UNIVERSITÀ DEGLI STUDI DI TORINO

Dipartimento di Scienze cliniche e biologiche

Dottorato di ricerca in Sistemi Complessi per le scienze della vita

Ciclo XXX

Fronte

Characterization of vaccine-induced CD8+ T cell responses through a single-cell gene expression analysis procedure

Tesi presentata da Fabiola Blengio
Tutors Prof. Duccio Medini
Dr. Emilio Siena

Coordinatore del Dottorato prof. Michele De Bortoli

Anni Accademici 2014/2015 – 2016/2017
Settore Scientifico Disciplinare BIO/11

UNIVERSITÀ DEGLI STUDI DI TORINO

Retro e lati

n. file contenuti 1
spazio su disco 690 mb

(Spazio riservato all'Ufficio:)

prot. n. _____ del ___ /___ /____

UNIVERSITA DEGLI STUDI DI TORINO

DOTTORATO DI RICERCA IN SISTEMI COMPLESSI PER LE SCIENZE DELLA VITA

XXX CICLO

TITOLO

Characterization of vaccine-induced CD8+ T cell responses through a single-cell gene expression analysis procedure

Author

Fabiola Blengio

GSK coordinator

Emilio Siena

University coordinator

Duccio Medini

Table of Contents

1 - Introduction

1.1 Aim and rationale of the study

Vaccination is one of the greatest achievements of modern medicine and consists of the injection of a biological preparation capable to induce active acquired immunity to a particular infectious disease[1]. Vaccine efficacy varies as a function of many variables, including the type of vaccine (e.g. live-attenuated or subunit vaccines), the adjuvant that is introduced into a vaccine to boost immune response, population demographic and geographic attributes. The ultimate vaccination outcome is represented by the degree of protection from the disease it can elicit. Our understanding of the immune mechanisms underlying vaccine's efficacy is still limited and we do not know why vaccines sometimes fail to protect a certain percentage of the recipient population.[2,3] Moreover, vaccines efficacy is known to vary considerably in the human population depending on several environmental factors and genetic predisposition.

For most vaccines, the most reliable correlate of efficacy is represented by the magnitude of antigen-specific antibody titers in blood after vaccination.[3,4] However, protection against organisms displaying a high level of antigenic variation, such as the HIV virus, or complex host-pathogen interaction biology, such as Plasmodium falciparum, requires the activation of multiple arms of the immune response and in these cases, the production of neutralizing serum antibodies may not be a suitable correlate of protection.[5,6] The lack of well-defined correlates of protection, which could unravel the mode of action of protective vaccines, has been in many cases a major impediment to the successful development of new and improved vaccines. This was primarily due to the low throughput of conventional analytical approaches that have been applied for profiling immunological responses to vaccination.

Recent years have seen the rising of system biology approaches initially using gene expression analysis of whole blood in vaccinated subjects.[7–9] Advances in DNA sequencing technology were developed to analyze immunoglobulin (Ig) and T-cell receptor (TCR) repertoires responding to vaccines. Such approaches hold the potential to discover new correlates of protection. Systems biology offers the unique possibility to analyze the complex network of immunological events after

vaccination. Single-cell approaches can be used to study immune response induced by vaccination and compare effects induced by different adjuvants.

Evaluation of the expression kinetics of key genes can predict the outcome of a vaccination. Thus, it is possible that the level of immune competency can be evaluated based on the profiling of immune-related molecules, cells and tissues. Such information before vaccination could be used to select the type and dosage of the vaccine, and information generated after vaccination would allow for the measurement of the T cell response and would enable us to predict whether a long-lasting memory T cell response is achievable and also to avoid unfruitful or even harmful consequences of vaccination.[2,10]

This thesis focused on the transcriptome profiling, for a selection of genes, of CD8+ T cells responses to heterologous vaccinations using the BioMark Fluidigm HD System. Subcellular visualization of RNA turnover can aid in gaining new insight into gene expression regulation. Expression of specific combinations of biomarkers gives the opportunity to identify new correlates of immunogenicity and further understanding the immune cellular mechanism of action following vaccination. These new single-cell approaches will redefine our cellular classification schemes, potentially revealing new and functionally distinct subsets of cells.

1.2 Vaccinology and vaccine science

Development of immune resistance to any infection is based on proper priming of immune cells after infection or after vaccination. Vaccines represent one of the most cost-effective interventions to prevent a primary infection and are administered to millions of people every year. Vaccines function by inducing protective cellular and humoral immune responses against the targeted pathogen and come in different forms, including live viruses, inactivated bacteria, polysaccharide and subunit vaccines. Vaccination results in a precisely synchronized perturbation of the immune system and, as such, represents a convenient mean to probe the human immune system. After vaccination cells are stimulated and activated in such a way to decrease subsequent infections form the targeted pathogen or to at least reduce the severity of the disease.[2,11–13]

1.2.1 Vaccines and correlates of protection

A central goal of vaccine research is to prospectively identify whether vaccination is able to confer protection from infection or disease. Correlates of protections can be used to develop and refine vaccines design and for predicting vaccine efficacy.

In order to be effective, a vaccine should be activated multiple arms of the immune system, including innate immunity (antigen presentation, the establishment of the immune-competent environment), adaptive immunity (activation of antigen specific T- and B- lymphocytes) and immune memory (long-lived memory B cells). In addition, antibodies can target the invading pathogen by either mediating complement attack or antibody-dependent cellular cytotoxicity. Ideally, a vaccine should also promote the generation of memory cytotoxic T lymphocytes, which in turn can limit the spread of infection by recognizing and killing infected cells or secreting specific antiviral cytokines. Vaccination should also elicit an immune memory, including the production of memory CD4+ T cells, which participate in the reduction, control and clearance of pathogens by producing cytokines that support activation and differentiation of B cells and CTL, while limiting the recruitment/expansion of regulatory T lymphocytes (Treg), which can suppress the activation of the immune response against vaccine's antigens. Finally, elicitation of long-term immune memory is also desirable.[2,4,5]

A limitation is represented by the fact that traditional methods used to measure vaccine immunogenicity are not the best option to predict vaccine efficacy. Tools to measure immune responses to a pathogen were represented by in vitro assays such as ELISA, neutralization and interferon γ release assays. These map in fine details the antigen recognized. However, response to a particular antigen does not always have a primary role in protection. Conversely, an antigen that is critical for activating a protective response may be not detected by the aforementioned in vitro assays.

New methods have emerged to assess vaccine-induced immune responses. One of the main interests has been the identification of molecular and cellular markers able to correlate and predict vaccine-induced protection. The primary surrogate of vaccine efficacy has traditionally been the antibody titer to vaccine antigens or the measurement of antibody function such as antiviral neutralizing activity. Moreover, the measurement of T-cell functionality, with or without antibody measurements, has

been used to assess vaccine efficacy (e.g. simultaneous measurement of intracellular cytokine production and cell phenotype using flow cytometry, binding of tetramers to cell surface receptors, measurement of epitope immunoreactivity using the ELISPOT). There is the need of new biomarkers that can evaluate T-cell functions (e.g. memory, helper, effector), as well as T-cell interactions with other cells of the immune system. It is important to develop methods that assess vaccine efficacy at the individual immune cell level rather than measuring the total immune response. To this end, a deeper knowledge on the molecular mechanisms of immune responses, the availability of high throughput genomic and proteomic technologies and the development of integrative computational analysis offer new approaches for modeling vaccine-induced immune responses and open the possibility to establish predictive signatures of effective responses.[6,7,14]

1.2.2 CD8+ T cells mediated immunity

T lymphocytes are key players of adaptive immune response. There are two main populations of T cells, defined by the expression of membrane glycoprotein CD4 ($CD4^+$ T cells) and CD8 ($CD8^+$ T cells).

$CD8^+$ T cells play a key role in adaptive immune response. After infection, antigen presenting cells activate $CD8^+$ T cells by presenting pathogen-derived peptides bound to major histocompatibility complex class I (MHC-I). CD8+ T cells proliferate and develop cytolytic functions and capacity for rapid cytokine production. Cytotoxic $CD8^+$ T cells have important roles in the clearance of intracellular pathogens and tumors.[15–17]

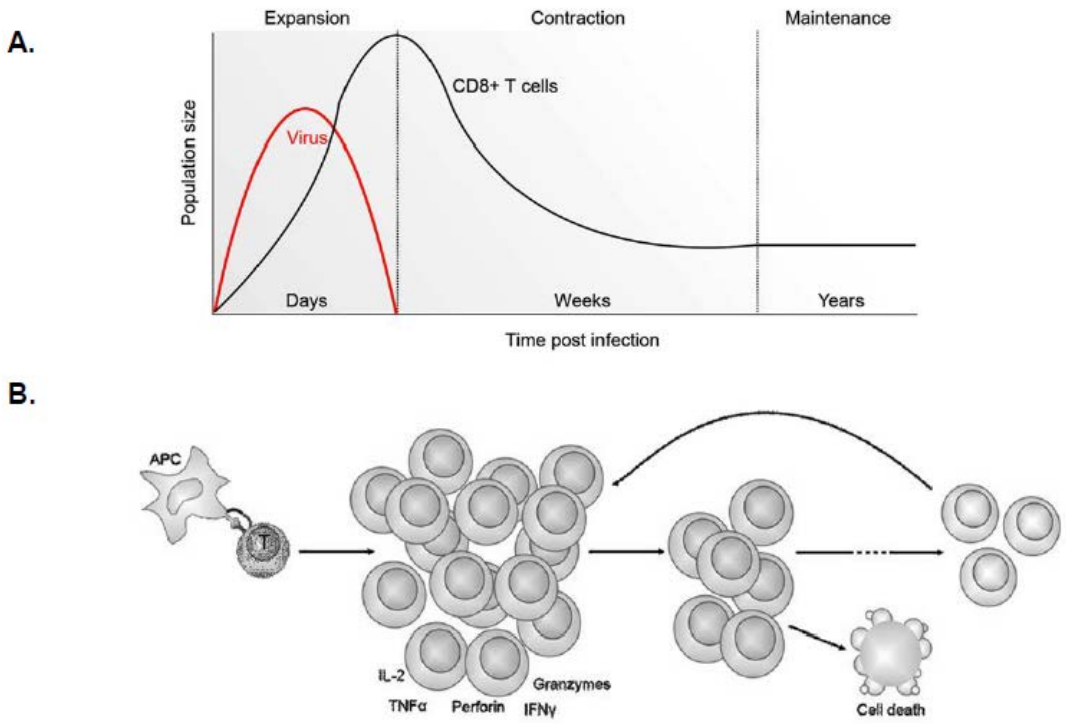


Figure 1: **CD8$^+$ T cells kinetics**

A. Kinetics of CD8$^+$ T cells responses after viral infection. **B.**Antigen presenting cells (APC) present antigens to CD8$^+$ T cells. Cells undergo on clonal expansion and differentiateinto effector CD8$^+$ T cells capable of lineage-specific effector functions, including the ability to secrete pro-inflammatory (TNF-α, IFN-γ) and cytotoxic (perforin, granzyme) molecules. Following viral clearance, the CD8$^+$ T cells undergo an extensive contraction phase, mediated by programed cell death. The remaining memory CD8$^+$ T cells can persist in the host for years.

