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.

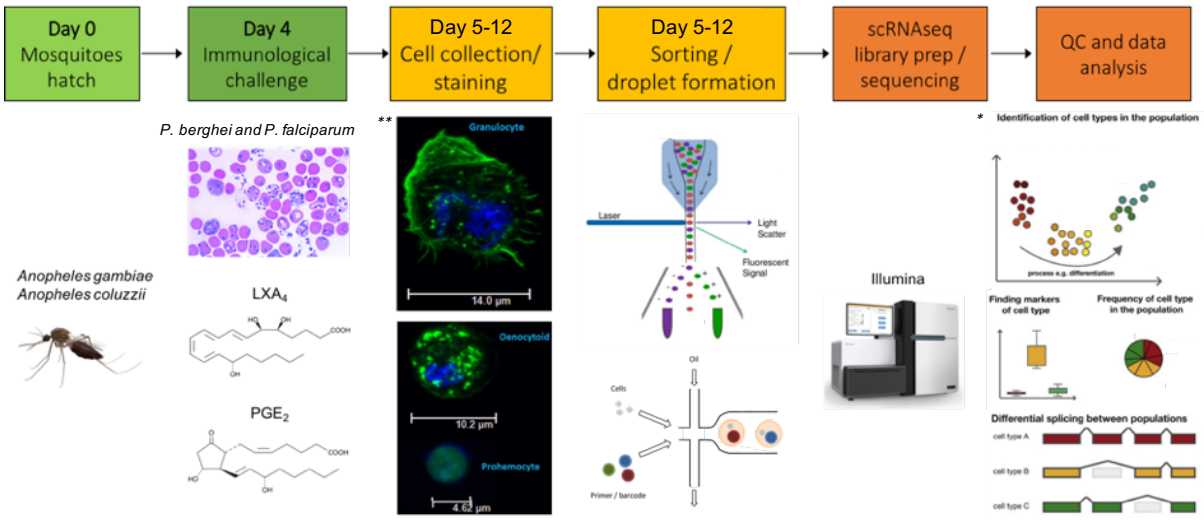

Fig. II.1 Experimental work-flow. At day 0 *A. gambiae* M-form (*A. coluzzi*) adult female mosquitoes hatch. After 3-4 days mosquitoes are challenged with *P. berghei* or *P. falciparum* infection or eicosanoid (lipoxin A_4 / prostaglandin E_2) injection. Samples are collected 1,2,3,7 days post-challenge. 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].

