a nitrated midgut basal lamina release microvesicles into the epithelial basal labyrinth and promote local complement activation, inducing parasite lysis[367]. An infection with *Plasmodium* primes mosquitoes to mount a stronger immune response to subsequent infections[370]. Primed mosquitoes release a

hemocyte differentiation factor (HDF) into the hemolymph[370], consisting of a complex of lipoxin A4 bound to the lipocalin carrier evoking [371]. HDF increases the proportion of circulating granulocytes, promotes microvesicle release, and enhances complement-mediated parasite lysis[367]. Enhanced immunity is lost if HDF synthesis is blocked[371]. Silencing the transcription factor LL3 also disrupts priming, and LL3 is expressed in hemocytes[372]. However, it is not clear whether LL3 is essential for HDF synthesis or for hemocytes to respond to HDF, nor which hemocyte subtypes express LL3.

In this chapter we shed light on the role of LL3 and of a subset of hemocytes (effector cells) in orchestrating the hemocyte responses to HDF. We also explored how the cell populations we identified in Chapter III change with blood feeding and infection, and we looked at whether the increase in circulating granulocytes is solely due to granulocyte proliferation and differentiation from prohemocytes or also to mobilization of sessile hemocytes. And finally, we assessed the transcriptomic changes brought about by blood feeding and *P. berghei* infection via both scRNA-seq and bulk RNAseq.

The immune response of *Anopheles gambiae* to malaria

1.1 Aims

1. To explore how the immune system of *Anopheles* mosquitos reacts to blood feeding and *P. berghei* or *P. falciparum* infection.
2. To determine cell types and states associated with malaria infection
3. To explore the role of LL3 in priming
4. To identify genes associated with immunity to malaria and anti-plasmodial effector states
5. To visualise how hemocyte populations change in time and space after *P. berghei* or *P. falciparum* infection, both on the surface of the body wall and the gut, and in circulation.

1.2 Colleagues

Dr. Ana Beatriz Barletta-Ferreira and the NIH imaging core prepared and imaged the isolated P. *berghei*-infected hemocyte RNA-FISH. Alvaro Molina Cruz and Gaspar Canepa grew the *P. falciparum* cultures. Rafael Cantera took care of the electron-microscopy. Jose Luis Ramirez performed the LL3 experiment. All other data and analyses presented is a result of my own work unless stated otherwise.

2 Methods

Most experiments in this thesis were performed in my other laboratory at the National Institutes of Health (NIH), and as such employed A. gambiae (G3 NIH strain) rather than the A. gambiae M-form (A. coluzzi) used for the scRNA-seq experiments.

2.1 *A. gambiae* mosquito rearing and *P. falciparum* infection

A. gambiae (G3 NIH strain) and *A. gambiae* M-form (*A. coluzzi*) were reared at 28 °C, 80% humidity, 12-hour light/dark cycle with standard laboratory procedures. The *P. falciparum* strains used were NF54 (WT *P. falciparum*), and NF54-Pfs47KO (Knock-out *P. falciparum*). They were maintained in O^+ human erythrocytes with RPMI 1640 medium with 25 mM HEPES, 50 mg/l hypoxanthine, 25 mM $NaHCO_3$, and 10% (v/v) heat-inactivated type O^+ human serum supplementation (Interstate Blood Bank, Inc., Memphis, TN) at 37°C and with a gas mixture of 5% O_2, 5% CO_2, and balance N_2. *P. falciparum* infections were done by diluting to 0.1% gametocytemia mature NF54 gametocytes. Mosquitoes were then allowed to feed with an artificial membrane feeder. NF54 with human red blood cells to 45% haematocrit was placed in warmed to 37C water-jacketed glass membrane feeders and mosquitoes allowed to feed for 20 minutes. Fed mosquitoes were then incubated at 26°C and 80% humidity. Infection levels (oocyst numbers) were checked by first dissecting midguts in 1× PBS and then staining them in 0.1% mercurochrome ahead of compound microscope visualization.

2.2 *A. gambiae* dsRNA micro-injections and LL3 knockdown

Two to three-day old female *A. gambiae* G3 mosquitoes were cold anesthetized and injected with 69 nl of 3 μg/μl dsRNA solution specific for LacZ, a bacterial gene not found in the genome of mosquitoes. dsRNA of LacZ is used as control during dsRNA-injection gene knockdown. A 218-bp fragment was amplified from LacZ gene cloned into pCRII-TOPO vector using M13 primers to add a T7 tail. For dsLL3 synthesis, a fragment was amplified with a T7 tail using the following primers as previously described:

T7-LL3 F -
TTAATACGACTCACTATAGGGAGAATGACTACCATCATAGTGACGAACCC
T7-LL3 R –
TTAATACGACTCACTATAGGGAGATTACACCATTATTAAATAAATAACACAACTT
GAG.
The PCR product, from LacZ and LL3, was used as a template for dsRNA synthesis with Megascript RNAi kit (ThermoFisher Scientific) according to the manufacturer's instructions.

2.3 Generation of naïve (-HDF) and challenged (+HDF) hemolymph and injection in LL3-knockdown recipients

Mosquitoes were infected with *P. berghei* and following blood feeding, the naive group was placed at 28^{o}C to prevent infection; while the challenged group was maintained at 21^{o}C for 48h, for normal infection to proceed. Subsequently, the challenged group was transferred to 28^{o}C to reduce parasite load. Hemolymph from naïve and challenged groups was collected at seven days post-infection and centrifuged at 4^{o}C, 10,000 rpms for 10 min. The cell-free supernatant was transferred to a new microcentrifuge tube and stored at -80^{o}C until its use. To evaluate the effects of LL3 depletion on the hemocyte's capacity to respond to HDF, 2-3-day old mosquitoes injected with dsRNA for LL3 or LacZ (control) were then injected with 138 nl of cell-free hemolymph from Naïve (- HDF) or Challenged (+ HDF) donors at 3 days post-silencing. Hemocyte differentiation was assessed in two independent experiments at four days post-hemolymph transfer.

2.4 Imaging

RNA-FISH performed as of Chapter II and III. EM performed by Dr. Rafael Cantera.

2.4.1 Image analysis

Acquisition as of Chapter III. For image analysis see below. Whole-mount RNA-FISH positive cells were manually counted by an observer blinded to experimental conditions using the

3DHISTECH CaseViewer 2.3 software (3DHISTECH, Budapest, Hungary). Body wall and gut areas were measured with the analysis tools of the same software. For *P. falciparum* RNA-FISH experiments of hemocytes in circulation, positive cells were counted automatically using the segmentation and thresholding features of the Leica LAS X 3D visualization and analysis software (Leica UK, Milton Keys, UK). For *P. berghei* RNA-FISH experiments of hemocytes in circulation, image processing was performed using Imaris 9.2.1 (Bitplane, Concord, MA, USA). Error bars represent 95% confidence intervals calculated with the T-test distribution on the number of samples obtained, and the standard deviation of the samples for each condition.

2.5 Bioinformatics

2.5.1 Bulk RNA-seq

Sequencing reads in CRAM format were fed into a personal BASH pipeline to convert cram files to fastq using biobam's bamtofastq program (Version 0.0.191). Forward and reverse fastq reads in paired mode were aligned to the *A. gambiae* AgamP4.3 reference genome using hisat2 (Version 2.0.4) and featureCounts (Version 1.5.1) with recommended settings. Count matrices were combined before downstream data processing and analysis within R version 3.5.3 (RStudio version 1.0.153). Downstream normalization, differential expression analysis and visualization were done with the R package DESeq2 (Version 1.18.1) [280], as of chapter III. P values for the differential expression analysis were adjusted for multiple testing using the Bonferroni correction. Genes were considered as differentially expressed if they had an adjusted P value < 0.001 (Wald T-test) and a $\log_2$ fold change > 2. Gene lists with vectorbase IDs were converted to gene annotations with g:Profiler [346]. g:Profiler utilises Ensembl as its primary data source and is anchored to its quarterly release cycle. g:GOSt was used to perform functional enrichment analysis on input gene lists to map the data onto enriched biological processes or pathways. In addition to Ensembl, also KEGG, Reactome, WikiPathways, miRTarBase, and TRANSFAC databases were used. Functional enrichment is evaluated with a cumulative hypergeometric test with g:SCS (Set Counts and Sizes) multiple testing correction (adjusted P value reported only < 0.05). Gene lists were ordered on log-fold changes. Each body part DE analysis was run separately, removing samples from all other body parts from

the matrices. The following model was then run for differential expression analysis, focused on treatment (*P. berghei* vs blood fed, and blood fed vs sugar fed). Experimental repeats, time, and effect of treatment at different time-point were considered as covariates:

```
ddsMat <- DESeqDataSetFromMatrix(countData = countdata, colData = coldata, design
= ~ time:treatment + experiment + time + treatment)
```

2.5.2 scRNA-seq

For details see methods Chapter II and III. For cell number normalization cells were first normalised to 10,000 total cells across all cell states in each condition. The percentage of cells in each cluster by condition was then calculated on the total normalised cells per cluster. Specific differential expression analyses were performed using the R package Seurat 3.0.2) [256, 277, 339]. Batches were integrated with a hybrid CCA / MNN strategy identifying 'anchors': cells with similar transcriptomes between conditions. Analysis of differentially expressed genes to identify marker genes for each cell population was performed based on the Wilcoxon rank-sum test (minimum Log Fold Change > 0.25, maximum adjusted P value 0.05). P values were adjusted for multiple testing using the Bonferroni correction. Differential expression analysis between sugar feeding, blood feeding, and *P. berghei* infection was performed with both Wilcoxon rank-sum test (minimum Log Fold Change > 0.25, maximum adjusted P value 0.05) and with a MAST (Model-based Analysis of Single-cell Transcriptomics) package adaptation for Seurat [279]. MAST is a differential expression analysis tool specifically developed for single-cell datasets, which employs a generalized linear model framework that considers cellular detection rate of genes as a covariate in the model. Volcano plots and labels were plotted with the Enhanced Volcano Plot package for R [373]. Gene lists containing vectorbase transcript IDs were converted to usable gene annotations, GO terms, and GO enrichment lists using g:Profiler for *A. gambiae* [346].

3 Results

3.1 Cell populations change with blood feeding and malaria infection

To explore how our immune populations change with infection and blood-feeding we first quantified the proportion of cells – as defined in our scRNAseq experiment – in each cell state by treatment. We found that samples infected with *P. berghei* showed an increased number of active type 1 and type 2 granulocytes. On the other hand, *P. berghei* treatment decreased the proportion of cells categorized as baseline granulocytes. Dividing cells and effector cells increased almost equally with both blood fed and infection, whereas secretory cells increased mostly with blood-feeding. Furthermore, oenocytoids increased with both blood feeding and *P. berghei* infection, whereas prohemocytes decreased. With regards to non-hemocytes, there was a higher number of baseline fat body cells with sugar feeding, whereas activated fat body appear to be enriched after *P. berghei* infection.

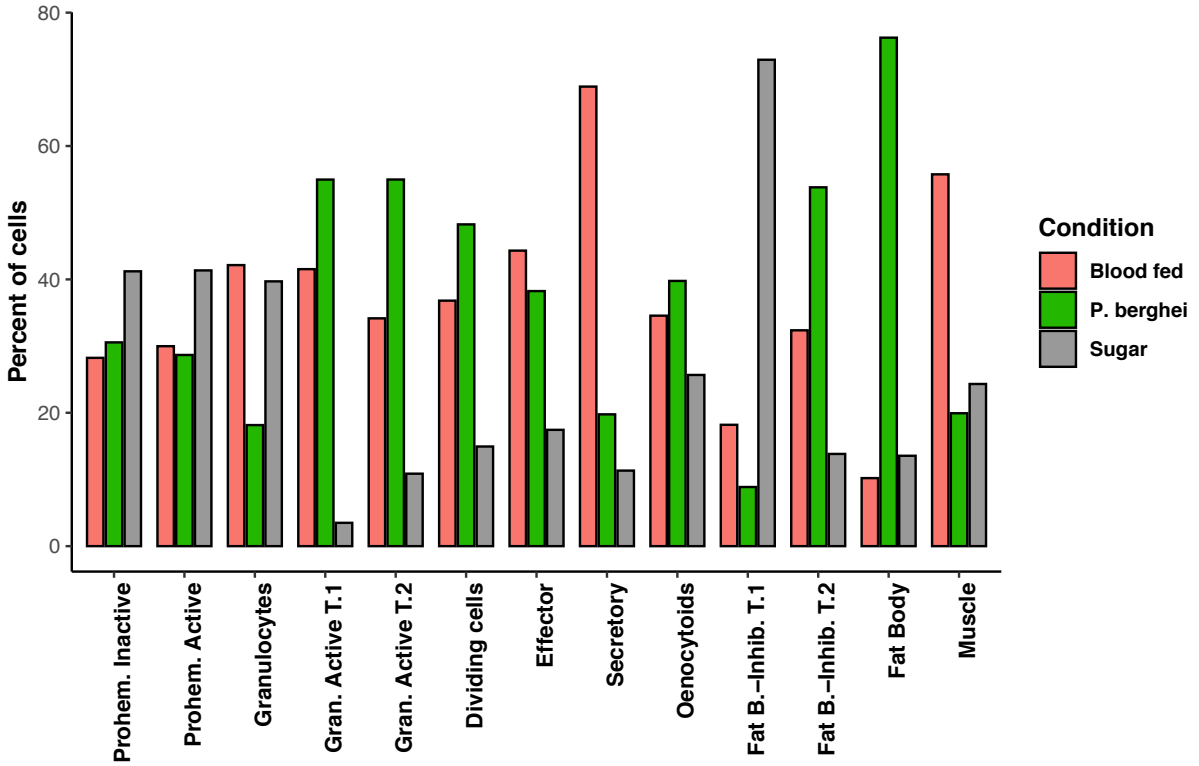


Fig. IV.1 Percentage of cells in each cluster by condition. Cell numbers first normalised to 10,000 total cells across all cell states in each condition. Then the percentage of cells in each cluster by condition was calculated on the total normalised cells per cluster.

Similarly, when we removed all cells that were not hemocytes from the calculations and again normalized the contribution of each cell state to the total number of remaining cells for each condition, we saw the same pattern, suggesting non-hemocytes were not skewing the calculations. The marked decrease in prohemocytes and correspondent increase in granulocytes appeared even more evident. The increased number of active granulocytes with malaria infection as compared to blood feeding or sugar feeding was also clearer. Conversely, secretory cells increased with blood feeding.

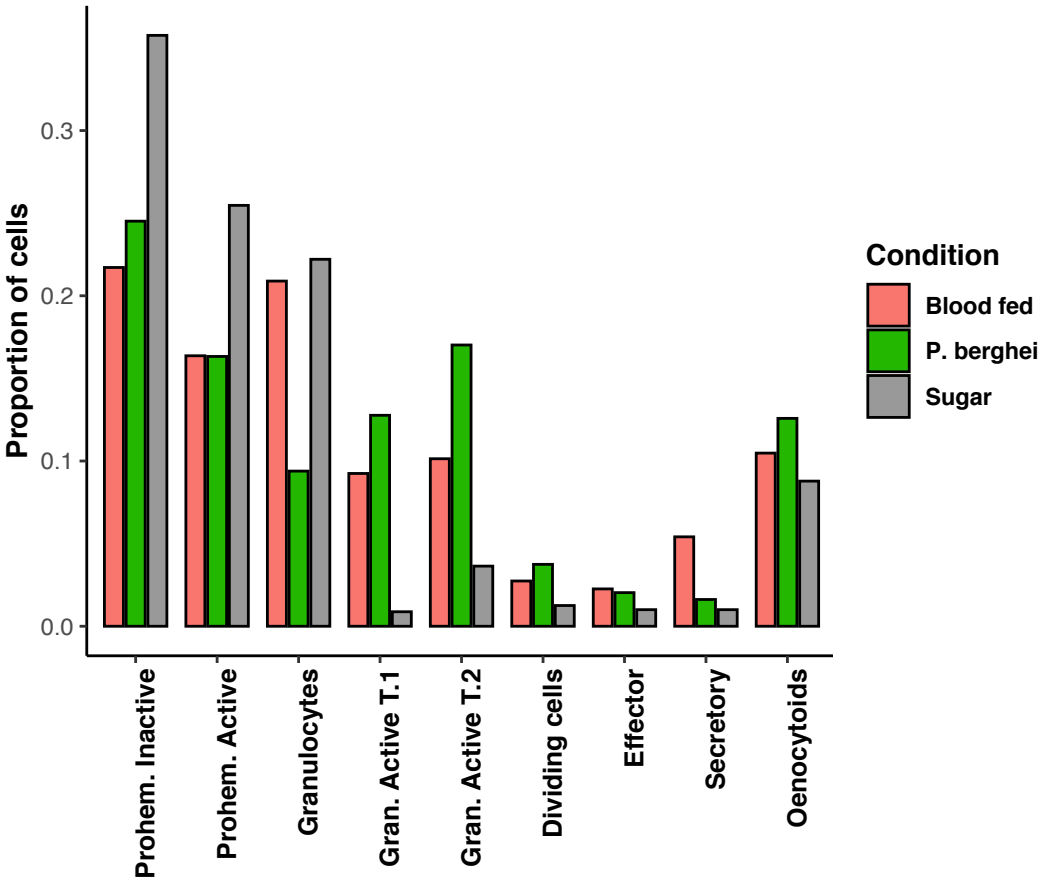


Fig. IV.2 Proportion of cells in each condition by cluster. Cell numbers first normalised to the percentage of cells in each condition that are hemocytes. Then the proportion by which each cluster contributed to the total number of remaining cells in each condition was recalculated.

Importantly, total cell numbers could be directly compared with manual hemocyte counts. The proportion of prohemocytes were 70.9% with blood feeding and 71.5% with *P. berghei* infection. Oenocytoids went from 23.2% with blood feeding to 17% with *P. berghei*. Granulocytes changed from 5.9% with blood feeding to 11.5% with *P. berghei* infection, which was statistically significant (P = 0.0073). The numbers from the manual hemocyte counting and single-cell RNAseq were largely in agreement. Prohemocyte counts were higher with manual counting, further suggesting a degree of similarity between prohemocytes and granulocytes. *P. berghei* infected mosquitoes were checked for infection [Fig. IV.3], and all mosquitoes were infected, with a median of 10 oocysts per midgut.

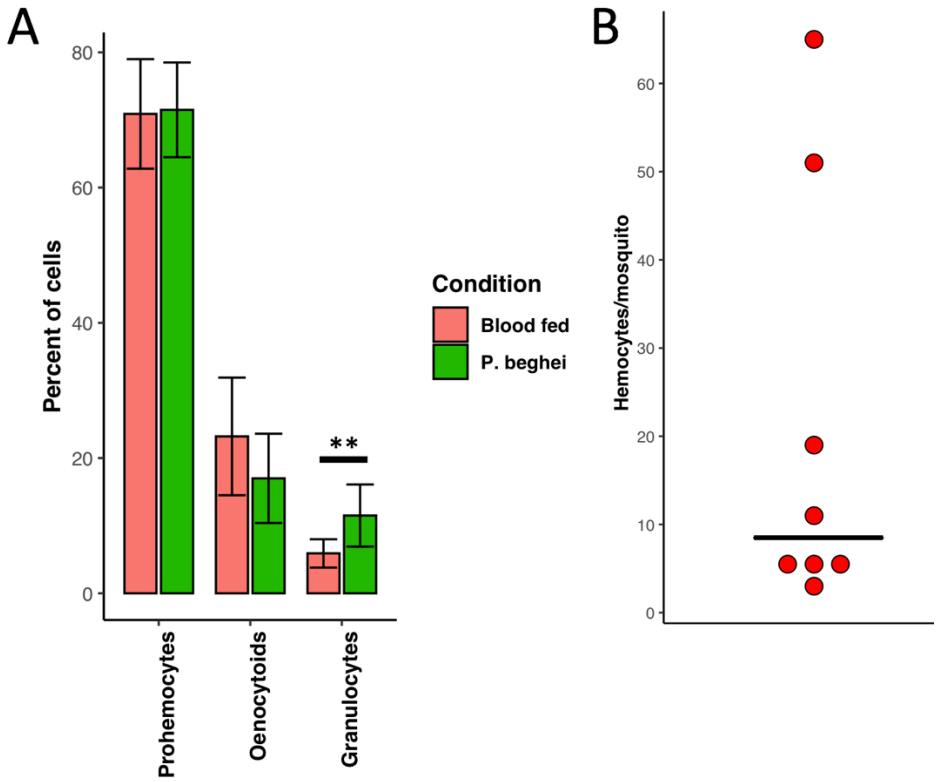


Fig. IV.3 Manual counting of hemocytes and oocysts. (A) 8 mosquitoes were dissected and hemocytes counted with hemocytometer. (B) 8 mosquitoes from the same experiment were dissected and oocysts checked with fluorescence microscopy for oocysts of GFP-CON *P. berghei*. Two repeats. Error bars are mean +/- standard deviation for the population for each sample. ** (P = 0.0073)

3.2 Transcription factor LL3 is required for hemocyte differentiation during immune priming

We have mentioned how the transcription factor LL3 can be detected in granulocytes from *Plasmodium*-infected *A. gambiae,* and that silencing LL3 expression disrupts priming[372]. And we have seen how *Plasmodium* infection and blood feeding leads to immune activation and granulocyte proliferation (which is mediated by HDF). However, it is not clear whether LL3 is essential for HDF synthesis or for hemocytes to respond to HDF. We found that LL3 is highly expressed in effector hemocytes and thus explored whether silencing LL3 affects the HDF response. Transfer of hemolymph from primed *A. gambiae* donors had HDF activity and elicited a strong priming response in control recipients injected with *lacZ* double stranded (ds) RNA, resulting in a prominent increase in circulating granulocytes, a modest increase in oenocytoids and a decrease in prohemocytes. This response was completely abolished when LL3 expression was silenced in the recipients by injection of dsLL3 RNA, indicating that LL3 and effector cells play a key role in orchestrating hemocyte responses to HDF [Fig. IV.4].

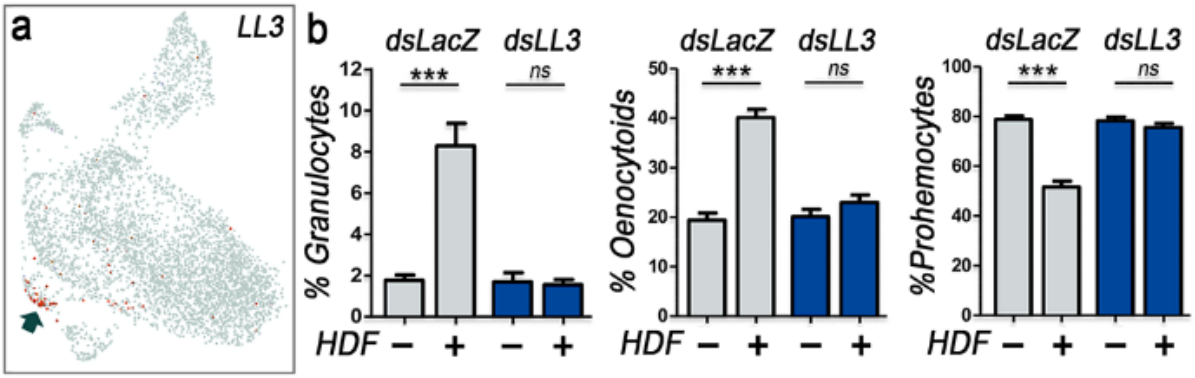


Fig. IV.4 LL3 is expressed in effector cells and required for hemocyte differentiation. (A) UMAP visualisation of all hemocytes by *LL3* expression. Red indicates cells with more than 1 UMI **(B)** Percentage of circulating granulocytes, oenocytoids and prohemocytes of *LL3*-silenced mosquitoes injected (+) or not (-) with HDF versus double-stranded *lacZ* RNA injected mosquitoes used as negative controls (**** P<0.0001, Unpaired t-test). Data are representative of two independent experiments (mean ± SEM).

3.3 Gene changes with blood-feeding and malaria infection

3.3.1 scRNA-seq

After looking at absolute changes in cell numbers we further probed the impact of *Plasmodium* infection on mosquito hemocytes by performing differential expression analysis on our scRNA-seq dataset with the Seurat Wilcoxon DE test as well as the Seurat implementation of MAST (see methods). As the MAST package produced similar numbers of significantly downregulated and upregulated genes when compared to the Wilcoxon-Rank Sum Test, we decided to use MAST for all DE analyses. The lists of positively regulated genes did not change by using MAST or Wilcoxon. Interestingly, when cells from day 7 post-infection were removed we found a heightened number of DE genes, suggesting *Plasmodium* largely modulates the immune system in the first three days of infection. However, prohemocytes saw the pattern reverse, with a higher DE gene count when including *P. berghei* day 7 prohemocytes in the analysis. Ooenocytoids did not seem to respond strongly to *P. berghei* infection at any time [Table IV.1]. Effector and secretory cells were too rare. Most DE genes were due to the activation of granulocytes, in agreement with the trajectory and differentiation analyses of the previous chapter. Genes upregulated included PPO2, 3, 4, 5, laminins, collagens and actins, CLIPB8, Tctp, Matuselah receptor 6, PGRPLD, LRIM6, Calreticulin, Cecropin 1, SCRC1. Downregulated genes included CLIPD1, fibrinogen, and fibronectins [Fig. IV5A].

Cluster	**DE genes – All days**	**DE genes - Day 1-3**
Fat Body	**75 (U:33-D: 42)**	**95 (U:45-D:50)**
Prohemocytes	**132(69-63)**	**10 (1-9)**
Granulocytes	**53 (23-30)**	**232 (119-113)**
All hemocytes w/o oenocytoids	**89 (39-50)**	**168 (108-60)**
All cells	**76 (36-40)**	**174 (99-75)**
Oenocytoids	**16 (2-14)**	**17 (16-1)**
Dividing cells	**1 (0-1)**	**3 (1-2)**

Table. IV.1. Summary of scRNA-seq *P. berghei* DE analyses. MAST package, filtered for absolute log fold change > 0.25 and Q-value <0.1. U = upregulated. D = downregulated.

Whereas granulocytes had 232 DE genes, of which 119 upregulated and 113 downregulated (days 1, 2, and 3 post *Plasmodium* infection), prohemocytes only had 10, of which 9 downregulated. However, when considering all time points the number of DE genes in prohemocytes was 132, of which 69 upregulated and 63 downregulated [Fig. IV.5B]. Many of the upregulated genes were shared between granulocytes and prohemocytes, including calreticulin, SPARC, Tctp, but there were some markers more specific to prohemocytes, including 14-3-3 protein epsilon, cofilin, FK506-binding protein 14, calmodulin, cellular nucleic acid binding protein, ARP, and PPO6.

Conversely, very few genes were differentially expressed in oenocytoids, including SPARC, and almost none in dividing cells and effector cells. Secretory cell had 26 upregulated genes during *P. berghei* infection. Top genes included all-*trans*- and 9-*cis*, SPARC, cathepsin F, ARP 2/3 complex, AKT, and PPO6. Fat body cells had 95 DE, of which 45 upregulated and 50 downregulated. Most downregulated genes in the fat body were not annotated, but among the few that were we could find phenoloxidase inhibitor protein, cathepsin L, autophagy related gene, or gelsolin, again indicating an increase in the immune activity of this organ.

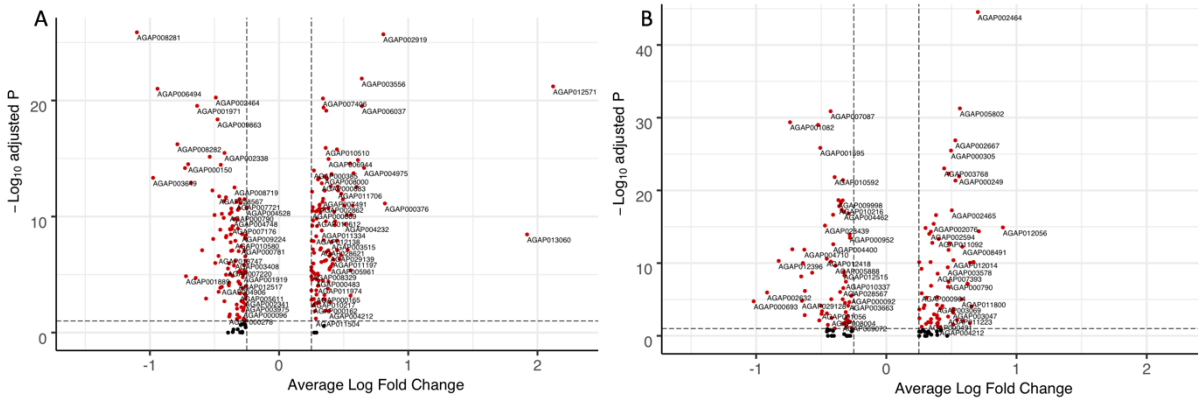


Figure IV.5 scRNA-seq DE gene analyses (A) Volcano plot of DE genes, granulocytes, day 1-3. (B) Volcano plot of DE genes, prohemocytes, days 1-7. DE with MAST package adapted for Seurat scRNA-seq pipeline. Filtered for absolute log fold change > 0.25 and Q-value <0.1.

3.3.2 Bulk RNA-seq

In parallel we performed the same analysis in our bulk RNAseq dataset, which had the same experimental design as the scRNA-seq experiments. In bulk hemocyte samples 65 genes were differentially upregulated (P-adjusted < 0.05 and Log_2 Fold >1) after *P. berghei* infection (day 1,2,3, and 7), including many immune effectors such as TEP1, APL1C, PGRPS3, PGRPS2, PGRPLB, PPO Inhibitor protein, CLIPs, SRPNs, and CTLs [Fig. IV.6].

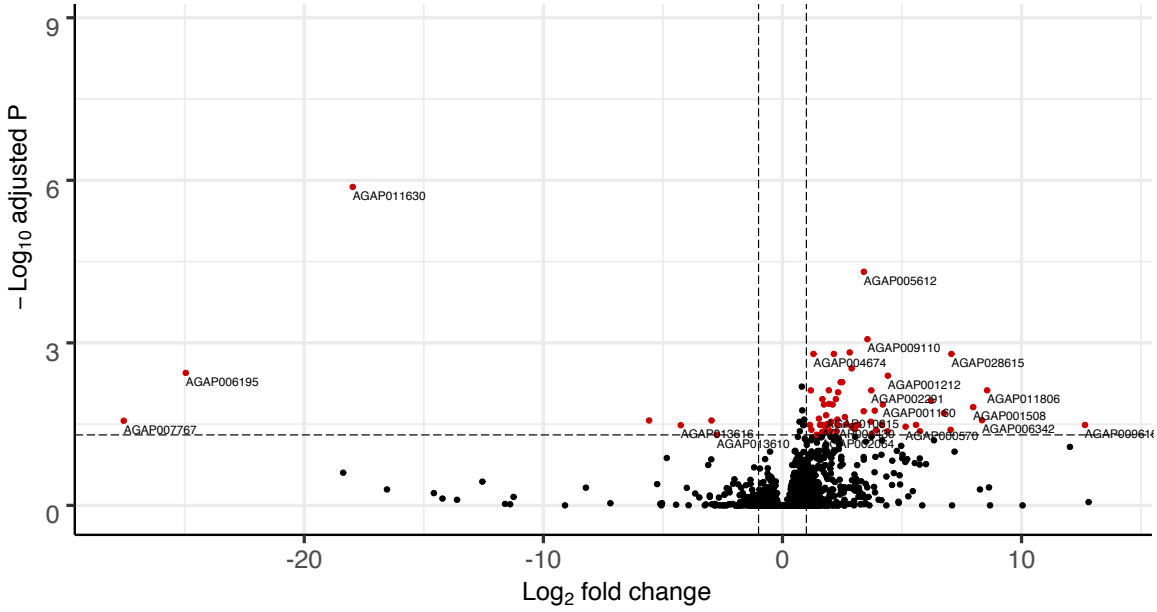


Fig. IV.6 Differential expression of *Anopheles* hemocytes – *P. berghei* vs blood-feeding. From a total of 12184 filtered genes, DE and upregulated genes during *Plasmodium* infection to the right, filtered for log2 fold change >1 and Q-value <0.05

Similarly, gut samples showed 502 upregulated genes during *P. berghei* infection, including a multitude of immune-related genes such as REL2 (IMD pathway signalling NF-kappaB Relish-like transcription factor), AGAP005933 (NFkappaB essential modulator), SCRC1, C-type lectins, APL1C, IAP2, PGRPLB, CLIPs, LRIMs, LL2 and LL3, TEP1, TEP4, TEP6, TEP14, Serpins, TRAF6 (TOLL pathway signalling TNF Receptor-Associated Factor), LYSC4, a PPO Inhibitor, and a Toll-interacting protein. TOLL1A, Ftz and Frizzled transcription factors, and DUOX were instead downregulated with infection [Fig. IV.7].