Source A,B: adapted from[6]

Three phases can be distinguished in $CD8^+$ T cells response: expansion, contraction and memory phases. Antigen-presenting cells, such as dendritic cells, will present those antigens to naïve $CD8^+$ T cells leading to their activation, differentiation, and expansion. Many of those pathogen specific $CD8^+$ T cells will then enter the blood and migrate to sites of infection. Following the elimination of the pathogen, $CD8^+$ T cells will undergo a massive contraction. Most effector $CD8^+$ T cells will die and only a small percentage (~5–10 %) will survive and form the memory $CD8^+$ T cell pool that will protect the individual from a secondary infection. Memory $CD8^+$ T cells can provide long time protection, fast immune response and viral clearance in case of secondary infections (Fig 1). After stimulation, effector antigen-specific $CD8^+$ T cells will expand and migrate to the site of infection, where they kill virus-infected cells expressing MHC-I/peptide complexes. Upon antigen-specific recognition, $CD8^+$ T cells release cytotoxic molecules at the site of cell contact and kill cells blocking further spread of the intracellular pathogens.[18]

Cytotoxic effector $CD8^+$ T cells are an essential component of immune response and are characterized by the production of cytotoxic molecules perforin and granzymes, as well as cytokines such as IFNγ, TNFα and IL2. High level of some transcriptional factors such as T-bet, Blimp-1, Id2, Gf-1 and XBP promote effector $CD8^+$ T cells terminal differentiation that give rise to short lived effector cells (SLEC) that express KLRG1 receptor. Effector cells that express markers of terminal differentiation but are able to survive to memory transition tend to maintain higher expression of cytotoxic molecules longer after pathogen clearance and can be considered effector memory (Tem) cells.[19–21]

After resolution of acute infection, most antigen-specific specific $CD8^+$ T cells undergo apoptosis while a small percentage of cells survive and give rise to a memory population. In general, memory $CD8^+$ T cells are long-lived populations and persist in the absence of antigen but maintain a distinct phenotype and elevated precursor frequency which is one way to distinguish them from the naïve $CD8^+$ T cell population. Memory T cells stimulated by pathogen reencounter respond consistently preventing host sickness. Memory T cells circulate in the blood or are found in tissues. A specific homing to the bone marrow have been described as a reservoir of antigen specific $CD8^+$ T cells.[22,23]

Two main subsets of memory $CD8^+$ T cells have been described. Central memory T cells (Tcm) defined as $IL7r^{high}$ $CD62L^{high}$ $CCR7^{high}$ which have a high and robust

recall rate with production of IL2 and effector memory T cells (Tem) IL7r^{low} CD62L^{low} CCR7^{high} considered to be fast producer of cytotoxic proteins. Tcm have a much higher proliferative potential than Tem. [24]

1.2.3 Current Influenza Vaccines

Influenza is a viral infection that affects mainly nose, throat, bronchi and, occasionally, lungs. In young and elderly population, and non-immunocompetent people, infection can lead to severe complications, including pneumonia and death. Vaccination represents on of the most effective interventions against influenza infection.[25,26]

Human influenza viruses (IVs) are enveloped virus, with a lipid bilayer encompassing an 8-stranded negative sense RNA genome that encode 12 distinct proteins. They comprise three distinct families: A, B and C. Influenza C viruses (ICV) generally cause a mild infection and are not considered a significant risk to population health. Influenza A (IAV) and B (IBV) are responsible for seasonal epidemics. IAV causes major infections and is the only subtype associated with pandemics. IAV are categorized based on their surface Haemagglutinin (HA) and Neuraminidase (NA) glycoproteins, which are embedded within the lipid bilayer envelope and coat the entire surface of the virion. The specific combination of HA and NA glycoproteins determines the subtype of a particular IAV. There are 18 distinct HA and 11 distinct NA glycoproteins. Accumulation of mutations in HA and NA glycoproteins can mediate evasion of pre-existing immunity. Pandemics may occur when a novel IAV strain or subtype typically generated by antigenic shift gains capacity to transmit in humans. IAV, influenza A virus is major cause of worldwide morbidity and mortality. T cells immunity is a key factor in limiting severity of disease particularly when antibody activity is ineffective.[11]

Many licensed influenza vaccines are available today (live attenuate, inactivated, subunit) with new vaccine candidate under development such as vectored, DNA or RNA vaccines.[27,28] Licensed inactivated influenza vaccines (IIV) include the hemagglutinin (HA) viral surface protein inducing strain specific antibody response that protect against closely related viruses. Seasonal influenza vaccines target each year three to four influenza virus strains, typically two influenza A (H1N1 and H3N2) and one or two influenza B strains (Victoria or Tamagata lineages). These are

delivered as inactivated split/subunit or live attenuated influenza vaccines.[29] Most adults have already an immunological memory against influenza antigens, meaning that the existing pool of memory T cells are able to expand appropriately against influenza infection or vaccine. However, adjuvants such as MF59 ™ are licensed to improve immunogenicity in the elderly and in infants, who do not develop a strong enough immune response against seasonal influenza vaccine[30–32]. MF59 ™ is an oil water emulsion that interacts with the antigens through electrostatic charges, and stimulates an immunogenic favorable environment for the easier uptake of the antigens at the site of injection[33]. MF59 ™ has been reported to increase the immune response against the seasonal and pandemic influenza vaccines. there are cases, however, where even adjuvanted vaccines are not capable of inducing a protective response.[26]

Despite the fact that production of neutralizing Ab, in response to vaccination, is critical to reduce the rate of infection, also the stimulation of a robust T cells response is important for viral clearance after infection.[34–36] It has been shown, both in humans and in animal models, that natural influenza virus infection confers protection through CD4 and CD8 T cells mediated immunity. Conversely, inactivated influenza vaccines induce antibody targeting the HA protein, while live-attenuated influenza vaccines (LAIV) was reported to rely more on cellular and mucosal immunity.

Adjuvanted IIV promote a strong HA-specific CD4 T cell helper response, which provide an helper function for the subsequent HA-specific antibody response. In addition, memory CD4 T cells also exert a direct effector function through the production of IFN-gamma and perforin.

These current vaccines suffer for some limitations. Firstly, traditionally available egg-based influenza vaccines follow a complex manufacturing process and might induce allergic response among some individuals. Vaccines need to be updated every year and they are delivered in the fall –winter period in the north hemisphere.[30,33,37]

1.2.4 T cells and influenza

Current IAV (Influenza A virus) vaccines elicit humoral immunity directed toward the surface HA and NA glycoproteins and are highly effective in the control of IAV infection. The antigenic drift causes rapid mutation of surface glycoproteins so

humoral immunity established against one IAV strain does not protect against subsequent infection caused by heterologous strains.

$CD8^+$ T cells rapidly recognize the more conserved internal proteins of IAV and thus have the potential to be strain-cross protective.[27,38]

Natural influenza infection confers protection against virus through $CD4^+$ and $CD8^+$ T cells mediated immunity. However, protective immunity induced by most inactivated vaccines (IIV) has been correlated with antibodies. Adjuvanted IIVs promote a strong HA-specific $CD4^+$ helper T cell response, which improves the cross-neutralization activity of HA-specific antibodies through the expansion of naïve B-cells. Memory $CD4^+$ T cells may also exert a direct effector function through the production of IFN gamma and perforin and the activation of innate response. Activated $CD4^+$ T cells play a pivotal role in supporting germinal center formation that results in affinity maturation and isotope switching. $CD4^+$ T cells provide help to B cells and $CD8^+$ T cells, IAV specific $CD4^+$ T cells also have the capacity to target directly IAV-infected cells, thereby contributing to the control and elimination of IAV infection. [34]

$CD8^+$ T cells kill virus-infected cells and they are important in IAV infection as they have the potential to protect against strains different than those included in the vaccine. While $CD4^+$ T cells provide secondary signals to optimize the mechanism of action, $CD8^+$ T cells express a range of effector genes, such as granzymes and perforins, which mediate their cytotoxic properties. In addition to cytotoxic mediators, $CD8^+$ T cells effector functions also consist in the production of a variety of pro- inflammatory cytokines.[34]

Activated CD4+ T cells are also key cells in supporting germinal center formation that results in affinity maturation and isotope switching. CD4 + T cells provide help to B cells and CD8 + T cells, IAV specific CD4 + T cells have also the ability to target directly IAV-infected cells and contribute to the control and elimination of IAV infection. CD4 + regulatory T cells (Treg) have been shown to limit effector CD8 + T cells and prevent tissue damage caused by prolonged CD8 + T cells response at later stage of infection.[34]

1.2.5 Self-amplifying messenger (SAM) vaccine platform

Messenger RNA-based vaccines have been investigated extensively in animal models of infectious and non-infectious disease. Like viral vectors and DNA

vaccines, mRNA vaccines induce both humoral and cellular immunity. In addition, antigen expression is transient avoiding T cells exhaustion that may occur with persistent antigen exposure.[39]

A novel RNA-based vaccine platform, called self-amplifying messenger (SAM), is based on a synthetic self-amplifying mRNA delivered by a lipid nanoparticle (LNP). The delivered messanger RNA replicons-based is processed by hosts cells and vaccine antigen is expressed without generating virus particles. SAM-based vaccines are able to elicit potent immune responses. They have been shown to be immunogenic in mice at low doses and to elicit antibody responses comparable to those induced by a licensed influenza vaccine.[25]

SAM main advantages include a completely synthetic process that do not require cell culture and can be automatized, a robust and generic process to generate vaccines against any influenza strain and a rapid progression from gene sequence to vaccine product.[40]

mRNA vaccines elicit both innate and adaptive immunity. Host's intracellular machinery interacts with mRNA delivered at multiple levels, as suggested by studies reporting that both mRNA and encoded antigen are detectable at inoculation site and in draining lymph nodes, shortly after immunization.[39]

Nucleic acids are formulated with two different delivery systems. A cationic nano-emulsion (CNE), consisting of hydrophobic/hydrophilic surfactant composed with squalene and DOTAP, a cationic lipid substance that allows the nucleic acid binding to the formulation. This formulation prevents RNAse mediated degradation and helps delivering the replicons to the target cell. mRNA is first transported to the cytoplasm by endocytosis. Replicons then engage the ribosomes in translation of the encoded antigen and replication of the replicon. This technology elicite broad, potent and protective immune response that are comparable to the viral delivery technology, but without the need of viral vectors. [25,40,41]

Ferrets immunized with SAM encoding H1 (SAM(H1)) vaccine showed a better response profile compared to animals immunized with HA alone or MF59 + HA. In addition, SAM(H1) immunization of BALB/c mice resulted in cytokine CD4 cells T-helper (Th1) phenotype (IFN-γ, TNF, IL-2), while protein and protein+MF59 induced cytokine CD4 cells with a Th2 phenotypes. Only SAM(H1) vaccinated mice induced cytotoxic CD8 T cells (CD107, IFN-γ, TNF, IL-2). [25]

SAM platform enhances a very potent and broad-based immune response due to antigen expression in host cells and to their intrinsic innate immune stimulating capabilities.[41]

1.3 Single-cell analysis approaches

1.3.1 Heterogeneity in biology

Traditionally biological experiments are performed on "bulk" samples containing populations of thousands of cells. Bulk samples average differences and, as a result, they cannot distinguish among different cell subtypes. It is known that a biological system is represented by different cell types that have different roles and functions.