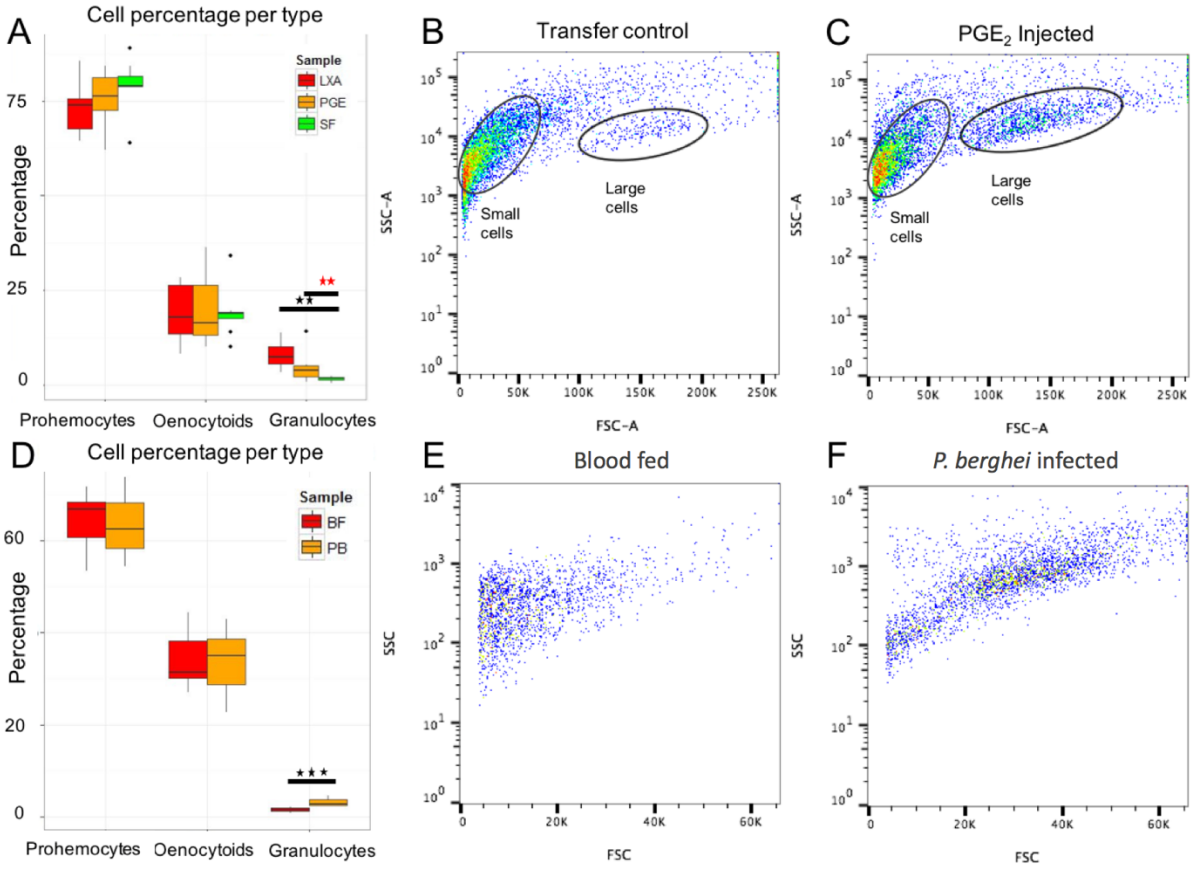

Fig. II.2 Effect of PGE_2, LXA_4 and *P. berghei* infection on hemocyte types. All events are CM-DiL positive. **(A,D)** Manual counting of cells by morphological differentiation into prohemocytes (small), oenocytoids (medium size and round), granulocytes (large and more complex) in PGE_2 and LXA_4 **(A)** 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.

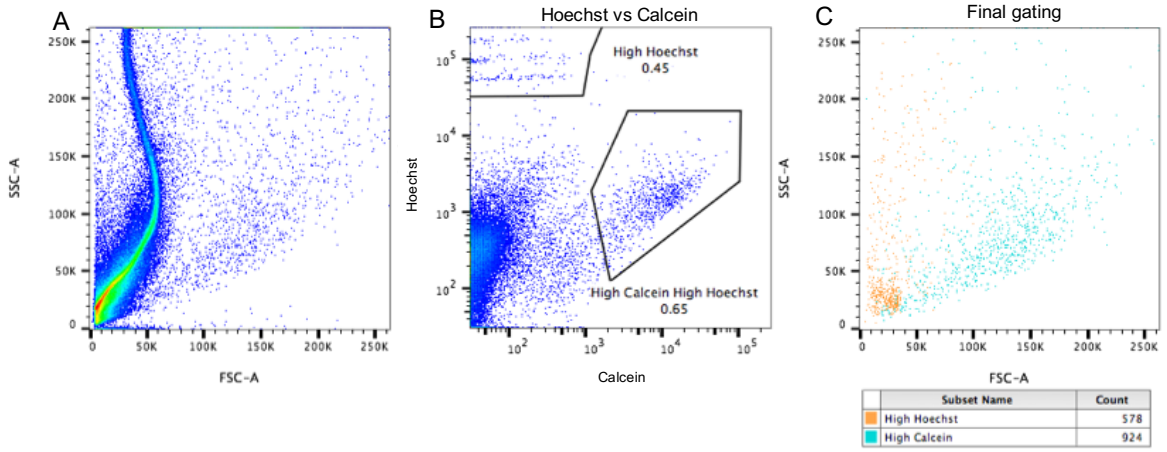

Fig. II.3 Sorting and flow cytometry analysis of hemocyte with Hoechst and calcein AM. **(A-B)** Pseudo-colouring refers to event density, going from light blue to green and finally red with higher events density. **(A)** Events on forward scatter (FSC, size) vs. side scatter (SSC, granularity). **(B)** 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.