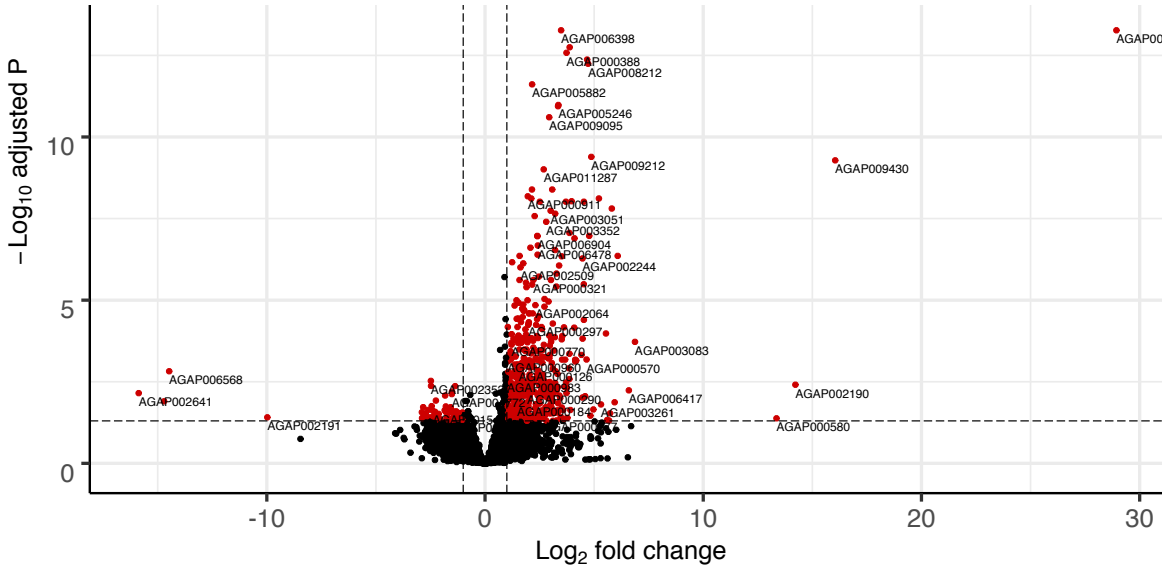


Fig. IV.7 Differential expression of *Anopheles* guts – *P. berghei* vs blood-feeding. From a total of 12726 filtered genes, DE and upregulated genes during *Plasmodium* infection to the right, filtered for log2 fold change >1 and Q-value <0.05

And finally, there were 10 upregulated genes and 10 downregulated genes in mosquito carcasses (fat body, ovaries, muscle, brain, skeleton, etc.) in response to *P. berghei* infection, showing a more subdued response compared to guts or hemocytes. Many DE genes are uncharacterized, but interestingly TRAF6 was upregulated in guts and carcasses [Fig. IV.8].

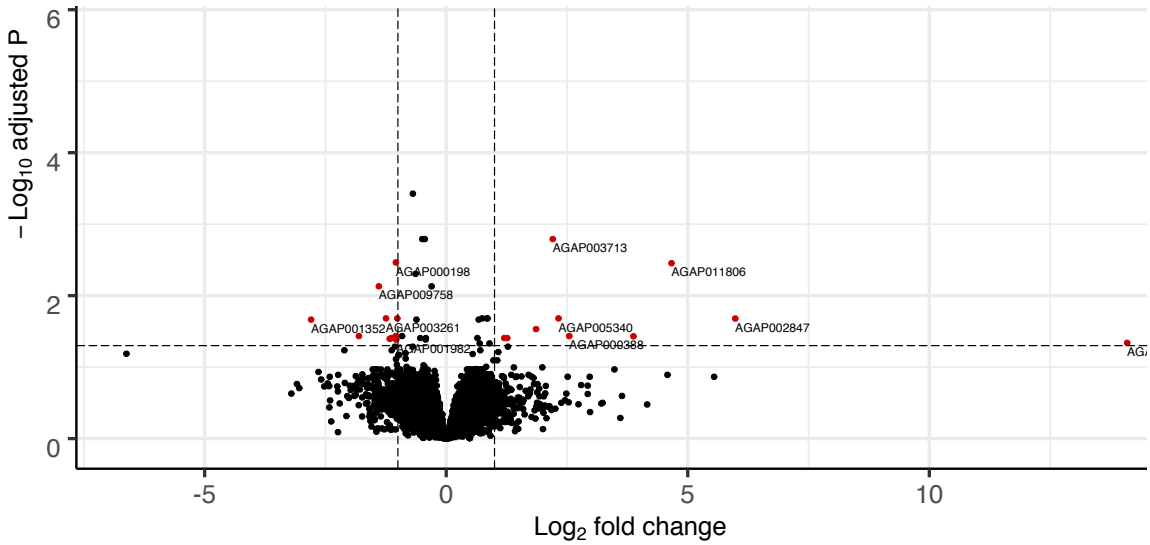


Fig. IV.8 Differential expression of *Anopheles* carcasses – *P. berghei* vs blood-feeding. From a total of 12952 filtered genes, DE and upregulated genes during *Plasmodium* infection to the right, filtered for log2 fold change >1 and Q-value <0.05

Conversely, blood feeding caused a tremendous rearrangement of the mosquito metabolism and transcriptional programming, in all tissues analyzed. When combining all samples together and performing a unified analysis 1731 genes were differentially expressed. Results were very similar when analyzing samples separately: 1778 DE genes if only looking at hemocytes, 1733 DE genes when analyzing the gut, and finally 1601 DE genes in the mosquito carcasses [Fig. IV.9]. A GO enrichment analysis showed that genes upregulated in sugar-fed conditions were involved in carbohydrate metabolism and transport. Conversely, mosquito carcasses were characterized by genes that involved in cellular reproduction, purine metabolism, DNA elongation and replication, lipid metabolism and transport (e.g. Vitellogenin). Many anti-microbial peptides and immune genes were also upregulated, such as defensins, gambicin, prophenoloxidases, Toll 1A, TNF receptor-associated factor 4, LRR [Fig. IV.10-11].

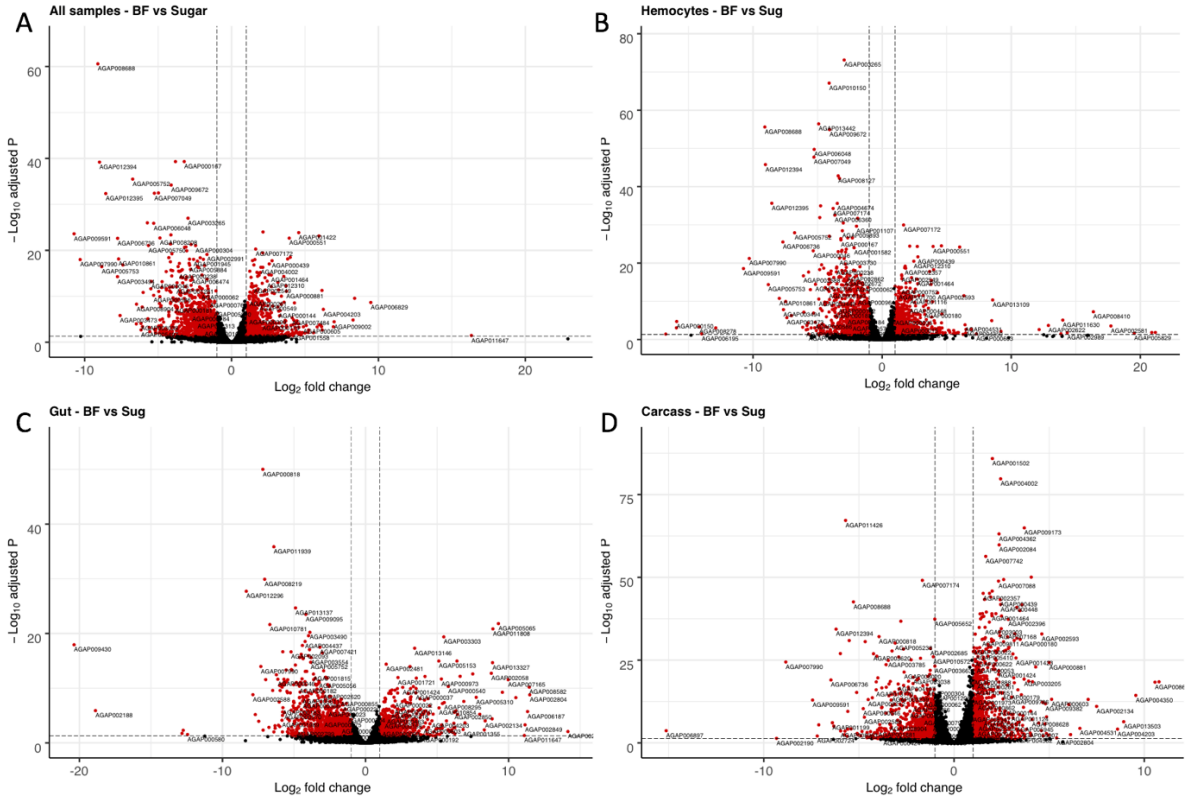


Fig. IV.9 Differential expression of *Anopheles* tissues – Blood feeding vs sugar feeding. From a total of 13048 filtered genes, DE and upregulated genes during *Plasmodium* infection to the right, filtered for log2 fold change >1 and Q-value <0.05 (A) All tissues combined (B) Hemocyte samples only (C) Gut samples only (D) Carcasses samples only.

GO:BP		stats	
Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$ 0 ≤16
carbohydrate transmembrane transport	GO:0034219	3.861×10^{-5}	
hexose transmembrane transport	GO:0008645	4.724×10^{-4}	
monosaccharide transmembrane transport	GO:0015749	4.724×10^{-4}	
glucose import	GO:0046323	4.724×10^{-4}	
glucose transmembrane transport	GO:1904659	4.724×10^{-4}	
carbohydrate transport	GO:0008643	5.897×10^{-4}	
oxidation-reduction process	GO:0055114	6.750×10^{-3}	
xanthine catabolic process	GO:0009115	1.784×10^{-2}	
xanthine metabolic process	GO:0046110	1.784×10^{-2}	

Fig. IV.10 GO Enrichment – Sugar samples. From a total of 13048 filtered genes, DE and upregulated genes during *Plasmodium* infection to the right, filtered for log2 fold change >2 and Q-value <0.05, imported into G:Profiler as of methods section.

GO:BP		stats	
Term name	Term ID	p_{adj}	$-\log_{10}(p_{adj})$ 0 ≤16
'de novo' IMP biosynthetic process	GO:0006189	1.619×10^{-5}	
IMP biosynthetic process	GO:0006188	9.468×10^{-5}	
IMP metabolic process	GO:0046040	3.230×10^{-4}	
alpha-amino acid metabolic process	GO:1901605	1.144×10^{-3}	
small molecule metabolic process	GO:0044281	3.386×10^{-3}	
alpha-amino acid catabolic process	GO:1901606	4.774×10^{-3}	
purine nucleoside monophosphate biosynthetic process	GO:0009127	6.411×10^{-3}	
purine ribonucleoside monophosphate biosynthetic pro...	GO:0009168	6.411×10^{-3}	
DNA strand elongation involved in DNA replication	GO:0006271	7.805×10^{-3}	
DNA strand elongation	GO:0022616	7.805×10^{-3}	
DNA replication	GO:0006260	1.102×10^{-2}	
cellular amino acid catabolic process	GO:0009063	1.410×10^{-2}	
purine nucleoside monophosphate metabolic process	GO:0009126	1.697×10^{-2}	
purine ribonucleoside monophosphate metabolic process	GO:0009167	1.697×10^{-2}	
leading strand elongation	GO:0006272	1.762×10^{-2}	
lipid transport	GO:0006869	1.769×10^{-2}	
lipid localization	GO:0010876	1.997×10^{-2}	
cellular amino acid metabolic process	GO:0006520	3.056×10^{-2}	
purine nucleobase biosynthetic process	GO:0009113	3.350×10^{-2}	

Fig. IV.11 GO Enrichment – Fat body samples. From a total of 13048 filtered genes, DE and upregulated genes during *Plasmodium* infection to the right, filtered for log2 fold change >2 and Q-value <0.05, imported into G:Profiler as of methods section.

3.4 *P. berghei* infection increases FBN-7+ circulating hemocytes

We then further probed the changes brought about by *P. berghei* infection in circulating hemocytes by collecting cells from infected and blood-fed mosquitoes, and doing RNA-FISH with the markers described for each cell type in Chapter III. We quantified the expression of key RNA-FISH markers in over 3200 hemocytes (two biological repeats), and found an increase in the number of cells positive for fibrinogen (FBN7) after *P. berghei* infection, as compared to blood feeding. This was true both for LRR8+ hemocytes, where FBN+ cells went from 18% to 77% [Fig. IV.12], and PPO4+ oenocytoids, where FBN+ cells increased from 22% to 66% [Fig. IV.13]. While FBN7 was originally chosen as the RNA-FISH marker for secretory cells, the expression seems to be upregulated in all morphological cell types upon infection [Fig. IV 12-13]. There were no changes in the other markers.

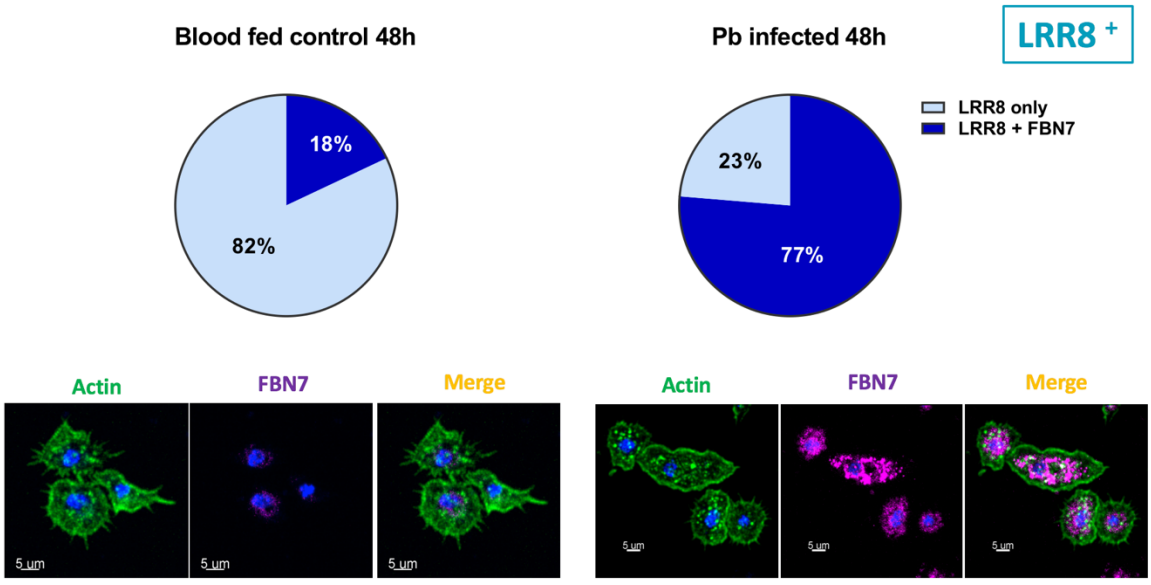


Fig. IV.12 *P. berghei* infection increases the number of FBN7+ hemocytes in circulation. From 3200 hemocytes, 2 biological repeats. LRR+ cells were considered.

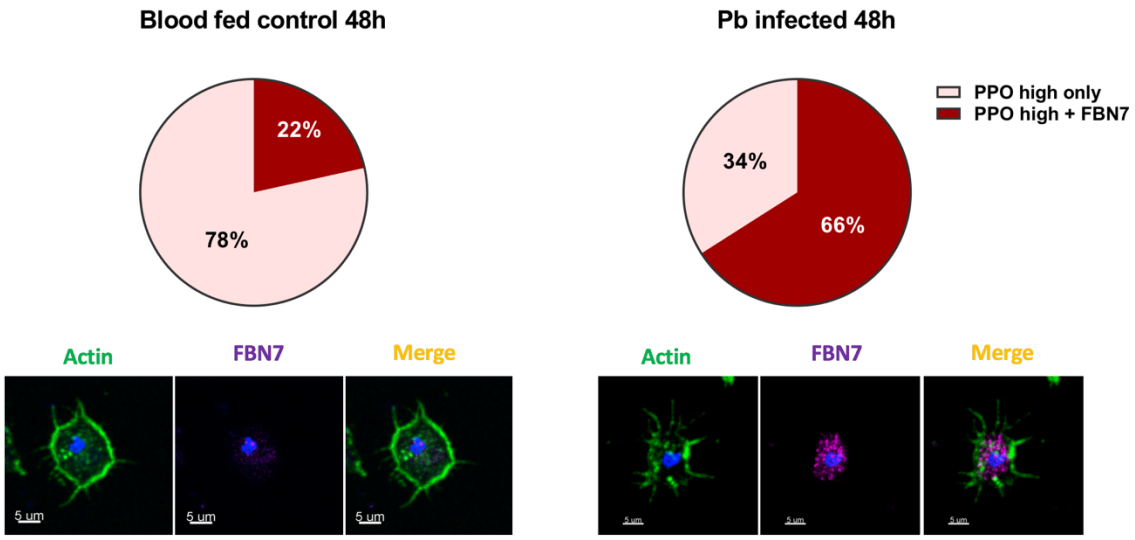


Fig. IV.13 *P. berghei* infection increases the number of FBN7+ oenocytoids in circulation. From 3200 hemocytes, 2 biological repeats. PPO4+ cells were considered.

3.5 *Plasmodium* recruits hemocytes from the fat body wall

Next, to understand how *Plasmodium* infection affects not only circulating but also sessile hemocytes (the hemocytes that are associated with mosquito organs and tissues) we did RNA-FISH on fat body walls and guts of blood-fed or *P. berghei* infected *A. gambiae* mosquitoes. There was a striking reduction in hemocytes per mm^2 of body wall (LRR8+ cells) in infected samples, from 286±76 to 90±34 /mm^2. Fibrinogen-CT+ (Secretory) and Transmembrane+ (Effector) cells remained constant: 6.7±6.1 vs 6.1±3.5 /mm^2 and 4.7±1.5 to 3.3±1.2 /mm^2 respectively. Oenocytoids were unchanged, 14.5±9.0 to 7.0±5.5 /mm^2, while total cells decreased from 312±85 to 106±41 /mm^2, largely due to LRR8+ cells [Fig. IV.14-15].

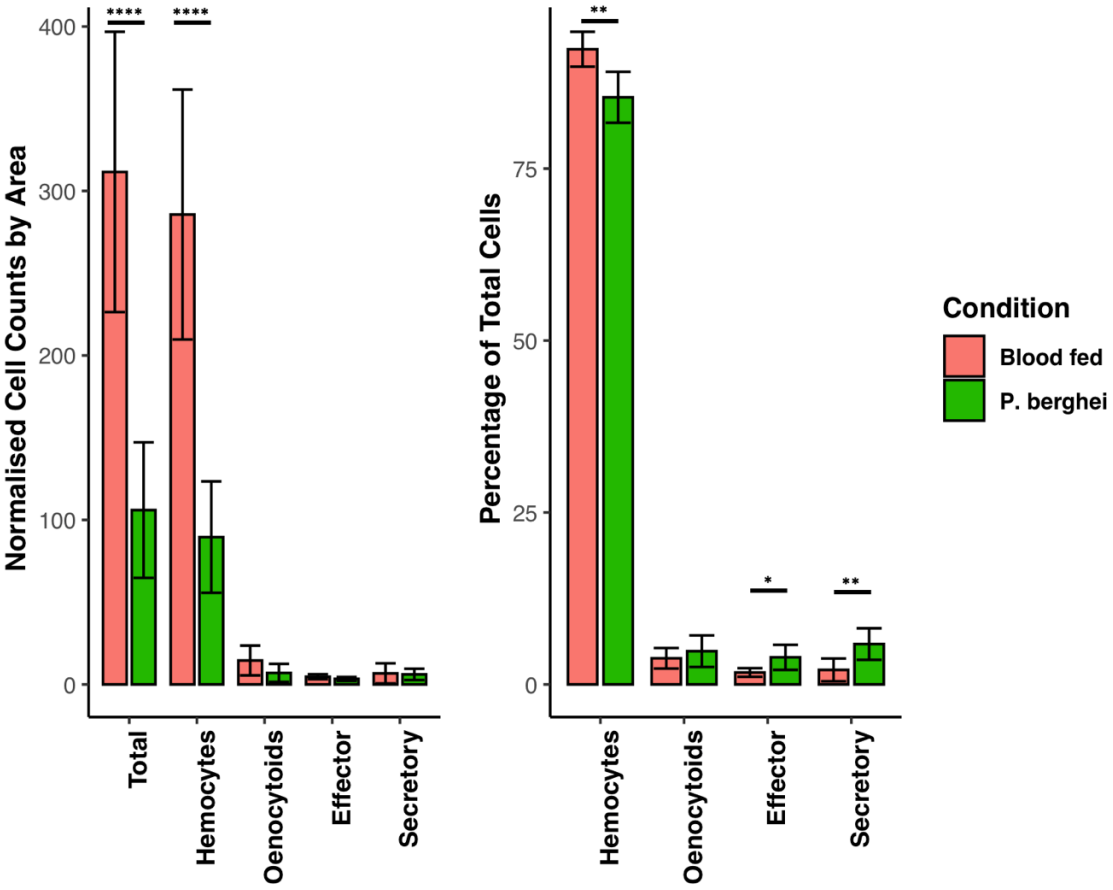


Fig. IV.14 Quantification of cell types on the body wall of *Anopheles* mosquitoes. 16 body walls of blood-fed and 12 of *P. berghei*-infected mosquitoes, followed by RNA-FISH. To the left cell counts normalized by mm^2 of body wall area. To the right percentages of each cell type from total RNA-FISH positive cells. Error bars indicate 95% Confidence Interval. **** ($P \leq 0.0001$), ** ($P \leq 0.01$), * ($P \leq 0.05$) – Welch T-Test. Three biological repeats.

Furthermore, while the percentage of LRR8+ cells also decreased from 92.4% (±2.5) to 85.4% (±3.7), and there was an increase in effector and secretory markers: from 1.7% (±0.6) to 3.9% (±1.8) and from 2.1% (±1.7) to 5.9% (±2.3) respectively, indicating recruitment of circulating hemocytes with infection. The percentage of oenocytoids instead remained unchanged, from to 3.8% (±1.5), to 4.8% (±2.3) [Fig. IV.14-15].

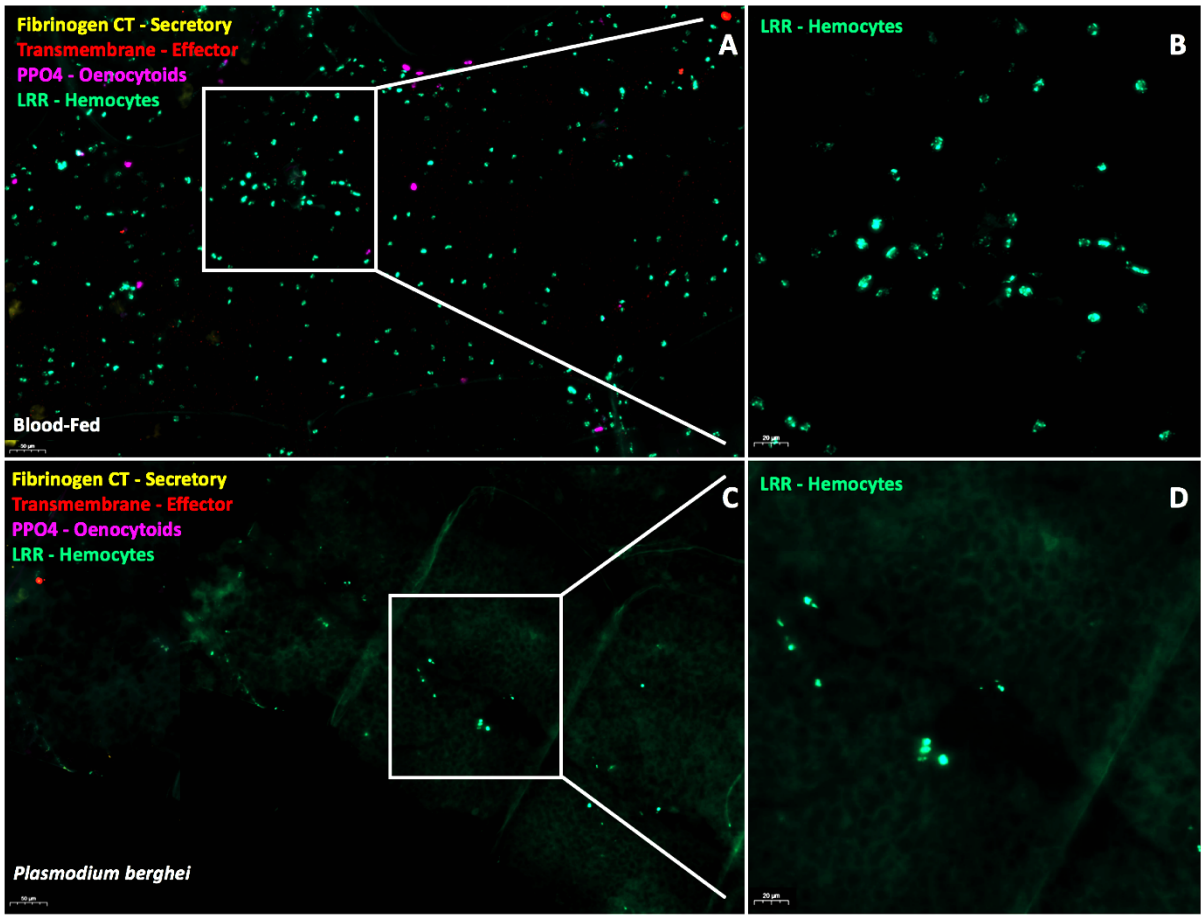


Fig. IV.15 Representative RNA-FISH image of *A. gambiae* mosquito body wall after infection. A total of 16 body walls of blood-fed *A. gambiae* and 12 of *A. gambiae* infected with *P. berghei* were processed following RNA-FISH protocol. (A-B) Blood-fed controls. (C-D) *P. berghei* infection (A, C) 20X magnification (B, D) 40X magnification.

Indeed, when we further explored the surface of the fat body of mosquitoes with electron microscopy we noticed granulocytes are only tenuously associated with the fat body, with immune cells only connected to the fat body by a few pseudopodia extending from the body wall. Our TEM experiment thus showed individual sessile hemocytes bathed by hemolymph and attached to the basal lamina of the tissues through pseudopods, indicative of a dynamic and potentially transient association, with granulocytes appearing ready to detach into circulation when responding to systemic stimuli such as *P. berghei* infection [Fig. IV 14-16].

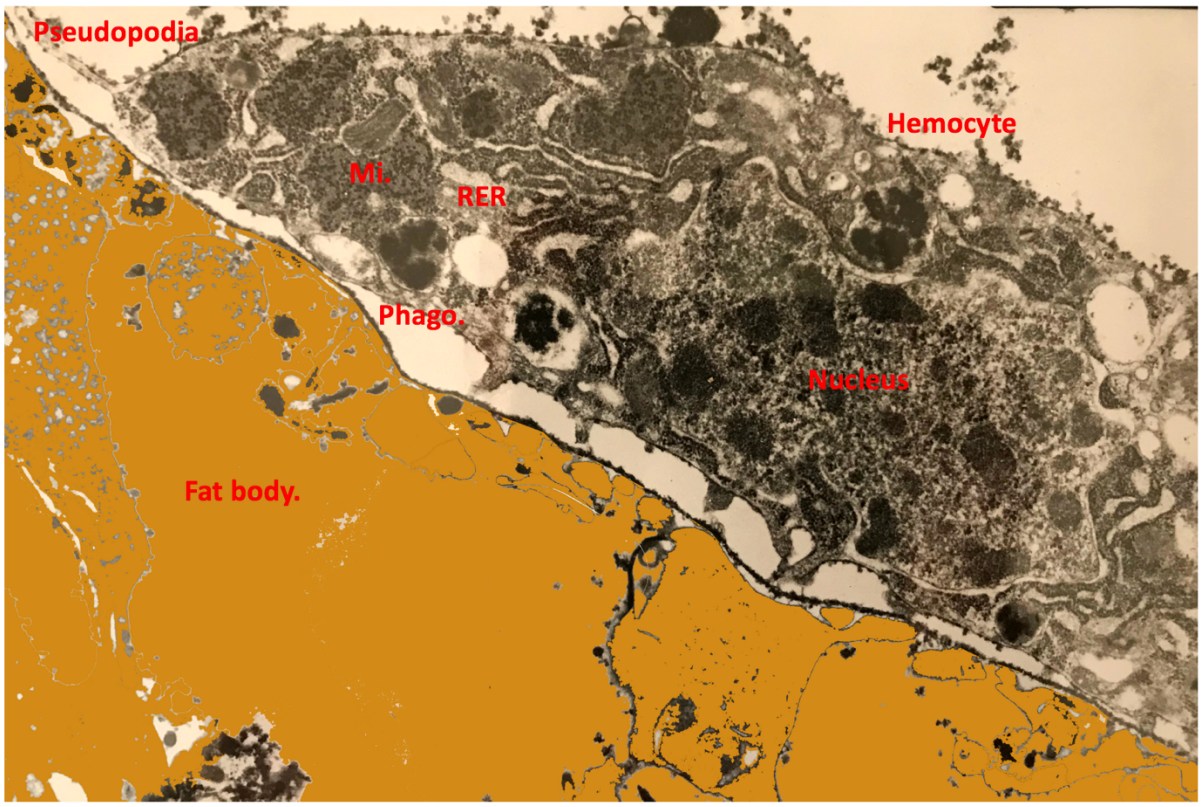


Fig. IV.16 Electron-microscopy image of granulocyte attached to fat body. Phago. = Phagosome. Mi. = Mitochondrion. RER = Rough Endoplasmic Reticulum. Fat body false-colored in yellow for orientation. Prepared by Rafael Cantera.