In recent years, it has become evident that cells are structured in a continuum state rather than strictly defined sub-populations. Even a well-defined population of cells consists of multiple subpopulations that work together in a complex landscape of conditions. Single-cell level analysis allows the characterization of various single cell state connected to each other through a landscape of cells in transition from a state to another. [42,43]

Single-cell diversity is driven by unequal distribution of cellular content during division, epigenetic modification of DNA, fluctuation of transcription and alternative splicing at protein level.

Heterogeneity allows cells to specialize in performing different functions with higher efficiency and the immune system is a remarkable example of this aspect. In fact, immune cells exist in a continuum of differentiation and activations' states rather than in distinct subsets. .[43]

Single-cell technologies provide the means to dissect heterogeneity. Both the huge cell-to-cell variability and the amount of data these new technologies have brought to light, constitute a great resource and at the same time a challenge for scientists. The parallel development of more sophisticated algorithms, faster computational approaches and new data visualization methods has already allowed scientists to gain new insights into immune system diversity.[44–46]

One of the first immunological studies focusing on single-cell analysis was published by Newell et al. in 2012[47]. The authors have shown that $CD8^+$ T cells could be classified into more than 200 types by single-cell mass cytometry analysis (CyTOF technology) and their data have been used to make the general hypothesis that

combinatorial diversity in functional capacity gives each $CD8^+$ T cell a whole immune potential. Newell et al have shown that an unsupervised simultaneous analysis of 25 phenotypic and functional properties of human $CD8^+$ lymphocytes generally agrees with previous classification schemes but also shows how these subsets represent nodes on a continuum of T cell phenotypes. These analyses also describe a memory cell phenotypic progression involving progressive gains and losses of surface markers and phenotypic capacities and they concluded that was important for defining the degree of T cell differentiation and exhaustion. In conclusion, these data showed that cytokine expression were not confined to specific subsets of CD8 T cells but it is much more combinatorial, allowing great flexibility in orchestrating an effective pathogen response.[47]

1.3.2 RNA at single cell level

One of the most important determinants of heterogeneity is represented by RNA. The transcriptome at single cell level is greatly informative both for cell function and phenotype. The transcription process in eukaryotic cells is characterized by short burst of activity followed by a period of quiescence in which the level of mRNA decays[48]. Gene expression in individual cells is not synchronized and gene expression noise is the results of cell-to-cell variation given by temporal fluctuations of mRNA numbers. Majority of genes is represented by very little mRNA transcripts, which increases the impact of stochastic events. Moreover, bursting kinetics appear to be gene specific and these random pulses of transcriptional activity contribute to the considerable heterogeneity of single cell. [49]

Small copy number and short life time of mRNA led to a great stochasticity and noise at single cell level determine the phenotype of the cell. Expression noise can be a disadvantage by affecting the precision of biological function but is also an advantage by enabling heterogeneous stress-response programs to environmental changes.[50]

Transcription at single cell level was previously studied using the bacteriophage coat proteins (MS2) that bind to specific RNA sequences. Single mRNA copies were visualized by mRNA-MS2-GFP complexes. This method revealed that the intermittent production of transcripts. In fact, short bursts (6 min) were observed followed by long periods (37 min) of inactivity[49]. Distribution of mRNA copies within a cell can be modeled by Poisson statistics, where the mean is equal to

variance. This model is based on the assumption that a gene can exist in two states: one where is OFF (activity is negligible) and one where is ON (certain probability of activation). The second state is where the transcription occurs (bimodal distribution). Additionally, probabilities of transcription initiation change at different phases of the cell cycle.[49,51]

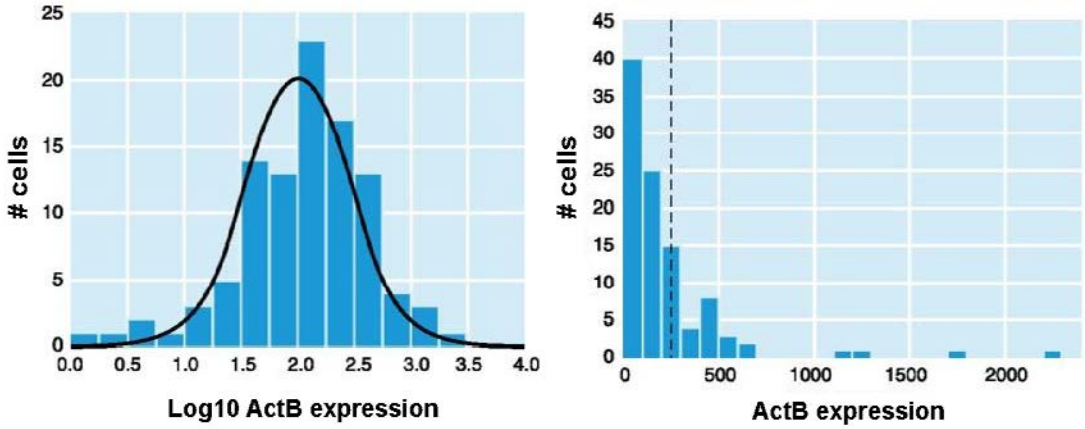


Figure 2: **mRNA at single-cell level**

Single-cell expression levels of ActB mRNA follow an approximate Gaussian distribution on a log scale (Left). A similar representation, in which the X axis has a linear scale, reveals a highly skewed distribution (Right). The dotted, vertical line corresponds to the mean expression level, illustrating that the mean expression level does not correlate well with the expression in the majority of the cells. Source: adapted from[48].

In steady state the probability distribution of mRNA molecules can be describes by a Poisson distribution.[52] A simple stochastic model that is widely used in analyzing bursting in gene expression is the random telegraph model that takes into account the switching of promoter between transcriptionally active (ON) and inactive (OFF) states.[52,53] This model considers the arrival of bursts as a Poisson process. Correspondingly, the waiting-time distribution between arrival of mRNA bursts is assumed to be exponential. [54]

Profiling the low amounts of mRNA contained within individual cell typically requires more than a million-fold amplification, which leads to severe non-linear distortions of relative transcript abundance and accumulation of nonspecific byproducts. Low starting amount also makes it more likely that a transcript will be "missed" during the initial reverse transcription step, and consequently not detected during next processing[55]. Moreover, a gene can be detected as medium/high expression in one cell but not detected in another cell. This phenomenon is called "drop-out" events and it was previously seen in single cell qPCR data by McDavid at al[56]. To overcome this problem, authors proposed a statistical model accounting for the fact that genes at the single-cell level can be on (and a continuous expression measure is recorded) or dichotomously off (and the recorded expression is zero). Based on this model, they derive a combined likelihood ratio test for differential expression that incorporates both the discrete and continuous components. Using an experiment that examines treatment-specific changes in expression, they have shown that their combined test is more powerful than either the continuous or dichotomous component alone, or a t-test on the zero-inflated data[56].

1.3.3 Overview of single-cell technologies

Single-cell approaches have a great potential for biological studies. The advent of single-cell genomics is an important turning point in the field of cell biology. This could lead to a big shift of experimental design in next years, with the potential to analyze the expression of every gene in the genome across thousands of single cells in a single experiment. Moreover, multi-omics profiling allows to dissect complex tissue and cellular lineage hierarchies in a data-driven manner, which complement the classical approaches.[57,58]

In the next paragraphs I will briefly introduce the most important current technologies.

Single- cell DNA sequencing

Single-cell sequencing will provide new understanding by studying genomes at single-cell level. [59,60]

Efficient isolation of single cells and genome amplification able to reach sufficient material are crucial step for high-quality sequencing data. Another crucial point is the data interpretation taking in account bias and errors that could have been introduced.

Single cell DNA sequencing requires three fundamental techniques that have been dramatically improved: (1) isolate a single cell, (2) amplify its genome efficiently and accurately, and (3) sequence the DNA.

Several studies sequenced and dissected cancer genomes to single-cell resolution, with the aim of understanding tumor development and progression of the disease. This revealed various chromosomal rearrangements, followed by distinct phases of clonal expansion during tumor evolution and metastasis[61]. Subsequent single-cell exome sequencing studies provided a detailed characterization of base mutations in specific genes in bladder[62], kidney[63], and hematopoietic neoplasms[64] . Interestingly, by sequencing daughter cells of a single mitotic division, the acquisition of new structure variations could be demonstrated for a breast cancer cell line[65].

Single-cell genome sequencing basically provide new insights into genomic instability and it will lead to a better understanding of the acquisition of genetic changes and the dissection of genetic content in individual cells.[61,65]

Single-cell transcriptomics methods

Assaying gene expression at the single-cell level represents a powerful, high-resolution tool for biological discovery. There are many biological questions that bulk experiments could not resolve. For instance, during early development there is only a small number of cells, each of them potentially having distinct functions and roles.[66,67] Moreover, bulk approaches may not provide insight into whether differences in expression between samples are driven by changes in the cellular composition or changes in the cellular phenotype.

Single-cell transcriptomics experiment allow a high-throughput profiling of gene expression that can potentially answer many biologically relevant questions and lead to unprecedented new discoveries in important areas of biology.[68–72]

The most important available methods for single-cell transcriptomic analysis are single-cell RNAseq, RNA-FISH and qPCR.

Single-cell RNA-Seq (scRNA-seq)

Single-cell RNA-Seq (scRNA-seq) provides the possibility to identify and characterize distinct cellular populations revealing heterogeneity in a specific tissue. Quantification of all transcripts expressed in a single cell revealing heterogeneity

given by the presence of masked rare subpopulations of cells. One of the major challenge in analysis of single cell sequencing data is represented by the noise due to the nature of single-cell transcriptomics.[73]

The aim of single-cell RNA-seq approaches is to study cell states at high resolution potentially revealing cell subtypes or gene expression dynamics that are masked in bulk population-averaged measurements. Currently published scRNA_seq protocols follow the same general workflow: single cells are isolated, lysed, RNA is captured for reverse transcription in cDNA, cDNA is pre-amplified and then used to prepare libraries for sequencing and downstream analysis. cDNA pre-amplification causes bias that limits the quantitative accuracy of scRNA-seq. Additionally, unique molecular identifiers can be used to barcode individual RNA molecules during the reverse transcription step, allowing direct transcript counting. Alternatively, exogenous RNA standards can be spiked in with cellular RNA to allow relative and absolute transcript counts.[60,74–76]

Existing scRNA-seq methods have low capture efficiency. Only a small fraction of each cells transcript is represented in the final sequencing libraries (approximately 10%) so it is unable to reliably detect low-abundance transcripts. Low amount of input material also leads to high levels of technical noise, which complicates data analysis and can mask underlying biological variation[75].