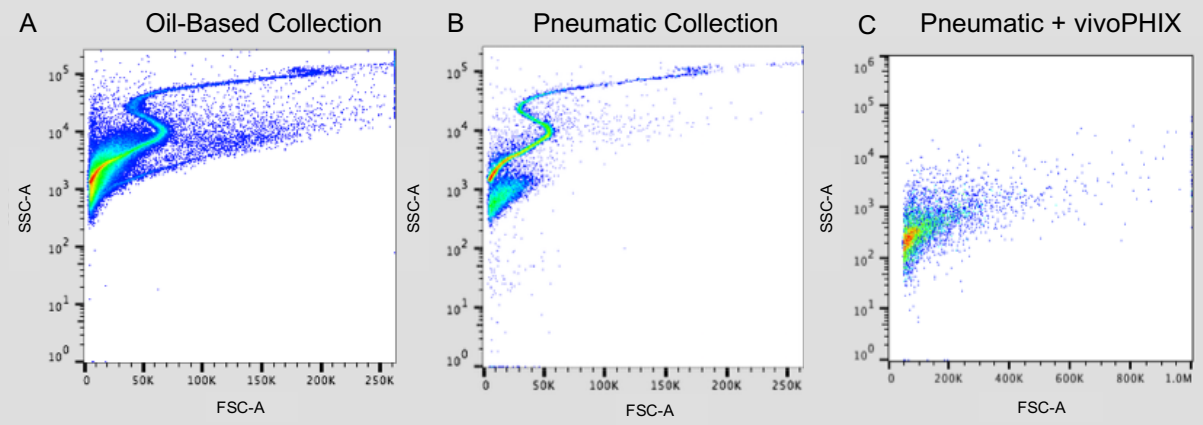

Fig. II.4 Hemocyte isolation optimisation with FACS. (A-C) Pseudo-colouring refers to event density, going from light blue to green and finally red with higher events **(A)** Standard oil-based collection **(B)** Custom oil-free pneumatic injector system and **(C)** Pneumatic injector plus cells fixed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

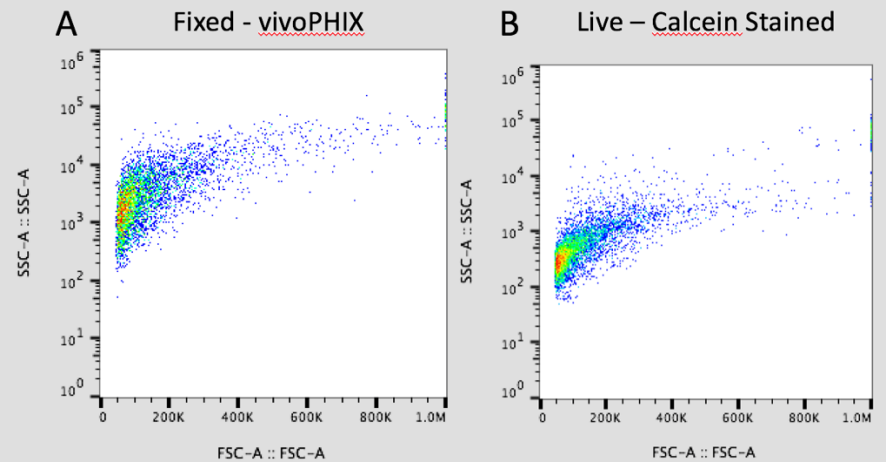

Fig. II.5 FACS of vivoPHIX vs. live calcein-stained hemocytes. (A-B) Pseudo-colouring refers to event density, going from light blue to green and finally red with higher density. Both samples were prepared with the pneumatic injector **(A)** Hemocytes fixed with vivoPHIX as of methods protocol, total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