3.6 Effect of *P. berghei* infection to hemocyte numbers in the gut

More hemocytes per mm^2 were attached to the body wall than to the gut of *Anopheles* mosquitoes that were either blood-fed or infected with *P. berghei*, with a total of 23±6.6 vs 15±3.9 cells/mm^2. LRR8+ cells were also present in lower amounts, with 19.4±6.6 to 12.1±3.8 cells/mm^2, which was close to a statistically significant decrease (P = 0.053, Welch T-test). Normalised Fibrinogen-CT+ (Secretory), Transmembrane+ (Effector), and oenocytoids cell numbers were all unchanged: 0.38±0.26 to 0.13±0.12 cells/mm^2, 1.87±1.06 to 0.87±0.64 cells/mm^2, and 1.55±0.9 to 1.66±0.56 cells/mm^2 respectively. Percentages were all unchanged, but oenocytoids increased from 7.8% (±4.3) to 15.0% (±7.4), which was almost statistically significant (P=0.083 Welch T-test) [Fig. IV.17].

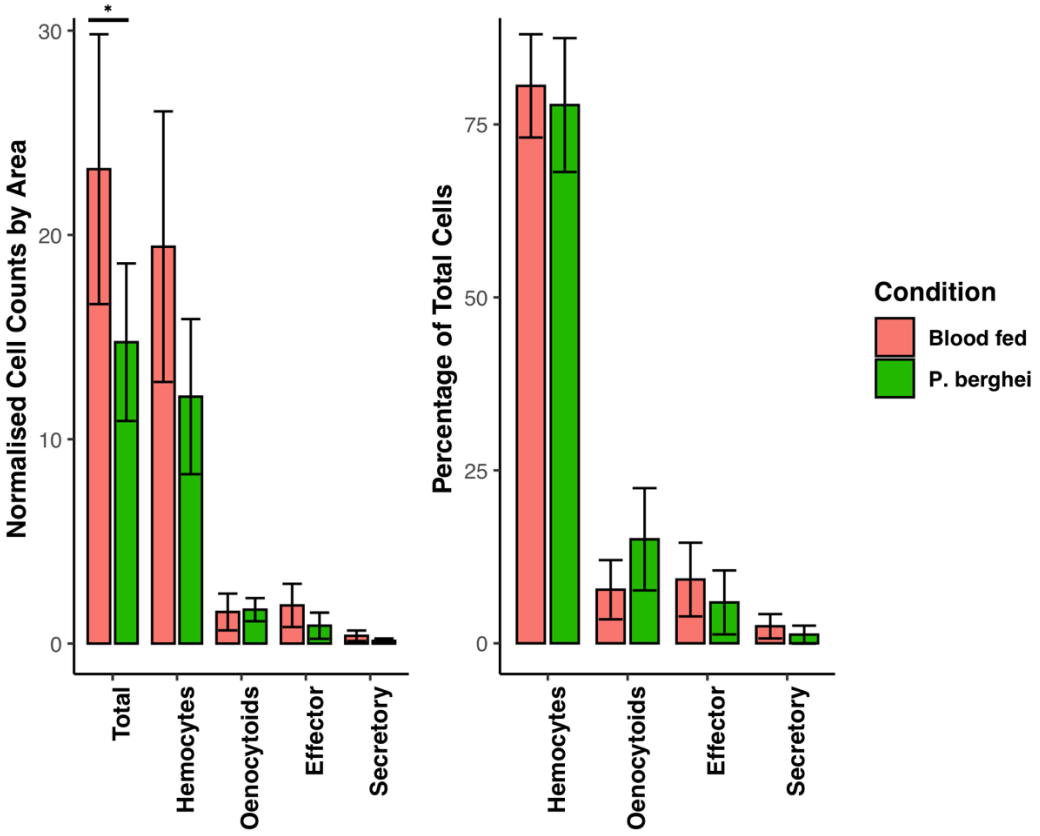


Fig. IV.17 Quantification of cell types on the gut of *Anopheles* mosquitoes. From 19 body walls of blood-fed mosquitoes and 17 of mosquitoes infected with *P. berghei*, followed by RNA-FISH as of methods. To the left cell counts normalized by mm^2 of body wall area. To the right percentages of each cell type from total RNA-FISH positive cells. Error bars indicate 95% Confidence Interval. * ($P \leq 0.05$) – Welch T-Test. Three biological repeats.

3.7 Effect of *P. falciparum* infection on sessile and motile hemocytes

To recapitulate our findings with a parasite more relevant to humans we repeated the experiments with *P. berghei,* but with *P. falciparum*. Similarly to *P. berghei* we found a trend towards decreased LRR8+ and total cell numbers attached to the body wall of mosquitoes with a Pfs47 knock-out *P. falciparum* which is unable to infect mosquitoes due to increase immune system clearance: 687 (±115) to 598 (±46) LRR8+ cells and 802 (±143) to 707 (±67) total cells per mm^2, albeit the decrease was not statistically significant (P = 0.145 and P = 0.128 respectively, Welch T-test). The trend seemed to be reversed by wild-type *P. falciparum* infection, with 712 (±152) LRR+ cells and 831 (±171) total cells per mm^2. All other normalized cell counts and percentages were the same except for an increase in oenocytoids, from 8.1% (±1.1) with blood-feeding to 10.9% (±3.1) with wild-type infection (P = 0.045, Welch T-test) and 10.5% (±2.2) with knock-out infection (P = 0.08, Welch T-test) [Fig. IV.18].

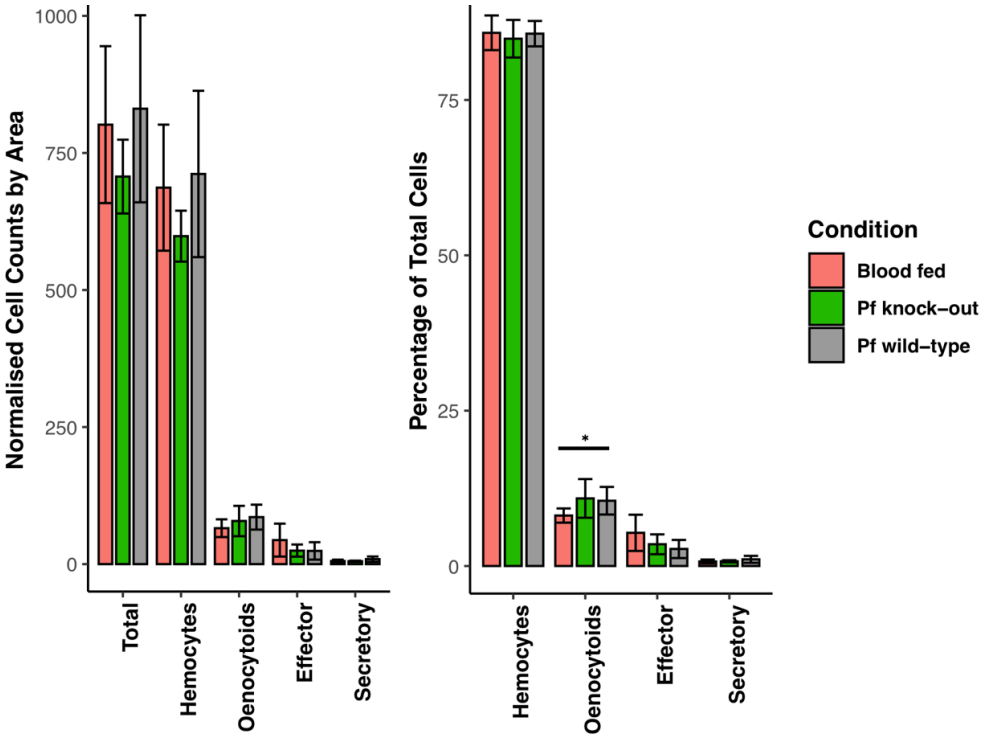


Fig. IV.18 RNA-FISH quantification of cell types on the body wall of *Anopheles*. From 6 blood-fed, 8 wild-type *P. falciparum*, and 8 Pfs47 KO *P. falciparum* mosquitoes. To the left cell counts normalized by mm^2 of body wall area. To the right % of cell type from total RNA-FISH positive cells. Error bars 95% CI. * ($P \leq 0.05$) – Welch T-Test. One biological repeat.

In the gut we again found a lower number of adherent hemocytes compared to carcasses. Cell numbers with blood-feeding were comparable: 23 (±6.6) in the *P. berghei* experiments and 34 (±20) in the *P. falciparum* experiments (per mm^2). The normalized counts showed no difference between *P. falciparum* infections and control, though confidence intervals were large due to low number of samples. We did find an increase in the percentage of LRR8+ cells (hemocytes) attached to the gut, which went from 52% (±19) with blood-feeding to 75% (±5.6) with wild-type infection (P = 0.013, Student T-test) and 69% (±11.5) with knock-out infection (P = 0.073, Welch T-test). At the same time, Transmembrane+ (Effector) cells decreased from 19.9% (±6.4) with blood-feeding to 7.3% (±5.4) with wild-type infection (P = 0.003, Student T-test) and 10.9% (±9.8) with Pfs47 knock-out infection (P = 0.060, Welch T-test) [Fig. IV.19].

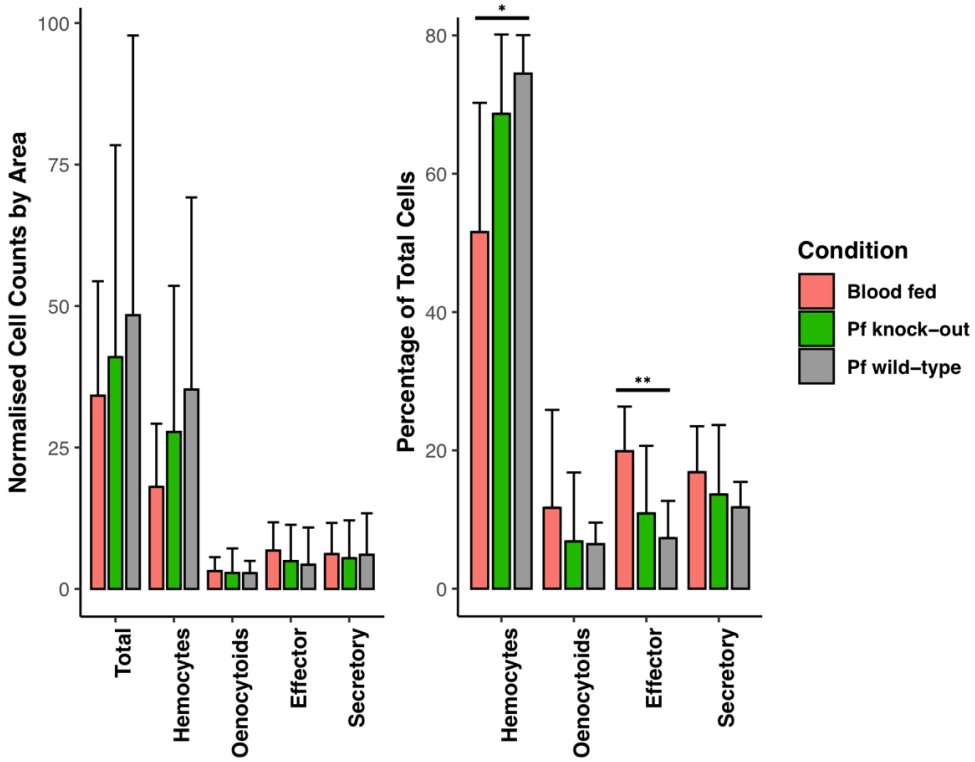


Fig. IV.19 RNA-FISH quantification of cell types on the gut of *Anopheles* mosquitoes. From 6 blood-fed, 6 wild-type *P. falciparum*, and 4 Pfs47 KO *P. falciparum* mosquitoes. To the left cell counts normalized by mm^2 of body wall area. To the right percentages of each cell type from total RNA-FISH positive cells. Error bars indicate 95% Confidence Interval. ** ($P \leq 0.01$), * ($P \leq 0.05$) – Welch T-Test. One biological repeat.

In order to recapitulate our findings with respect to the increase in FBN7 hemocytes after *P. berghei* infection we also looked at how FBN7+ hemocytes changed in circulation after *P. falciparum* infection. Again, we found that infection with the wild type *P. falciparum* significantly increased ($P < 0.0001$, Welch T-test) expression of FBN7 in hemocytes. Interestingly, the increase is abrogated by Pfs47 knock-out *P. falciparum* ($P = 0.02$, Welch T-test). FBN7+ cells went from 48.3% (±9.4) with blood-feeding to 83.6% (±11.3) with wild-type infection and 59.8% (±21.8) with knock-out infection. No other changes were observed [Fig. IV.20].

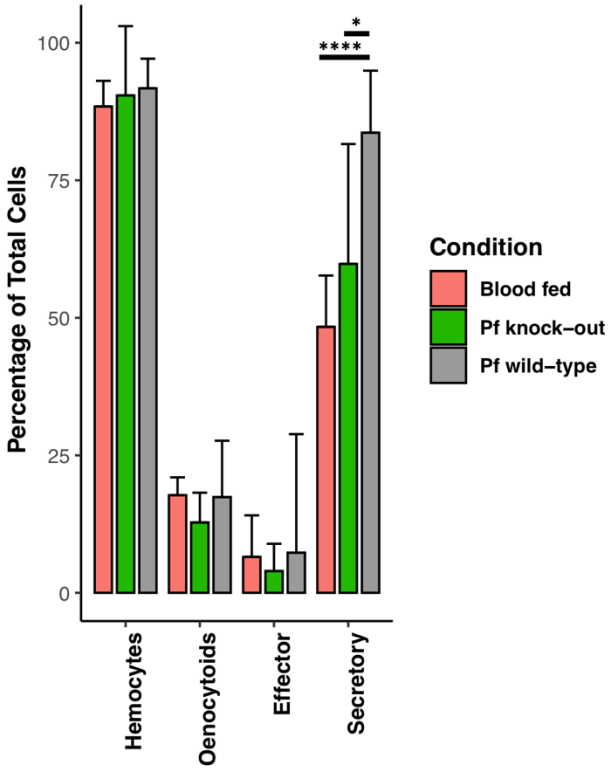


Fig. IV.20 Quantification of *Anopheles* hemocytes. From a total of 1066 blood-fed cells, 966 wild-type *P. falciparum* cells, and 1694 of Pfs47 knock-out cells (8 mosquitoes per repeat), followed by RNA-FISH as of methods. Error bars, 95% Confidence Interval. **** ($P \leq 0.0001$), * ($P \leq 0.05$) – Welch T-Test. Two biological repeat, two technical repeats each.

4 Discussion

In this chapter we looked at how the *A. gambiae* immune system responds to *Plasmodium* infection. First, we quantified the hemocyte cell clusters obtained via scRNA-seq in control (sugar-fed), and challenged conditions (blood feeding and infection). Both caused a decrease in the proportion of cells identified as prohemocytes. The proportion of inactive, baseline granulocytes was the same in sugar-fed and blood-fed mosquitoes, whereas infection caused a large increase in the number of active granulocytes (both type 1 and 2). This suggests a recruitment of baseline granulocytes to more active granulocytes states. Other terminal cells such as effector cells, oenocytoids, and rapidly dividing cells increased by the same amount with both blood-feeding and infection. Secretory cells are an exception and were mostly detected in blood-fed samples. Either secretory cells are up-regulated with blood feeding as a way to pre-empt invasions of bacteria, or our result was a spurious technical artefact due to mosquito manipulation. These are rare cells, and muscle cells were also detected more in blood-fed samples. It is possible secretory cells could be associated with wing muscles or cardio-vascular tissues. Fat body cells also increased with *P. berghei* infection. We hypothesize the increase is due to mosquito-wide immune responses causing cells to dislodge more easily. However, mobile fat body cells have recently been hypothesized to be genuine cells in circulation in the mosquito hemolymph. Indeed, at least some of the large number of fat droplets we observed via FACS in the mosquito hemolymph (Chapter II) could have been cells. If so, these cells could represent true, yet completely unexplored biology that will need to be investigated further.

We confirmed the number of hemocytes estimated by scRNA-seq with manual hemocyte counts, and they are reassuringly consistent. When the mosquito immune system is not activated, granulocytes are rare, whereas with activation (blood feeding and especially *Plasmodium* infection) large granulocytes can increase to double-digit percentages. We already described how prohemocytes and granulocytes lie in a continuum of transcriptomic similarity, as likely alternative cell stages along the same cell developmental trajectory. The relative absence of activated type 1 and type 2 granulocytes, effector cells, and dividing cells under

baseline conditions, and their increase with infection indicate these cells are what we identify as granulocytes via manual morphological microscopic analyses. Indeed, correlative microscopy experiments confirmed it. The only discrepancy between scRNA-seq and microscopy hemocyte numbers was for oenocytoids (~10% in the scRNA-seq dataset vs ~20% with microscopy quantification). Either these cells are more difficult to isolate and sequence, or not all cells that we previously morphologically identified as oenocytoids are in fact so. Correlative microscopy showed similarities between mature oenocytoids and maturing granulocytes, suggesting at least a partial overlap between these cell types, making manual morphological quantification more challenging.

Importantly, we discovered that the transcription factor LL3 is both highly and specifically expressed in a subset of hemocytes, the effector granulocytes. It had been separately shown that LL3 is expressed in hemocytes, and that silencing LL3 disrupt the mosquito's priming response[372]. However, the precise role of LL3 is priming was not clear. We found that systemically silencing LL3 abrogates the ability of mosquitoes to respond to an immune challenge, and that granulocyte numbers don't increase. Our results suggest that LL3, and the effector cells that express it, play an important role in coordinating hemocytes. However, we were not able to specifically target effector cells with our silencing because no appropriate experimental system existed. Our research will however provide the scientific community with the knowledge required to synthesise antibodies to specific hemocyte cell types. We hope future adoptive transfer experiments will fully elucidate the role of LL3 and effector cells in mosquito immune memory.

We then explored how *P. berghei* infection changes the transcriptomic landscape of mosquitoes. We performed DE analyses on both of our bulk and scRNA-seq datasets. In both, a large number of transcripts were differentially regulated in hemocytes in response to malaria infection. The effects of malaria infection peaked between one- and three-days post-infection, especially for granulocytes. When hemocytes obtained seven days post-infection were included in DE analyses the number of differentially expressed genes decreased, except for prohemocytes. Granulocytes thus appear to be the first cell type to respond, and prohemocytes

the last, consistently with the hypothesis that prohemocytes identified transcriptomically via scRNA-seq can include both morphological prohemocytes as well as some inactivated granulocytes.

There were many transcripts of interest upregulated in hemocytes upon infection. PGRPLD for example has been shown to increase *Plasmodium* infection prevalence upon knock-down [374]. CLIPB8 is required for prophenoloxidase activation and melanization of invading pathogens, and is highly upregulated by *P. berghei* infection [375]. Translationally-controlled tumor protein homolog (TCTP) has been shown to be an opsonin in silkworm and other hemocytes but has never before been implicated in the response against malaria, and could be a novel anti-plasmodial effector molecule [376]. The matuselah receptor 6 was also upregulated. A paralogue in *Drosophila* has been linked with effective immune responses and increased longevity [377]. Protein homolog 5 was upregulated with infection and has epigenetic functions. Memory in mosquitoes has long been thought to be mediated by epigenetic changes, and Cbx5 could be involved. Interestingly, an FK506-binding protein was also upregulated, especially in prohemocytes. This transcript is an orthologue of *Drosophila* FKBP39, is expressed throughout the life of *Drosophila* flies, binds DNA, is localized mostly in the nucleus, and could be a novel transcription factor important in immunity against *Plasmodium* [378]. In addition, calreticulin was also upregulated. Calreticulin has been shown to mediate phagocytosis and encapsulation in *Anopheles* mosquitoes as well as *Drosophila* [379, 380]. Another interesting transcript is the transcription factor BTF homologue 4, called "bicaudal" in *Drosophila*. During development bicaudal mutations have been shown to abort the establishment of the head. In addition, no hemocytes develop in these mutants. BTF4 could be important for hemocyte replication and development also in *Anopheles* mosquitoes [381]. Finally, GTP-binding nuclear protein Ran is an orthologue of *Drosophila* Ran, and is also involved in mitosis and cell division. In shrimp hemocytes it instead regulates phagocytosis, and in our dataset Ran was particularly upregulated in prohemocytes.

Most transcripts differentially expressed with infection in prohemocytes were also upregulated in granulocytes, but there were some interesting transcripts unique to

prohemocytes. For example, 'cellular nucleic acid binding protein' is an orthologue of a human protein that is involved in steroid signalling and proliferation. 14-3-3 protein epsilon is highly conserved in vertebrates and has been shown in lower organisms to promote anti-microbial hemocyte function. In addition, cofilin is involved in actin reorganisation and together with the actin-related protein (ARP) complex have important roles in immune responses and cytokinesis. Furthermore, granulocytes upregulated PPO2, 3, 4 and 5, while prohemocytes upregulated PPO6. And lastly, FK506-binding protein 14 expression in *Drosophila* has been shown to be activate Notch signalling and control lineage specification towards crystal cells (equivalent to oenocytoids) [382]. Whether late prohemocyte activation leads to oenocytoids differentiation remains an open question.

Interestingly, oenocytoids did not appear to strongly respond to infection. There were only ~16-17 DE transcripts, and no significant changes in the number of oenocytoids. Although the frequency of FBN7+ oenocytoids increased just as much as that FBN7+ granulocytes all in all our data suggests oenocytoids are not crucial mediators of anti-plasmodial defenses. On the other hand, fat body cells did respond, especially in the first few days (up to day 3) after infection. However, many up and downregulated genes in the fat body are not annotated, and much work remains to be done to elucidate what these genes do. Among annotated genes we saw downregulation of PPO inhibitor protein, LYSC4, and APG8, consistent with heightened immune activation. It has recently been proposed floating fat body cells may serve true immune functions, and in our hands the number of fat body cells did increase considerably after *P. berghei* infection. We hope future experiments will elucidate what these cells are and what function they possess in health and disease. Few genes were upregulated in dividing cells, effector cells, or secretory cells. However, secretory cells interestingly upregulated all-*trans*- and 9-*cis*-retinoic acids production. Both are important gene expression regulators and have essential roles in immunity, including cell proliferation and differentiation. Secretory cells could be releasing signals to activate nearby immune cells in response to infection. This hypothesis will also need to be tested in future experiments.

The immune response of *Anopheles gambiae* to malaria

Comparably fewer differentially expressed genes were detected with bulk RNAseq of hemocytes after *Plasmodium* infection, likely due to a dilution effect from a majority of non-responding cells. Among the genes that were upregulated we detected LL1 (LITAF-1) - a gene also upregulated in effector hemocytes and an LPS-responsive transcription factor - as well as TEP1, the key effector molecule for the early anti-*Plasmodium* responses. Furthermore, three peptidoglycan recognition proteins (PGRPs) were among the top differentially expressed genes. They are known immune regulators. PGRPS2 and PGRPS3 participate in antiparasitic defences, and PGRPLB was shown promote mosquito permissiveness to *P. falciparum* by deactivating the Imd pathway [374]. In addition, PPO inhibitor protein and CLIPA1 were upregulated. GO enrichment analyses found an enrichment of immune-related genes, confirming that hemocytes actively respond to malaria parasite infection.

In the gut on the other hand we found hundreds of upregulated genes during *P. berghei* infection, suggesting gut-intrinsic as well as hemocyte-mediated responses. Of all the tissues considered, the mosquito gut featured the highest number of DE genes. This is not surprising. *Plasmodium* ookinetes and oocysts mostly interact with the gut within the timeframe of our experiments. It would interest to confirm whether this is also true weeks after infection, as sporozoites move into salivary glands. The gut upregulated many important immune-related genes, including REL2/Imd. The REL2/Imd pathway is one of the two key nuclear factor-κB (NF-κB) immune pathways responsible for controlling *Plasmodium* infection, and the only pathway shown to mediate *P. falciparum* control, in addition to killing of viruses and Gram-negative bacteria [383]. Rel2 activation is mediated by transmembrane peptidoglycan recognition proteins (PGRPs). Then, the IKK-γ subunit of the IKK complex phosphorylates the Relish transcription factor, and IAP2 serves as an activator. IKK-γ, IAP2 and PGRPs were all upregulated in our hemocyte and gut samples with infection. We conclude that *P. berghei* activates the Rel2/Imd pathway in our mosquitoes. Rel1/Toll is the other key anti-plasmodial pathway, responsible for killing Gram-positive bacteria, fungi, and *P. berghei* [383]. Interestingly, 3 of the 4 most upregulated transcripts with infection were odorant binding proteins (OBP12, 38, 39), suggesting a role in immunity for this family of proteins. Ornithine decarboxylase was also highly upregulated both in the gut and in hemocytes. This protein is

thought to control macrophage activation and limit active macrophage M1 formation, suggesting hemocyte inhibition [384].

Unexpectedly, we could only detect 10 upregulated and 10 downregulated genes in mosquito carcasses, suggesting an absent or diluted immune response, and confirming the importance of using single cell rather than bulk approaches. We did nevertheless identify TRAF6 as an immune-related gene upregulated in the gut after infection. (TNFR)-associated factor 6 (TRAF6) is an adaptor protein first identified to mediate IL-1 receptor (IL-1R) NFκB activation. TRAF6 is now known to mediate TNFR, toll-like receptor (TLR), and tumor growth factor-β receptors (TGFβR) signaling to activate NFκB, MAPK, PI3K, and IRF, serving as a master regulator of immunity [385]. To our knowledge this is the first time TRAF6 has been involved in anti-plasmodial immunity.

Conversely, when we analysed the response of mosquitoes to blood feeding vis-à-vis sugar feeding we observed a complete rearrangement of the mosquito metabolism. Thousands of genes were differentially expressed, and our GO analysis showed the expected downregulation of sugar and xanthine metabolism and transport transcripts after blood-feeding (glucose, hexose, simple and complex sugar movement and catabolism), whereas transcripts involved in lipid metabolism and transport were upregulated. Blood-feeding turns the mosquito into a biosynthetic factory, upregulating transcripts involved in IMP and purine biosynthesis, amino acid metabolism, protein glycosylation, protein folding, and ER / membrane protein targeting. In addition, blood-feeding signals the start of the reproductive life-cycle of mosquitoes, with the production of fertile eggs. In our GO analyses the most significantly upregulated transcripts were indeed those involved in DNA replication initiation, DNA processing, (such as DNA geometric change and unwinding), DNA integrity checkpoints, mitotic checkpoints, and mitotic cell cycle.

We then explored the spatial dynamics of the *A. gambiae* immune response to *Plasmodium*. We leveraged RNA-FISH to observe how infection changed the proportion of sessile, tissue-resident hemocytes in the guts and carcasses of mosquitoes, and of motile

hemocytes in circulation. There were two major changes. First, hemocytes (likely granulocytes due to large cellular size and high LRR8 expression) were recruited in large numbers from the body wall of mosquitoes into the circulation. The marked increase in type 1 and type 2 granulocytes observed with infection would thus appear not to be solely caused by the heightened activation, replication, and differentiation of existing circulating hemocytes, but also by the recruitment of a reservoir of hemocytes attached to the body wall of mosquitoes. Indeed, when the interaction between hemocytes and fat body was probed in more detail by electron microscopy, we observed that the connection between these cells is tenuous at best, with hemocytes almost "walking" on the fat body, connected only through a few pseudopodia spreading from the central body of the cell, and thus readily dislodged. While the *P. falciparum* results are not conclusive due to the low number of samples, we observed a decreased number of LRR8+ cells attached to the midgut of mosquitoes infected with a *P. falciparum* Pfs47 knock-out line that is susceptible to immune killing by the mosquito immune system, whereas wild-type *P. falciparum* infection did not seem to change the number of attached hemocytes. Wild-type *P. falciparum* is able to masque its presence from the mosquito, and not activate immune responses, whereas *P. berghei* is not able to do so. Blocking hemocyte recruitment in the circulation could be one of the ways through which *P. falciparum* is able to survive in the mosquito.

The second main finding - a large increase in the number of hemocytes expressing fibrinogen (FBN7), a marker of secretory cells - was more surprising. Very few cells were positive for fibrinogen-CT (FBN7) in our scRNA-seq dataset. Conversely, the vast majority of motile hemocytes in circulation (upwards of 80%) becomes positive for FBN7 when mosquitoes are infected with *P. berghei*. The result holds true also for wild type *P. falciparum*. A Pfs47 *P. falciparum* knock-out, incapable of developing past the oocyte stage because of successful targeting and killing by the mosquito immune system, does not elicit the same response. Instead the number of FBN7+ cells is equal to that of blood feeding. More work will need to be done to fully understand the role of FBN7 and FBN7+ cells after *Plasmodium* infection. Other scRNA-seq markers of secretory cells included anti-microbial peptides,

Cathepsin L (perforin activator) - which in vertebrates promotes NK cell cytotoxicity - and Cathepsin F, which in invertebrates is involved in MHC II antigen presentation and Th1-immune responses [386]. An intriguing speculation is that the malaria parasite is able to skew the immune response towards an microbicidal state, akin to M2 state of macrophages in vertebrates, while at the same time blocking – and protecting itself from – the more potent phagocytic M1-like granulocyte activity. However, while Cathepsin-L is indeed downregulated in activated, IFN-γ treated, M1 human macrophages [387], these cells also upregulate ferritin, which is a conserved inflammatory response of activated M1 macrophages. Furthermore, secretory cells are also characterized by high levels of CLIPB4, important to control *P. berghei* infection. In addition, lineage tracing showed secretory and effector cells to be relatively similar. As such, it is also possible that fibrinogen-CT(FBN7) is instead upregulated as a general non-specific response to the gut wounds that are a result of *Plasmodium* ookinetes' midgut penetration, but that the expression is so low that our scRNA-seq library preparation was not able to capture it.

Finally, with RNA-FISH we saw a decrease in the number of effector hemocytes attached to the mosquito gut during *P. falciparum* WT infection. Infection with Pfs47 KO *P. falciparum* instead showed the number of effector cells to increase back to normal. If our hypothesis is correct and effector cells are modulating the immune system to respond to malaria infection then our results consistently suggest WT *P. falciparum* is able to blunt mosquito immune responses, and thus limit the number of anti-plasmodial effector cells. The identity of these cells will need to be explored further. These are large, rare cells, bigger than normal granulocytes, and could thus be the equivalent of *Drosophila*'s lamellocytes, which until now were not thought to exist in *Anopheles*.

Final summary & discussion

As we have seen, *Anopheline* mosquitoes are responsible for transmitting *Plasmodium* parasites to humans, and are the causative agent for over 219 million cases of malaria, and over 400,000 deaths annually[388]. However, the mosquito's immune system is far from being a passive bystander, and can limit *Plasmodium* infection in several ways[364, 365]. Hemocytes, the mosquito's equivalent of our white blood cells, are key players in these defensive responses, both through direct killing and through their role in complement activation and consequent parasite lysis. Infection with *Plasmodium* leads to a heightened state of immune activation in mosquitoes called priming. Primed mosquitoes are then able to mount a stronger immune response to subsequent infections. This response has been shown to be due to the release of a hemocyte differentiation factor (HDF) into the hemolymph[370], which is sufficient to increase the proportion of circulating granulocytes (the active subtype), as well as to promote hemocyte microvesicle release and subsequent complement activation[367].