Quality control analysis is an important step that includes filtering for read quality and eliminating cells with overall low library size. Spike in can be used to model technical variability and examine relative variability in cell size for a unique molecular identifier. One important challenge is asses a proper normalization of the data.[76,77]

The applications of scRNA-seq are extensive and a proper data analysis approach is essential. Computational analysis of scRNA-seq data is a big challenge. Many tools for quantifying gene expression report the amount of reads that are associated to a gene by normalizing. After the preprocessing analysis, batch effect, dropout effect and amplifications bias are peformed. Furthermore, dimensionality reduction algorithms were used to globally visualize data, reveal cells subpopulations and infer their trajectory.[60,78] Clustering methods can be used directly on single-cell expression data to group cells by transcriptome similarity and to detect the underlying population structure in an unsupervised manner. Cell subgroups identified from such analysis can often been matched to known cell types via previously established

marker genes. Furthermore, structural analysis of single-cell data has also led to the discovery of novel cells subtypes and to the identification of new marker genes for known cell types[60]. Already, scRNA-seq has been applied for the characterization of intra-tumoral heterogeneity[79] , the detection of variation among cell states within a homogeneous population (such as differences in cell cycle stage or differential signaling responses to an outside stimulus)[80], the study of cellular transitions between different states by inferring cells trajectories[43,44]

Single-cell RNA FISH

Counting RNA molecules is always been performed by fluorescence in situ hybridization (FISH) in which a fluorescent nucleic acid probe is hybridized to fixed, lysed cells on a specific support. Fluorescence is detected via high-resolution microscopy. The amount of probe hybridized is determined by comparison to a standard dilution series of probes. This allows total count mRNA molecules per cell.[49] FISH provides an orthogonal methods of quantifying transcript levels and is often used to validate results from scRNA-seq data. Single-cell FISH preserves the spatial context of transcript and can localize molecules at subcellular resolution. Single-cell RNA FISH could supplement the global transcriptomic snapshot of scRNA-seq with information on the spatial dynamics of selected transcripts. A limitation is done by overlap between fluorophores that limits the number of transcript that could be simultaneously assayed. [42]

Single-cell qPCR – Microfluidic System (Biomark-Fluidigm)

BioMark has been released by Fluidigm Corporation[81]. It is a multiplexed RT-qPCR technique that provide a high-throughput multiplexing by a microfluidic architecture that combine samples and prime-probe sets into 9216 PCR reactions.[81] Different sample types and probing chemistry (TaqMan, EvaGreen) can been chosen for many applications. The platform is fast with an automated workflow and can be used as a validation technique or fast throughput screening. The single microfluidic device is based on PCR reactions in nanoliter volumes and a small amount of sample is required. Samples are load on a dynamic array together with the primer probe system and by using integrated fluidics circuit (IFC) they are mixed on a chip controlled by pressurized valves.[18]

Two main chemistry systems can be used: EvaGreen and Taqman. EvaGreen is a DNA-binding dye flexible and inexpensive. Drawbacks are represented by intercalation of all double stranded DNA and this could be led to false positive detection (primer-dimers, wrong amplicon) and a proper melting curve analysis is required. Taqman probing system is expensive with highly sensitive and gene-specific. Taqman probes are labeled with fluorescent reporter dye on the 5' and no-fluorescent quencher (NFQ) on the 3'. They are flanked by upstream and downstream primer pair that generates PCR products and hybridize to a complementary region of cDNA. Intensity of fluorescence expression is proportional to the number of molecules. The fluorescence is converted by software into Quantification cycles (Cq) and can be further analyzed. [81]

The Biomark microfluidic system was applied in different biological fields in the past years.

One of the first papers based on Biomark technology applied to single cell gene expression analysis was the one published by Guo et all in 2010 in Developmental Cell. In their study, the authors analyzed the expression dynamics of 48 genes on hundreds of cells to monitor the development 8-cell stage to 64-cell stage in mouse. The analysis was focused on transcription factors which are drivers of cellular fate. This approach has provided a rich dataset and subsequent analyses have focused on genes differentially expressed between cells. Key observation of their work is that there is an inner cell-specific upregulation of Sox2 and that temporal differences in Sox2 create differences in inner cells formations and, consequently, a high degree of heterogeneity. Finally, the authors highlighted if differential expression of transcript in single cells were the result of the presence of different subpopulations or caused by stochastic noise.[82]

To understand key factors that characterize iPCS (induced pluripotent stem cells) differentiation Bugamin et al. in 2012 performed single-cell microfluidic qPCR analysis to profile the expression of 48 genes in single cells from early, intermediate and fully programmed iPSCS demonstrating that cells at different stages can be separated in two defined populations with high variation of genes. In conclusion, the authors have shown that differential expression of genes at single cell level reveal heterogeneity between sister cells during early phase of the reprogramming process.[83]

In 2011 Flatz et al. applied single CD8+ T cells gene expression profiling to understand qualitatively differences between cells elicited by different prime boost vaccine regimens combinations. Antigen specific CD8 T cells stimulated by three prime-boost vector combinations encoding HIV env antigen were compared at single cell transcriptional level. Authors defined by single-cell gene expression profiling specific subset of CM and EM CD8 T cell differentially induced by different vaccine. Authors performed also a microarray analysis to validate single cell gene expression data.

Flatz et al. used this approach to better understand how different vaccines induce cells in different way and to find new approaches that can be used to discriminate immune cells.

The authors identified he smallest set of genes that could allow classification of cells elicited by different vaccine. Frequency of central memory T cells $Eomes^{+}$ vary between different vaccines. Flatz at al concluded that a single-cell transcriptional approach is able to resolve heterogeneity within cells population that a bulk approach cannot. Finally, they concluded that this approach could facilitate the design and evaluation of vaccines and enable a better understanding of protective immunity.[70]

Arsenio et al. in 2014 traced CD8 T cells during infection with the aim to discover gene that give early signal to cells to follow different fates. Tracing individual lymphocytes give them the possibility to discover genes that have effect on cells path.

Single CD8 T cells has been collected at different time points after infection with recombinant Listeria Monocytogenes. CD8+ T cells subsets were isolated at various time points post-infection: division1, days 3, 5 and 7 post-infection, short lived effector cells (TSLE), memory precursor cell (Tmp), day 45 central memory cells (Tcm/Cd62l hi) and day 45 effector memory cells (Tem / Cd62l low). A fundamental point that authors highlight was that this experimental approach would not be possible by bulk analysis. Using a classifier, they discovered genes which early expression decided different fates. Finally, authors concluded that the differential expression of Il2ra may reflect one of the earliest molecular determinants influencing memory vs effector cells fate[71]

In 2015 McHeyzer-Williams et al. applied single-cells qPCR analysis for mapping GC B cell fate within the clonal progeny of memory B cells. They developed a high resolution molecular strategy for monitoring antigen-specific differentiation in vivo.

Authors have shown that the vaccine boost elicited robust secondary germinal center reactions in a large cohort of switched memory B cells. PCA of single-cell gene expression segregate a subset of GC-associated activities in putative LZ (light zone) and DZ (dark zone) sub compartments. Analysis reveal that Cd83 and Pol segregates the secondary germinal center transcriptional program in 4 cyclic stages of GC activity. Changes in the expression of Cd83 and Polh at a single-cell level may distinguish an evolutionary mechanism that can rapidly reinitiate GC-specific transcriptional program and can rapidly remodel antibody repertoires of preexisting memory B cells.[72]

Single-cell proteomic methods

A variety of single cell proteomic tools have been developed. These tools can be distinguished into two main categories: large number of parameters measured across thousands of single cells at a given time point and monitoring some parameters in the same cells over time. One of the best methods for single cells proteomic analysis is represented by CyTOF. CyTOF (Cytometry Time of Flight) is a single-cell mass spectrometry-flow cytometry hybrid instrument that replaces fluorophores with stable mass reporters. Each antibody is tagged with a unique stable metal isotope, such as lanthanide, and the read out for each antibody can be correlated of level of antigen associated with individual cell. The mass cytometry technology is characterized by minimal background and consequently less spread around zero. [47]

The experimental workflow in CyTOF analysis is represented by cells labeled with mass-tagged antibodies are nebulized into droplets, ionized, and atomized by argon plasma. The resulting ion cloud passes through a mass filter where transition metal reporters are quantified by a time-of-flight mass spectrometry.[84] This platform offers an increased signal resolution with the use of many isotopes.

A typical single cell proteomic dataset can be formularized as a table where each row denotes a single-cell measurement and each column denotes a measured protein level across the single cells. Data are transformed into inverse hyperbolic sine function to compress values around zero, resulting in a more coherent negative population when marker of interest are not detected.

The distribution of a protein level as tabulated across many single cells is termed fluctuation of that protein that reveals the inherent heterogeneity of the cell

population. The biaxial plot of two proteins (right) can be used to identify specific subpopulations or extract protein–protein correlations.[84,85]

One of the first algorithms developed to analyze mass cytometry data was spanning tree progression analysis of density normalized events (SPADE). This algorithm uses hierarchical, agglomerative clustering after performing density-dependent down sampling. The resulting clusters can be displayed into a spanning tree or connected graphs. SPADE analysis identified distinct population clusters in each sample [86,87]

viSNE is another tool for cytometry data analysis that employees t-stochastic neighbor embedding (t-SNE) in mapping individual cells. viSNE provide a 2D view of the cells that are arranged in a way that approximates high-dimensional phenotypic similarity. Cells are grouped based on variability of the markers considered in the analysis. Markers that are highly variables between cells polarize cellular subsets. Data analysis in cytometry remains largely manual, supervised and focused on large change in magnitude of expression.[88,89]

Single-cell epigenomic methods

Analysis of the epigenome at single-cell level allow the understanding of mechanism of epigenetic regulation by filling the gap between microscopy examination and modern bulk genomics. Single cells need to be isolated with high precision and minimal epigenetic perturbation. Then chromosomal and DNA templated must be profiled with high recovery rate and minimal loss of material. This approach can provide valuable insight into the dynamics of DNA methylation and into identification of subpopulations with distinct methylation patterns.[90]

One of the most used single-cell DNA methylation profile is bisulfite sequencing that can be applied to small populations of cells or niches. Applications to bigger populations are feasible but require increase sequencing throughput. Alternatively approaches for characterizing DNA methylation are represented by single-cell restriction analysis of methylation (SCRAM), methylation-sensitive restriction analysis and single-cell quantitative PCR that are combined to facilitate profiling of methylation sates across cells. [91]

Single-cell epigenomics can allow flexible classification of cells based on prior knowledge or de novo identifications of subpopulations. This can also be used for studying correlation between the epigenetic state of unlinked loci. Single-cell epigenomics can improve understanding of epigenetic mechanism and their intricate

relationships with gene regulation, such as the causality between gene expression and epigenetic mechanism and their intricate relationship with gene regulation.[92]

1.4 Overview on single-cell transcriptomics data analysis

Single-cell genomics approach need the application of statistical and computational methods to extract meaningful information. Each cell can be represented as a point in a high-dimensional expression space. Classical approach is represented by the measure the distances between all pairs of cells, then grouping them into neighborhoods based on mutual proximity. Given that distance between points become more and more similar as the dimension of the space they reside within increases cells become equidistant and it is difficult to cluster them.[42]

Principal components analysis is the most widely used approach, and it has already proven effective in a number of single-cell genomics studies[82,83,93,94] .By projecting cells down to the first two principal components, each cell it is represented by a point in two dimensions instead of higher number. Reducing dimensionality might enable grouping the cells by type, which might be required for the other downstream analyses. The problem is that two dimensions are not sufficient to capture valuable information in single-cell experiments. In fact, a cell-state transition typically involves changes in hundreds or even thousands of genes. Inferring regulatory networks from single cell genomics allow the identification of master regulator of the transition[44].