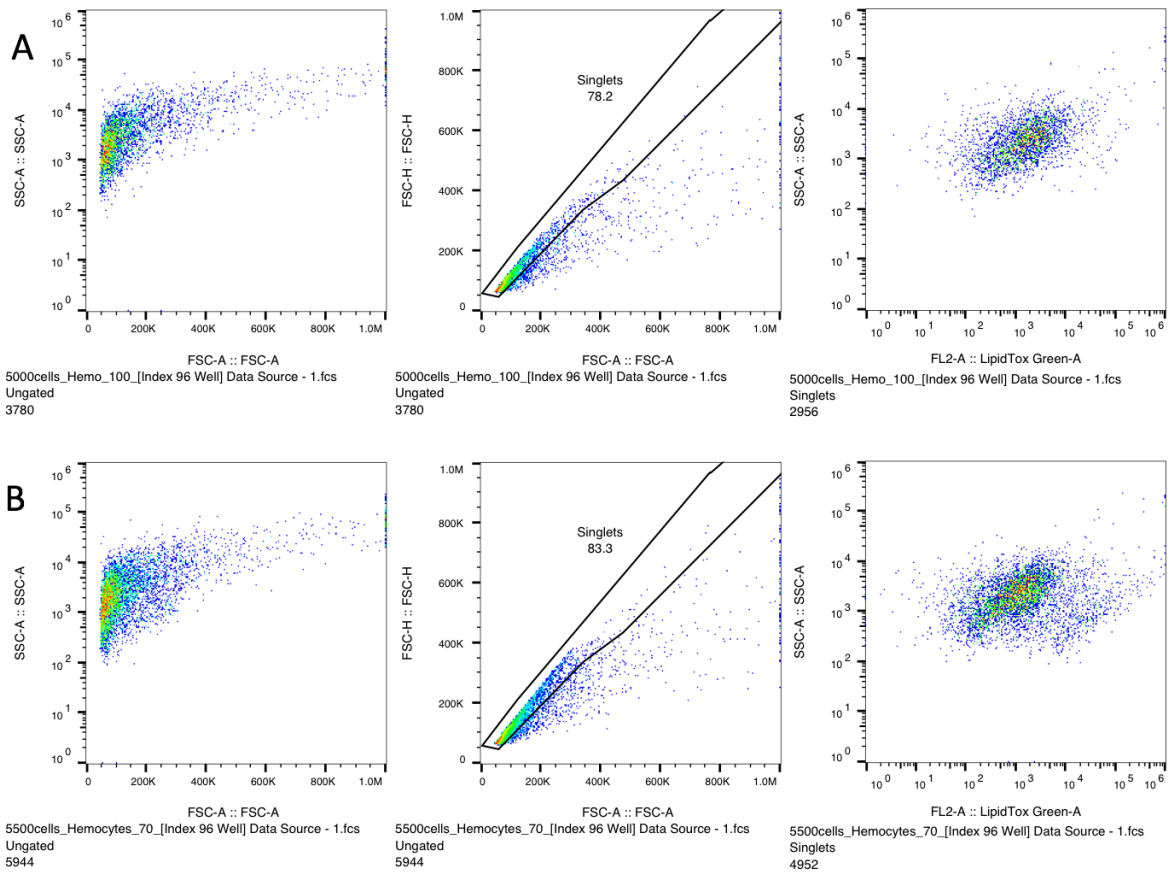

Fig. II.6 FACS analyses of vivoPHIX fixed cells stained with LipidTox show few adipocytes with 100% ethanol spin-down. To the left, side scatter area (granularity) vs. forward-scatter area (granularity). In the middle, forward scatter height (size) vs. side-scatter area (granularity), where the straight-line 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).