When I started working on this project three hemocyte types, both circulating and sessile, had been described in *Anopheles gambiae* based on cellular morphology[389]. Granulocytes (10-20 μm) are the main phagocytic cells, while oenocytoids are smaller round cells 8-12 μm in diameter responsible for producing melanin, crucial for wound healing and pathogen encapsulation. And lastly, small round prohemocytes (4-6 μm) were thought of as the precursor cells of the other two cell subtypes. However, the full functional diversity of mosquito hemocytes was unexplored, their developmental trajectories were completely unknown, and it was unclear how much morphologically similar hemocytes are also functionally equivalent.

In this thesis I used single cell RNA sequencing (scRNA-seq) to profile the transcriptomes of 8506 hemocytes of *Anopheles gambiae* and *Aedes aegypti,* two important mosquito vectors. Blood feeding, infection with malaria parasites and other immune challenges revealed a previously unknown functional diversity of hemocytes, with different types of

granulocytes expressing distinct and evolutionarily conserved subsets of effector genes. A new cell type, which we term effector granulocytes, was defined in *Anopheles* by a unique transmembrane protein marker (TM7318) and high expression of LPS-Induced TNF-alpha transcription factor 3 (LL3). Knock-down experiments indicated that LL3 mediates hemocyte differentiation during immune priming. We identified two main hemocyte lineages and differentiation pathways in prohemocytes and granulocytes and found evidence of proliferating granulocyte populations. We discovered new hemocyte populations and markers of immune activation for each. We validated our analysis with RNA in-situ hybridization to integrate the transcriptional profiles with morphological analysis of circulating hemocytes and highlighted the mobilization of sessile hemocytes into circulation upon infection. And a comparison of *Anopheles* and *Aedes* hemocytes showed differences and similarities between these two mosquito species.

scRNA-seq revealed new types of hemocytes

First, circulating hemocytes were collected from adult *A. gambiae* M form (*A. coluzzii*) females that were either kept on a sugar meal or fed on a healthy or a *Plasmodium berghei*-infected mouse. Transcriptomes from 5,383 cells (collected 1, 3, and 7 days after feeding) revealed nine major cell clusters. Two clusters originated from adipose tissue and one from muscle tissue. Baseline fat body cells expressed several immune-modulatory genes such as CLIPs, LRIMs, lectins, and SRPNS, while active fat body cells expressed high levels of vitellogenin after blood feeding, a canonical marker of the mosquito fat body[390]. Based on their unique transcriptional profiles, we then identified six hemocyte clusters, including known cells such as oenocytoids, with high mRNAs levels of prophenoloxidases (e.g. PPO4 and PPO9), and granulocytes. We then selected hemocyte specific genes markers for our hemocyte lineages by combining scRNA-seq expression data with parallel bulk RNAseq data from different tissues. Oenocytoids contained low levels of leucine-repeat protein 8 (LRR8) mRNA, whereas granulocytes had an inverse pattern (low or absent PPO4 and high LLR8 levels).

In situ hybridization using these markers confirmed the oenocytoid-like round morphology (with few granules and pseudopodia) of circulating PPO4^{high}-LLR8^{low} cells, while PPO4^{low}-LLR8^{high} cells looked like prohemocytes and granulocytes (prominent pseudopodia and abundant granules). These two cell types appear to lie in a continuum of transcriptomic similarity, and while they shared many markers such as SPARC, cathepsin-L and LRR8, they differed in the amount of UMIs (73% fewer in prohemocytes). Dividing granulocytes were characterised by shared granulocyte markers, as well as their own unique subset of markers including cyclin B, aurora kinase and other mitotic markers. Secretory and effector hemocytes were both are negative for PPO4 and LLR8^{low}, but while effector hemocytes were large cells (25-40 μm) that expressed high levels of an uncharacterized transmembrane protein AGAP007318 (TM7318) and LPS-induced TNF-alpha transcription factor 3 (LL3), secretory hemocytes were smaller cells negative for both markers and instead expressing antimicrobial peptides such as defensin 1 and cecropins 1.

Mosquito hemocyte lineages in *Anopheles*

To investigate hemocyte differentiation dynamics, we then re-clustered the *Anopheles* cellular transcriptomes at higher resolution and performed lineage tree reconstruction with partition-based graph abstraction (PAGA) and found that proliferating cell were connected with the main granulocyte population which in turn was linked to effector cells and secretory antimicrobial granulocytes and prohemocytes. Sub-clusters within the major granulocyte populations reflected transcriptional responses to feeding and *Plasmodium* infection. Our findings were confirmed with both diffusion maps and slingshot analyses. They suggest the existence of a proliferative cell population that can replenish the pool of granulocytes, which can then differentiate further into more specialized regulatory or end-stage cells represented by effector and secretory cells without the need for further proliferation. Our data suggests that granulocyte proliferation and prohemocyte differentiation both appear to contribute to the observed increase in granulocyte numbers after blood feeding. However, the placement of prohemocytes in the granulocyte lineage tree should be considered tentative due to the few markers uniquely characterising prohemocytes. Prohemocytes are proposed to be precursors to both granulocytes and oenocytoids but the latter were transcriptionally disconnected from other

hemocyte subtypes, and we did not observe transcriptional markers of cell proliferation in oenocytoids. This suggests oenocytoids may represent a wholly separate lineage with origins either in larval stages or in other adult tissues. Alternatively, oenocytoids could derive from granulocytes, but due to a low differentiation rate we may not have captured intermediate cells.

To assess which of the newly discovered cell types are shared between anopheline and culicine mosquitoes, we also analysed the single-cell transcriptome of 3123 cells from *Ae. aegypti*, a vector for several relevant arboviruses including dengue. As with *Anopheles*, a dimensional reduction plot showed both canonical hemocytes and other cell types such as fat body cells and muscle cells. Our cross-species analysis revealed conserved transcriptome signatures for oenocytoids (99% and 77% correlation with *Anopheles* oenocytoids), and granulocyte subtypes, including antimicrobial cells (94% with secreting hemocytes), and proliferating granulocytes (87% with dividing granulocytes in *Anopheles*). Granulocytes and prohemocytes were again positioned on a continuum of transcriptomic similarity, with four different cell states, including a proliferating S-phase granulocyte cluster without a clear *Anopheles* equivalent. Granulocyte cells instead expressed laminins, leucine-rich repeat proteins, scavenger receptors, Toll receptor 5, and the transcription factor Rel2. Conversely, *Anopheles* effector granulocytes lack a counterpart in *Aedes*, and furthermore the main gene marker appear to be restricted to African and Asian *Anopheles*, suggesting effector hemocytes could be unique to a subgroup of *Anopheles* mosquitoes.

Transcription factor LL3 is required for hemocyte differentiation during priming

LL3 had been previously detected in granulocytes in response to *Plasmodium* infection. And LL3 silencing had been shown to abolish the *Anopheles* priming response[372]. Since we found LL3 to be a specific marker of effector granulocyte we explored whether silencing LL3 affects the ability of hemocytes to respond to HDF and we found the priming response to be completely abolished when LL3 expression was silenced in mosquitoes that had received HDF, suggesting LL3 and effector granulocytes play an important role in orchestrating the hemocyte priming response.

Blood-feeding and *Plasmodium* infection trigger granulocyte activation and mobilization

Finally, we explored how the cell types we described respond to malaria infection. It had been previously shown that the proportion of circulating granulocytes was low (1-3%) under normal conditions but increased after *Plasmodium* infection[370]. However, it was still unknown whether the increase was due to proliferation and differentiation of circulating progenitor cells, or mobilization of sessile hemocytes. Our transmission electron microscopy of individual sessile granulocytes attached to the basal lamina of the tissues through were indicative of a dynamic and potentially transient association. To explore that possibility, we used whole tissue mount *in situ* hybridization to find most sessile hemocytes to be $PPO4^{low}/LLR8^{high}$ granulocytes whereas oenocytoids, effector, and secretory cells are rare. Importantly, we found a dramatic reduction of sessile $PPO4^{low}/LLR8^{high}$ granulocytes in response to *Plasmodium* infection and no significant difference in the numbers of sessile oenocytoids, effector, or secretory cells.

Lastly, in circulating hemocytes both *P. berghei* and *P. falciparum* infection induced a significant increase in the proportion of FBN7 positive cells, indicating that this is a general marker of hemocyte immune activation. Combined, our results suggest that hemocyte recruitment from the body wall, granulocyte activation and proliferation, and prohemocyte differentiation can all contribute to boost circulating granulocyte numbers upon immune challenge.

Final considerations and outlook

Our knowledge of cellular immunity in vertebrates relies critically on understanding the functional diversity of cell types, their developmental trajectories and their trafficking dynamics. This thesis represents significant progress towards this understanding for two invertebrate immune systems that limit the vectorial capacity of mosquitoes for deadly human diseases such as malaria and Dengue. We confirmed the existence of oenocytoids and granulocytes and with gene markers we related cellular morphology to a more comprehensive molecular characterization of these cells. Unlike current thinking in the field we show

prohemocytes and granulocytes are closely related cells, and furthermore discovered the transcriptional profiles and molecular markers of novel hemocyte subtypes (effector, dividing, and secreting granulocytes), as well as fat body cells with immune-modulatory functions.

We defined two main hemocyte lineages in *A. gambiae*: the oenocytoid lineage, and the prohemocyte-granulocyte lineage. The latter can be further split into three sub lineages leading to differentiated effector, antimicrobial, or dividing granulocytes. We unearthed their precise molecular diversity and showed them to be largely conserved between distantly related mosquito genera and as such presumably of functional importance. However, we were not able to find effector granulocytes in *Aedes* mosquitoes. These cells have a unique, large morphology, and specifically express LL3. Silencing this transcription factor provided tentative evidence for a regulatory role of these cells in immune priming. However, we cannot rule out that other hemocyte types or mosquito tissues express LL3 and might thus be directly affected by LL3 silencing.

The cell-type markers and FISH probes we identified and validated will allow investigators to probe the immune functions of effector granulocytes and other specialized hemocyte types in detail. We leave open the question of the developmental origin of oenocytoids, but we identify two potential origins for the expansion of circulating granulocytes after immune challenge (blood feeding or *Plasmodium* infection). One is the mobilization of sessile granulocytes from the body wall, the other is a pool of proliferating, oligopotent granulocytes. Whether prohemocytes, which are transcriptionally related but less responsive than granulocytes, can transform into granulocytes and whether they can enter the cell cycle, also remains to be discovered.

In summary, the cell-type-specific marker genes, reference transcriptomes, and companion website (https://hemocytes.cellgeni.sanger.ac.uk/) from our study provide the first atlas of medically relevant invertebrate immune cells at single cell resolution and will serve as a resource for the field, providing a starting point for the type of lineage tracing and functional experiments which, in vertebrates, are resolving the developmental origins and functions of diverse immune cell populations.

References

1. WHO | World malaria report 2018. In: WHO. http://www.who.int/malaria/publications/world-malaria-report-2018/en/. Accessed 14 Jan 2019

2. Crompton PD, Moebius J, Portugal S, et al (2014) Malaria Immunity in Man and Mosquito: Insights into Unsolved Mysteries of a Deadly Infectious Disease. Annu Rev Immunol 32:157–187. https://doi.org/10.1146/annurev-immunol-032713-120220

3. Hillyer JF (2010) Mosquito immunity. Adv Exp Med Biol 708:218–238

4. Castillo JC, Robertson AE, Strand MR (2006) Characterization of hemocytes from the mosquitoes Anopheles gambiae and Aedes aegypti. Insect Biochem Mol Biol 36:891–903. https://doi.org/10.1016/j.ibmb.2006.08.010

5. WHO | Malaria. In: WHO. https://www.who.int/ith/diseases/malaria/en/. Accessed 8 Jul 2019

6. Coetzee M, Hunt RH, Wilkerson R, et al (2013) Anopheles coluzzii and Anopheles amharicus , new members of the Anopheles gambiae complex. Zootaxa 3619:246–274. https://doi.org/10.11646/zootaxa.3619.3.2

7. Greenwood B, Marsh K, Snow R (1991) Why do some African children develop severe malaria? Parasitol Today 7:277–281. https://doi.org/10.1016/0169-4758(91)90096-7

8. Langhorne J, Ndungu FM, Sponaas A-M, Marsh K (2008) Immunity to malaria: more questions than answers. Nat Immunol 9:725–732. https://doi.org/10.1038/ni.f.205

9. Marsh K, Kinyanjui S (2006) Immune effector mechanisms in malaria. Parasite Immunol 28:51–60. https://doi.org/10.1111/j.1365-3024.2006.00808.x

10. WHO | World Malaria Report 2014. In: WHO. http://www.who.int/malaria/publications/world_malaria_report_2014/report/en/. Accessed 2 Sep 2015

11. S Clinical Trials Partnership R (2015) Efficacy and safety of RTS,S/AS01 malaria vaccine with or without a booster dose in infants and children in Africa: final results of a phase 3, individually randomised, controlled trial. The Lancet 386:31–45. https://doi.org/10.1016/S0140-6736(15)60721-8

12. Miller LH, Ackerman HC, Su X, Wellems TE (2013) Malaria biology and disease pathogenesis: insights for new treatments. Nat Med 19:156–167. https://doi.org/10.1038/nm.3073

13. WHO | Global report on insecticide resistance in malaria vectors: 2010–2016. In: WHO. http://www.who.int/malaria/publications/atoz/9789241514057/en/. Accessed 24 Aug 2019

14. Smith DL, McKenzie FE, Snow RW, Hay SI (2007) Revisiting the Basic Reproductive Number for Malaria and Its Implications for Malaria Control. PLoS Biol 5:e42. https://doi.org/10.1371/journal.pbio.0050042

15. Gonçalves D, Hunziker P (2016) Transmission-blocking strategies: the roadmap from laboratory bench to the community. Malar J 15:. https://doi.org/10.1186/s12936-016-1163-3

16. Scholte E-J, Knols BGJ, Samson RA, Takken W (2004) Entomopathogenic fungi for mosquito control: A review. J Insect Sci 4:

17. Ramirez JL, Garver LS, Brayner FA, et al (2014) The role of hemocytes in Anopheles gambiae antiplasmodial immunity. J Innate Immun 6:119–128. https://doi.org/10.1159/000353765

18. Otto TD, Gilabert A, Crellen T, et al (2018) Genomes of all known members of a Plasmodium subgenus reveal paths to virulent human malaria. Nat Microbiol 3:687. https://doi.org/10.1038/s41564-018-0162-2

19. Kwiatkowski DP (2005) How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. Am J Hum Genet 77:171–192

20. Elmariah H, Garrett ME, De Castro LM, et al (2014) Factors associated with survival in a contemporary adult sickle cell disease cohort. Am J Hematol 89:530–535. https://doi.org/10.1002/ajh.23683

21. Gardner K, Douiri A, Drasar E, et al (2016) Survival in adults with sickle cell disease in a high-income setting. Blood 128:1436–1438. https://doi.org/10.1182/blood-2016-05-716910

22. The impact of parasitic diseases and their control, with an emphasis on malaria and Africa | POPLINE.org. https://www.popline.org/node/418298. Accessed 8 Jul 2019

23. Molina-Cruz A, Canepa GE, Kamath N, et al (2015) Plasmodium evasion of mosquito immunity and global malaria transmission: The lock-and-key theory. Proc Natl Acad Sci U S A 112:15178–15183. https://doi.org/10.1073/pnas.1520426112

24. Cowman AF, Healer J, Marapana D, Marsh K (2016) Malaria: Biology and Disease. Cell 167:610–624. https://doi.org/10.1016/j.cell.2016.07.055

25. Tavares J, Formaglio P, Thiberge S, et al (2013) Role of host cell traversal by the malaria sporozoite during liver infection. J Exp Med 210:905–915. https://doi.org/10.1084/jem.20121130

26. Ishino T, Yano K, Chinzei Y, Yuda M (2004) Cell-passage activity is required for the malarial parasite to cross the liver sinusoidal cell layer. PLoS Biol 2:E4. https://doi.org/10.1371/journal.pbio.0020004

27. Risco-Castillo V, Topçu S, Marinach C, et al (2015) Malaria Sporozoites Traverse Host Cells within Transient Vacuoles. Cell Host Microbe 18:593–603. https://doi.org/10.1016/j.chom.2015.10.006

28. Bhanot P, Schauer K, Coppens I, Nussenzweig V (2005) A surface phospholipase is involved in the migration of plasmodium sporozoites through cells. J Biol Chem 280:6752–6760. https://doi.org/10.1074/jbc.M411465200

29. Coppi A, Tewari R, Bishop JR, et al (2007) Heparan sulfate proteoglycans provide a signal to Plasmodium sporozoites to stop migrating and productively invade host cells. Cell Host Microbe 2:316–327. https://doi.org/10.1016/j.chom.2007.10.002

30. Rodrigues CD, Hannus M, Prudêncio M, et al (2008) Host scavenger receptor SR-BI plays a dual role in the establishment of malaria parasite liver infection. Cell Host Microbe 4:271–282. https://doi.org/10.1016/j.chom.2008.07.012

31. Herrera R, Anderson C, Kumar K, et al (2015) Reversible Conformational Change in the Plasmodium falciparum Circumsporozoite Protein Masks Its Adhesion Domains. Infect Immun 83:3771–3780. https://doi.org/10.1128/IAI.02676-14

32. Sturm A, Amino R, van de Sand C, et al (2006) Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. Science 313:1287–1290. https://doi.org/10.1126/science.1129720

33. Weiss GE, Gilson PR, Taechalertpaisarn T, et al (2015) Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during Plasmodium falciparum invasion of erythrocytes. PLoS Pathog 11:e1004670. https://doi.org/10.1371/journal.ppat.1004670

34. Holder AA (1994) Proteins on the surface of the malaria parasite and cell invasion. Parasitology 108 Suppl:S5-18

35. Tham W-H, Healer J, Cowman AF (2012) Erythrocyte and reticulocyte binding-like proteins of Plasmodium falciparum. Trends Parasitol 28:23–30. https://doi.org/10.1016/j.pt.2011.10.002

36. Chen L, Lopaticki S, Riglar DT, et al (2011) An EGF-like protein forms a complex with PfRh5 and is required for invasion of human erythrocytes by Plasmodium falciparum. PLoS Pathog 7:e1002199. https://doi.org/10.1371/journal.ppat.1002199

37. Reddy KS, Amlabu E, Pandey AK, et al (2015) Multiprotein complex between the GPI-anchored CyRPA with PfRH5 and PfRipr is crucial for Plasmodium falciparum erythrocyte invasion. Proc Natl Acad Sci U S A 112:1179–1184. https://doi.org/10.1073/pnas.1415466112

38. Crosnier C, Bustamante LY, Bartholdson SJ, et al (2011) Basigin is a receptor essential for erythrocyte invasion by Plasmodium falciparum. Nature 480:534–537. https://doi.org/10.1038/nature10606

39. Volz JC, Yap A, Sisquella X, et al (2016) Essential Role of the PfRh5/PfRipr/CyRPA Complex during Plasmodium falciparum Invasion of Erythrocytes. Cell Host Microbe 20:60–71. https://doi.org/10.1016/j.chom.2016.06.004

40. Besteiro S, Dubremetz J-F, Lebrun M (2011) The moving junction of apicomplexan parasites: a key structure for invasion. Cell Microbiol 13:797–805. https://doi.org/10.1111/j.1462-5822.2011.01597.x

41. Riglar DT, Richard D, Wilson DW, et al (2011) Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. Cell Host Microbe 9:9–20. https://doi.org/10.1016/j.chom.2010.12.003

42. Kafsack BFC, Rovira-Graells N, Clark TG, et al (2014) A transcriptional switch underlies commitment to sexual development in malaria parasites. Nature 507:248–252. https://doi.org/10.1038/nature12920

43. Joice R, Nilsson SK, Montgomery J, et al (2014) Plasmodium falciparum transmission stages accumulate in the human bone marrow. Sci Transl Med 6:244re5-244re5. https://doi.org/10.1126/scitranslmed.3008882

44. Aly ASI, Vaughan AM, Kappe SHI (2009) Malaria Parasite Development in the Mosquito and Infection of the Mammalian Host. Annu Rev Microbiol 63:195–221. https://doi.org/10.1146/annurev.micro.091208.073403

45. Bennink S, Kiesow MJ, Pradel G (2016) The development of malaria parasites in the mosquito midgut. Cell Microbiol 18:905–918. https://doi.org/10.1111/cmi.12604

46. Kawamoto F, Alejo-Blanco R, Fleck SL, Sinden RE (1991) Plasmodium berghei: ionic regulation and the induction of gametogenesis. Exp Parasitol 72:33–42

47. Billker O, Shaw MK, Margos G, Sinden RE (1997) The roles of temperature, pH and mosquito factors as triggers of male and female gametogenesis of Plasmodium berghei in vitro. Parasitology 115 (Pt 1):1–7

48. Billker O, Lindo V, Panico M, et al (1998) Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. Nature 392:289–292. https://doi.org/10.1038/32667

49. Garcia GE, Wirtz RA, Barr JR, et al (1998) Xanthurenic acid induces gametogenesis in Plasmodium, the malaria parasite. J Biol Chem 273:12003–12005. https://doi.org/10.1074/jbc.273.20.12003

50. Sologub L, Kuehn A, Kern S, et al (2011) Malaria proteases mediate inside-out egress of gametocytes from red blood cells following parasite transmission to the mosquito. Cell Microbiol 13:897–912. https://doi.org/10.1111/j.1462-5822.2011.01588.x

51. Muhia DK, Swales CA, Deng W, et al (2001) The gametocyte-activating factor xanthurenic acid stimulates an increase in membrane-associated guanylyl cyclase activity in the human malaria parasite Plasmodium falciparum. Mol Microbiol 42:553–560

52. McRobert L, Taylor CJ, Deng W, et al (2008) Gametogenesis in malaria parasites is mediated by the cGMP-dependent protein kinase. PLoS Biol 6:e139. https://doi.org/10.1371/journal.pbio.0060139

53. Billker O, Dechamps S, Tewari R, et al (2004) Calcium and a calcium-dependent protein kinase regulate gamete formation and mosquito transmission in a malaria parasite. Cell 117:503–514. https://doi.org/10.1016/s0092-8674(04)00449-0

54. Martin SK, Jett M, Schneider I (1994) Correlation of phosphoinositide hydrolysis with exflagellation in the malaria microgametocyte. J Parasitol 80:371–378

55. Raabe AC, Wengelnik K, Billker O, Vial HJ (2011) Multiple roles for Plasmodium berghei phosphoinositide-specific phospholipase C in regulating gametocyte activation and differentiation. Cell Microbiol 13:955–966. https://doi.org/10.1111/j.1462-5822.2011.01591.x

56. Sebastian S, Brochet M, Collins MO, et al (2012) A Plasmodium calcium-dependent protein kinase controls zygote development and transmission by translationally activating repressed mRNAs. Cell Host Microbe 12:9–19. https://doi.org/10.1016/j.chom.2012.05.014

57. Alano P, Read D, Bruce M, et al (1995) COS cell expression cloning of Pfg377, a Plasmodium falciparum gametocyte antigen associated with osmiophilic bodies. Mol Biochem Parasitol 74:143–156

58. Ponzi M, Sidén-Kiamos I, Bertuccini L, et al (2009) Egress of Plasmodium berghei gametes from their host erythrocyte is mediated by the MDV-1/PEG3 protein. Cell Microbiol 11:1272–1288. https://doi.org/10.1111/j.1462-5822.2009.01331.x

59. Talman AM, Lacroix C, Marques SR, et al (2011) PbGEST mediates malaria transmission to both mosquito and vertebrate host. Mol Microbiol 82:462–474. https://doi.org/10.1111/j.1365-2958.2011.07823.x

60. Wirth CC, Glushakova S, Scheuermayer M, et al (2014) Perforin-like protein PPLP2 permeabilizes the red blood cell membrane during egress of Plasmodium falciparum gametocytes. Cell Microbiol 16:709–733. https://doi.org/10.1111/cmi.12288

61. Janse CJ, Ponnudurai T, Lensen AH, et al (1988) DNA synthesis in gametocytes of Plasmodium falciparum. Parasitology 96 (Pt 1):1–7

62. van Dijk MR, Janse CJ, Thompson J, et al (2001) A central role for P48/45 in malaria parasite male gamete fertility. Cell 104:153–164. https://doi.org/10.1016/s0092-8674(01)00199-4

63. van Schaijk BCL, van Dijk MR, van de Vegte-Bolmer M, et al (2006) Pfs47, paralog of the male fertility factor Pfs48/45, is a female specific surface protein in Plasmodium falciparum. Mol Biochem Parasitol 149:216–222. https://doi.org/10.1016/j.molbiopara.2006.05.015

64. Pradel G, Hayton K, Aravind L, et al (2004) A multidomain adhesion protein family expressed in Plasmodium falciparum is essential for transmission to the mosquito. J Exp Med 199:1533–1544. https://doi.org/10.1084/jem.20031274

65. Janse CJ, van der Klooster PF, van der Kaay HJ, et al (1986) DNA synthesis in Plasmodium berghei during asexual and sexual development. Mol Biochem Parasitol 20:173–182

66. Reininger L, Billker O, Tewari R, et al (2005) A NIMA-related protein kinase is essential for completion of the sexual cycle of malaria parasites. J Biol Chem 280:31957–31964. https://doi.org/10.1074/jbc.M504523200

67. Reininger L, Tewari R, Fennell C, et al (2009) An essential role for the Plasmodium Nek-2 Nima-related protein kinase in the sexual development of malaria parasites. J Biol Chem 284:20858–20868. https://doi.org/10.1074/jbc.M109.017988

68. Ukegbu CV, Cho J-S, Christophides GK, Vlachou D (2015) Transcriptional silencing and activation of paternal DNA during Plasmodium berghei zygotic development and transformation to oocyst. Cell Microbiol 17:1230–1240. https://doi.org/10.1111/cmi.12433

69. Yuda M, Iwanaga S, Shigenobu S, et al (2009) Identification of a transcription factor in the mosquito-invasive stage of malaria parasites. Mol Microbiol 71:1402–1414. https://doi.org/10.1111/j.1365-2958.2009.06609.x

70. Kaneko I, Iwanaga S, Kato T, et al (2015) Genome-Wide Identification of the Target Genes of AP2-O, a Plasmodium AP2-Family Transcription Factor. PLoS Pathog 11:e1004905. https://doi.org/10.1371/journal.ppat.1004905

71. Tomas AM, Margos G, Dimopoulos G, et al (2001) P25 and P28 proteins of the malaria ookinete surface have multiple and partially redundant functions. EMBO J 20:3975–3983. https://doi.org/10.1093/emboj/20.15.3975

72. Aikawa M, Carter R, Ito Y, Nijhout MM (1984) New observations on gametogenesis, fertilization, and zygote transformation in Plasmodium gallinaceum. J Protozool 31:403–413

73. Sinden RE, Hartley RH, Winger L (1985) The development of Plasmodium ookinetes in vitro: an ultrastructural study including a description of meiotic division. Parasitology 91 (Pt 2):227–244

74. Vlachou D, Zimmermann T, Cantera R, et al (2004) Real-time, in vivo analysis of malaria ookinete locomotion and mosquito midgut invasion. Cell Microbiol 6:671–685. https://doi.org/10.1111/j.1462-5822.2004.00394.x

75. Hirai M, Arai M, Kawai S, Matsuoka H (2006) PbGCbeta is essential for Plasmodium ookinete motility to invade midgut cell and for successful completion of parasite life cycle in mosquitoes. J Biochem (Tokyo) 140:747–757. https://doi.org/10.1093/jb/mvj205

76. Moon RW, Taylor CJ, Bex C, et al (2009) A cyclic GMP signalling module that regulates gliding motility in a malaria parasite. PLoS Pathog 5:e1000599. https://doi.org/10.1371/journal.ppat.1000599

77. Ishino T, Orito Y, Chinzei Y, Yuda M (2006) A calcium-dependent protein kinase regulates Plasmodium ookinete access to the midgut epithelial cell. Mol Microbiol 59:1175–1184. https://doi.org/10.1111/j.1365-2958.2005.05014.x

78. Siden-Kiamos I, Ecker A, Nybäck S, et al (2006) Plasmodium berghei calcium-dependent protein kinase 3 is required for ookinete gliding motility and mosquito midgut invasion. Mol Microbiol 60:1355–1363. https://doi.org/10.1111/j.1365-2958.2006.05189.x

79. Lehane MJ (1997) Peritrophic matrix structure and function. Annu Rev Entomol 42:525–550. https://doi.org/10.1146/annurev.ento.42.1.525

80. Vinetz JM, Dave SK, Specht CA, et al (1999) The chitinase PfCHT1 from the human malaria parasite Plasmodium falciparum lacks proenzyme and chitin-binding domains and displays unique substrate preferences. Proc Natl Acad Sci U S A 96:14061–14066. https://doi.org/10.1073/pnas.96.24.14061

81. Vinetz JM, Valenzuela JG, Specht CA, et al (2000) Chitinases of the avian malaria parasite Plasmodium gallinaceum, a class of enzymes necessary for parasite invasion of the mosquito midgut. J Biol Chem 275:10331–10341. https://doi.org/10.1074/jbc.275.14.10331

82. Tsai YL, Hayward RE, Langer RC, et al (2001) Disruption of Plasmodium falciparum chitinase markedly impairs parasite invasion of mosquito midgut. Infect Immun 69:4048–4054. https://doi.org/10.1128/IAI.69.6.4048-4054.2001

83. Dessens JT, Beetsma AL, Dimopoulos G, et al (1999) CTRP is essential for mosquito infection by malaria ookinetes. EMBO J 18:6221–6227. https://doi.org/10.1093/emboj/18.22.6221

84. Yuda M, Sakaida H, Chinzei Y (1999) Targeted disruption of the plasmodium berghei CTRP gene reveals its essential role in malaria infection of the vector mosquito. J Exp Med 190:1711–1716. https://doi.org/10.1084/jem.190.11.1711

85. Li F, Templeton TJ, Popov V, et al (2004) Plasmodium ookinete-secreted proteins secreted through a common micronemal pathway are targets of blocking malaria transmission. J Biol Chem 279:26635–26644. https://doi.org/10.1074/jbc.M401385200

86. Wirth CC, Bennink S, Scheuermayer M, et al (2015) Perforin-like protein PPLP4 is crucial for mosquito midgut infection by Plasmodium falciparum. Mol Biochem Parasitol 201:90–99. https://doi.org/10.1016/j.molbiopara.2015.06.005

87. Yuda M, Yano K, Tsuboi T, et al (2001) von Willebrand Factor A domain-related protein, a novel microneme protein of the malaria ookinete highly conserved throughout Plasmodium parasites. Mol Biochem Parasitol 116:65–72

88. Dessens JT, Sidén-Kiamos I, Mendoza J, et al (2003) SOAP, a novel malaria ookinete protein involved in mosquito midgut invasion and oocyst development. Mol Microbiol 49:319–329

89. Kadota K, Ishino T, Matsuyama T, et al (2004) Essential role of membrane-attack protein in malarial transmission to mosquito host. Proc Natl Acad Sci U S A 101:16310–16315. https://doi.org/10.1073/pnas.0406187101

90. Kariu T, Ishino T, Yano K, et al (2006) CelTOS, a novel malarial protein that mediates transmission to mosquito and vertebrate hosts. Mol Microbiol 59:1369–1379. https://doi.org/10.1111/j.1365-2958.2005.05024.x