Unfortunately, computational methods for inferring regulatory networks have been hamstrung by two major issues. The first is due to dimensionality: because there are so many possible gene-gene interactions, even large-scale experiments lack enough data to reliably predict which gene interact. The second arises due to averaging, which destroys the crucial source of variation that any algorithm needs to accurately reconstruct gene regulatory network for expression data. A big challenge is represented by the fact that single-cell experiment generate more data than the conventional RNA-seq or microarray study generate, and thus far more information than any existing network algorithm has ever been provided.[42,44]

Several algorithm and analysis frameworks were developed in the last years. Here, I will summarize an overview of the most important ones.

- MIMOSA (Mixture Models for Single Cells Assays), was developed by Finak at al. [95] to identify biomarker (or combination of biomarker) differentially expressed between two biological condition in single cell assays.

MIMOSA is based on mixtures of beta-binomials and rigorously analyzes count data derived from ICS assays but was mainly developed for univariate analysis of cell subsets, such as cells expressing a single function or a specified combination of functions. Cell counts were modeled by a binomial (or multinomial in multivariate case) distribution, and information is shared across samples through a prior distribution on the proportions. MIMOSA use dichotomized data (cells are positive or negative) After thresholding, a Boolean matrix of N cells x K biomarkers was obtained and 2^k putative cells subsets were formed. When k is large there is a combinatorial explosion of the number of subsets, and many of these might be small or even empty.[96]

A common statistical problem is to identify subjects for whom the proportion of cells expressing a specific combination is significantly different, between two experimental conditions. Samples were tested separately; no information is shared across samples. In order to discriminate between responders and no-responders a priori is written as a mixture of two beta distributions.

MIMOSA can be applied also to Biomark's data. In the case of qPCR based single cell gene expression technology, gene are recorded as expressed or not at single cell level.

- HURDLE MODEL was applied by McDavid at all to improve the detection of changes in single cell gene expression by testing both the frequency of expression and mean over the cell expressing the gene. Data analyzed were obtained from profiling 333 genes in 930 cells across three different cells line. [97]The dichotomous characteristic of the data prevented the use of typical tools for linear modeling and analysis of variance and the computational framework that they developed overcome this problem. Application of Hurdle model allowed the observation of a bimodality in single-cell gene expression wherein the expression of abundant genes is either positive or undetectable within individual cells. Hurdle model framework improves the detection of cell-cycle genes by identification of phase-dependent patterns even if G2 an M phases were clearly distinct. This framework also can be used to estimate single cells co-expression networks.

- LRT is a statistical framework published by McDavid et al [98]for single-cell qPCR analysis (microfluidic platform). The likelihood ratio test (LRT) is a statistical test of the goodness-of-fit between two models. LRT accounts for the fact that genes at single-cell level can be on (and continuous expression is measured). Mc David at al proposed a discrete/continuous model for single-cell expression data based on a mixture of a point mass at zero and a log-normal distribution. In details, three parameters characterize the expression distribution: the mean and the standard deviation (for the continuous part) and the Bernoulli probability of expression (for the discrete part). Using this model, they derived a likelihood ratio test (LRT) that can simultaneously test for changes in mean expression (conditional on the gene being expressed) and in the percentage of expressed cells and dichotomously off (and the recorded expression is zero). This model can be applied to various experimental questions.

- MONOCLE it is an unsupervised algorithm for pseudo temporal ordering of cells published by Trapnell at all [44]. Monocle was applied to a skeletal muscle differentiation model and unveiled dynamics and novel regulatory factors. The algorithm project the expression profile of each cells as a point in a Euclidean space, with one dimension for each gene. Then reduce dimensionality of this space using Independent Component Analysis. Third, construct a minimum spanning tree (MST) on the cells. Fourth, it founds the longest path through the MST that correspond to the longest sequence of transcriptionally similar cells. Finally, Monocle uses this sequence to produce a trajectory of an individual cell during differentiation. As cell progress, they may diverge along two or more separate paths. Monocle order cells by progress through differentiation and can reconstruct branched of biological process. Monocle decomposed myoblast differentiation into a two-phase trajectory and isolated a branch of non-differentiating cells. The first phase was composed by cells that proliferate actively (CDK1 positives) while the second mainly consisted of cells positive for markers of cells differentiations (MYOG). This pseudo-time ordering events are masked in bulk experiments. [44]

1.5 Common application of single cell transcriptomic analysis

1.5.1 Deconvolution of heterogeneous cell populations

Clustering by single-cell can reveal subpopulations structure and identification of cell subtypes and rare cell species.[60] Single-cell transcriptomics has been found to be very effective in provide discovery of novel cells subtypes. Examples were represented by some interesting work in which novel CD4 T cells subpopulations were discovered.[99–101]

1.5.2 Trajectory analysis of cell states transitions

Single-cell RNA sequencing can be applied to time-series experiments and cell developmental trajectories can be found. Dynamic processes analyzed could be represented by differentiations or signaling responses to external stimulus. In recent years, some computational suite such as Monocle were developed to enable branching, enable identification of lineage-specific gene expression and key genes that drive branching trajectories. Previous work reconstructed iPCS to iCM differentiation trajectory by measuring 96 genes by single-cell qPCR in 1900 cells obtained during 6 first days of differentiation. It was found that at day 2 a major lineage branching took place and after that cells were committed to a specific lineage. Cells fate decision can be reduced to HAND1-SOX17 transcriptional circuit and cKIT distribution. [42,60,102]

1.5.3 Dissecting transcription mechanism

Single-cell gene expression profiles can be compared to study transcriptional bursting and to model the kinetics of stochastic gene expression. The final aim is to understand which are the mechanisms that modulate burst size and frequency of mRNA transcription. [60,78,103]

1.5.4 Network inference

Variability between cells can be used to infer regulatory networks. Genes can be clustered by expression profile to identify modules of co-regulated genes, and information about gene-gene relationships can be used to infer gene regulatory networks or subnetworks. RNA-seq data are very noisy and separation of biological variation from background remains a problem. Another limitation is given by the fact

that the number of parameters (genes and gene interactions) exceeds the number of samples (cells). So, it becomes necessary simplifying model on the basis of prior knowledges.[60,79,104]. Comparative co-expression analysis between single-cells and bulk samples reflect distinct co-regulatory networks. The co-expressed genes in bulk analysis tend to have the same biological functions, whereas interchromosomal and protein-protein interactions are highly enriched in single-cell samples.[105]

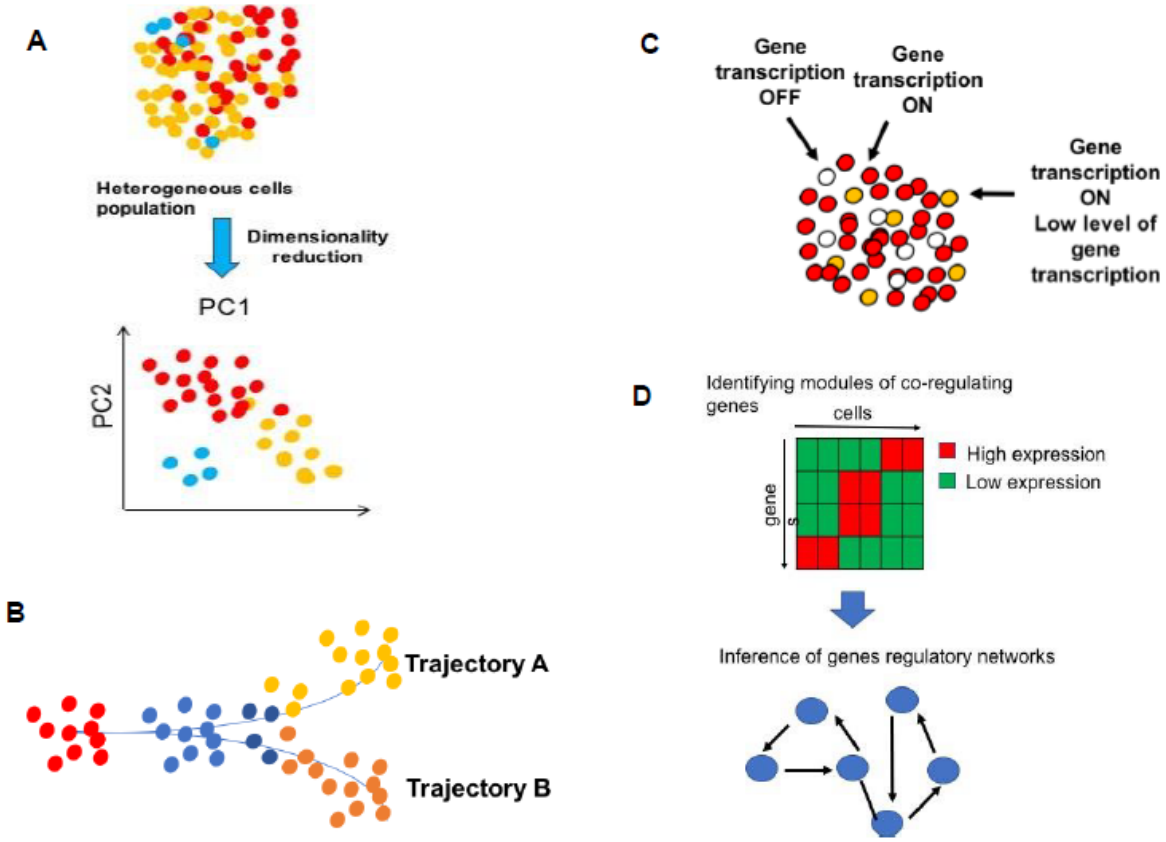


Figure 3: **Application single-cell analysis**

A. Deconvolution of heterogeneous cells population. **B.** Trajectory analysis of cell states transitions. **C.** Dissecting transcription mechanism. **D.** Network inference. Source A,B,C,D: adapted from Liu et al. [60]

1.6 Overview on single-cell transcriptomic studies

Single-cell approaches have been applied to different fields with the aim to dissect heterogeneity in immune response, tumor evolution and stem cell differentiation.