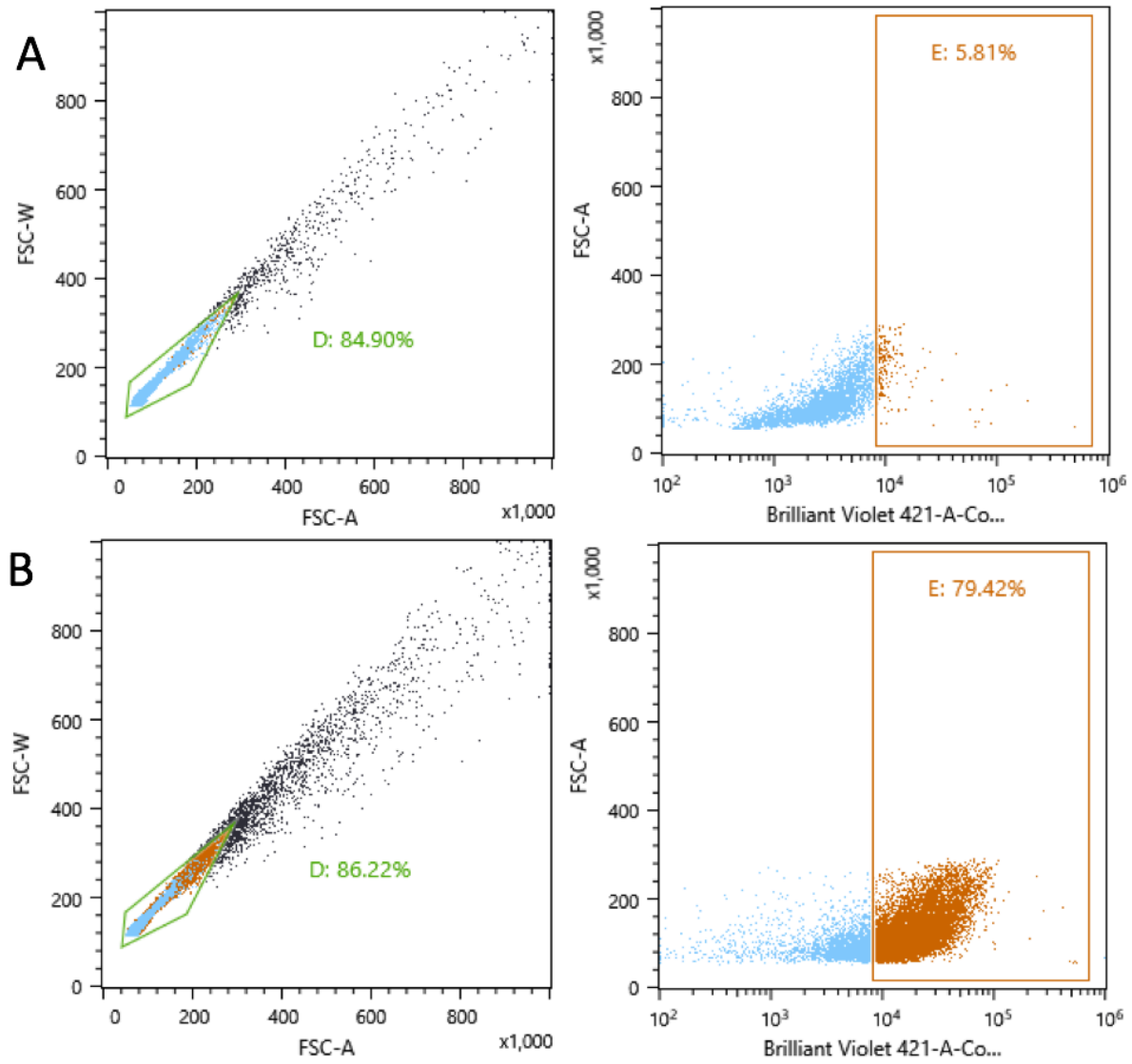

Fig. II.7 FACS of vivoPHIX double-fixed and Hoechst 33342 stained hemocytes. Cells fixed with vivoPHIX followed by secondary fixation with acetic acid and stained in water for 15 minutes with NucBlue Reagent as of methods section **(A)** Non-stained control. To the left, forward scatter (size) vs side-scatter (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].

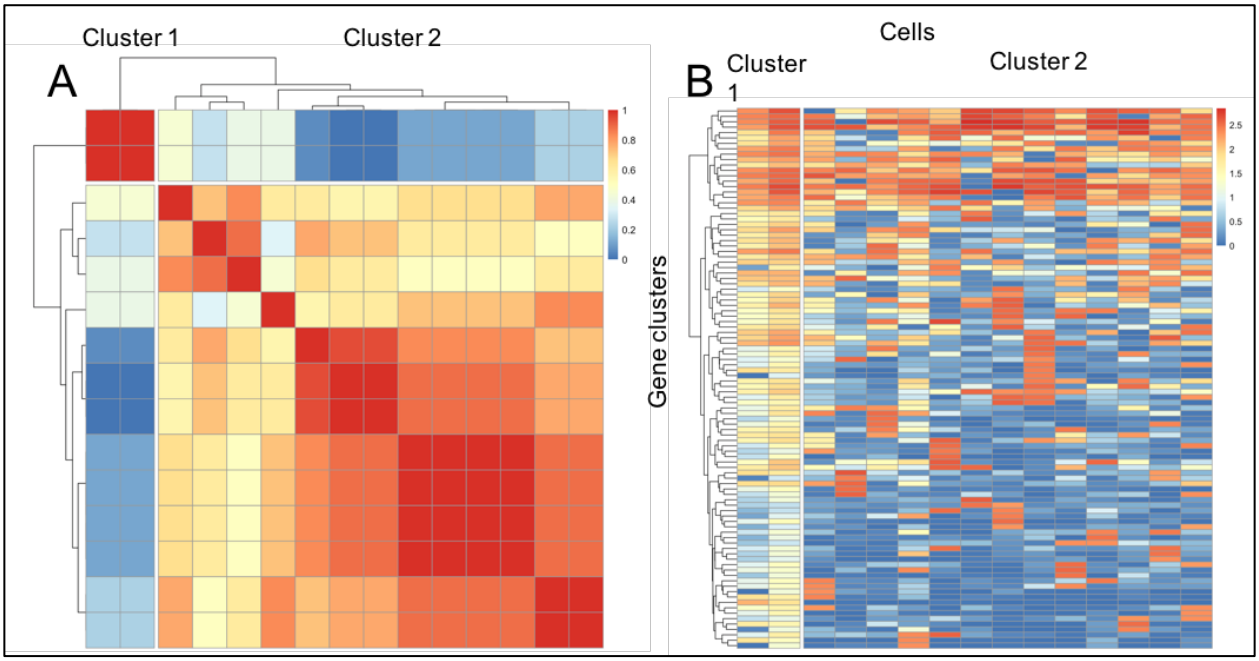

Fig. II.8 scRNA-seq with Smart-seq2: hemocytes cluster into two main groups. All data shown after QC and normalization. **(A)** Cell clustering with Single-Cell Consensus Clustering (SC3) [289]. Matrix shows percentage of times cells were assigned to the same cluster by different parameter combinations, 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.