91. Ecker A, Bushell ESC, Tewari R, Sinden RE (2008) Reverse genetics screen identifies six proteins important for malaria development in the mosquito. Mol Microbiol 70:209–220. https://doi.org/10.1111/j.1365-2958.2008.06407.x

92. Adini A, Warburg A (1999) Interaction of Plasmodium gallinaceum ookinetes and oocysts with extracellular matrix proteins. Parasitology 119:331–336. https://doi.org/10.1017/S0031182099004874

93. Mahairaki V, Voyatzi T, Sidén-Kiamos I, Louis C (2005) The Anopheles gambiae gamma1 laminin directly binds the Plasmodium berghei circumsporozoite- and TRAP-related protein (CTRP). Mol Biochem Parasitol 140:119–121. https://doi.org/10.1016/j.molbiopara.2004.11.012

94. Limviroj W, Yano K, Yuda M, et al (2002) IMMUNO-ELECTRON MICROSCOPIC OBSERVATION OF PLASMODIUM BERGHEI CTRP LOCALIZATION IN THE MIDGUT OF THE VECTOR MOSQUITO ANOPHELES STEPHENSI. J Parasitol 88:664–672. https://doi.org/10.1645/0022-3395(2002)088[0664:IEMOOP]2.0.CO;2

95. Srinivasan P, Fujioka H, Jacobs-Lorena M (2008) PbCap380, a novel oocyst capsule protein, is essential for malaria parasite survival in the mosquito. Cell Microbiol 10:1304–1312. https://doi.org/10.1111/j.1462-5822.2008.01127.x

96. Sasaki H, Sekiguchi H, Sugiyama M, Ikadai H (2017) Plasmodium berghei Cap93, a novel oocyst capsule-associated protein, plays a role in sporozoite development. Parasit Vectors 10:399. https://doi.org/10.1186/s13071-017-2337-8

97. Moran P, Caras IW (1994) Requirements for glycosylphosphatidylinositol attachment are similar but not identical in mammalian cells and parasitic protozoa. J Cell Biol 125:333–343. https://doi.org/10.1083/jcb.125.2.333

98. Trueman HE, Raine JD, Florens L, et al (2004) FUNCTIONAL CHARACTERIZATION OF AN LCCL–LECTIN DOMAIN CONTAINING PROTEIN FAMILY IN PLASMODIUM BERGHEI. J Parasitol 90:1062–1071. https://doi.org/10.1645/GE-3368

99. Sinden RE (1974) Excystment by sporozoites of malaria parasites. Nature 252:314. https://doi.org/10.1038/252314a0

100. Ménard R, Sultan AA, Cortes C, et al (1997) Circumsporozoite protein is required for development of malaria sporozoites in mosquitoes. Nature 385:336. https://doi.org/10.1038/385336a0

101. Aly ASI, Matuschewski K (2005) A malarial cysteine protease is necessary for Plasmodium sporozoite egress from oocysts. J Exp Med 202:225–230. https://doi.org/10.1084/jem.20050545

102. Sultan AA, Thathy V, Frevert U, et al (1997) TRAP Is Necessary for Gliding Motility and Infectivity of Plasmodium Sporozoites. Cell 90:511–522. https://doi.org/10.1016/S0092-8674(00)80511-5

103. Wang Q, Fujioka H, Nussenzweig V (2005) Exit of Plasmodium Sporozoites from Oocysts Is an Active Process That Involves the Circumsporozoite Protein. PLOS Pathog 1:e9. https://doi.org/10.1371/journal.ppat.0010009

104. Lasonder E, Janse CJ, Gemert G-J van, et al (2008) Proteomic Profiling of Plasmodium Sporozoite Maturation Identifies New Proteins Essential for Parasite Development and Infectivity. PLOS Pathog 4:e1000195. https://doi.org/10.1371/journal.ppat.1000195

105. Sidjanski SP, Vanderberg JP, Sinnis P (1997) Anopheles stephensi salivary glands bear receptors for region I of the circumsporozoite protein of Plasmodium falciparum. Mol Biochem Parasitol 90:33–41. https://doi.org/10.1016/S0166-6851(97)00124-2

106. Sinden RE, Matuschewski K (2005) The Sporozoite. Mol Approaches Malar 169–190. https://doi.org/10.1128/9781555817558.ch9

107. Matuschewski K, Nunes AC, Nussenzweig V, Ménard R (2002) Plasmodium sporozoite invasion into insect and mammalian cells is directed by the same dual binding system. EMBO J 21:1597–1606. https://doi.org/10.1093/emboj/21.7.1597

108. Ghosh AK, Devenport M, Jethwaney D, et al (2009) Malaria Parasite Invasion of the Mosquito Salivary Gland Requires Interaction between the Plasmodium TRAP and the Anopheles Saglin Proteins. PLOS Pathog 5:e1000265. https://doi.org/10.1371/journal.ppat.1000265

109. Okulate MA, Kalume DE, Reddy R, et al (2007) Identification and molecular characterization of a novel protein Saglin as a target of monoclonal antibodies affecting salivary gland infectivity of Plasmodium sporozoites. Insect Mol Biol 16:711–722. https://doi.org/10.1111/j.1365-2583.2007.00765.x

110. Thompson J, Fernandez-Reyes D, Sharling L, et al (2007) Plasmodium cysteine repeat modular proteins 1–4: complex proteins with roles throughout the malaria parasite life cycle. Cell Microbiol 9:1466–1480. https://doi.org/10.1111/j.1462-5822.2006.00885.x

111. Kariu T, Yuda M, Yano K, Chinzei Y (2002) MAEBL Is Essential for Malarial Sporozoite Infection of the Mosquito Salivary Gland. J Exp Med 195:1317–1323. https://doi.org/10.1084/jem.20011876

112. Kappe SHI, Buscaglia CA, Bergman LW, et al (2004) Apicomplexan gliding motility and host cell invasion: overhauling the motor model. Trends Parasitol 20:13–16. https://doi.org/10.1016/j.pt.2003.10.011

113. Pimenta PF, Touray M, Miller L (1994) The Journey of Malaria Sporozoites in the Mosquito Salivary Gland. J Eukaryot Microbiol 41:608–624. https://doi.org/10.1111/j.1550-7408.1994.tb01523.x

114. Mahy BWJ (2004) Vector-borne diseases. In: Microbe-Vector Interact. Vector-Borne Dis. /core/books/microbevector-interactions-in-vectorborne-diseases/vectorborne-diseases/91DB7C46CCE13320BBA5CBCC37CDA140. Accessed 11 Jul 2019

115. Pennacchio F, Strand MR (2005) Evolution of developmental strategies in parasitic hymenoptera. Annu Rev Entomol 51:233–258. https://doi.org/10.1146/annurev.ento.51.110104.151029

116. Vega FE, Kaya HK (2012) Insect Pathology. Academic Press

117. Hillyer JF (2016) Insect immunology and hematopoiesis. Dev Comp Immunol 58:102–118. https://doi.org/10.1016/j.dci.2015.12.006

118. Hillyer JF, Strand MR (2014) Mosquito hemocyte-mediated immune responses. Curr Opin Insect Sci 3:14–21. https://doi.org/10.1016/j.cois.2014.07.002

119. Publication : USDA ARS. https://www.ars.usda.gov/research/publications/publication/?seqNo115=271050. Accessed 11 Jul 2019

120. Pedrini N, Crespo R, Juárez MP (2007) Biochemistry of insect epicuticle degradation by entomopathogenic fungi. Comp Biochem Physiol Part C Toxicol Pharmacol 146:124–137. https://doi.org/10.1016/j.cbpc.2006.08.003

121. Cirimotich CM, Dong Y, Clayton AM, et al (2011) Natural Microbe-Mediated Refractoriness to Plasmodium Infection in Anopheles gambiae. Science 332:855–858. https://doi.org/10.1126/science.1201618

122. McGreevy PB, Bryan JH, Oothuman P, Kolstrup N (1978) The lethal effects of the cibarial and pharyngeal armatures of mosquitoes on microfilariae. Trans R Soc Trop Med Hyg 72:361–368. https://doi.org/10.1016/0035-9203(78)90128-1

123. Siva-Jothy MT, Moret Y, Rolff J (2005) Insect Immunity: An Evolutionary Ecology Perspective. In: Simpson SJ (ed) Advances in Insect Physiology. Academic Press, pp 1–48

124. Pinto SB, Kafatos FC, Michel K (2008) The parasite invasion marker SRPN6 reduces sporozoite numbers in salivary glands of Anopheles gambiae. Cell Microbiol 10:891–898. https://doi.org/10.1111/j.1462-5822.2007.01091.x

125. Gupta L, Molina-Cruz A, Kumar S, et al (2009) The STAT Pathway Mediates Late-Phase Immunity against Plasmodium in the Mosquito Anopheles gambiae. Cell Host Microbe 5:498–507. https://doi.org/10.1016/j.chom.2009.04.003

126. Lim J, Gowda DC, Krishnegowda G, Luckhart S (2005) Induction of Nitric Oxide Synthase in Anopheles stephensi by Plasmodium falciparum: Mechanism of Signaling and the Role of Parasite Glycosylphosphatidylinositols. Infect Immun 73:2778–2789. https://doi.org/10.1128/IAI.73.5.2778-2789.2005

127. Luckhart S, Vodovotz Y, Cui L, Rosenberg R (1998) The mosquito Anopheles stephensi limits malaria parasite development with inducible synthesis of nitric oxide. Proc Natl Acad Sci 95:5700–5705

128. Waterhouse RM, Kriventseva EV, Meister S, et al (2007) Evolutionary Dynamics of Immune-Related Genes and Pathways in Disease-Vector Mosquitoes. Science 316:1738–1743. https://doi.org/10.1126/science.1139862

129. Lowenberger CA, Smartt CT, Bulet P, et al (1999) Insect immunity: molecular cloning, expression, and characterization of cDNAs and genomic DNA encoding three isoforms of insect defensin in Aedes aegypti. Insect Mol Biol 8:107–118

130. Richman AM, Bulet P, Hetru C, et al (1996) Inducible immune factors of the vector mosquito Anopheles gambiae: biochemical purification of a defensin antibacterial peptide and molecular cloning of preprodefensin cDNA. Insect Mol Biol 5:203–210

131. Warr E, Aguilar R, Dong Y, et al (2007) Spatial and sex-specific dissection of the Anopheles gambiae midgut transcriptome. BMC Genomics 8:37. https://doi.org/10.1186/1471-2164-8-37

132. Kim W, Koo H, Richman AM, et al (2004) Ectopic expression of a cecropin transgene in the human malaria vector mosquito Anopheles gambiae (Diptera: Culicidae): effects on susceptibility to Plasmodium. J Med Entomol 41:447–455. https://doi.org/10.1603/0022-2585-41.3.447

133. Christensen BM, Li J, Chen C-C, Nappi AJ (2005) Melanization immune responses in mosquito vectors. Trends Parasitol 21:192–199. https://doi.org/10.1016/j.pt.2005.02.007

134. Neafsey DE, Waterhouse RM, Abai MR, et al (2015) Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 Anopheles mosquitoes. Science 347:1258522. https://doi.org/10.1126/science.1258522

135. Ahmed A, Martín D, Manetti AGO, et al (1999) Genomic structure and ecdysone regulation of the prophenoloxidase 1 gene in the malaria vector Anopheles gambiae. Proc Natl Acad Sci 96:14795–14800. https://doi.org/10.1073/pnas.96.26.14795

136. Müller HM, Dimopoulos G, Blass C, Kafatos FC (1999) A hemocyte-like cell line established from the malaria vector Anopheles gambiae expresses six prophenoloxidase genes. J Biol Chem 274:11727–11735. https://doi.org/10.1074/jbc.274.17.11727

137. Osta MA, Christophides GK, Kafatos FC (2004) Effects of mosquito genes on Plasmodium development. Science 303:2030–2032. https://doi.org/10.1126/science.1091789

138. An C, Budd A, Kanost MR, Michel K (2011) Characterization of a regulatory unit that controls melanization and affects longevity of mosquitoes. Cell Mol Life Sci CMLS 68:1929–1939. https://doi.org/10.1007/s00018-010-0543-z

139. Baxter RHG, Chang C-I, Chelliah Y, et al (2007) Structural basis for conserved complement factor-like function in the antimalarial protein TEP1. Proc Natl Acad Sci U S A 104:11615–11620. https://doi.org/10.1073/pnas.0704967104

140. Levashina EA, Moita LF, Blandin S, et al (2001) Conserved Role of a Complement-like Protein in Phagocytosis Revealed by dsRNA Knockout in Cultured Cells of the Mosquito, Anopheles gambiae. Cell 104:709–718. https://doi.org/10.1016/S0092-8674(01)00267-7

141. Povelones M, Bhagavatula L, Yassine H, et al (2013) The CLIP-Domain Serine Protease Homolog SPCLIP1 Regulates Complement Recruitment to Microbial Surfaces in the Malaria Mosquito Anopheles gambiae. PLOS Pathog 9:e1003623. https://doi.org/10.1371/journal.ppat.1003623

142. Yassine H, Kamareddine L, Chamat S, et al (2014) A serine protease homolog negatively regulates TEP1 consumption in systemic infections of the malaria vector Anopheles gambiae. J Innate Immun 6:806–818. https://doi.org/10.1159/000363296

143. Riehle MM, Xu J, Lazzaro BP, et al (2008) Anopheles gambiae APL1 Is a Family of Variable LRR Proteins Required for Rel1-Mediated Protection from the Malaria Parasite, Plasmodium berghei. PLOS ONE 3:e3672. https://doi.org/10.1371/journal.pone.0003672

144. Fraiture M, Baxter RHG, Steinert S, et al (2009) Two Mosquito LRR Proteins Function as Complement Control Factors in the TEP1-Mediated Killing of Plasmodium. Cell Host Microbe 5:273–284. https://doi.org/10.1016/j.chom.2009.01.005

145. Bartholomay LC, Michel K (2018) Mosquito Immunobiology: The Intersection of Vector Health and Vector Competence. Annu Rev Entomol 63:145–167. https://doi.org/10.1146/annurev-ento-010715-023530

146. King JG, Hillyer JF (2013) Spatial and temporal in vivo analysis of circulating and sessile immune cells in mosquitoes: hemocyte mitosis following infection. BMC Biol 11:55. https://doi.org/10.1186/1741-7007-11-55

147. Bryant WB, Michel K (2014) Blood feeding induces hemocyte proliferation and activation in the African malaria mosquito, Anopheles gambiae Giles. J Exp Biol 217:1238–1245. https://doi.org/10.1242/jeb.094573

148. Castillo J, Brown MR, Strand MR (2011) Blood Feeding and Insulin-like Peptide 3 Stimulate Proliferation of Hemocytes in the Mosquito Aedes aegypti. PLOS Pathog 7:e1002274. https://doi.org/10.1371/journal.ppat.1002274

149. Bryant WB, Michel K (2016) Anopheles gambiae hemocytes exhibit transient states of activation. Dev Comp Immunol 55:119–129. https://doi.org/10.1016/j.dci.2015.10.020

150. King JG, Hillyer JF (2012) Infection-Induced Interaction between the Mosquito Circulatory and Immune Systems. PLoS Pathog 8:e1003058. https://doi.org/10.1371/journal.ppat.1003058

151. Michel K, Suwanchaichinda C, Morlais I, et al (2006) Increased melanizing activity in Anopheles gambiae does not affect development of Plasmodium falciparum. Proc Natl Acad Sci U S A 103:16858–16863. https://doi.org/10.1073/pnas.0608033103

152. Christensen BM, Forton KF (1986) Hemocyte-mediated melanization of microfilariae in Aedes aegypti. J Parasitol 72:220–225

153. Yassine H, Kamareddine L, Osta MA (2012) The Mosquito Melanization Response Is Implicated in Defense against the Entomopathogenic Fungus Beauveria bassiana. PLOS Pathog 8:e1003029. https://doi.org/10.1371/journal.ppat.1003029

154. Cheng G, Liu Y, Wang P, Xiao X (2016) Mosquito Defense Strategies against Viral Infection. Trends Parasitol 32:177–186. https://doi.org/10.1016/j.pt.2015.09.009

155. Parikh GR, Oliver JD, Bartholomay LC (2009) A haemocyte tropism for an arbovirus. J Gen Virol 90:292–296. https://doi.org/10.1099/vir.0.005116-0

156. Castillo JC, Ferreira ABB, Trisnadi N, Barillas-Mury C (2017) Activation of mosquito complement antiplasmodial response requires cellular immunity. Sci Immunol 2:eaal1505. https://doi.org/10.1126/sciimmunol.aal1505

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

158. Shen Z, Jacobs-Lorena M (1998) A Type I Peritrophic Matrix Protein from the Malaria VectorAnopheles gambiae Binds to Chitin CLONING, EXPRESSION, AND

CHARACTERIZATION. J Biol Chem 273:17665–17670. https://doi.org/10.1074/jbc.273.28.17665

159. Kumar S, Molina-Cruz A, Gupta L, et al (2010) A Peroxidase/Dual Oxidase System Modulates Midgut Epithelial Immunity in Anopheles gambiae. Science 327:1644–1648. https://doi.org/10.1126/science.1184008

160. Shen Z, Dimopoulos G, Kafatos FC, Jacobs-Lorena M (1999) A cell surface mucin specifically expressed in the midgut of the malaria mosquito Anopheles gambiae. Proc Natl Acad Sci 96:5610–5615. https://doi.org/10.1073/pnas.96.10.5610

161. Kumar S, Gupta L, Han YS, Barillas-Mury C (2004) Inducible Peroxidases Mediate Nitration of Anopheles Midgut Cells Undergoing Apoptosis in Response to Plasmodium Invasion. J Biol Chem 279:53475–53482. https://doi.org/10.1074/jbc.M409905200

162. Oliveira G de A, Lieberman J, Barillas-Mury C (2012) Epithelial Nitration by a Peroxidase/NOX5 System Mediates Mosquito Antiplasmodial Immunity. Science 335:856–859. https://doi.org/10.1126/science.1209678

163. Ramphul UN, Garver LS, Molina-Cruz A, et al (2015) Plasmodium falciparum evades mosquito immunity by disrupting JNK-mediated apoptosis of invaded midgut cells. Proc Natl Acad Sci 112:1273–1280. https://doi.org/10.1073/pnas.1423586112

164. Han YS, Barillas-Mury C (2002) Implications of Time Bomb model of ookinete invasion of midgut cells. Insect Biochem Mol Biol 32:1311–1316. https://doi.org/10.1016/S0965-1748(02)00093-0

165. Blandin S, Shiao S-H, Moita LF, et al (2004) Complement-Like Protein TEP1 Is a Determinant of Vectorial Capacity in the Malaria Vector Anopheles gambiae. Cell 116:661–670. https://doi.org/10.1016/S0092-8674(04)00173-4

166. Blandin SA, Marois E, Levashina EA (2008) Antimalarial Responses in Anopheles gambiae: From a Complement-like Protein to a Complement-like Pathway. Cell Host Microbe 3:364–374. https://doi.org/10.1016/j.chom.2008.05.007

167. Povelones M, Waterhouse RM, Kafatos FC, Christophides GK (2009) Leucine-Rich Repeat Protein Complex Activates Mosquito Complement in Defense Against Plasmodium Parasites. Science 324:258–261. https://doi.org/10.1126/science.1171400

168. Jaramillo-Gutierrez G, Rodrigues J, Ndikuyeze G, et al (2009) Mosquito immune responses and compatibility between Plasmodium parasites and anopheline mosquitoes. BMC Microbiol 9:154. https://doi.org/10.1186/1471-2180-9-154

169. Blandin SA, Wang-Sattler R, Lamacchia M, et al (2009) Dissecting the Genetic Basis of Resistance to Malaria Parasites in Anopheles gambiae. Science 326:147–150. https://doi.org/10.1126/science.1175241

170. Garver LS, de Almeida Oliveira G, Barillas-Mury C (2013) The JNK Pathway Is a Key Mediator of Anopheles gambiae Antiplasmodial Immunity. PLoS Pathog 9:e1003622. https://doi.org/10.1371/journal.ppat.1003622

171. Molina-Cruz A, DeJong RJ, Charles B, et al (2008) Reactive Oxygen Species Modulate Anopheles gambiae Immunity against Bacteria and Plasmodium. J Biol Chem 283:3217–3223. https://doi.org/10.1074/jbc.M705873200

172. Kumar S, Christophides GK, Cantera R, et al (2003) The role of reactive oxygen species on Plasmodium melanotic encapsulation in Anopheles gambiae. Proc Natl Acad Sci 100:14139–14144. https://doi.org/10.1073/pnas.2036262100

173. Oliveira JHM, Gonçalves RLS, Oliveira GA, et al (2011) Energy metabolism affects susceptibility of Anopheles gambiae mosquitoes to Plasmodium infection. Insect Biochem Mol Biol 41:349–355. https://doi.org/10.1016/j.ibmb.2011.02.001

174. Mitri C, Jacques J-C, Thiery I, et al (2009) Fine Pathogen Discrimination within the APL1 Gene Family Protects Anopheles gambiae against Human and Rodent Malaria Species. PLoS Pathog 5:e1000576. https://doi.org/10.1371/journal.ppat.1000576

175. Garver LS, Dong Y, Dimopoulos G (2009) Caspar Controls Resistance to Plasmodium falciparum in Diverse Anopheline Species. PLoS Pathog 5:e1000335. https://doi.org/10.1371/journal.ppat.1000335

176. Cohuet A, Osta MA, Morlais I, et al (2006) Anopheles and Plasmodium: from laboratory models to natural systems in the field. EMBO Rep 7:1285–1289. https://doi.org/10.1038/sj.embor.7400831

177. Molina-Cruz A, Garver LS, Alabaster A, et al (2013) The Human Malaria Parasite Pfs47 Gene Mediates Evasion of the Mosquito Immune System. Science 340:984–987. https://doi.org/10.1126/science.1235264

178. Molina-Cruz A, DeJong RJ, Ortega C, et al (2012) Some strains of Plasmodium falciparum, a human malaria parasite, evade the complement-like system of Anopheles gambiae mosquitoes. Proc Natl Acad Sci 109:E1957–E1962. https://doi.org/10.1073/pnas.1121183109

179. Manske M, Miotto O, Campino S, et al (2012) Analysis of Plasmodium falciparum diversity in natural infections by deep sequencing. Nature 487:375–379. https://doi.org/10.1038/nature11174

180. Anthony TG, Polley SD, Vogler AP, Conway DJ (2007) Evidence of non-neutral polymorphism in Plasmodium falciparum gamete surface protein genes Pfs47 and Pfs48/45. Mol Biochem Parasitol 156:117–123. https://doi.org/10.1016/j.molbiopara.2007.07.008

181. Molina-Cruz A, Canepa GE, Barillas-Mury C (2017) Plasmodium P47: a key gene for malaria transmission by mosquito vectors. Curr Opin Microbiol 40:168–174. https://doi.org/10.1016/j.mib.2017.11.029

182. Dong Y, Dimopoulos G (2009) Anopheles Fibrinogen-related Proteins Provide Expanded Pattern Recognition Capacity against Bacteria and Malaria Parasites. J Biol Chem 284:9835–9844. https://doi.org/10.1074/jbc.M807084200

183. Ukegbu CV, Giorgalli M, Yassine H, et al (2017) Plasmodium berghei P47 is essential for ookinete protection from the Anopheles gambiae complement-like response. Sci Rep 7:6026. https://doi.org/10.1038/s41598-017-05917-6

184. Smith RC, Jacobs-Lorena M (2015) Malaria parasite Pfs47 disrupts JNK signaling to escape mosquito immunity. Proc Natl Acad Sci 112:1250–1251. https://doi.org/10.1073/pnas.1424227112

185. Smith RC, Barillas-Mury C, Jacobs-Lorena M (2015) Hemocyte differentiation mediates the mosquito late-phase immune response against Plasmodium in Anopheles gambiae. Proc Natl Acad Sci U S A 112:E3412-3420. https://doi.org/10.1073/pnas.1420078112

186. Smith RC, Eappen AG, Radtke AJ, Jacobs-Lorena M (2012) Regulation of Anti-Plasmodium Immunity by a LITAF-like Transcription Factor in the Malaria Vector Anopheles gambiae. PLoS Pathog 8:e1002965. https://doi.org/10.1371/journal.ppat.1002965

187. Nsango SE, Pompon J, Xie T, et al (2013) AP-1/Fos-TGase2 Axis Mediates Wounding-induced Plasmodium falciparum Killing in Anopheles gambiae. J Biol Chem 288:16145–16154. https://doi.org/10.1074/jbc.M112.443267

188. Lemaitre B, Hoffmann J (2007) The Host Defense of Drosophila melanogaster. Annu Rev Immunol 25:697–743. https://doi.org/10.1146/annurev.immunol.25.022106.141615

189. Imler J-L (2014) Overview of Drosophila immunity: A historical perspective. Dev Comp Immunol 42:3–15. https://doi.org/10.1016/j.dci.2013.08.018

190. Lawrence PO (2008) Hemocytes of Insects: Their Morphology and Function. In: Capinera JL (ed) Encyclopedia of Entomology. Springer Netherlands, pp 1787–1790

191. Severo MS, Landry JJM, Lindquist RL, et al (2018) Unbiased classification of mosquito blood cells by single-cell genomics and high-content imaging. Proc Natl Acad Sci 115:E7568–E7577. https://doi.org/10.1073/pnas.1803062115

192. Kwon H, Smith RC (2019) Chemical depletion of phagocytic immune cells in Anopheles gambiae reveals dual roles of mosquito hemocytes in anti-Plasmodium immunity. Proc Natl Acad Sci 116:14119–14128. https://doi.org/10.1073/pnas.1900147116

193. Rodrigues J, Brayner FA, Alves LC, et al (2010) Hemocyte Differentiation Mediates Innate Immune Memory in Anopheles gambiae Mosquitoes. Science 329:1353–1355. https://doi.org/10.1126/science.1190689

194. Ramirez JL, de Almeida Oliveira G, Calvo E, et al (2015) A mosquito lipoxin/lipocalin complex mediates innate immune priming in Anopheles gambiae. Nat Commun 6:. https://doi.org/10.1038/ncomms8403

195. Simões ML, Dimopoulos G A mosquito mediator of parasite-induced immune priming. Trends Parasitol. https://doi.org/10.1016/j.pt.2015.07.004

196. Stanley D (2006) PROSTAGLANDINS AND OTHER EICOSANOIDS IN INSECTS: Biological Significance. Annu Rev Entomol 51:25–44. https://doi.org/10.1146/annurev.ento.51.110104.151021

197. Hoxmeier JC, Thompson BD, Broeckling CD, et al (2015) Analysis of the metabolome of Anopheles gambiae mosquito after exposure to Mycobacterium ulcerans. Sci Rep 5:9242. https://doi.org/10.1038/srep09242

198. Hwang J, Park Y, Kim Y, et al (2013) AN ENTOMOPATHOGENIC BACTERIUM, Xenorhabdus nematophila, SUPPRESSES EXPRESSION OF ANTIMICROBIAL PEPTIDES CONTROLLED BY TOLL AND IMD PATHWAYS BY BLOCKING EICOSANOID BIOSYNTHESIS. Arch Insect Biochem Physiol 83:151–169. https://doi.org/10.1002/arch.21103

199. Moreno-García M, Recio-Tótoro B, Claudio-Piedras F, Lanz-Mendoza H (2014) Injury and immune response: applying the danger theory to mosquitoes. Plant Biot Interact 5:451. https://doi.org/10.3389/fpls.2014.00451

200. Choi Y-J, Fuchs JF, Mayhew GF, et al (2012) Tissue-enriched expression profiles in Aedes aegypti identify hemocyte-specific transcriptome responses to infection. Insect Biochem Mol Biol 42:729–738. https://doi.org/10.1016/j.ibmb.2012.06.005

201. Mestas J, Hughes CCW (2004) Of Mice and Not Men: Differences between Mouse and Human Immunology. J Immunol 172:2731–2738. https://doi.org/10.4049/jimmunol.172.5.2731

202. Pinto SB, Lombardo F, Koutsos AC, et al (2009) Discovery of Plasmodium modulators by genome-wide analysis of circulating hemocytes in Anopheles gambiae. Proc Natl Acad Sci 106:21270–21275. https://doi.org/10.1073/pnas.0909463106

203. Baton LA, Robertson A, Warr E, et al (2009) Genome-wide transcriptomic profiling of Anopheles gambiae hemocytes reveals pathogen-specific signatures upon bacterial challenge and Plasmodium berghei infection. BMC Genomics 10:257. https://doi.org/10.1186/1471-2164-10-257

204. Thomas T, De TD, Sharma P, et al (2016) Hemocytome: deep sequencing analysis of mosquito blood cells in Indian malarial vector Anopheles stephensi. Gene 585:177–190. https://doi.org/10.1016/j.gene.2016.02.031

205. Lombardo F, Ghani Y, Kafatos FC, Christophides GK (2013) Comprehensive Genetic Dissection of the Hemocyte Immune Response in the Malaria Mosquito Anopheles gambiae. PLoS Pathog 9:e1003145. https://doi.org/10.1371/journal.ppat.1003145

206. Guttman M, Garber M, Levin JZ, et al (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nat Biotechnol 28:503–510. https://doi.org/10.1038/nbt.1633

207. Treutlein B, Brownfield DG, Wu AR, et al (2014) Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 509:371–375. https://doi.org/10.1038/nature13173

208. Shalek AK, Satija R, Shuga J, et al (2014) Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. Nature 510:363–369. https://doi.org/10.1038/nature13437

209. Kolodziejczyk AA, Kim JK, Svensson V, et al (2015) The Technology and Biology of Single-Cell RNA Sequencing. Mol Cell 58:610–620. https://doi.org/10.1016/j.molcel.2015.04.005

210. Stegle O, Teichmann SA, Marioni JC (2015) Computational and analytical challenges in single-cell transcriptomics. Nat Rev Genet 16:133–145. https://doi.org/10.1038/nrg3833

211. Trapnell C, Cacchiarelli D, Grimsby J, et al (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32:381–386. https://doi.org/10.1038/nbt.2859

212. Bengtsson M, Ståhlberg A, Rorsman P, Kubista M (2005) Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res 15:1388–1392. https://doi.org/10.1101/gr.3820805

213. Chang HH, Hemberg M, Barahona M, et al (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. Nature 453:544–547. https://doi.org/10.1038/nature06965

214. Chen X, Teichmann SA, Meyer KB (2018) From Tissues to Cell Types and Back: Single-Cell Gene Expression Analysis of Tissue Architecture. Annu Rev Biomed Data Sci 1:29–51. https://doi.org/10.1146/annurev-biodatasci-080917-013452

215. Picelli S (2017) Single-cell RNA-sequencing: The future of genome biology is now. RNA Biol 14:637–650. https://doi.org/10.1080/15476286.2016.1201618

216. Tang F, Barbacioru C, Wang Y, et al (2009) mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382. https://doi.org/10.1038/nmeth.1315

217. Svensson V, Vento-Tormo R, Teichmann SA (2018) Exponential scaling of single-cell RNA-seq in the past decade. Nat Protoc 13:599–604. https://doi.org/10.1038/nprot.2017.149

218. Eberwine J, Kim J (2015) Cellular Deconstruction: Finding Meaning in Individual Cell Variation. Trends Cell Biol 25:569–578. https://doi.org/10.1016/j.tcb.2015.07.004

219. Islam S, Zeisel A, Joost S, et al (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods 11:163–166. https://doi.org/10.1038/nmeth.2772