1.6.1 Heterogeneity in immune response

Single-cell transcriptomics studies have been applied to different immune cells populations. Innate and adaptive immune response depends on the proper utilization and regulation of cellular heterogeneity. Heterogeneity reflects the extreme flexibility and plasticity of immune system. In fact, evolution has selected a way to recognize a multitude of pathogens through genomic diversity of cells.[58,106]

Single-cell transcriptomics data must be visualized in a biologically meaningful way remaining robust to the high level of noise that is present in single-cell data.

Single-cell sequencing applied to CD4 T cells revealed the existence of a subpopulation of Th2 that is distinct from the rest of population by the expression of a specific enzyme (Cyp11a1) that is at the basis of steroid biosynthesis. Data obtained by single-cell transcriptomics profiling were validated by purifications of these cells by antibody directed to these new markers.[101,107]

1.6.2 Cancer evolution

Tumor is formed from a heterogeneous mass of cells with different somatic mutations and different differentiated states. Heterogeneity could be a consequence of somatic mutations and may it be itself an important or even essential contributor to tumor evolution. Single cell analysis is necessary to understand the role that heterogeneity play in cancer evolution. [108]

Single-cell sequencing is likely to profoundly impact cancer diagnostics and prognosis through the detection of rare tumor cell or through the monitoring of circulating tumor cells. Single-cell analysis could also investigate tumor subpopulations and delineate differences between primary and metastatic tumor.

A final goal of the deep understanding of cancer heterogeneity will be the contribution to develop therapeutic decisions.[108] It has been shown that human colon cancer multi-lineage differentiation represent a key source of cancer cell heterogeneity[94]. Single-cell transcriptomics approach in mouse model of leukemia identified two different sub-populations of leukemic cells, each characterized by

different co-expressed genes[109]. Another study using single-cell RNA-seq analysis in five primary glioblastomas revealed that established subtypes classifiers are variably expressed across cells within a tumor suggesting prognostic use of intratumoral heterogeneity [110]

1.6.3 Stem cells differentiation

Pluripotent stem cells (PSCs) are characterized by unlimited capacity of self-renew and the potential to differentiate into all three germ layers of the developing embryo. Single-cell transcriptomics on PSCs has provided new understanding in cellular variation and subpopulation structures. A great number of computational approaches have been designed to infer cell-cycle stages, reconstruct gene regulatory networks, characterize transcriptional stage, differentiate trajectories and understand sources of transcriptional heterogeneity. [83,111]

Single-cell RNA-seq analysis applied to mouse embryonic stem cells demonstrated that they exist in a dynamic equilibrium between states that show different differentiation fates. In particular, a model applied to these data confirmed kinetics of RNA polymerase II binding and chromatin modifications. This different chromatin state of genes affects transcriptional bursting.[78]

Analysis of induced pluripotent stem cells reveals considerable variation in gene expression between early versus later time points. Single-cell gene expression profiling has been used to distinguish cells into two groups that follow different fates in the pluripotent circuitry.[94] Single-cell transcriptomics applied to neural stem cells (NSCs) isolated from adult mice identified rare intermediate cells in the continuum of NCS lineage and machine learning approach reveled subpopulations that were experimental validated.[93]

1.7 Single-cell approach in vaccine research

The ability of single cell studies to evaluate population heterogeneity can be used to study the response to new vaccines and compare the effect of vaccines adjuvants. Moreover, single-cell analysis is useful when seeking insight into how unique subsets of cells may correlate with outcomes and when a small group of cells are essential for conferring protection (*i.e* antigen-specific CD8 T cells or B cells). Single-cell technologies will allow the analysis of such rare subsets of cells that

contribute in minor component of the total measurement and the implicit averaging of parameters in these measurements also masks the specific phenotypic state of these cells. Single-cell heterogeneity is informative. However, this information is generally lost in assays that measure cell mixtures. In vaccine research, there is a need to assess immunogenicity of a vaccine and single-cell approaches give the possibility to do that[10]. Upon vaccination, antigen present in the vaccine is taken up and presented to CD4 or CD8 T cells via antigen presenting cells. T cells that recognize antigen become activated, produce cytokines that potentiate the immune response and proliferate and persist in the immune system providing memory that can more rapidly recognize the same antigen in the future. This vaccine memory subsets it's analyzed by measuring antigen specific cytokine production in response to stimulation (ICS). Antigen-specific subpopulations represent a small fraction of the total number of CD4 and CD8 T-cells and a large number of cells must be collected (50000-100000 T cells) to ensure that they can be reliable detected. Then each cell is classified as positive or negative for each cytokine based on fixed threshold. Cell counts are compared between antigen-stimulated and unstimulated samples from a subject to identify differences. Responders are subjects whose T cells respond to stimulations and the response in vaccines is defined specific if it raised after vaccination and it was not present at pre-vaccine time point.[4,7,17,23]

MIMOSA (Mixture Model for Single Cell Assays)[95], one of the first framework for single-cell analysis developed to overcome a common problem in the analysis of vaccine research data that is represented by the identification biomarkers (or combinations) that are differentially expressed before/after vaccination, where expression is defined as the proportion of cells expressing the biomarker or combination in the cell subset of interest. MIMOSA application showed that multiple subsets can be modeled simultaneously and small biological could be detected.[95]

One of the first work of single-cell gene expression analysis applied on CD8 T cells elicited by different vaccines regimes, proved that this approach could allow the discovery of new subpopulations of cells differentially elicited by different immunizations.[70] Authors have shown that single cell gene expression analysis is a great additional value compared to bulk analysis. Their approach allowed the qualitative discrimination of CD8 T-cell responses from three vaccines that could not be resolved using conventional assays.

1.8 Significance of the study and main objectives

The aim of this thesis was to perform single-cell gene and protein expression profiling of antigen specific CD8 + T cells elicited by two different vaccine platforms (RNA- and adjuvanted protein-based vaccines) to characterize different sub-populations uniquely regulated in each vaccine.

CD8 T cells play a key role in response after vaccination. The quality rather than magnitude of T cells response is important for determining the outcome of infection or response to vaccination. Our attempts to design and evaluate CD8-T-cell vaccines are confounded by our inability to resolve major qualitative differences between T cells elicited by different vaccines. Being able to resolve differences between T cells elicited by different vaccines is therefore critically important.

Recent advances in high-throughput single-cell gene expression profiling enabled their utilization in diverse research fields such as immunology, stem cell reprogramming, neuronal development and cancer biology. These advances, coupled with computational modeling approaches, enabled us to investigate, on a level of molecular detail not previously possible.[60,71,72,105,106,112,113]

This approach allowed us to analyze the qualitative differences of CD8 T-cell responses elicited by the two vaccine formulations with a resolution that was precluded by more conventional assays. Two different vaccine formulations (RNA-based and protein adjuvanted - based) were used to immunize BALB/c mice in different prime-boost combinations. Single antigen specific CD8 + T cells were sorted and analyzed for the simultaneous expression of 96 markers by microfluidic single-cell qPCR and 32 markers by single-cell mass cytometry. Single cell gene expression data were acquired by BioMark microfluidic qPCR platform and expression data from 1,300 single cells were retained for in-depth analyses.

Here we describe a strategy that we used to explore the molecular identities of individual single cells and dissect transcriptional heterogeneity. We applied an analysis workflow given by the combination of principal component analysis, differential expression analysis and genes co-expression analysis. The goal of this analysis was to find consistent transcriptional patterns and reveal distinct subsets of cells. We identified previously unrecognized subsets of antigen specific CD8 + T cells based upon analysis of gene-expression patterns within single cells and show that they were differentially induced by the two vaccines.

We also identified profiles of vaccine induced CD8 T cell response that provide insight into molecular basis of immunological memory following vaccination and identify potential biomarkers for prediction of vaccine efficacy. The next goal will be mapping this specific subset of cells at protein level by mass cytometry data analysis.

In this study, the analysis of gene transcription within single immune cells has allowed the qualitative discrimination of CD8+ T-cell responses from three vaccines that could not be resolved using conventional assays. These new single cell approaches will undoubtedly refine our cellular classification schemes, revealing new and functionally distinct subset of cells.

Overall, our observations provide a compelling argument for the integration of single-cell approaches into future studies of immune cell fate specification.

2 – Methodology

2.1 Characterization and single-cell sorting of antigen specific CD8 T cells

This study focused on the analysis of $HA_{533-541}$-specific CD8 T cells induced by two different influenza vaccine platforms. Seasonal MF59-adjuvanted Monovalent Influenza Vaccine (aMIV) (A/California/07/2009/(H1N1)) – corresponding to the 7th isolate of an H1N1 subtype virus isolated in human in California in 2009) and RNA-based SAM encoding H1 from the same virus and formulated with CNE56 were compared.

2.1.1 Formulation of MF59-adjuvanted H1N1 and CNE56-adjuvanted SAM

H1N1/MF59 formulation

Monovalent H1N1 subunit and MF59 adjuvant were for laboratory use only and were not final commercial products intended for human use. MF59 is a trade mark for Novartis, used under license by GSK group. Production of Influenza A/California/7/2009 (H1N1) monovalent vaccine was performed as follow: virions were chemically inactivated and disrupted by detergent; subunit H1N1 proteins were purified and were prepared alone (MIV) or with 50% (vol:vol) of oil-in-water MF59 nano-emulsion (aMIV) per mouse dose.

MF59 nano emulsion was composed by polysorbate 80 (Tween), sorbitan trioleate 85 (Span) and squalene. The vaccine formulation was prepared by mixing each component (water for injection, PBS, MF59, antigen) in sequential order.

RNA/CNE56 formulation

RNA was prepared as previously reported.[33] Briefly, the H1 gene was amplified from cDNA of influenza virus A/California/7/2009 (H1N1) and cloned into an optimized replicon construct. DNA plasmid encoding H1 replicon was amplified in Escherichia Coli and purified. DNA was linearized immediately downstream of the 3' end of the SAM sequence by endonuclease restriction digestion. The linearized DNA templates were purified and transcribed into RNA using MEGAscript T7 Kit (Life Technologies). RNA was capped, purified and suspended in nuclease-free water. RNA was formulated with CNE56 (Tween80 (0,5%), Span85 (0,5%), DoTap (0,4%),

squalene (4,3%)), added in equal volume. RNA/CNE56 formulations were prepared fresh for each immunization.