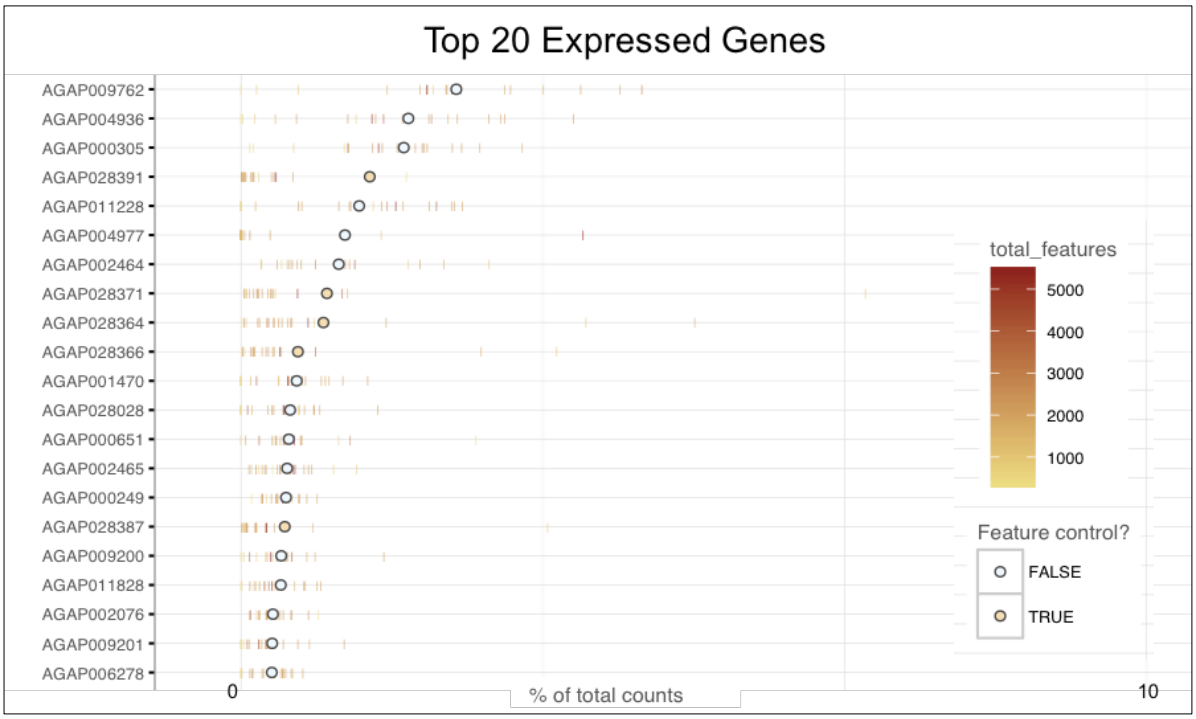

Fig. II.9 scRNA-seq from Smart-seq2: top expressed genes. Note the relatively flat distribution which typically indicates good coverage. Yellow (feature control) are mitochondrial genes. Proportion of reads mapping to mitochondrial genes may are useful for identifying low-quality cells as in broken cells 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.