220. Depledge DP, Srinivas KP, Sadaoka T, et al (2019) Direct RNA sequencing on nanopore arrays redefines the transcriptional complexity of a viral pathogen. Nat Commun 10:754. https://doi.org/10.1038/s41467-019-08734-9

221. Islam S, Kjällquist U, Moliner A, et al (2011) Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. Genome Res 21:1160–1167. https://doi.org/10.1101/gr.110882.110

222. Brennecke P, Anders S, Kim JK, et al (2013) Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods 10:1093–1095. https://doi.org/10.1038/nmeth.2645

223. Jaitin DA, Kenigsberg E, Keren-Shaul H, et al (2014) Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. Science 343:776–779. https://doi.org/10.1126/science.1247651

224. Klein AM, Mazutis L, Akartuna I, et al (2015) Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells. Cell 161:1187–1201. https://doi.org/10.1016/j.cell.2015.04.044

225. Macosko EZ, Basu A, Satija R, et al (2015) Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell 161:1202–1214. https://doi.org/10.1016/j.cell.2015.05.002

226. Bose S, Wan Z, Carr A, et al (2015) Scalable microfluidics for single-cell RNA printing and sequencing. Genome Biol 16:120. https://doi.org/10.1186/s13059-015-0684-3

227. Cao J, Packer JS, Ramani V, et al (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. Science 357:661–667. https://doi.org/10.1126/science.aam8940

228. Rosenberg AB, Roco CM, Muscat RA, et al (2018) Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. Science 360:176–182. https://doi.org/10.1126/science.aam8999

229. Gross A, Schoendube J, Zimmermann S, et al (2015) Technologies for Single-Cell Isolation. Int J Mol Sci 16:16897–16919. https://doi.org/10.3390/ijms160816897

230. Hwang B, Lee JH, Bang D (2018) Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50:96. https://doi.org/10.1038/s12276-018-0071-8

231. Chen G, Ning B, Shi T (2019) Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front Genet 10:. https://doi.org/10.3389/fgene.2019.00317

232. Nguyen QH, Pervolarakis N, Nee K, Kessenbrock K (2018) Experimental Considerations for Single-Cell RNA Sequencing Approaches. Front Cell Dev Biol 6:. https://doi.org/10.3389/fcell.2018.00108

233. Tung P-Y, Blischak JD, Hsiao CJ, et al (2017) Batch effects and the effective design of single-cell gene expression studies. Sci Rep 7:. https://doi.org/10.1038/srep39921

234. van den Brink SC, Sage F, Vértesy Á, et al (2017) Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat Methods 14:935–936. https://doi.org/10.1038/nmeth.4437

235. Attar M, Sharma E, Li S, et al (2018) A practical solution for preserving single cells for RNA sequencing. Sci Rep 8:2151. https://doi.org/10.1038/s41598-018-20372-7

236. Chen J, Cheung F, Shi R, et al (2018) PBMC fixation and processing for Chromium single-cell RNA sequencing. J Transl Med 16:. https://doi.org/10.1186/s12967-018-1578-4

237. Ziegenhain C, Vieth B, Parekh S, et al (2017) Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell 65:631-643.e4. https://doi.org/10.1016/j.molcel.2017.01.023

238. Sasagawa Y, Nikaido I, Hayashi T, et al (2013) Quartz-Seq: a highly reproducible and sensitive single-cell RNA sequencing method, reveals non-genetic gene-expression heterogeneity. Genome Biol 14:R31. https://doi.org/10.1186/gb-2013-14-4-r31

239. Picelli S, Faridani OR, Björklund ÅK, et al (2014) Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9:171–181. https://doi.org/10.1038/nprot.2014.006

240. Picelli S, Björklund ÅK, Faridani OR, et al (2013) Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat Methods 10:1096–1098. https://doi.org/10.1038/nmeth.2639

241. Hashimshony T, Wagner F, Sher N, Yanai I (2012) CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell Rep 2:666–673. https://doi.org/10.1016/j.celrep.2012.08.003

242. Kivioja T, Vähärautio A, Karlsson K, et al (2012) Counting absolute numbers of molecules using unique molecular identifiers. Nat Methods 9:72–74. https://doi.org/10.1038/nmeth.1778

243. Hashimshony T, Senderovich N, Avital G, et al (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol 17:77. https://doi.org/10.1186/s13059-016-0938-8

244. Islam S, Kjällquist U, Moliner A, et al (2012) Highly multiplexed and strand-specific single-cell RNA 5’ end sequencing. Nat Protoc 7:813–828. https://doi.org/10.1038/nprot.2012.022

245. Zhang X, Li T, Liu F, et al (2019) Comparative Analysis of Droplet-Based Ultra-High-Throughput Single-Cell RNA-Seq Systems. Mol Cell 73:130-142.e5. https://doi.org/10.1016/j.molcel.2018.10.020

246. Zheng GXY, Terry JM, Belgrader P, et al (2017) Massively parallel digital transcriptional profiling of single cells. Nat Commun 8:14049. https://doi.org/10.1038/ncomms14049

247. Single-Cell RNA Sequencing with Drop-Seq | Springer Nature Experiments. https://experiments.springernature.com/articles/10.1007/978-1-4939-9240-9_6. Accessed 12 Jul 2019

248. Gierahn TM, Ii MHW, Hughes TK, et al (2017) Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat Methods 14:395–398. https://doi.org/10.1038/nmeth.4179

249. Seq-Well: A Sample-Efficient, Portable Picowell Platform for Massively Parallel Single-Cell RNA Sequencing | Springer Nature Experiments. https://experiments.springernature.com/articles/10.1007/978-1-4939-9240-9_8. Accessed 12 Jul 2019

250. Cao J, Spielmann M, Qiu X, et al (2019) The single-cell transcriptional landscape of mammalian organogenesis. Nature 566:496–502. https://doi.org/10.1038/s41586-019-0969-x

251. Cusanovich DA, Daza R, Adey A, et al (2015) Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. Science 348:910–914. https://doi.org/10.1126/science.aab1601

252. Vitak SA, Torkenczy KA, Rosenkrantz JL, et al (2017) Sequencing thousands of single-cell genomes with combinatorial indexing. Nat Methods 14:302–308. https://doi.org/10.1038/nmeth.4154

253. Ramani V, Deng X, Qiu R, et al (2017) Massively multiplex single-cell Hi-C. Nat Methods 14:263–266. https://doi.org/10.1038/nmeth.4155

254. Mulqueen RM, Pokholok D, Norberg SJ, et al (2018) Highly scalable generation of DNA methylation profiles in single cells. Nat Biotechnol 36:428–431. https://doi.org/10.1038/nbt.4112

255. Pollen AA, Nowakowski TJ, Shuga J, et al (2014) Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat Biotechnol 32:1053–1058. https://doi.org/10.1038/nbt.2967

256. Satija R, Farrell JA, Gennert D, et al (2015) Spatial reconstruction of single-cell gene expression data. Nat Biotechnol 33:495–502. https://doi.org/10.1038/nbt.3192

257. Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. Genome Biol 19:15. https://doi.org/10.1186/s13059-017-1382-0

258. McCarthy DJ, Campbell KR, Lun ATL, Wills QF (2017) Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinforma Oxf Engl 33:1179–1186. https://doi.org/10.1093/bioinformatics/btw777

259. Bray NL, Pimentel H, Melsted P, Pachter L (2016) Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol 34:525–527. https://doi.org/10.1038/nbt.3519

260. Patro R, Duggal G, Love MI, et al (2017) Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 14:417–419. https://doi.org/10.1038/nmeth.4197

261. Kim D, Pertea G, Trapnell C, et al (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36. https://doi.org/10.1186/gb-2013-14-4-r36

262. Dobin A, Gingeras TR (2015) Mapping RNA-seq Reads with STAR. Curr Protoc Bioinforma 51:11.14.1-19. https://doi.org/10.1002/0471250953.bi1114s51

263. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nat Methods 12:357–360. https://doi.org/10.1038/nmeth.3317

264. Engström PG, Steijger T, Sipos B, et al (2013) Systematic evaluation of spliced alignment programs for RNA-seq data. Nat Methods 10:1185–1191. https://doi.org/10.1038/nmeth.2722

265. Everaert C, Luypaert M, Maag JLV, et al (2017) Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. Sci Rep 7:1559. https://doi.org/10.1038/s41598-017-01617-3

266. Teng M, Love MI, Davis CA, et al (2016) A benchmark for RNA-seq quantification pipelines. Genome Biol 17:74. https://doi.org/10.1186/s13059-016-0940-1

267. Huang M, Wang J, Torre E, et al (2018) SAVER: gene expression recovery for single-cell RNA sequencing. Nat Methods 15:539. https://doi.org/10.1038/s41592-018-0033-z

268. Jiang L, Schlesinger F, Davis CA, et al (2011) Synthetic spike-in standards for RNA-seq experiments. Genome Res 21:1543–1551. https://doi.org/10.1101/gr.121095.111

269. Ilicic T, Kim JK, Kolodziejczyk AA, et al (2016) Classification of low quality cells from single-cell RNA-seq data. Genome Biol 17:. https://doi.org/10.1186/s13059-016-0888-1

270. Vallejos CA, Risso D, Scialdone A, et al (2017) Normalizing single-cell RNA sequencing data: challenges and opportunities. Nat Methods 14:565–571. https://doi.org/10.1038/nmeth.4292

271. Bacher R, Kendziorski C (2016) Design and computational analysis of single-cell RNA-sequencing experiments. Genome Biol 17:63. https://doi.org/10.1186/s13059-016-0927-y

272. Katayama S, Töhönen V, Linnarsson S, Kere J (2013) SAMstrt: statistical test for differential expression in single-cell transcriptome with spike-in normalization. Bioinforma Oxf Engl 29:2943–2945. https://doi.org/10.1093/bioinformatics/btt511

273. Buettner F, Natarajan KN, Casale FP, et al (2015) Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. Nat Biotechnol 33:155–160. https://doi.org/10.1038/nbt.3102

274. Bacher R, Chu L-F, Leng N, et al (2017) SCnorm: robust normalization of single-cell RNA-seq data. Nat Methods 14:584–586. https://doi.org/10.1038/nmeth.4263

275. Hafemeister C, Satija R (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. bioRxiv 576827. https://doi.org/10.1101/576827

276. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8:118–127. https://doi.org/10.1093/biostatistics/kxj037

277. Butler A, Hoffman P, Smibert P, et al (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 36:411–420. https://doi.org/10.1038/nbt.4096

278. Haghverdi L, Lun ATL, Morgan MD, Marioni JC (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol 36:421–427. https://doi.org/10.1038/nbt.4091

279. Finak G, McDavid A, Yajima M, et al (2015) MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol 16:278. https://doi.org/10.1186/s13059-015-0844-5

280. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

281. Ritchie ME, Phipson B, Wu D, et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47–e47. https://doi.org/10.1093/nar/gkv007

282. Büttner M, Miao Z, Wolf FA, et al (2019) A test metric for assessing single-cell RNA-seq batch correction. Nat Methods 16:43. https://doi.org/10.1038/s41592-018-0254-1

283. Maaten L van der, Hinton G (2008) Visualizing Data using t-SNE. J Mach Learn Res 9:2579–2605

284. Becht E, McInnes L, Healy J, et al (2018) Dimensionality reduction for visualizing single-cell data using UMAP. Nat Biotechnol. https://doi.org/10.1038/nbt.4314

285. Ding J, Condon A, Shah SP (2018) Interpretable dimensionality reduction of single cell transcriptome data with deep generative models. Nat Commun 9:2002. https://doi.org/10.1038/s41467-018-04368-5

286. Andrews TS, Hemberg M (2018) M3Drop: Dropout-based feature selection for scRNASeq. Bioinforma Oxf Engl. https://doi.org/10.1093/bioinformatics/bty1044

287. Andrews TS, Hemberg M (2018) Identifying cell populations with scRNASeq. Mol Aspects Med 59:114–122. https://doi.org/10.1016/j.mam.2017.07.002

288. Soneson C, Robinson MD (2018) Bias, robustness and scalability in single-cell differential expression analysis. Nat Methods 15:255–261. https://doi.org/10.1038/nmeth.4612

289. Kiselev VY, Kirschner K, Schaub MT, et al (2016) SC3 - consensus clustering of single-cell RNA-Seq data. bioRxiv 036558. https://doi.org/10.1101/036558

290. Seyednasrollah F, Rantanen K, Jaakkola P, Elo LL (2016) ROTS: reproducible RNA-seq biomarker detector-prognostic markers for clear cell renal cell cancer. Nucleic Acids Res 44:e1. https://doi.org/10.1093/nar/gkv806

291. Angerer P, Haghverdi L, Büttner M, et al (2016) destiny: diffusion maps for large-scale single-cell data in R. Bioinforma Oxf Engl 32:1241–1243. https://doi.org/10.1093/bioinformatics/btv715

292. Chen L, Zheng S (2018) BCseq: accurate single cell RNA-seq quantification with bias correction. Nucleic Acids Res 46:e82. https://doi.org/10.1093/nar/gky308

293. Xu C, Su Z (2015) Identification of cell types from single-cell transcriptomes using a novel clustering method. Bioinforma Oxf Engl 31:1974–1980. https://doi.org/10.1093/bioinformatics/btv088

294. Kharchenko PV, Silberstein L, Scadden DT (2014) Bayesian approach to single-cell differential expression analysis. Nat Methods 11:740–742. https://doi.org/10.1038/nmeth.2967

295. Grün D, Lyubimova A, Kester L, et al (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. Nature 525:251–255. https://doi.org/10.1038/nature14966

296. Miao Z, Deng K, Wang X, Zhang X (2018) DEsingle for detecting three types of differential expression in single-cell RNA-seq data. Bioinforma Oxf Engl 34:3223–3224. https://doi.org/10.1093/bioinformatics/bty332

297. Marco E, Karp RL, Guo G, et al (2014) Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proc Natl Acad Sci U S A 111:E5643-5650. https://doi.org/10.1073/pnas.1408993111

298. Qiu X, Hill A, Packer J, et al (2017) Single-cell mRNA quantification and differential analysis with Census. Nat Methods 14:309–315. https://doi.org/10.1038/nmeth.4150

299. Zeisel A, Muñoz-Manchado AB, Codeluppi S, et al (2015) Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science 347:1138–1142. https://doi.org/10.1126/science.aaa1934

300. Delmans M, Hemberg M (2016) Discrete distributional differential expression (D3E)--a tool for gene expression analysis of single-cell RNA-seq data. BMC Bioinformatics 17:110. https://doi.org/10.1186/s12859-016-0944-6

301. Fan J, Salathia N, Liu R, et al (2016) Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. Nat Methods 13:241–244. https://doi.org/10.1038/nmeth.3734

302. Vu TN, Wills QF, Kalari KR, et al (2016) Beta-Poisson model for single-cell RNA-seq data analyses. Bioinforma Oxf Engl 32:2128–2135. https://doi.org/10.1093/bioinformatics/btw202

303. Lin P, Troup M, Ho JWK (2017) CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. Genome Biol 18:59. https://doi.org/10.1186/s13059-017-1188-0

304. Žurauskienė J, Yau C (2016) pcaReduce: hierarchical clustering of single cell transcriptional profiles. BMC Bioinformatics 17:140. https://doi.org/10.1186/s12859-016-0984-y

305. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Oxf Engl 26:139–140. https://doi.org/10.1093/bioinformatics/btp616

306. Ji Z, Ji H (2016) TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis. Nucleic Acids Res 44:e117. https://doi.org/10.1093/nar/gkw430

307. Pierson E, Yau C (2015) ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. Genome Biol 16:241. https://doi.org/10.1186/s13059-015-0805-z

308. Frazee AC, Pertea G, Jaffe AE, et al (2015) Ballgown bridges the gap between transcriptome assembly and expression analysis. Nat Biotechnol 33:243–246. https://doi.org/10.1038/nbt.3172

309. Street K, Risso D, Fletcher RB, et al (2018) Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics 19:477. https://doi.org/10.1186/s12864-018-4772-0

310. Lummertz da Rocha E, Rowe RG, Lundin V, et al (2018) Reconstruction of complex single-cell trajectories using CellRouter. Nat Commun 9:892. https://doi.org/10.1038/s41467-018-03214-y

311. Saelens W, Cannoodt R, Todorov H, Saeys Y (2019) A comparison of single-cell trajectory inference methods. Nat Biotechnol 37:547–554. https://doi.org/10.1038/s41587-019-0071-9

312. Aibar S, González-Blas CB, Moerman T, et al (2017) SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14:1083–1086. https://doi.org/10.1038/nmeth.4463

313. Chan TE, Stumpf MPH, Babtie AC (2017) Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. Cell Syst 5:251-267.e3. https://doi.org/10.1016/j.cels.2017.08.014

314. Wang ET, Sandberg R, Luo S, et al (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470–476. https://doi.org/10.1038/nature07509

315. Welch JD, Hu Y, Prins JF (2016) Robust detection of alternative splicing in a population of single cells. Nucleic Acids Res 44:e73. https://doi.org/10.1093/nar/gkv1525

316. Huang Y, Sanguinetti G (2017) BRIE: transcriptome-wide splicing quantification in single cells. Genome Biol 18:123. https://doi.org/10.1186/s13059-017-1248-5

317. Song Y, Botvinnik OB, Lovci MT, et al (2017) Single-Cell Alternative Splicing Analysis with Expedition Reveals Splicing Dynamics during Neuron Differentiation. Mol Cell 67:148-161.e5. https://doi.org/10.1016/j.molcel.2017.06.003

318. See P, Lum J, Chen J, Ginhoux F (2018) A Single-Cell Sequencing Guide for Immunologists. Front Immunol 9:. https://doi.org/10.3389/fimmu.2018.02425

319. Franke-Fayard B, Trueman H, Ramesar J, et al (2004) A Plasmodium berghei reference line that constitutively expresses GFP at a high level throughout the complete life cycle. Mol Biochem Parasitol 137:23–33. https://doi.org/10.1016/j.molbiopara.2004.04.007

320. McCarthy D, Wills Q, Campbell K scater: Single-cell analysis toolkit for gene expression data in R. In: R Package Version 1113. https://github.com/davismcc/scater. Accessed 8 Sep 2016

321. Hartigan JA, Wong MA (1979) Algorithm AS 136: A K-Means Clustering Algorithm. J R Stat Soc Ser C Appl Stat 28:100–108. https://doi.org/10.2307/2346830

322. Volohonsky G, Terenzi O, Soichot J, et al (2015) Tools for Anopheles gambiae Transgenesis. G3 GenesGenomesGenetics 5:1151–1163. https://doi.org/10.1534/g3.115.016808

323. Sandiford SL, Dong Y, Pike A, et al (2015) Cytoplasmic Actin Is an Extracellular Insect Immune Factor which Is Secreted upon Immune Challenge and Mediates Phagocytosis and Direct Killing of Bacteria, and Is a Plasmodium Antagonist. PLOS Pathog 11:e1004631. https://doi.org/10.1371/journal.ppat.1004631

324. Lombardo F, Christophides GK (2016) Novel factors of Anopheles gambiae haemocyte immune response to Plasmodium berghei infection. Parasit Vectors 9:. https://doi.org/10.1186/s13071-016-1359-y

325. Robert Hooke (1665) Micrographia, or some physiological descriptions of minute bodies made by magnifying glasses, with observations and inquiries thereupon. By R. Hooke. Printed by JoMartyn, and JaAllestry, printers to the Royal Society, London

326. Hyman AH, Simons K (2011) The new cell biology: Beyond HeLa cells. Nature 480:34. https://doi.org/10.1038/480034a

327. Reynolds A (2008) Amoebae as Exemplary Cells: The Protean Nature of an Elementary Organism. J Hist Biol 41:307–337. https://doi.org/10.1007/s10739-007-9142-8

328. Fontana F (1781) Traité sur le vénin de la vipère, sur les poisons américaines

329. Schleiden: Arch Anat Physiol - Google Scholar. https://scholar.google.com/scholar_lookup?title=&journal=Arch.%20Anat.%20Physiol.%20Wiss.%20Med.&volume=13&pages=137-176&publication_year=1838&author=Schleiden%2CMJ. Accessed 24 Jul 2019

330. Harris H (2000) The Birth of the Cell. Yale University Press

331. Brown R (1833) XXXV. On the Organs and Mode of Fecundation in Orchideæ and Asclepiadeæ. Trans Linn Soc Lond 16:685–738. https://doi.org/10.1111/j.1095-8339.1829.tb00158.x

332. Russell RJ (1984) The growth of biological thought: Diversity, evolution, and inheritance. Ethol Sociobiol 5:63–64. https://doi.org/10.1016/0162-3095(84)90038-4

333. Mazzarello P (1999) A unifying concept: the history of cell theory. Nat Cell Biol 1:E13. https://doi.org/10.1038/8964

334. GOLGI C (1898) Intorno all struttura delle cellule nervose. Arch Ital Biol 30:60–71

335. Ramón y Cajal S (1909) Histologie du système nerveux de l'homme & des vertébrés. Paris : Maloine

336. Mazzarello P (1996) La struttura nascosta: La vita di Camillo Golgi. Cisalpino, Bologna

337. Trapnell C (2015) Defining cell types and states with single-cell genomics. Genome Res 25:1491–1498. https://doi.org/10.1101/gr.190595.115

338. Giraldo-Calderón GI, Emrich SJ, MacCallum RM, et al (2015) VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic Acids Res 43:D707-713. https://doi.org/10.1093/nar/gku1117

339. Stuart T, Butler A, Hoffman P, et al (2019) Comprehensive Integration of Single-Cell Data. Cell 177:1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

340. Haghverdi L, Büttner M, Wolf FA, et al (2016) Diffusion pseudotime robustly reconstructs lineage branching. Nat Methods 13:845–848. https://doi.org/10.1038/nmeth.3971

341. Qiu X, Mao Q, Tang Y, et al (2017) Reversed graph embedding resolves complex single-cell trajectories. Nat Methods 14:979–982. https://doi.org/10.1038/nmeth.4402

342. Manno GL, Soldatov R, Zeisel A, et al (2018) RNA velocity of single cells. Nature 560:494. https://doi.org/10.1038/s41586-018-0414-6

343. (2019) Stochastic Single Cell RNA Velocity. Contribute to theislab/scvelo development by creating an account on GitHub. Theis Lab

344. Wolf FA, Hamey FK, Plass M, et al (2019) PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. Genome Biol 20:59. https://doi.org/10.1186/s13059-019-1663-x

345. trajectory-based differential expression analysis for sequencing data. https://statomics.github.io/tradeSeq/index.html. Accessed 7 Sep 2019

346. Raudvere U, Kolberg L, Kuzmin I, et al (2019) g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res 47:W191–W198. https://doi.org/10.1093/nar/gkz369

347. Ribeiro JMC, Topalis P, Louis C (2004) AnoXcel: an Anopheles gambiae protein database. Insect Mol Biol 13:449–457. https://doi.org/10.1111/j.0962-1075.2004.00503.x

348. Smith RC, King JG, Tao D, et al (2016) Molecular Profiling of Phagocytic Immune Cells in Anopheles gambiae Reveals Integral Roles for Hemocytes in Mosquito Innate Immunity. Mol Cell Proteomics MCP 15:3373–3387. https://doi.org/10.1074/mcp.M116.060723

349. Zhao YO, Kurscheid S, Zhang Y, et al (2012) Enhanced survival of Plasmodium-infected mosquitoes during starvation. PloS One 7:e40556. https://doi.org/10.1371/journal.pone.0040556

350. Kramerova IA, Kramerov AA (1999) Mucinoprotein is a universal constituent of stable intercellular bridges in Drosophila melanogaster germ line and somatic cells. Dev Dyn Off Publ Am Assoc Anat 216:349–360. https://doi.org/10.1002/(SICI)1097-0177(199912)216:4/5<349::AID-DVDY4>3.0.CO;2-X

351. Smiley ST, King JA, Hancock WW (2001) Fibrinogen stimulates macrophage chemokine secretion through toll-like receptor 4. J Immunol Baltim Md 1950 167:2887–2894. https://doi.org/10.4049/jimmunol.167.5.2887

352. Dong Y, Manfredini F, Dimopoulos G (2009) Implication of the Mosquito Midgut Microbiota in the Defense against Malaria Parasites. PLOS Pathog 5:e1000423. https://doi.org/10.1371/journal.ppat.1000423

353. Volz J, Müller H-M, Zdanowicz A, et al (2006) A genetic module regulates the melanization response of Anopheles to Plasmodium. Cell Microbiol 8:1392–1405. https://doi.org/10.1111/j.1462-5822.2006.00718.x

354. Dong Y, Aguilar R, Xi Z, et al (2006) Anopheles gambiae Immune Responses to Human and Rodent Plasmodium Parasite Species. PLOS Pathog 2:e52. https://doi.org/10.1371/journal.ppat.0020052

355. Kajla MK, Shi L, Li B, et al (2011) A New Role for an Old Antimicrobial: Lysozyme c-1 Can Function to Protect Malaria Parasites in Anopheles Mosquitoes. PLoS ONE 6:. https://doi.org/10.1371/journal.pone.0019649

356. Nakhleh J, Christophides GK, Osta MA (2017) The serine protease homolog CLIPA14 modulates the intensity of the immune response in the mosquito Anopheles gambiae. J Biol Chem 292:18217–18226. https://doi.org/10.1074/jbc.M117.797787

357. Michel K, Budd A, Pinto S, et al (2005) Anopheles gambiae SRPN2 facilitates midgut invasion by the malaria parasite Plasmodium berghei. EMBO Rep 6:891–897. https://doi.org/10.1038/sj.embor.7400478

358. Ramirez JL, Garver LS, Dimopoulos G (2009) Challenges and approaches for mosquito targeted malaria control. Curr Mol Med 9:116–130

359. Zieler H, Keister DB, Dvorak JA, Ribeiro JM (2001) A snake venom phospholipase A(2) blocks malaria parasite development in the mosquito midgut by inhibiting ookinete association with the midgut surface. J Exp Biol 204:4157–4167

360. Ito J, Ghosh A, Moreira LA, et al (2002) Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. Nature 417:452–455. https://doi.org/10.1038/417452a

361. Hammond A, Galizi R, Kyrou K, et al (2016) A CRISPR-Cas9 Gene Drive System Targeting Female Reproduction in the Malaria Mosquito vector Anopheles gambiae. Nat Biotechnol 34:78–83. https://doi.org/10.1038/nbt.3439

362. Yoshida S, Ioka D, Matsuoka H, et al (2001) Bacteria expressing single-chain immunotoxin inhibit malaria parasite development in mosquitoes. Mol Biochem Parasitol 113:89–96

363. Gomes FM, Barillas-Mury C (2018) Infection of anopheline mosquitoes with Wolbachia: Implications for malaria control. PLoS Pathog 14:. https://doi.org/10.1371/journal.ppat.1007333

364. Luckhart S, Vodovotz Y, Cui L, Rosenberg R (1998) The mosquito Anopheles stephensi limits malaria parasite development with inducible synthesis of nitric oxide. Proc Natl Acad Sci U S A 95:5700–5705. https://doi.org/10.1073/pnas.95.10.5700

365. Blandin S, Shiao S-H, Moita LF, et al (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector Anopheles gambiae. Cell 116:661–670. https://doi.org/10.1016/s0092-8674(04)00173-4

366. Ramirez JL, Garver LS, Brayner FA, et al (2014) The role of hemocytes in Anopheles gambiae antiplasmodial immunity. J Innate Immun 6:119–128. https://doi.org/10.1159/000353765

367. Castillo JC, Ferreira ABB, Trisnadi N, Barillas-Mury C (2017) Activation of mosquito complement antiplasmodial response requires cellular immunity. Sci Immunol 2:. https://doi.org/10.1126/sciimmunol.aal1505

368. Kumar S, Gupta L, Han YS, Barillas-Mury C (2004) Inducible peroxidases mediate nitration of anopheles midgut cells undergoing apoptosis in response to Plasmodium invasion. J Biol Chem 279:53475–53482. https://doi.org/10.1074/jbc.M409905200

369. Oliveira G de A, Lieberman J, Barillas-Mury C (2012) Epithelial nitration by a peroxidase/NOX5 system mediates mosquito antiplasmodial immunity. Science 335:856–859. https://doi.org/10.1126/science.1209678

370. Rodrigues J, Brayner FA, Alves LC, et al (2010) Hemocyte differentiation mediates innate immune memory in Anopheles gambiae mosquitoes. Science 329:1353–1355. https://doi.org/10.1126/science.1190689

371. Ramirez JL, de Almeida Oliveira G, Calvo E, et al (2015) A mosquito lipoxin/lipocalin complex mediates innate immune priming in Anopheles gambiae. Nat Commun 6:7403. https://doi.org/10.1038/ncomms8403

372. Smith RC, Barillas-Mury C, Jacobs-Lorena M (2015) Hemocyte differentiation mediates the mosquito late-phase immune response against Plasmodium in Anopheles gambiae. Proc Natl Acad Sci U S A 112:E3412-3420. https://doi.org/10.1073/pnas.1420078112

373. Blighe K (2019) Publication-ready volcano plots with enhanced colouring and labeling: kevinblighe/EnhancedVolcano

374. Gendrin M, Turlure F, Rodgers FH, et al (2017) The Peptidoglycan Recognition Proteins PGRPLA and PGRPLB Regulate Anopheles Immunity to Bacteria and Affect Infection by Plasmodium. J Innate Immun 9:333–342. https://doi.org/10.1159/000452797

375. Zhang X, An C, Sprigg K, Michel K (2016) CLIPB8 is part of the prophenoloxidase activation system in Anopheles gambiae mosquitoes. Insect Biochem Mol Biol 71:106–115. https://doi.org/10.1016/j.ibmb.2016.02.008

376. Wang F, Hu C, Hua X, et al (2013) Translationally Controlled Tumor Protein, a Dual Functional Protein Involved in the Immune Response of the Silkworm, Bombyx mori. PLoS ONE 8:. https://doi.org/10.1371/journal.pone.0069284

377. Sung EJ, Ryuda M, Matsumoto H, et al (2017) Cytokine signaling through Drosophila Mthl10 ties lifespan to environmental stress. Proc Natl Acad Sci U S A 114:13786–13791. https://doi.org/10.1073/pnas.1712453115

378. Ghartey-Kwansah G, Li Z, Feng R, et al (2018) Comparative analysis of FKBP family protein: evaluation, structure, and function in mammals and Drosophila melanogaster. BMC Dev Biol 18:7. https://doi.org/10.1186/s12861-018-0167-3

379. Rodríguez M del C, Martínez-Barnetche J, Alvarado-Delgado A, et al (2007) The surface protein Pvs25 of Plasmodium vivax ookinetes interacts with calreticulin on the midgut apical surface of the malaria vector Anopheles albimanus. Mol Biochem Parasitol 153:167–177. https://doi.org/10.1016/j.molbiopara.2007.03.002

380. Zhang G, Schmidt O, Asgari S (2006) A calreticulin-like protein from endoparasitoid venom fluid is involved in host hemocyte inactivation. Dev Comp Immunol 30:756–764. https://doi.org/10.1016/j.dci.2005.11.001

381. Markesich DC, Gajewski KM, Nazimiec ME, Beckingham K (2000) bicaudal encodes the Drosophila beta NAC homolog, a component of the ribosomal translational machinery*. Dev Camb Engl 127:559–572

382. Duvic B, Hoffmann JA, Meister M, Royet J (2002) Notch Signaling Controls Lineage Specification during Drosophila Larval Hematopoiesis. Curr Biol 12:1923–1927. https://doi.org/10.1016/S0960-9822(02)01297-6