2.1.2 BALB/c immunization and preparation of spleens

Female BALB/c mice, age, 6-8 weeks, were immunized at day 0 and 56 intramuscular in the quadriceps muscle of each hind leg with 50 µl of vaccine formulation per leg with PBS, aMIV and SAM(H1). On experimental day 10 post first immunization (d10p1), week 5 post first immunization (w5p1), day 10 post second immunization (d10p2) and week 6 post second immunization (w6p1), 1-6 mice were sacrificed, and spleen harvested in optimized medium (Fig 4).

Spleens were processed to a single-cell suspension by dissociation through a 70 µm mesh filter.

Ethical statement: all mouse studies were performed at GSK Vaccines S.r.l. Animal Research Center in compliance with all the current Italian laws on the care and use of animals in experimentation (Legislative decree 116/92), and with the Company Animal Welfare Policy and Standards. Protocols were approved by the Italian Ministry of Health (authorizations 249/2011-B and 22/2015-PR).

Group	Antigen	Adjuvant	Route	#mice/ time
PBS	PBS	-	I.M	2
**aMIV	H1N1	MF59	I.M.	2
**SAM(H1)	H1	CNE	I.M.	2

* **aMIV: Adjuvanted pandemic Monovalent Influenza Vaccine (A/California/2009 H1N1)**

** **SAM (Self Amplifying mRNA) encoding H1 (A/California/2009 H1N1)**

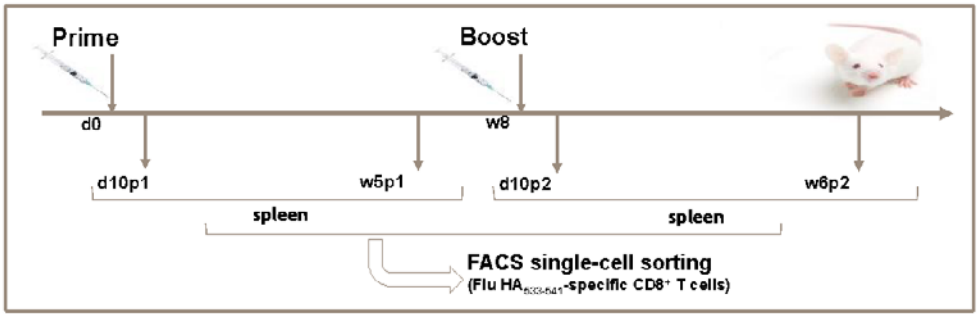


Figure 4: **Immunization scheme.**

2.1.3 Ex vivo MHC-I $HA_{533-541}$-pentamer staining and sorting of single cell for RT-qPCR

For the detection, $HA_{533-541}$-specific CD8 T cells were stained with a recombinant H-2kd restricted MHC-I pentamer loaded with $HA_{533-541}$ peptide and bound to a phycoerythrin (PE) –labeled streptavidin targeting the TCR of $HA_{533-541}$-specific CD8 T cells.[14]

MHC-$HA_{533-541}$-pentamer titration

MHC-$HA_{533-541}$-pentamers were tested at different dosages on splenocytes of naïve and SAM(H1) vaccinated mice at d10p2. Cells were stained in PBS, washed and incubated with H-$2K^d$ restricted MHC-I HA-pentamer ($HA_{533-541}$(IYSTVASSL)) or the control H-$2K^d$ HIV gag pentamer (HIV gag $_{199-207}$ (AMQMLKETI)). Anti CD8 allophycocyanin (APC) (BD Bioscences), anti-CD14 (FITC), anti-CD19 (FITC), anti-CD335 (FITC) and anti-F4/80 (FITC) (eBiosciences) were further added. After wash, sample were run on a LSR-II special order FACS analyzer (BD Bioscience). $HA_{533-541}$-$pentamer^+$ CD8 T cells were identified by applying the following gating strategy: live cells, morphology, singlets, lineage markers ($CD14^-$, $CD19^-$, $CD335^-$, $F4/80^-$), $CD8^+$ and $HA_{533-541}$-$pentamer^+$ or HIV gag $_{199-207}$-$pentamer^+$. Samples were analyzed using FlowJo (software version 9.8.3).

Single-cell sorting

Single-cell sorting was performed using FACSAria II flow cytometer. Single-cell sorting was performed on $HA_{533-541}$-$pentamer^+$ CD8 T cells in splenocytes from PBS, aMIV or SAM(H1)-vaccinated BALB/c mice at different time point after immunization. Single-cells were sorted using a 70 μm ceramic nozzle (BD Bioscience) and an acquisition rate of 8000 events per second.

The sorting plate layout was designed in order that cells were sorted into 96-well plate in the middle of the well. $HA_{533-541}$-$pentamer^+$ CD8 T from vaccinated mice were single cell sorted as lineage marker negative ($CD14^-$, $CD19^-$, $CD335^-$, $F4/80^-$), $CD8^+$ and $HA_{533-541}$-$pentamer^+$, while $CD8^+$ cells from PBS-treated mice were single-cell sorted as lineage marker negative ($CD14^-$, $CD19^-$, $CD335^-$, $F4/80^-$), and $CD8^+$. Cells were deposited into a 96-well sorting plate (1cell/well) containing 5 μl of nuclease-free water with 1mg/ml BSA (Life Technologies) and 1 U/well RNasin (Frementas) per well. Each plate contained one single cell well with no cell as

negative control. Two plates for each vaccine groups were single-cell sorted at each time point and were immediately centrifuged, freeze-dried and stored at -80°Cnuntil multiplexing qPCR was performed. Flow cytometry-based experiments were analyzed with FlowJow (version 9.8.3) and GraphPad Prism (version 6.05).

2.1.4 Multiplexing RT-qPCR of $HA_{533-541}$-pentamer^{+} CD8 T

Transcript abundance relative to 96 genes was assessed from individual antigen-specific CD8 T cells using the IFC qPCR Fluidigm System. For genes detection, a TaqMan based system was applied. TaqMan probes consist of a fluorophore(FAM) covalently attached to the 5'-end of the oligonucleotide probe and a quencher (MGB-NFQ) at the 3'-end. The quencher molecule quenches the fluorescence emitted by the fluorophore when excited by the cycler's light source via FRET (Förster Resonance Energy Transfer). As long as the fluorophore and the quencher are in proximity, quenching inhibits any fluorescence signals.

TaqMan probes are designed such that they anneal within a DNA region amplified by a specific set of primers. As the Taq polymerase extends the primer and synthesizes the nascent strand (again, on a single-strand template, but in the direction opposite to that shown in the diagram, i.e. from 3' to 5' of the complementary strand), the 5' to 3' exonuclease activity of the Taq polymerase degrades the probe that has annealed to the template. Degradation of the probe releases the fluorophore from it and breaks the close proximity to the quencher, thus relieving the quenching effect and allowing fluorescence of the fluorophore. Hence, fluorescence detected in the quantitative PCR thermal cycler is directly proportional to the fluorophore released and the amount of DNA template present in the PCR. cDNA from mRNA templates were prepared by adding 2 µl per well of oligo(dT) primers (25ng/µl final), random hexamers (50ng/µl final), dNTPs (1mM final) and nuclease-free water. Plates were placed in thermocycler (Biometra) and cDNA was synthetized following protocol. Wells were analyzed for the presence of CD3 cDNA to check cell sorting quality. Only wells positive for CD3 were further retain for subsequent analysis.

One of the challenges in single-cell transcriptomics is given by the low quantity of the mRNA per single-cell. To mitigate this problem cDNAs samples were pre-amplified in order to increase amount of starting material. Gene-specific primers were used in a multiplex reaction following manufacturer instructions. BioMark Real-Time PCR system (Fluidigm) was used to perform single-cell gene expression

profiling. The 96.96 Dynamic Arrays were loaded into the IFC controller HX for priming. Microfluidic architecture does the work of combining sample and primer-probe sets. Pre-amplified cDNA and TaqMan gene assays were loaded in separate wells and then mixed in the IFC.

2.2 Single-cell gene expression analysis workflow

2.2.1 Pre-processing

Pre-processing of single-cell RT-qPCR data involved multiple steps including data arrangement, false positives elimination, missing and off scale data corrections and linear transformation of relative quantities of the transcripts.

The design of the chip generates each combination of the 96 genes and 96 enriched cDNA libraries producing 9216 separate PCR reactions. After each cycle, the fluorescence is read. The cycle (or interpolated fraction thereof) at which the fluorescence crosses a pre-determined threshold is recorded, defined as the '*ct*' value. The fluorescence reporter signal was normalized to the signal from the passive reference ROX based on the background signals is collected before the onset of the experiment. In the standard BioMark protocol the background signal was collected at ambient temperature. ROX is negatively uncharged at neutral ph that does not interact with DNA.

The Curve Quality Treshold is a qualitative tool designed by Fluidigm which provide a measure of the quality of each amplification curve for DNA binding dye qPCR detection chemistry. In this analysis, each individual amplification curve is compared to a mathematically ideal exponential curve and given a quality score between 0 and 1 (0 is a flat line and 1 is a perfect sigmoidal curve). The algorithm also takes in account the linearity of the baseline, the delta Rn of the final product (Rn=emitted fluorescence/ROX) (delta Rn is the normalization of Rn obtained by subtracting the baseline obtained in the initial cycle of PCR where there is a little change in fluorescence signal), the actual level of fluorescence, the slope of the amplification plot, and the return to linearity after exponential curve growth.

$$Baseline = y\text{-}intercept + slope\ X\ (amplification\ cycle)$$

The curve quality threshold is a qualitative tool designed by Fluidigm which provide a measure of the quality of each amplification curve for DNA biding dye.[114]

Amplification curves with atypical shapes are not processed correctly and they also indicate sample specific problem such as enzymatic inhibition. Wells that present this kind of problem were removed from analysis.

Raw data that passed Fluidigm software quality control were used for subsequent analysis. The resulting single-cell qPCR data (Cq values) were exported as csv files and subsequently organized in Microsoft Excel spreadsheets with single-cell samples in rows and genes in columns. The amplification curves of all PCR reactions were first manually analyzed to remove anomalous curves and false positives.

Because of lognormal distribution of transcripts among cells, even high expressed genes will have rather few transcripts in most cells. When single cells analysis is performed, it is important to use a workflow that minimizes losses.

Data should not be normalized given that frequently used normalization schemes are not directly applicable in single-cell gene expression experiments. Indeed, the individual cell is the atomic unit of normalization and the amount of starting material naturally measured in number of cells per reaction. The dichotomous nature of single-cell expression does not allow direct application of traditional normalization approaches[98].

Fast most intuitive way to compare expression data for single cells is compared data as measured. It has been suggested to do not normalize to any kind of housekeeping genes or presumed reference genes given that the mRNA burst kinetics enhanced uncorrelated variations between randomly selected genes. An option could be performing global normalization i.e. normalize the mean of expression of all genes, but it could introduce bias.