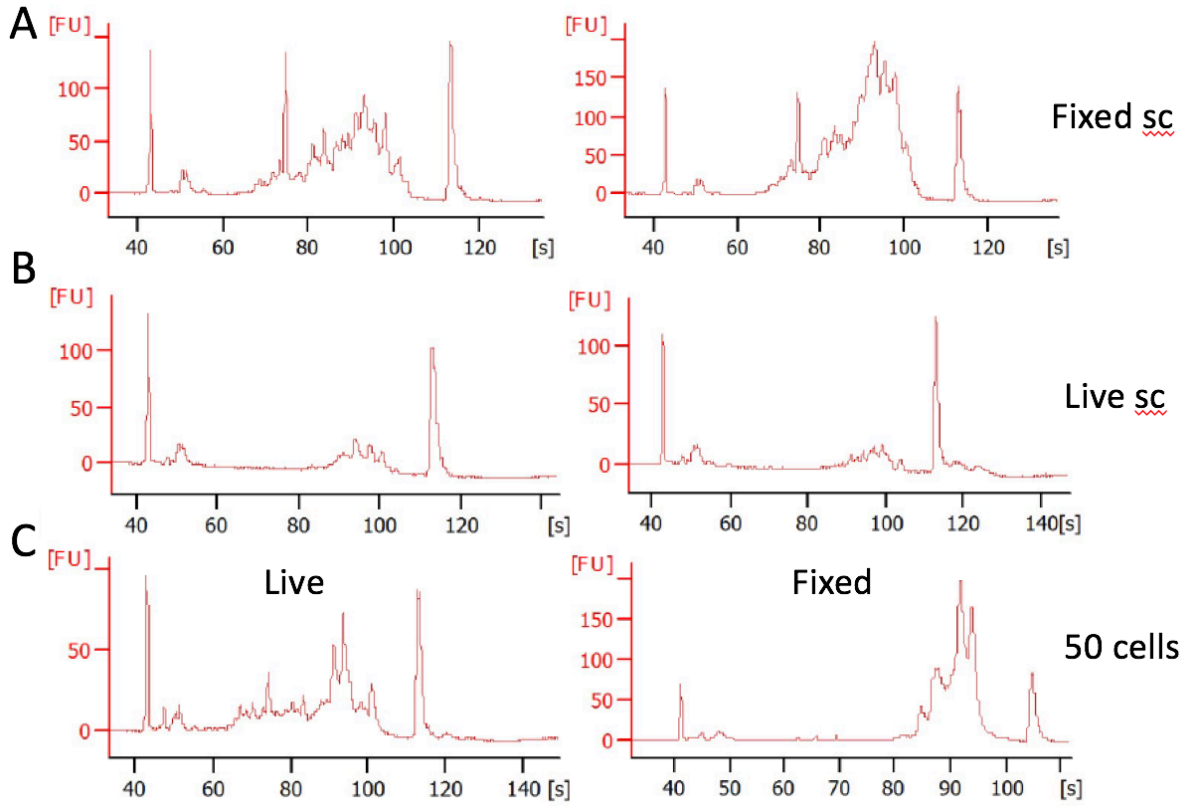

Fig. II.10 Bioanalyser traces from fixed and live hemocytes after RT. Heightened [FU] readings indicate higher amounts of cDNA **(A)** To the left and right representative examples of cDNA traces from vivoPHIX-fixed single cells **(B)** To the left and right representative examples of cDNA traces of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

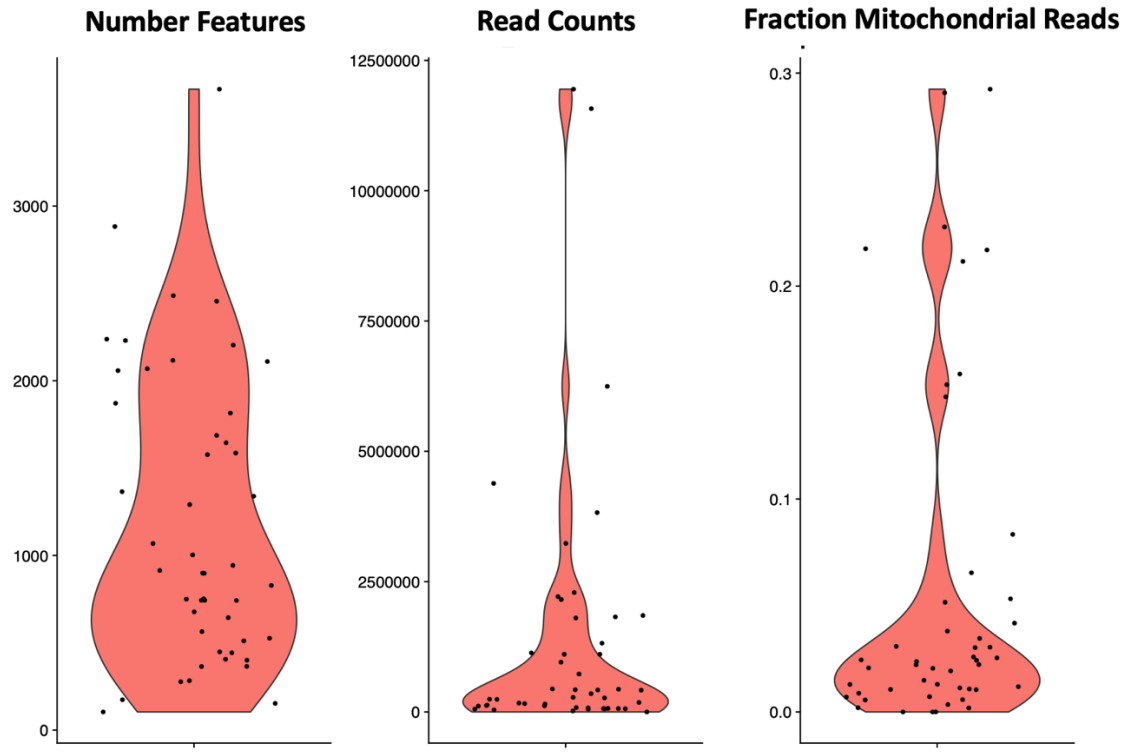

Fig. II.11 scRNAseq QC (Smart-seq2) with Seurat. Violin plots showing QC metrics for both Smart-seq2 libraries combined. To the left total number of features per cell. In the middle total number of reads per cell and to the right the ratio of total reads in mitochondrial genes.