383. Zakovic S, Levashina EA (2017) NF-κB-Like Signaling Pathway REL2 in Immune Defenses of the Malaria Vector Anopheles gambiae. Front Cell Infect Microbiol 7:. https://doi.org/10.3389/fcimb.2017.00258

384. Hardbower DM, Asim M, Luis PB, et al (2017) Ornithine decarboxylase regulates M1 macrophage activation and mucosal inflammation via histone modifications. Proc Natl Acad Sci U S A 114:E751–E760. https://doi.org/10.1073/pnas.1614958114

385. Walsh MC, Lee J, Choi Y (2015) Tumor necrosis factor receptor associated factor 6 (TRAF6) regulation of development, function, and homeostasis of the immune system. Immunol Rev 266:72–92. https://doi.org/10.1111/imr.12302

386. Perišić Nanut M, Sabotič J, Jewett A, Kos J (2014) Cysteine Cathepsins as Regulators of the Cytotoxicity of NK and T Cells. Front Immunol 5:. https://doi.org/10.3389/fimmu.2014.00616

387. Beers C, Honey K, Fink S, et al (2003) Differential Regulation of Cathepsin S and Cathepsin L in Interferon γ–treated Macrophages. J Exp Med 197:169–179. https://doi.org/10.1084/jem.20020978

388. WHO World malaria report 2019

389. Castillo JC, Robertson AE, Strand MR (2006) Characterization of hemocytes from the mosquitoes Anopheles gambiae and Aedes aegypti. Insect Biochem Mol Biol 36:891–903. https://doi.org/10.1016/j.ibmb.2006.08.010

390. Attardo GM, Hansen IA, Raikhel AS (2005) Nutritional regulation of vitellogenesis in mosquitoes: implications for anautogeny. Insect Biochem Mol Biol 35:661–675. https://doi.org/10.1016/j.ibmb.2005.02.013

Appendix

Seurat Marker Genes - Cell Types

gene	p_val_adj	avg_logFC	pct.1	pct.2	cluster
AGAP012100	5.21E-87	0.325	0.973	0.975	0
AGAP002464	9.33E-75	0.471	0.954	0.896	0
AGAP011828	1.00E-71	0.498	0.833	0.699	0
AGAP010163	2.29E-68	0.322	0.95	0.961	0
AGAP000305	6.01E-58	0.383	0.877	0.702	0
AGAP004936	5.04E-50	0.428	0.791	0.617	0
AGAP007740	4.04E-45	0.258	0.964	0.966	0
AGAP002422	2.74E-41	0.656	0.61	0.54	0
AGAP011119	1.73E-40	0.421	0.744	0.622	0
AGAP002465	1.54E-36	0.421	0.815	0.767	0
AGAP012990	7.94E-33	0.291	0.858	0.89	0
AGAP010591	3.65E-28	0.259	0.944	0.943	0
AGAP029054	3.37E-22	0.340	0.73	0.625	0
AGAP005611	4.83E-21	0.393	0.668	0.669	0
AGAP004887	1.66E-19	0.254	0.872	0.916	0
AGAP011424	2.37E-19	0.263	0.81	0.846	0
AGAP009998	1.93E-17	0.271	0.845	0.894	0
AGAP009324	5.68E-16	0.285	0.734	0.788	0
AGAP005602	1.30E-13	0.285	0.133	0.258	0
AGAP012893	1.25E-11	0.264	0.137	0.255	0
AGAP008693	1.29E-10	0.288	0.173	0.307	0
AGAP028028	1.23E-08	0.341	0.562	0.541	0
AGAP004017	1.98E-07	0.375	0.532	0.525	0
AGAP013186	4.14E-07	0.292	0.179	0.299	0
AGAP008165	1.45E-06	0.351	0.175	0.287	0
AGAP028064	3.02E-06	0.371	0.499	0.482	0
AGAP004431	8.93E-06	0.345	0.156	0.258	0
AGAP008432	4.85E-05	0.285	0.163	0.262	0
AGAP003493	0.0001717	0.310	0.193	0.305	0
AGAP005685	0.00090264	0.569	0.462	0.538	0
AGAP002685	0.00908519	0.276	0.259	0.386	0
AGAP008762	0.01102749	0.355	0.206	0.309	0
AGAP004993	0.0210001	0.346	0.454	0.462	0
AGAP000044	0.04556246	0.274	0.266	0.391	0

AGAP011228	2.12E-189	0.746	0.986	0.753	1
AGAP007312	7.96E-162	0.799	0.769	0.353	1
AGAP004936	1.16E-142	0.597	0.923	0.586	1
AGAP006278	3.23E-137	0.667	0.864	0.525	1
AGAP000651	2.72E-136	0.714	0.784	0.394	1
AGAP004017	8.90E-129	0.591	0.815	0.411	1
AGAP004164	1.58E-125	0.705	0.444	0.126	1
AGAP028028	1.70E-121	0.594	0.824	0.437	1
AGAP004016	2.29E-119	0.557	0.69	0.294	1
AGAP006367	2.62E-118	0.869	0.331	0.075	1
AGAP028064	3.88E-118	0.559	0.782	0.369	1
AGAP007176	4.37E-117	0.674	0.361	0.089	1
AGAP010580	5.37E-117	0.630	0.521	0.178	1
AGAP013509	1.04E-116	0.673	0.615	0.247	1
AGAP011940	3.35E-116	0.543	0.412	0.111	1
AGAP005820	6.22E-115	0.549	0.639	0.253	1
AGAP009201	7.87E-114	0.687	0.769	0.404	1
AGAP000305	5.44E-112	0.535	0.941	0.698	1
AGAP000927	2.44E-111	0.495	0.701	0.307	1
AGAP011119	2.70E-111	0.511	0.921	0.566	1
AGAP010548	3.40E-110	0.682	0.704	0.348	1
AGAP004918	3.24E-107	0.749	0.549	0.219	1
AGAP000044	6.26E-106	0.479	0.603	0.238	1
AGAP008165	1.06E-104	0.462	0.466	0.155	1
AGAP002878	9.21E-104	0.489	0.693	0.313	1
AGAP028439	2.13E-103	0.499	0.657	0.278	1
AGAP002594	2.27E-103	0.502	0.737	0.356	1
AGAP001064	1.33E-102	0.549	0.651	0.292	1
AGAP005929	1.80E-102	0.528	0.642	0.281	1
AGAP000268	5.33E-102	0.468	0.454	0.148	1
AGAP000964	8.74E-102	0.591	0.646	0.289	1
AGAP001470	8.97E-102	0.541	0.477	0.163	1
AGAP008822	1.06E-101	0.488	0.551	0.218	1
AGAP004993	5.82E-100	0.586	0.722	0.352	1
AGAP003759	1.36E-99	0.604	0.285	0.063	1
AGAP011197	5.45E-98	0.496	0.543	0.209	1
AGAP005914	6.61E-96	0.461	0.489	0.177	1
AGAP001712	8.24E-94	0.526	0.399	0.13	1

AGAP002016	3.56E-93	0.593	0.371	0.114	1
AGAP007088	1.78E-92	0.513	0.589	0.254	1
AGAP003493	3.84E-92	0.438	0.478	0.175	1
AGAP007045	1.16E-91	0.490	0.395	0.128	1
AGAP003475	7.65E-91	0.495	0.507	0.195	1
AGAP012056	4.56E-90	0.523	0.493	0.19	1
AGAP008908	6.85E-90	0.441	0.435	0.151	1
AGAP009193	8.50E-90	0.729	0.27	0.065	1
AGAP012916	2.04E-89	0.750	0.354	0.109	1
AGAP009194	3.38E-89	0.907	0.312	0.089	1
AGAP003088	1.06E-88	0.509	0.484	0.186	1
AGAP001813	1.48E-88	0.465	0.636	0.292	1
AGAP011859	7.21E-88	0.479	0.327	0.09	1
AGAP001387	1.05E-87	0.375	0.661	0.298	1
AGAP029054	4.65E-87	0.463	0.88	0.577	1
AGAP011974	1.23E-86	0.512	0.551	0.233	1
AGAP007721	1.41E-86	0.463	0.63	0.292	1
AGAP000092	4.56E-86	0.452	0.506	0.203	1
AGAP012893	6.71E-86	0.423	0.397	0.134	1
AGAP009584	6.50E-85	0.557	0.375	0.126	1
AGAP010608	3.16E-84	0.454	0.404	0.14	1
AGAP006921	3.74E-84	0.544	0.439	0.165	1
AGAP002338	1.47E-83	0.324	0.614	0.268	1
AGAP004916	1.63E-83	0.870	0.254	0.061	1
AGAP007629	2.73E-83	0.470	0.555	0.238	1
AGAP012334	1.94E-81	0.432	0.561	0.245	1
AGAP000790	3.64E-81	0.474	0.424	0.152	1
AGAP012014	4.76E-80	0.470	0.464	0.184	1
AGAP000881	2.88E-79	0.432	0.368	0.123	1
AGAP010510	3.75E-79	0.356	0.462	0.176	1
AGAP005618	5.44E-79	0.452	0.379	0.133	1
AGAP029139	6.34E-79	0.496	0.473	0.194	1
AGAP011334	4.03E-78	0.465	0.462	0.184	1
AGAP002685	1.61E-77	0.426	0.555	0.25	1
AGAP004948	1.69E-77	0.410	0.352	0.115	1
AGAP001264	3.78E-77	0.441	0.614	0.29	1
AGAP011076	3.88E-76	0.343	0.471	0.189	1
AGAP010445	5.75E-76	0.371	0.386	0.137	1

AGAP007543	1.48E-75	0.398	0.375	0.131	1
AGAP001306	8.23E-75	0.336	0.311	0.095	1
AGAP009156	3.92E-74	0.461	0.296	0.087	1
AGAP003721	4.24E-74	0.355	0.405	0.151	1
AGAP002759	4.33E-74	0.507	0.274	0.078	1
AGAP003057	8.04E-73	0.876	0.292	0.091	1
AGAP004484	1.79E-72	0.499	0.343	0.12	1
AGAP004247	2.64E-72	0.271	0.561	0.253	1
AGAP000385	6.63E-72	0.428	0.42	0.161	1
AGAP006389	6.66E-72	0.480	0.281	0.083	1
AGAP008503	9.58E-72	0.379	0.368	0.133	1
AGAP010298	2.60E-71	0.441	0.33	0.112	1
AGAP007609	2.71E-71	0.429	0.363	0.129	1
AGAP009651	3.20E-71	0.303	0.435	0.168	1
AGAP007619	8.07E-71	0.389	0.305	0.094	1
AGAP003578	8.46E-71	0.337	0.438	0.171	1
AGAP004533	2.46E-70	0.407	0.781	0.466	1
AGAP028157	3.57E-70	0.505	0.288	0.088	1
AGAP002415	5.52E-70	0.342	0.325	0.106	1
AGAP010347	5.92E-70	0.347	0.728	0.394	1
AGAP002459	9.64E-70	0.455	0.315	0.103	1
AGAP004431	1.24E-69	0.295	0.397	0.147	1
AGAP004395	2.97E-69	0.320	0.536	0.242	1
AGAP001718	3.07E-68	0.318	0.421	0.167	1
AGAP011050	7.16E-68	0.367	0.429	0.173	1
AGAP003183	2.90E-67	0.387	0.269	0.079	1
AGAP000358	6.80E-67	0.303	0.575	0.271	1
AGAP000290	2.32E-66	0.365	0.278	0.084	1
AGAP010224	2.66E-66	0.387	0.287	0.091	1
AGAP005528	4.30E-66	0.340	0.337	0.12	1
AGAP012261	4.92E-66	0.362	0.295	0.095	1
AGAP000439	7.13E-66	0.339	0.317	0.107	1
AGAP000852	8.89E-66	0.401	0.342	0.124	1
AGAP000698	1.04E-65	0.408	0.303	0.102	1
AGAP002599	2.15E-65	0.397	0.634	0.334	1
AGAP011098	3.84E-65	0.357	0.352	0.128	1
AGAP001174	1.08E-64	0.680	0.274	0.087	1
AGAP008693	8.81E-64	0.268	0.439	0.183	1

AGAP010174	1.01E-63	0.409	0.285	0.092	1
AGAP003430	2.53E-63	0.370	0.259	0.077	1
AGAP013511	3.11E-63	0.480	0.254	0.075	1
AGAP003722	3.84E-63	0.319	0.325	0.114	1
AGAP000911	3.88E-63	0.334	0.308	0.104	1
AGAP011984	3.14E-62	0.310	0.261	0.078	1
AGAP007643	7.58E-62	0.293	0.341	0.127	1
AGAP007583	2.00E-61	0.299	0.343	0.128	1
AGAP006459	4.33E-61	0.324	0.628	0.323	1
AGAP001151	6.27E-61	0.313	0.484	0.219	1
AGAP000448	8.24E-61	0.395	0.304	0.105	1
AGAP003553	1.05E-60	0.338	0.356	0.136	1
AGAP008895	1.44E-60	0.362	0.289	0.098	1
AGAP028154	5.52E-60	0.341	0.444	0.195	1
AGAP011824	5.52E-60	0.336	0.392	0.159	1
AGAP004743	2.06E-59	0.336	0.506	0.236	1
AGAP009670	2.33E-58	0.318	0.489	0.227	1
AGAP011284	2.38E-58	0.345	0.343	0.132	1
AGAP008280	2.40E-58	0.300	0.429	0.184	1
AGAP009944	6.72E-58	0.358	0.335	0.128	1
AGAP002413	1.37E-57	0.266	0.37	0.15	1
AGAP010957	2.05E-57	0.372	0.382	0.157	1
AGAP004917	1.13E-56	0.320	0.308	0.111	1
AGAP000414	3.90E-56	0.352	0.256	0.083	1
AGAP000780	4.04E-56	0.271	0.336	0.131	1
AGAP003016	1.07E-55	0.414	0.29	0.103	1
AGAP001069	2.56E-55	0.294	0.31	0.116	1
AGAP010133	2.74E-55	0.351	0.278	0.096	1
AGAP010010	3.37E-55	0.324	0.305	0.112	1
AGAP003069	3.84E-55	0.329	0.407	0.176	1
AGAP001919	8.87E-55	0.270	0.376	0.155	1
AGAP007393	1.16E-54	0.313	0.472	0.219	1
AGAP011828	1.42E-54	0.366	0.901	0.688	1
AGAP003549	1.82E-54	0.307	0.321	0.122	1
AGAP005935	3.85E-54	0.281	0.346	0.139	1
AGAP002473	7.59E-54	0.289	0.352	0.144	1
AGAP001701	7.37E-53	0.326	0.411	0.185	1
AGAP003588	4.89E-52	0.275	0.301	0.115	1

AGAP000167	5.05E-52	0.305	0.366	0.154	1
AGAP009271	1.76E-51	0.329	0.458	0.22	1
AGAP006130	2.48E-51	0.304	0.278	0.101	1
AGAP002113	7.97E-51	0.366	0.272	0.1	1
AGAP010257	1.11E-50	0.255	0.943	0.756	1
AGAP001911	2.55E-50	0.291	0.912	0.658	1
AGAP006037	5.39E-49	0.312	0.721	0.44	1
AGAP005611	5.72E-49	0.309	0.858	0.591	1
AGAP000669	1.23E-48	0.263	0.315	0.128	1
AGAP007406	2.59E-48	0.285	0.909	0.71	1
AGAP003556	3.12E-48	0.326	0.709	0.431	1
AGAP006670	4.06E-48	0.269	0.329	0.134	1
AGAP010310	6.53E-48	0.311	0.333	0.142	1
AGAP028143	3.97E-47	0.257	0.322	0.131	1
AGAP000249	8.61E-47	0.271	0.782	0.484	1
AGAP004960	1.05E-46	0.329	0.287	0.114	1
AGAP005960	1.86E-46	0.255	0.94	0.768	1
AGAP011706	5.25E-45	0.266	0.703	0.417	1
AGAP008816	1.33E-44	0.314	0.289	0.117	1
AGAP029078	3.74E-43	0.315	0.635	0.373	1
AGAP005432	2.43E-42	0.275	0.259	0.101	1
AGAP002667	5.04E-41	0.263	0.755	0.479	1
AGAP007050	5.15E-37	0.268	0.258	0.108	1
AGAP001805	1.78E-35	0.252	0.795	0.537	1
AGAP004212	2.45E-35	0.261	0.375	0.186	1
AGAP001826	9.20E-29	0.350	0.336	0.178	1
AGAP010968	0	2.461	0.475	0.037	2
AGAP013060	0	1.976	0.66	0.092	2
AGAP012571	0	1.944	0.783	0.173	2
AGAP008011	0	1.903	0.479	0.035	2
AGAP003473	2.70E-303	3.031	0.847	0.268	2
AGAP003474	1.54E-298	2.451	0.994	0.952	2
AGAP005888	1.20E-295	1.828	0.959	0.533	2
AGAP008004	7.26E-291	2.368	0.894	0.368	2
AGAP004674	1.01E-278	2.010	0.381	0.02	2
AGAP009527	2.92E-272	2.044	0.606	0.1	2
AGAP006275	3.50E-259	2.270	0.643	0.134	2
AGAP013365	3.69E-234	2.218	0.609	0.123	2

AGAP011792	3.46E-232	1.856	0.367	0.027	2
AGAP002632	2.63E-230	2.413	0.729	0.23	2
AGAP005563	2.27E-227	2.423	0.491	0.071	2
AGAP002518	1.19E-225	2.038	0.518	0.085	2
AGAP007663	4.06E-225	1.820	0.424	0.046	2
AGAP028065	5.09E-218	1.722	0.546	0.095	2
AGAP007385	1.74E-217	1.810	0.459	0.062	2
AGAP028406	7.48E-216	2.130	0.395	0.04	2
AGAP004742	6.88E-213	1.666	0.439	0.054	2
AGAP011791	1.44E-205	1.745	0.34	0.027	2
AGAP012394	1.22E-198	2.093	0.372	0.038	2
AGAP012352	9.80E-198	1.811	0.562	0.119	2
AGAP008127	4.60E-194	1.464	0.265	0.012	2
AGAP005334	1.03E-192	1.652	0.398	0.048	2
AGAP009859	2.97E-192	2.001	0.509	0.097	2
AGAP008227	3.20E-191	1.813	0.489	0.085	2
AGAP013005	1.39E-190	1.916	0.407	0.054	2
AGAP007711	1.53E-190	1.853	0.494	0.092	2
AGAP003250	2.01E-188	1.553	0.308	0.023	2
AGAP011788	2.10E-188	1.519	0.262	0.013	2
AGAP002890	4.00E-188	1.538	0.337	0.031	2
AGAP028463	7.71E-180	1.621	0.506	0.099	2
AGAP007456	1.88E-176	1.610	0.315	0.028	2
AGAP011369	3.35E-176	2.362	0.723	0.304	2
AGAP007453	8.69E-165	1.551	0.357	0.045	2
AGAP005335	1.33E-164	1.443	0.342	0.04	2
AGAP006911	1.19E-162	1.580	0.367	0.05	2
AGAP009146	3.08E-162	1.439	0.281	0.024	2
AGAP011319	4.49E-159	1.572	0.367	0.051	2
AGAP013348	4.97E-155	1.589	0.277	0.024	2
AGAP008012	2.95E-153	1.272	0.272	0.023	2
AGAP007454	2.11E-149	1.308	0.27	0.024	2
AGAP005889	2.78E-148	2.422	0.449	0.098	2
AGAP003168	1.13E-144	1.809	0.484	0.116	2
AGAP005175	3.23E-144	1.520	0.391	0.068	2
AGAP011604	1.44E-143	1.537	0.324	0.044	2
AGAP006821	7.65E-142	1.598	0.498	0.126	2
AGAP005009	8.47E-142	1.556	0.368	0.062	2

AGAP003350	4.83E-139	1.709	0.277	0.03	2
AGAP028491	3.25E-138	1.188	0.262	0.025	2
AGAP002503	2.75E-136	1.402	0.284	0.033	2
AGAP009176	2.23E-132	1.808	0.398	0.078	2
AGAP001713	1.09E-130	1.933	0.405	0.085	2
AGAP012966	1.05E-128	1.229	0.252	0.025	2
AGAP008061	2.87E-125	1.619	0.394	0.083	2
AGAP008225	1.85E-120	1.098	0.284	0.038	2
AGAP028370	2.63E-117	1.547	0.799	0.512	2
AGAP028164	2.25E-115	1.226	0.27	0.036	2
AGAP006348	3.48E-114	1.446	0.275	0.039	2
AGAP028373	3.84E-112	1.278	0.583	0.207	2
AGAP001065	3.03E-110	1.378	0.454	0.124	2
AGAP003580	5.09E-110	1.307	0.348	0.07	2
AGAP008051	1.29E-107	1.668	0.392	0.096	2
AGAP028386	1.03E-104	1.368	0.746	0.453	2
AGAP028366	4.43E-98	1.315	0.886	0.712	2
AGAP013755	3.36E-96	1.685	0.317	0.066	2
AGAP008141	9.12E-93	1.309	0.307	0.064	2
AGAP005620	1.71E-91	1.066	0.267	0.046	2
AGAP028371	1.90E-87	1.482	0.792	0.577	2
AGAP028380	4.60E-87	1.412	0.742	0.446	2
AGAP005662	1.12E-86	1.336	0.419	0.133	2
AGAP028391	1.83E-86	1.199	0.979	0.973	2
AGAP028387	2.63E-84	1.517	0.732	0.445	2
AGAP009648	4.20E-68	0.991	0.285	0.07	2
AGAP028389	5.81E-68	1.143	0.506	0.215	2
AGAP028360	1.59E-58	1.209	0.491	0.229	2
AGAP012956	2.42E-57	2.292	0.454	0.209	2
AGAP028364	5.11E-56	1.225	0.829	0.661	2
AGAP007460	2.06E-54	1.143	0.372	0.143	2
AGAP007939	3.39E-54	0.799	0.267	0.074	2
AGAP005327	2.92E-53	1.016	0.561	0.297	2
AGAP007119	8.29E-52	0.983	0.281	0.086	2
AGAP028393	1.68E-45	1.118	0.531	0.29	2
AGAP013400	1.35E-44	1.095	0.472	0.241	2
AGAP006548	4.43E-43	0.620	0.445	0.201	2
AGAP004203	1.05E-41	1.652	0.268	0.092	2

AGAP001826	1.11E-39	0.759	0.428	0.193	2
AGAP001889	6.97E-39	1.404	0.429	0.215	2
AGAP004790	5.35E-37	0.640	0.699	0.484	2
AGAP028382	1.23E-34	1.174	0.471	0.267	2
AGAP001760	5.14E-33	0.619	0.265	0.098	2
AGAP007201	1.06E-30	0.519	0.262	0.098	2
AGAP004710	2.08E-30	0.703	0.672	0.501	2
AGAP002564	4.93E-20	0.510	0.338	0.176	2
AGAP011833	3.28E-16	0.618	0.294	0.156	2
AGAP004616	6.41E-16	0.428	0.598	0.433	2
AGAP004653	1.80E-15	0.586	0.404	0.259	2
AGAP000109	6.20E-15	0.458	0.324	0.178	2
AGAP000897	5.84E-14	0.647	0.301	0.176	2
AGAP004400	6.17E-13	0.268	0.728	0.606	2
AGAP011317	8.54E-13	0.488	0.518	0.365	2
AGAP000260	3.02E-12	0.287	0.514	0.354	2
AGAP002020	1.05E-10	0.382	0.362	0.232	2
AGAP002630	2.59E-09	0.257	0.484	0.336	2
AGAP008724	5.92E-09	0.307	0.529	0.398	2
AGAP007621	1.35E-07	0.331	0.593	0.482	2
AGAP000851	4.83E-07	0.298	0.546	0.419	2
AGAP006630	1.37E-06	0.250	0.441	0.311	2
AGAP004031	8.46E-06	0.331	0.305	0.204	2
AGAP008955	0.02805942	0.255	0.287	0.208	2
AGAP004978	0	4.470	0.812	0.123	3
AGAP011223	0	4.449	0.84	0.113	3
AGAP006258	0	4.365	0.789	0.128	3
AGAP004977	0	4.055	0.98	0.338	3
AGAP012616	0	3.962	0.828	0.077	3
AGAP012851	0	3.830	0.742	0.024	3
AGAP006570	0	3.669	0.734	0.107	3
AGAP006743	0	3.490	0.628	0.029	3
AGAP000162	0	3.471	0.802	0.06	3
AGAP000679	0	3.160	0.984	0.356	3
AGAP012000	0	2.843	0.425	0.016	3
AGAP004981	0	2.530	0.632	0.052	3
AGAP006914	2.15E-277	2.281	0.421	0.024	3
AGAP006726	6.84E-261	2.213	0.566	0.067	3

AGAP011107	8.60E-247	2.331	0.904	0.325	3
AGAP003490	5.41E-201	2.083	0.462	0.056	3
AGAP001884	4.96E-175	3.125	0.544	0.107	3
AGAP004376	4.10E-163	1.995	0.35	0.036	3
AGAP008000	1.66E-141	1.696	0.658	0.203	3
AGAP002227	4.99E-141	1.674	0.532	0.118	3
AGAP002317	3.88E-131	3.195	0.387	0.062	3
AGAP011158	3.61E-93	2.113	0.344	0.067	3
AGAP028022	2.66E-90	1.168	0.344	0.067	3
AGAP006228	1.10E-84	1.224	0.278	0.046	3
AGAP004775	7.26E-83	1.398	0.55	0.199	3
AGAP005865	1.06E-81	1.304	0.45	0.129	3
AGAP004918	6.50E-77	1.208	0.673	0.278	3
AGAP007572	8.27E-55	0.943	0.513	0.202	3
AGAP002076	8.85E-54	0.808	0.818	0.565	3
AGAP005775	1.02E-47	1.122	0.444	0.177	3
AGAP000249	5.73E-42	0.751	0.8	0.547	3
AGAP008883	5.14E-39	0.816	0.28	0.085	3
AGAP006766	4.06E-37	0.654	0.699	0.44	3
AGAP010188	3.83E-30	0.700	0.258	0.086	3
AGAP011054	2.58E-26	0.703	0.515	0.278	3
AGAP006614	1.11E-25	0.544	0.599	0.368	3
AGAP004916	1.36E-20	0.638	0.26	0.102	3
AGAP006670	4.37E-18	0.543	0.354	0.174	3
AGAP001903	1.04E-16	0.444	0.288	0.133	3
AGAP000385	1.33E-16	0.559	0.401	0.219	3
AGAP003878	1.75E-13	0.693	0.409	0.248	3
AGAP005327	6.20E-13	0.569	0.495	0.315	3
AGAP011092	8.24E-13	0.390	0.742	0.549	3
AGAP003767	8.47E-10	0.339	0.556	0.382	3
AGAP009944	5.52E-09	0.358	0.313	0.175	3
AGAP029078	1.43E-08	0.349	0.611	0.433	3
AGAP011824	1.09E-07	0.667	0.346	0.214	3
AGAP008432	4.91E-07	0.396	0.344	0.213	3
AGAP008909	3.36E-06	0.263	0.292	0.172	3
AGAP004991	7.09E-06	0.332	0.305	0.186	3
AGAP006937	0.00418714	0.255	0.272	0.176	3
AGAP009584	0.00899526	0.338	0.286	0.189	3

AGAP004203	2.94E-162	2.999	0.779	0.096	4
AGAP007940	9.56E-127	2.767	0.718	0.109	4
AGAP006548	1.20E-126	2.566	0.913	0.214	4
AGAP002593	6.61E-114	2.098	0.43	0.035	4
AGAP001065	8.30E-105	2.552	0.765	0.15	4
AGAP004700	3.30E-100	2.240	0.376	0.03	4
AGAP010046	4.33E-88	2.513	0.289	0.019	4
AGAP009173	7.86E-83	2.190	0.376	0.037	4
AGAP001116	1.29E-81	1.947	0.443	0.054	4
AGAP002198	2.09E-76	2.051	0.456	0.062	4
AGAP007455	5.49E-75	1.797	0.356	0.036	4
AGAP004954	6.02E-71	1.664	0.275	0.022	4
AGAP004534	2.34E-70	2.142	0.262	0.02	4
AGAP012352	1.91E-69	1.812	0.685	0.163	4
AGAP012966	2.85E-63	1.523	0.369	0.046	4
AGAP005327	2.86E-48	1.548	0.779	0.319	4
AGAP001423	1.01E-46	1.988	0.43	0.085	4
AGAP011504	1.41E-45	1.823	0.544	0.143	4
AGAP028494	2.69E-45	1.638	0.336	0.052	4
AGAP002378	1.83E-41	1.987	0.423	0.092	4
AGAP009182	5.60E-41	1.793	0.409	0.086	4
AGAP007453	7.04E-37	2.203	0.376	0.078	4
AGAP005861	1.37E-35	1.664	0.51	0.151	4
AGAP008761	1.21E-33	1.781	0.55	0.186	4
AGAP004616	4.48E-33	1.227	0.799	0.445	4
AGAP004550	6.52E-33	1.179	0.315	0.059	4
AGAP007456	3.72E-32	1.644	0.309	0.058	4
AGAP004742	6.99E-32	1.395	0.403	0.096	4
AGAP011319	1.06E-29	1.381	0.369	0.084	4
AGAP003473	1.19E-29	0.671	0.738	0.332	4
AGAP029139	3.41E-29	1.385	0.624	0.265	4
AGAP001826	3.90E-29	1.493	0.591	0.213	4
AGAP000179	2.41E-28	1.786	0.289	0.058	4
AGAP005979	2.88E-28	1.426	0.349	0.081	4
AGAP004790	3.76E-28	1.245	0.785	0.504	4
AGAP012515	6.77E-27	1.165	0.832	0.623	4
AGAP003474	1.59E-26	1.056	0.96	0.957	4
AGAP007460	3.84E-26	1.503	0.483	0.164	4

AGAP004819	1.94E-25	1.546	0.45	0.145	4
AGAP009859	2.95E-25	1.186	0.463	0.142	4
AGAP000180	1.38E-23	1.470	0.349	0.092	4
AGAP005335	1.88E-22	1.166	0.309	0.073	4
AGAP007768	8.21E-22	0.962	0.779	0.556	4
AGAP004653	5.46E-21	1.317	0.57	0.27	4
AGAP028370	8.59E-21	0.877	0.785	0.542	4
AGAP028143	9.56E-20	1.152	0.47	0.178	4
AGAP028386	1.38E-19	1.010	0.765	0.484	4
AGAP013400	1.49E-19	1.270	0.564	0.263	4
AGAP005334	7.33E-19	1.163	0.322	0.087	4
AGAP007492	2.85E-18	1.292	0.369	0.118	4
AGAP007120	2.90E-18	0.800	0.812	0.611	4
AGAP028463	4.11E-18	1.072	0.423	0.145	4
AGAP003534	5.82E-18	1.152	0.275	0.07	4
AGAP002245	5.44E-16	0.894	0.705	0.439	4
AGAP012418	6.75E-16	0.905	0.772	0.605	4
AGAP007621	7.31E-16	0.947	0.711	0.491	4
AGAP028373	9.34E-16	1.027	0.544	0.248	4
AGAP013365	1.44E-15	0.586	0.477	0.178	4
AGAP002277	5.40E-15	1.125	0.443	0.185	4
AGAP008724	7.71E-15	0.988	0.658	0.408	4
AGAP004890	1.27E-14	1.534	0.409	0.169	4
AGAP004710	2.39E-14	0.915	0.732	0.518	4
AGAP000439	2.83E-14	1.387	0.396	0.161	4
AGAP004400	7.46E-14	0.781	0.779	0.617	4
AGAP007927	8.42E-14	0.600	0.819	0.72	4
AGAP005888	8.77E-14	0.536	0.812	0.582	4
AGAP010337	9.87E-14	0.862	0.732	0.546	4
AGAP007385	1.80E-13	0.977	0.329	0.107	4
AGAP000260	3.81E-12	0.950	0.611	0.368	4
AGAP011833	1.67E-11	1.178	0.389	0.168	4
AGAP008004	2.02E-11	0.924	0.651	0.43	4
AGAP011484	2.73E-11	0.766	0.698	0.453	4
AGAP007711	9.22E-11	0.880	0.356	0.138	4
AGAP012717	1.08E-10	1.057	0.336	0.129	4
AGAP008727	1.23E-10	0.771	0.631	0.384	4
AGAP000165	1.68E-10	0.901	0.295	0.104	4