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.

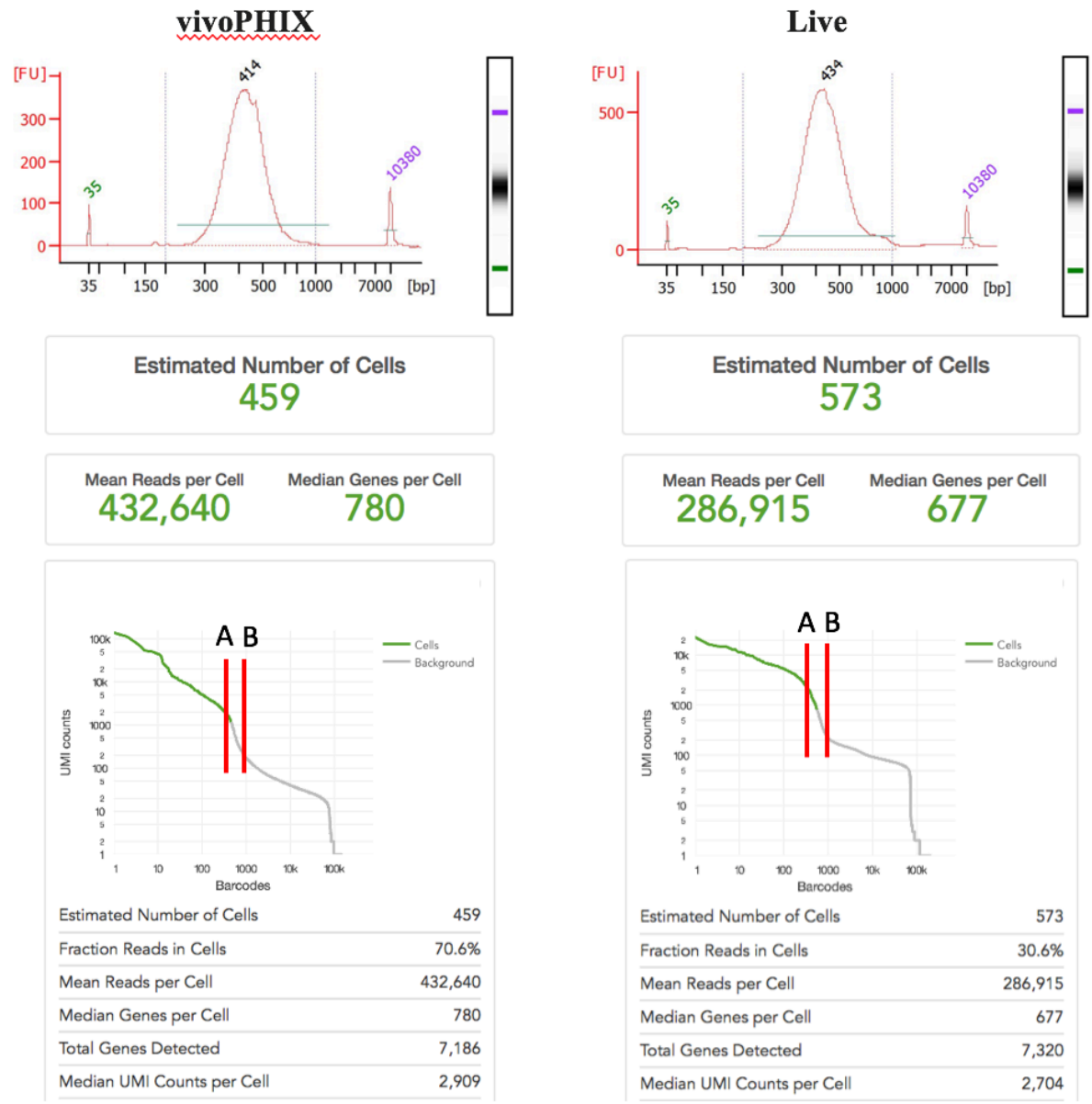

Fig. II.12 Library traces and Cell Ranger statistics: vivoPHIX vs live hemocytes. To the left vivoPHIX sample metrics, to the right the live calcein sorted cells. At the top Bioanalyser profiles for the libraries, and below Cell Ranger pipeline results. Marked as **(A)** are the cell Ranger default settings 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.