AGAP007087	4.72E-10	0.433	0.98	0.991	4
AGAP012188	1.37E-09	0.649	0.711	0.499	4
AGAP011131	3.91E-09	0.762	0.678	0.481	4
AGAP009833	4.62E-09	1.071	0.483	0.266	4
AGAP011832	4.79E-09	1.336	0.376	0.181	4
AGAP013005	5.37E-09	1.021	0.268	0.095	4
AGAP006879	1.51E-08	0.650	0.711	0.559	4
AGAP003538	1.68E-08	0.358	0.893	0.915	4
AGAP005581	1.81E-08	1.187	0.362	0.168	4
AGAP010606	3.81E-08	1.033	0.302	0.124	4
AGAP010177	3.91E-08	1.120	0.255	0.092	4
AGAP002470	5.79E-08	1.002	0.268	0.101	4
AGAP028380	2.02E-07	0.583	0.658	0.479	4
AGAP002499	6.32E-07	1.088	0.362	0.181	4
AGAP002520	1.60E-06	0.766	0.255	0.098	4
AGAP001595	2.41E-06	0.487	0.819	0.806	4
AGAP006630	3.23E-06	0.708	0.517	0.323	4
AGAP001138	4.18E-06	0.817	0.584	0.404	4
AGAP000851	5.33E-06	0.661	0.604	0.431	4
AGAP009527	6.87E-06	0.450	0.356	0.16	4
AGAP009564	9.85E-06	0.606	0.51	0.305	4
AGAP002630	4.70E-05	0.599	0.544	0.35	4
AGAP028366	7.84E-05	0.285	0.819	0.733	4
AGAP008491	9.05E-05	0.779	0.443	0.271	4
AGAP011896	0.00025959	0.334	0.832	0.813	4
AGAP003768	0.00028217	0.334	0.906	0.842	4
AGAP003599	0.00047237	0.790	0.45	0.287	4
AGAP010386	0.00059786	0.965	0.262	0.126	4
AGAP004296	0.00075248	0.634	0.43	0.263	4
AGAP000109	0.00112229	0.850	0.342	0.193	4
AGAP009491	0.00130174	0.601	0.611	0.485	4
AGAP004235	0.00300418	1.044	0.336	0.195	4
AGAP007574	0.0036735	0.718	0.315	0.168	4
AGAP009072	0.00599111	0.542	0.416	0.253	4
AGAP004743	0.00724895	0.632	0.47	0.309	4
AGAP009865	0.00736454	0.677	0.255	0.124	4
AGAP003900	0.01079693	0.693	0.282	0.147	4
AGAP028389	0.01431724	0.505	0.403	0.249	4

AGAP010464	0.0145139	0.766	0.295	0.162	4
AGAP010517	0.01821295	0.796	0.322	0.192	4
AGAP028360	0.02371424	0.451	0.409	0.259	4
AGAP003864	0.04423235	0.455	0.49	0.331	4
AGAP005363	0	1.730	0.45	0.003	5
AGAP004962	0	1.527	0.412	0.004	5
AGAP007855	4.72E-295	1.583	0.427	0.007	5
AGAP013736	8.53E-285	1.076	0.313	0.002	5
AGAP005019	2.01E-274	2.028	0.557	0.018	5
AGAP003550	3.62E-271	1.303	0.321	0.003	5
AGAP006671	1.30E-267	1.118	0.298	0.002	5
AGAP006105	5.29E-230	1.019	0.282	0.003	5
AGAP004963	7.99E-223	0.989	0.252	0.002	5
AGAP004239	1.13E-212	1.284	0.282	0.003	5
AGAP004753	3.00E-189	1.310	0.305	0.006	5
AGAP011219	1.17E-188	1.380	0.313	0.006	5
AGAP007706	1.48E-176	1.128	0.275	0.005	5
AGAP029045	1.31E-153	1.297	0.305	0.009	5
AGAP008420	5.52E-151	1.078	0.29	0.008	5
AGAP028483	2.60E-149	1.042	0.275	0.007	5
AGAP008753	1.59E-83	1.710	0.718	0.132	5
AGAP005778	7.04E-64	1.403	0.45	0.061	5
AGAP012334	1.61E-49	1.455	0.832	0.324	5
AGAP012202	2.52E-45	1.775	0.504	0.107	5
AGAP006238	2.80E-43	0.872	0.252	0.027	5
AGAP010929	1.89E-39	1.235	0.504	0.113	5
AGAP007569	4.70E-38	0.724	0.336	0.051	5
AGAP000531	9.61E-30	0.712	0.298	0.05	5
AGAP009444	4.92E-24	0.820	0.374	0.09	5
AGAP004164	2.37E-23	0.836	0.603	0.208	5
AGAP001219	7.15E-22	0.799	0.382	0.097	5
AGAP028034	1.45E-21	1.266	0.595	0.218	5
AGAP001387	4.70E-20	0.808	0.779	0.394	5
AGAP011859	2.20E-19	0.657	0.489	0.15	5
AGAP001321	2.67E-19	0.676	0.298	0.067	5
AGAP001701	2.96E-19	0.618	0.634	0.24	5
AGAP003561	3.22E-19	0.574	0.282	0.062	5
AGAP028149	3.80E-19	1.213	0.71	0.32	5

AGAP007867	7.96E-19	0.821	0.527	0.186	5
AGAP006818	6.26E-17	0.850	0.282	0.067	5
AGAP000651	1.75E-16	0.813	0.863	0.498	5
AGAP003936	2.09E-16	0.823	0.336	0.091	5
AGAP003069	1.76E-15	0.517	0.611	0.233	5
AGAP007397	1.84E-15	0.811	0.305	0.082	5
AGAP012028	2.63E-15	0.590	0.389	0.119	5
AGAP011050	7.62E-15	0.765	0.573	0.239	5
AGAP011346	9.56E-15	0.757	0.298	0.08	5
AGAP013296	1.88E-14	0.440	0.313	0.086	5
AGAP005136	2.35E-14	0.394	0.328	0.091	5
AGAP013511	2.51E-14	0.647	0.382	0.12	5
AGAP008344	3.30E-14	0.433	0.275	0.069	5
AGAP012397	1.20E-13	0.487	0.336	0.098	5
AGAP006733	1.30E-13	0.393	0.305	0.084	5
AGAP002459	1.64E-13	0.497	0.45	0.157	5
AGAP006367	1.68E-13	0.668	0.42	0.142	5
AGAP009192	2.69E-13	0.499	0.328	0.095	5
AGAP003016	5.82E-13	0.567	0.427	0.15	5
AGAP008045	6.07E-13	0.493	0.26	0.066	5
AGAP002387	6.75E-13	0.482	0.267	0.07	5
AGAP002338	1.16E-12	0.709	0.718	0.359	5
AGAP013008	1.79E-12	0.508	0.305	0.088	5
AGAP002032	1.91E-12	0.652	0.305	0.089	5
AGAP028143	1.95E-12	0.528	0.473	0.179	5
AGAP005749	2.47E-12	0.642	0.344	0.109	5
AGAP005314	2.50E-12	0.563	0.282	0.079	5
AGAP010510	3.01E-12	0.578	0.58	0.251	5
AGAP004317	6.72E-12	0.573	0.359	0.117	5
AGAP007088	1.13E-11	0.546	0.695	0.342	5
AGAP007699	1.21E-11	0.549	0.542	0.23	5
AGAP009944	1.42E-11	0.536	0.466	0.181	5
AGAP010718	1.47E-11	0.600	0.26	0.07	5
AGAP009194	1.69E-11	0.450	0.412	0.147	5
AGAP005540	1.72E-11	0.448	0.298	0.089	5
AGAP010682	1.89E-11	0.526	0.29	0.085	5
AGAP005929	2.03E-11	0.632	0.702	0.377	5
AGAP004395	2.88E-11	0.531	0.641	0.319	5

AGAP001501	3.17E-11	0.488	0.305	0.094	5
AGAP009687	6.72E-11	0.374	0.298	0.09	5
AGAP011197	1.01E-10	0.602	0.641	0.297	5
AGAP012056	1.10E-10	0.441	0.603	0.269	5
AGAP002265	1.40E-10	0.369	0.359	0.121	5
AGAP009271	2.07E-10	0.493	0.603	0.28	5
AGAP003135	3.19E-10	0.462	0.328	0.107	5
AGAP007024	3.48E-10	0.483	0.45	0.179	5
AGAP009201	3.87E-10	0.597	0.817	0.501	5
AGAP002470	4.06E-10	0.337	0.313	0.1	5
AGAP002706	4.93E-10	0.380	0.405	0.149	5
AGAP008249	8.80E-10	0.460	0.374	0.137	5
AGAP010440	1.70E-09	0.376	0.267	0.081	5
AGAP003021	2.50E-09	0.298	0.305	0.098	5
AGAP011076	2.63E-09	0.402	0.565	0.263	5
AGAP008299	2.85E-09	0.604	0.282	0.092	5
AGAP009075	2.96E-09	0.371	0.336	0.118	5
AGAP008393	2.98E-09	0.515	0.298	0.097	5
AGAP007505	4.17E-09	0.409	0.374	0.139	5
AGAP010580	5.55E-09	0.366	0.595	0.269	5
AGAP006937	5.86E-09	0.426	0.435	0.178	5
AGAP028064	9.22E-09	0.512	0.786	0.481	5
AGAP003462	9.50E-09	0.489	0.29	0.098	5
AGAP007574	1.36E-08	0.397	0.412	0.166	5
AGAP002415	1.42E-08	0.481	0.412	0.163	5
AGAP004916	1.68E-08	0.490	0.321	0.112	5
AGAP011473	1.85E-08	0.508	0.374	0.143	5
AGAP002520	2.09E-08	0.422	0.29	0.098	5
AGAP004917	2.46E-08	0.391	0.412	0.162	5
AGAP002061	2.69E-08	0.396	0.29	0.099	5
AGAP000439	2.75E-08	0.396	0.405	0.162	5
AGAP009209	3.02E-08	0.454	0.29	0.099	5
AGAP008909	3.87E-08	0.304	0.427	0.177	5
AGAP003057	4.31E-08	0.524	0.366	0.144	5
AGAP003134	5.72E-08	0.273	0.374	0.144	5
AGAP010225	5.86E-08	0.489	0.321	0.113	5
AGAP002087	8.01E-08	0.322	0.412	0.166	5
AGAP005523	1.05E-07	0.366	0.26	0.086	5

AGAP004079	1.11E-07	0.444	0.298	0.106	5
AGAP011634	1.39E-07	0.272	0.374	0.146	5
AGAP002140	1.49E-07	0.386	0.282	0.097	5
AGAP004484	1.56E-07	0.376	0.427	0.179	5
AGAP003162	1.72E-07	0.279	0.275	0.092	5
AGAP001998	1.72E-07	0.332	0.328	0.121	5
AGAP010253	1.82E-07	0.335	0.267	0.09	5
AGAP000044	1.85E-07	0.423	0.641	0.336	5
AGAP008612	2.22E-07	0.272	0.321	0.118	5
AGAP004668	2.28E-07	0.515	0.282	0.099	5
AGAP011444	2.39E-07	0.357	0.344	0.132	5
AGAP001069	2.63E-07	0.374	0.405	0.166	5
AGAP011190	2.74E-07	0.279	0.351	0.136	5
AGAP000881	2.87E-07	0.376	0.435	0.188	5
AGAP011733	3.26E-07	0.438	0.313	0.116	5
AGAP028439	3.38E-07	0.466	0.679	0.38	5
AGAP001064	4.04E-07	0.383	0.702	0.388	5
AGAP012014	4.60E-07	0.364	0.542	0.258	5
AGAP029139	5.82E-07	0.393	0.557	0.268	5
AGAP007120	5.93E-07	0.388	0.885	0.61	5
AGAP005432	6.58E-07	0.336	0.359	0.141	5
AGAP002521	6.87E-07	0.325	0.275	0.095	5
AGAP010445	6.90E-07	0.373	0.458	0.203	5
AGAP001505	7.03E-07	0.324	0.26	0.089	5
AGAP002931	7.21E-07	0.336	0.45	0.196	5
AGAP008632	7.50E-07	0.286	0.366	0.146	5
AGAP007583	8.22E-07	0.274	0.427	0.184	5
AGAP010347	9.26E-07	0.399	0.771	0.484	5
AGAP001364	9.95E-07	0.327	0.29	0.106	5
AGAP011940	1.38E-06	0.393	0.435	0.192	5
AGAP009193	1.42E-06	0.330	0.313	0.119	5
AGAP011531	1.50E-06	0.342	0.305	0.114	5
AGAP006040	1.52E-06	0.348	0.282	0.103	5
AGAP000546	1.69E-06	0.346	0.298	0.111	5
AGAP013336	1.76E-06	0.350	0.359	0.147	5
AGAP003919	1.81E-06	0.486	0.313	0.121	5
AGAP008008	1.85E-06	0.446	0.618	0.311	5
AGAP008345	2.08E-06	0.337	0.282	0.104	5

AGAP012135	2.25E-06	0.256	0.282	0.102	5
AGAP013102	2.57E-06	0.320	0.29	0.108	5
AGAP005410	2.59E-06	0.261	0.351	0.141	5
AGAP007629	2.72E-06	0.407	0.626	0.322	5
AGAP012524	2.82E-06	0.318	0.641	0.332	5
AGAP002020	2.88E-06	0.322	0.511	0.243	5
AGAP007575	2.95E-06	0.320	0.321	0.124	5
AGAP000669	3.16E-06	0.252	0.412	0.176	5
AGAP008163	3.30E-06	0.558	0.397	0.177	5
AGAP008693	3.57E-06	0.271	0.527	0.25	5
AGAP006944	3.92E-06	0.279	0.359	0.147	5
AGAP007721	4.15E-06	0.352	0.695	0.382	5
AGAP011907	4.34E-06	0.634	0.29	0.11	5
AGAP005820	4.82E-06	0.457	0.649	0.357	5
AGAP004839	4.94E-06	0.337	0.328	0.132	5
AGAP000448	5.36E-06	0.257	0.374	0.157	5
AGAP013291	5.73E-06	0.348	0.29	0.11	5
AGAP002413	6.27E-06	0.271	0.45	0.208	5
AGAP000949	6.45E-06	0.388	0.305	0.118	5
AGAP009738	7.14E-06	0.384	0.336	0.137	5
AGAP010331	7.26E-06	0.367	0.298	0.116	5
AGAP009584	8.03E-06	0.361	0.42	0.192	5
AGAP001068	9.65E-06	0.337	0.321	0.127	5
AGAP002917	1.12E-05	0.380	0.26	0.096	5
AGAP013509	1.15E-05	0.393	0.634	0.346	5
AGAP004212	1.22E-05	0.414	0.489	0.234	5
AGAP007614	1.35E-05	0.257	0.267	0.099	5
AGAP007393	1.43E-05	0.357	0.557	0.285	5
AGAP009305	1.77E-05	0.461	0.298	0.119	5
AGAP010957	1.80E-05	0.502	0.443	0.217	5
AGAP004993	1.87E-05	0.453	0.725	0.452	5
AGAP006670	1.96E-05	0.361	0.412	0.185	5
AGAP006607	2.06E-05	0.280	0.351	0.15	5
AGAP008837	4.33E-05	0.276	0.328	0.137	5
AGAP013228	4.57E-05	0.307	0.313	0.129	5
AGAP009907	5.37E-05	0.351	0.26	0.098	5
AGAP009202	5.55E-05	0.357	0.344	0.15	5
AGAP000290	6.33E-05	0.354	0.321	0.135	5

AGAP004852	7.24E-05	0.289	0.305	0.126	5
AGAP003759	7.33E-05	0.276	0.305	0.123	5
AGAP011800	9.16E-05	0.340	0.427	0.21	5
AGAP010464	9.34E-05	0.260	0.359	0.161	5
AGAP009651	0.00010774	0.393	0.481	0.239	5
AGAP011228	0.00010962	0.280	0.962	0.817	5
AGAP004296	0.00012429	0.307	0.511	0.261	5
AGAP003851	0.00014452	0.361	0.634	0.364	5
AGAP028157	0.00014811	0.289	0.328	0.141	5
AGAP003360	0.00021573	0.281	0.29	0.12	5
AGAP001718	0.00023109	0.273	0.473	0.235	5
AGAP004721	0.00023432	0.262	0.656	0.371	5
AGAP004743	0.00025339	0.379	0.565	0.308	5
AGAP001311	0.00034795	0.304	0.282	0.117	5
AGAP002606	0.00036802	0.482	0.267	0.11	5
AGAP004960	0.00040251	0.304	0.344	0.16	5
AGAP003553	0.00046293	0.366	0.397	0.195	5
AGAP008816	0.00048233	0.266	0.351	0.162	5
AGAP004606	0.00051606	0.338	0.389	0.185	5
AGAP028617	0.00060721	0.307	0.282	0.118	5
AGAP000852	0.00081411	0.269	0.382	0.182	5
AGAP002935	0.00083071	0.272	0.351	0.163	5
AGAP003367	0.0009167	0.254	0.45	0.225	5
AGAP004918	0.001039	0.348	0.542	0.308	5
AGAP000927	0.00163827	0.277	0.679	0.414	5
AGAP001976	0.00181993	0.331	0.252	0.105	5
AGAP002759	0.0019539	0.438	0.29	0.131	5
AGAP004026	0.00223784	0.428	0.252	0.105	5
AGAP010608	0.00243062	0.274	0.42	0.211	5
AGAP000414	0.00253615	0.295	0.29	0.129	5
AGAP003721	0.00414231	0.324	0.42	0.219	5
AGAP009152	0.00910534	0.278	0.351	0.175	5
AGAP007312	0.01111189	0.311	0.718	0.467	5
AGAP002422	0.0143814	0.267	0.832	0.56	5
AGAP028028	0.02248838	0.251	0.802	0.543	5
AGAP009526	1.72E-104	2.864	0.736	0.117	6
AGAP006181	1.12E-97	2.622	0.579	0.069	6
AGAP003939	5.44E-83	2.675	0.562	0.078	6

AGAP001622	2.17E-72	2.641	0.76	0.189	6
AGAP003778	1.13E-70	2.418	0.504	0.071	6
AGAP001569	6.19E-66	2.280	0.479	0.067	6
AGAP004161	8.04E-64	2.322	0.736	0.196	6
AGAP002358	3.84E-58	2.334	0.446	0.065	6
AGAP008311	2.87E-50	2.093	0.273	0.026	6
AGAP004790	5.28E-46	1.919	0.909	0.503	6
AGAP011484	9.81E-46	2.081	0.86	0.451	6
AGAP003474	4.98E-41	0.920	1	0.956	6
AGAP000693	6.27E-39	1.360	0.521	0.126	6
AGAP012418	3.01E-32	1.539	0.868	0.604	6
AGAP004400	9.69E-32	1.549	0.86	0.616	6
AGAP006879	2.41E-31	1.531	0.851	0.556	6
AGAP011131	5.24E-31	1.621	0.793	0.479	6
AGAP003473	3.70E-29	0.703	0.769	0.334	6
AGAP003586	2.91E-28	2.010	0.306	0.056	6
AGAP009491	8.13E-28	1.367	0.793	0.481	6
AGAP004710	1.10E-27	1.390	0.835	0.516	6
AGAP000260	2.87E-26	1.687	0.702	0.367	6
AGAP002245	1.84E-24	1.521	0.744	0.439	6
AGAP007768	2.47E-24	1.444	0.793	0.557	6
AGAP000851	1.18E-23	1.628	0.711	0.43	6
AGAP005888	2.27E-22	0.551	0.868	0.582	6
AGAP010337	3.17E-22	1.430	0.777	0.546	6
AGAP008724	2.11E-21	1.515	0.694	0.409	6
AGAP012188	2.26E-20	1.522	0.736	0.5	6
AGAP008727	3.88E-19	1.418	0.669	0.385	6
AGAP007621	6.59E-19	1.439	0.736	0.491	6
AGAP004616	2.88E-18	1.425	0.694	0.449	6
AGAP013092	3.53E-18	1.365	0.744	0.541	6
AGAP000849	8.95E-17	1.532	0.62	0.36	6
AGAP013365	3.44E-15	1.115	0.471	0.18	6
AGAP009537	4.37E-15	1.470	0.587	0.319	6
AGAP012515	6.18E-15	1.031	0.769	0.626	6
AGAP010672	6.85E-13	1.554	0.455	0.208	6
AGAP002564	8.62E-13	1.552	0.438	0.191	6
AGAP002632	2.09E-10	0.892	0.545	0.289	6
AGAP011159	7.14E-10	1.452	0.537	0.324	6

AGAP008004	2.50E-09	0.564	0.678	0.431	6
AGAP013060	6.42E-09	0.930	0.397	0.161	6
AGAP012533	8.92E-06	1.258	0.438	0.262	6
AGAP012571	1.16E-05	0.685	0.463	0.248	6
AGAP011829	7.90E-05	1.244	0.463	0.308	6
AGAP001138	0.00013625	0.981	0.529	0.406	6
AGAP003538	0.00016288	0.336	0.917	0.914	6
AGAP001889	0.00036591	0.735	0.421	0.239	6
AGAP009859	0.00067777	0.952	0.306	0.147	6
AGAP009564	0.00095546	1.170	0.446	0.307	6
AGAP003328	0.0023294	1.253	0.331	0.184	6
AGAP003864	0.00380745	1.136	0.455	0.333	6
AGAP003599	0.00393763	1.207	0.421	0.289	6
AGAP013400	0.00861525	1.033	0.405	0.268	6
AGAP009527	0.01766886	0.972	0.306	0.163	6
AGAP007841	0.01977527	1.122	0.43	0.311	6
AGAP008955	0.02020887	1.158	0.347	0.215	6
AGAP007297	1	0.969	0.322	0.256	6
AGAP007347	7.26E-217	4.377	0.908	0.084	7
AGAP005848	6.18E-105	2.456	0.392	0.026	7
AGAP011294	2.59E-69	1.857	0.283	0.02	7
AGAP000694	2.91E-63	2.455	0.267	0.02	7
AGAP000376	1.50E-51	2.140	0.758	0.237	7
AGAP011197	1.33E-40	1.780	0.775	0.294	7
AGAP005888	2.24E-37	2.574	0.933	0.581	7
AGAP000693	1.49E-32	2.856	0.492	0.127	7
AGAP005612	8.23E-23	2.085	0.317	0.071	7
AGAP010816	1.11E-17	1.344	0.342	0.094	7
AGAP004721	3.78E-12	0.936	0.658	0.371	7
AGAP028463	1.74E-10	1.596	0.367	0.147	7
AGAP010225	1.62E-09	1.061	0.317	0.113	7
AGAP012708	1.75E-08	0.611	0.842	0.579	7
AGAP004918	6.89E-08	0.786	0.558	0.308	7
AGAP009849	3.22E-07	1.212	0.267	0.098	7
AGAP005149	7.34E-07	1.161	0.442	0.234	7
AGAP007721	1.11E-06	0.714	0.625	0.384	7
AGAP003851	1.37E-06	0.745	0.6	0.366	7
AGAP007087	1.10E-05	0.356	1	0.99	7

AGAP002338	1.90E-05	0.860	0.558	0.364	7
AGAP004710	3.32E-05	0.482	0.75	0.518	7
AGAP028439	0.0001014	0.664	0.583	0.383	7
AGAP011802	0.00012866	0.328	1	0.983	7
AGAP007619	0.00015509	0.976	0.317	0.151	7
AGAP001718	0.00152987	0.814	0.4	0.237	7
AGAP004743	0.00217728	0.679	0.492	0.31	7
AGAP001971	0.00326132	0.502	0.692	0.557	7
AGAP011824	0.00491245	0.790	0.392	0.222	7
AGAP004790	0.02712352	0.251	0.708	0.508	7
AGAP011228	0.03268744	0.318	0.983	0.816	7
AGAP008163	0.03889787	0.800	0.325	0.179	7
AGAP007318	0	3.649	0.788	0.021	8
AGAP009053	6.97E-212	3.014	0.541	0.016	8
AGAP028208	3.96E-195	2.728	0.341	0.005	8
AGAP009051	1.62E-177	1.973	0.365	0.008	8
AGAP007320	4.28E-175	1.529	0.294	0.004	8
AGAP001002	2.27E-129	3.813	0.424	0.018	8
AGAP001652	9.61E-107	2.220	0.612	0.05	8
AGAP003319	6.01E-95	2.148	0.494	0.035	8
AGAP011226	1.25E-92	1.942	0.424	0.026	8
AGAP005209	1.06E-73	1.818	0.471	0.041	8
AGAP010759	2.84E-65	2.023	0.694	0.105	8
AGAP009098	2.29E-58	1.554	0.259	0.015	8
AGAP003352	7.88E-53	1.353	0.376	0.035	8
AGAP002379	8.34E-42	1.548	0.271	0.023	8
AGAP002878	1.78E-32	1.539	0.918	0.415	8
AGAP006367	2.30E-30	1.494	0.6	0.142	8
AGAP010548	4.91E-29	1.652	0.859	0.444	8
AGAP000964	6.85E-27	1.504	0.835	0.385	8
AGAP028157	1.64E-21	1.062	0.529	0.139	8
AGAP003088	6.13E-21	1.358	0.682	0.265	8
AGAP004247	5.34E-20	1.442	0.765	0.335	8
AGAP001174	1.10E-18	1.266	0.494	0.135	8
AGAP000651	3.16E-18	1.172	0.859	0.501	8
AGAP012916	9.91E-18	1.045	0.565	0.173	8
AGAP013511	1.95E-17	1.086	0.459	0.121	8
AGAP013027	2.78E-17	0.650	0.259	0.042	8

AGAP005246	2.59E-16	0.959	0.435	0.115	8
AGAP003057	3.68E-16	1.012	0.494	0.144	8
AGAP007176	2.02E-14	1.025	0.506	0.162	8
AGAP010133	4.46E-14	1.110	0.471	0.143	8
AGAP009156	7.35E-13	0.900	0.459	0.143	8
AGAP000321	2.98E-12	0.732	0.424	0.126	8
AGAP012396	9.79E-12	1.004	0.6	0.242	8
AGAP011334	1.37E-11	0.929	0.612	0.258	8
AGAP028617	2.11E-11	0.883	0.4	0.118	8
AGAP011958	3.75E-11	0.616	0.424	0.127	8
AGAP000235	1.80E-10	0.704	0.318	0.082	8
AGAP002473	2.34E-10	0.901	0.529	0.199	8
AGAP007721	3.55E-10	0.818	0.753	0.384	8
AGAP012056	3.93E-10	0.986	0.612	0.272	8
AGAP010175	1.15E-09	0.939	0.271	0.066	8
AGAP005933	1.44E-09	0.707	0.294	0.076	8
AGAP003475	2.17E-09	0.764	0.635	0.279	8
AGAP000385	4.73E-09	0.721	0.565	0.23	8
AGAP001823	5.17E-09	0.670	0.294	0.079	8
AGAP006745	7.05E-09	0.703	0.341	0.1	8
AGAP005931	1.70E-08	0.762	0.388	0.129	8
AGAP003722	1.83E-08	0.861	0.459	0.17	8
AGAP007583	1.84E-08	0.591	0.494	0.185	8
AGAP010307	2.17E-08	0.794	0.365	0.118	8
AGAP009305	5.33E-08	0.769	0.365	0.119	8
AGAP008905	5.35E-08	0.608	0.365	0.118	8
AGAP003408	1.47E-07	0.565	0.365	0.12	8
AGAP007864	3.68E-07	0.760	0.282	0.082	8
AGAP000927	3.73E-07	0.676	0.741	0.415	8
AGAP028154	4.13E-07	0.704	0.576	0.262	8
AGAP007941	5.85E-07	0.637	0.259	0.072	8
AGAP009110	7.10E-07	0.650	0.271	0.077	8
AGAP007049	8.42E-07	0.550	0.329	0.106	8
AGAP008895	1.56E-06	0.694	0.4	0.149	8
AGAP013331	2.48E-06	0.703	0.271	0.08	8
AGAP010233	3.06E-06	0.623	0.271	0.081	8
AGAP001285	7.36E-06	0.617	0.341	0.121	8
AGAP009464	8.32E-06	0.556	0.306	0.101	8

AGAP001616	9.23E-06	0.513	0.271	0.082	8
AGAP010298	1.54E-05	0.528	0.424	0.171	8
AGAP005121	1.83E-05	0.730	0.259	0.08	8
AGAP006204	0.00011402	0.352	0.306	0.105	8
AGAP005160	0.00014013	0.825	0.306	0.113	8
AGAP002569	0.00016278	0.585	0.282	0.098	8
AGAP005608	0.00026338	0.405	0.271	0.091	8
AGAP006389	0.00026475	0.607	0.341	0.137	8
AGAP007643	0.00031647	0.609	0.412	0.185	8
AGAP011140	0.00033321	0.424	0.294	0.106	8
AGAP001306	0.00091632	0.631	0.365	0.154	8
AGAP007209	0.0009847	0.576	0.294	0.108	8
AGAP004431	0.00146403	0.447	0.459	0.215	8
AGAP007157	0.0017956	0.409	0.929	0.781	8
AGAP003021	0.00241962	0.546	0.271	0.101	8
AGAP011054	0.00291066	0.570	0.553	0.296	8
AGAP008094	0.0033246	0.689	0.259	0.095	8
AGAP001711	0.0044172	0.481	0.365	0.164	8
AGAP008908	0.00451939	0.526	0.459	0.229	8
AGAP007609	0.00513771	0.395	0.424	0.193	8
AGAP011119	0.00641057	0.374	0.929	0.664	8
AGAP004709	0.0071901	0.432	0.294	0.117	8
AGAP009883	0.00777433	0.311	0.282	0.109	8
AGAP009647	0.00805217	0.315	0.553	0.284	8
AGAP012717	0.01025701	0.434	0.318	0.132	8
AGAP000565	0.01034011	0.301	0.306	0.124	8
AGAP011798	0.01111463	0.573	0.259	0.099	8
AGAP001976	0.01712058	0.340	0.271	0.106	8
AGAP002931	0.01969398	0.335	0.412	0.199	8
AGAP004852	0.02508642	0.352	0.306	0.128	8
AGAP002599	0.03220097	0.435	0.671	0.417	8
AGAP001971	0.03253498	0.395	0.812	0.556	8
AGAP002113	0.03696752	0.500	0.329	0.147	8
AGAP003477	0.04175418	0.310	0.259	0.103	8
AGAP005462	0.04202719	0.413	0.259	0.104	8