If cells studied are of the same type and expected to express common markers failure to record a Cq value in those cases is probably due to failure in that particular reaction chamber. Profiling is initially performed for all genes many of which will be not responsive. Removing nonresponsive marker will improve separation in multivariate analysis[115,116].

Missing data handling

A typical goal of gene expression experiments is to search for differential expression across groups. The zero-inflation of expression in Biomark's experiments introduces problems for testing differential representation of cell subsets characterized by expression patterns, as well.

Missing data are common in scientific experiments data analysis and are a classical problem in statistics. Missing data (NA) were caused for two differently reasons: the reaction chamber contains mRNA molecules, but the reaction failed or there is no template in the reaction well. these two cases should be handled differently. It's difficult understand whether missing data are due to technical failures or expression levels below to limit of detection. Variability in mRNA content at single cell level complicated the handling of missing value. Single-cell gene expression profiling is performed without replicates to maximize the number of cells analyzed. Moreover, protocols are optimized to reduce risk for technical failures. A pragmatic approach is to assume that all missing data are due to few molecules and high frequency of missing data is expected due to low expression levels and fluctuations over time.

Missing data were considered as transcript expression below limit of detection and a 0 value was added in each case on NA detection.[115,117]

False positives detection

A supervised pre-processing of raw data obtained from BioMark system was performed. Firstly, cells that "failed" the quality control of the software reading were filtered out. Cells where no cd8a/cd3 signal was detected were removed from the analysis, as well as cells were positive signal for cd4/cd19 was detected. Gene assays where no positive signals were observed in any sample were filtered out. Moreover, an empirical cut off was set to cell sample expressing at least 10% of genes.

Cq_{cutoff} determination and outlier detection

Very high Cq values are not reliable even if a correct PCR product is formed. It could be caused by failed amplifications of single molecules targets in the initial cycles, delayed amplification due to competing reactions or enzymatic inhibition. High Cq values should be discarded. A useful approach is to delete all Cq-values above a certain threshold Cq_{cutoff} which will be the same for all assays. A reasonable Cq_{cutoff} can be chosen by inspecting control charts such as box and whisker plot

provides an overview of the spread of the genes' expressions. Potential and extreme outliers have expression levels outside 1.5*I*QR and 3*I*QR, respectively (IQR = Quartile 3–1). Quartile 1 and 3 represent the bottom and top of the box and are the 25th and 75th percentile of the gene expression values.[116]

Control charts are expected to be symmetric because of the lognormal distribution of transcript among individual cells. Asymmetry in control charts may indicates that copy number of a given transcript is too low to be reliably detected. However, genes should not be eliminated due to their expression characteristics. Control charts can be calculated and inspected for different Cq_{cutoff} values guiding for the selection of appropriate cutoff. A pratical approach is to set all missing data to -1 in log scale, corresponding to 0.5 molecules in RQ. For example, a Cq_{cutoff}=27 was applied and and data converted into RQ. Remaining missing data were assigned RQ=0.5 Low abundance genes were identified, for which missing values were reassigned a Cq value of 40. Any gene whose average expression was within a cutoff of 3SD of the mean Cq value for the two chosen genes was included. All cells expressing less than half of these genes were excluded. Another method is data exclusion based on median standard deviation cutoff all missing data values were initially set to Cq 40, and the mean Ct and number of missing data points were calculated for all genes. The second and third highest expressed genes in the data set were selected and their mean Ct and standard deviation calculated. The limit of detection was set to Cq 37. All data above LOD (also 999 value) were replaced with 37. The LOD Ct values was subtracted from all other Ct according to the Log2EX method (Log2EX=LOD Ct-Ct).[115,116,118]

After inspection of quality control charts, we noticed that dataset didn't show Cq values higher than 29. Given that, we calculated cutoff as mean of all max Cq values recorded for genes measured as previously reported. We set Cqcutoff=26.

Statistical considerations

Single-cell transcriptomic analysis can be used to characterize variation in gene expression levels at high resolution. However, the sources of experimental noise are not yet well understood

To get an overview of the data, it common to calculate some basic statistics for all genes studied, including the number or fraction of cells expressing each gene, and the mean and standard deviation of all the genes expression.

Mean and standard deviation could be calculated in logarithmic scale since the underlying distribution is lognormal.

It is a good idea to visualize the distributions of the different transcripts among the cells either as frequency histograms or violin plot. [48,116,118]

Expression index computation

The standard assumptions of qPCR-based assays apply to the Fluidigm technology, consider that the cycle threshold (*cq*) is inversely proportional to the log of fluorescence. The fluorescence is directly proportional to the starting concentration of mRNA. [119] The Fluidigm instrument returns the cycle threshold (*cq*); previous work have shown that it more useful to work with the complement of *cq* [71,98].

$$\log E_{g,c} = \mathrm{Cq}_{\mathrm{cutoff}} - Cq_{g,c}$$

Where $\mathrm{Cq}_{\mathrm{cutoff}}$ is the maximum Cq, 26 in our case. Assuming that all reactions are in the exponential amplification phase, this quantity should be directly proportional to the log-abundance of mRNA, plus an intercept term corresponding to the number of cycles it takes for the minimally detectable quantity of mRNA to cross threshold. If the fluorescence does not cross the threshold after 40 cycles, then the Fluidigm instrument records a value of N/A, and we say that the gene is *not detected*. We considered undetected genes as unexpressed genes. This assumption is supported by the idea that transcription of mRNA is thought to occur in bursts of activity followed by quiescence.[49,54] As a consequence, we treated the undetected genes as unexpressed genes, and we set the corresponding Cq value to - ∞ corresponding Cq value to - ∞ so that the mRNA abundance is zero (i.e. $2^{-logEg,c}$=0). The log expression of each gene *g* was computed as follows: $\log E_{g,c} = 40 - Cq_{g,c}$ where *c* is the cell and $Cq_{g,c}$ is the Cq value obtained from the BioMark (Fluidigm).

The expression index (EI) quantification of gene expression was computed as the product of the proportion of cells expressing a given gene times the average gene expression value in these cells. Finally, differences in the proportion of cells expressing a given gene, across different groups or conditions, were tested for statistical significance using a Fisher's exact test followed by correction for multiple testing using Benjamini-Hochberg Procedure.[70–72]

Autoscaling and mean centering

It was previously demonstrated that giving to all genes the same weight in the analysis, makes them equally important in the data processing[120]. In fact, multivariate methods (PCA, Hierarchical clustering) consider the expression levels of all genes and if no scaling is applied the more expressed genes will have higher weights in the analysis and it will dominate the analysis. Data autoscaling gives to all genes the same weight. This is done by calculating z score for each gene by subtracting the mean expression and divide by its SD (a Z score of 2 indicates that a gene in a particular sample is overexpressed by two SD to its mean expression in all samples.[116,118]

Problem in autoscaling arise when panel includes genes that are not responsive, and their expression show only random variation. If the number of non-responsive genes is large the data quality may be compromised by auto scaling. An option is only mean centering the data. Mean centering data is performed by subtracting the average expression of each gene, but not dividing by standard deviation. [115]

Preprocessing steps not suitable for single cell analysis

Some steps of classical qPCR analysis are not used for single cell transcriptomic data. In this case we don't use reference genes, given that any gene has constant steady state level of transcripts. Global normalization based on averaged expression of all transcript can be performed (or mean center data) but this could cause some complications because the rather small number of genes typically analyzed per cell and the ambivalence in the handling of missing data.[116]

2.2.2 Principal component analysis

Principal component analysis is a technique that reduces dimensionality of the data, by maintaining most of the variation in the dataset[121]. The reduction is performed by the identification of the directions (called principal components) that explain the maximal variation in the dataset.[122] Plotting the samples allow the visualization of similarities and differences between them by the identification of different groups in which they are separated. [123]

In the principal component analysis, the original variables are transformed by linear combination in new variables which are correlated with each other displaying significant patterns in the data[122]. The first principal component is the direction along which the samples show the largest variation. The second one is the direction, orthogonal to the first component, along which the samples show the second largest variation. [123]

The input for principal component analysis is a data matrix X in which column represent variables and rows correspond to the samples. The final number of principal components correspond to the number of variables.[122]

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system. The method creates a new set of variables called principal components which are calculated by the linear combination of the original variables.[122,124]

The principal components are stored in the "PCA loading matrix" which can be interpreted as a rotation matrix.[125] Geometrically, PCA is equivalent to a rotation of the original data space in which the new axes are represented by the principal components. The first component, PC 1, represent the direction of highest variance. The second component, PC 2, is the direction that maximizes the remaining variance in the orthogonal subspace complementary to the first component. The first and second components together explain the two-dimensional plane of highest variance.[126] PCA transforms the data into a new lower-dimensional subspace in the new coordinate system the first axis corresponds to the first principal component, which is the component that explains the greatest amount of the variance in the data. The second principal component must be orthogonal to the first principal component, it does its best to capture the variance in the data that is not captured by the first principal component.[83]

The main step of PCA is the computation of the weighted coefficients that are used for the linear combination of the variables.[122] The classical way is to calculate the eigenvectors of covariance matrix between variables, $cov_{ij} = \frac{1}{n-1}\sum_{m=1}^{n} (x_{im} - \bar{x}_i)(x_{jm} - \bar{x}_j)$ where $\bar{x}$ is the mean of all variables, n is the number of samples, x_{im} is the value of variable i in object m and , x_{jm} is the value of variable j in object

m . The eigenvectors are sorted by their corresponding eigenvalues and are orthogonal to each other.[124,126]

In detail, PCA transform a d-dimensional sample vector $x=(x_1,x_2,\ldots,x_d)^T$ into a usually lower dimensional vector $y=(y_1, y_2, \ldots y_k)^T$, where *d* is the number of variables and *k* is the number of selected components. The transformation is defined by the k x d matrix V, such that

$$y = Vx$$

Each row-vector of matrix y contains scores of a new variable yj defined as principal component (PC). The principal component j is a linear combination of all original variables, weighted by the elements of the corresponding transformation vector *vj = (vj1,vj2,...,vjd)*

$$yj = \sum_{i=1}^{d} vjixi = vj1x1 + vj2x2 + \cdots + vjdxd$$

n is the number of samples and d is the number of original variables. The weights v_{ji} described the contribution of all original variables x_i to the *j* component.[124,126]

Geometrically, PCA is equivalent to a rotation of the original data space. The new axes are the principal components. The vector v_j gives the direction of the j^{th} principal component (PC j) in the original data space. The first component, PC 1, represented by the variable y_1, is in the direction of highest variance. The second component, PC 2, is the direction that maximizes the remaining variance in the orthogonal subspace complementary to the first component. The first and second components together explain the two-dimensional plane of highest variance. [122,123,125]

To globally visualize the data, we used principal component analysis. As previously reported[60,71,82], PCA was applied to reduce dimensionality of the data by finding linear combinations of the original data ranked by their importance. The data are represented by a gene expression space is n dimensions, where n is the number of genes analyzed and each point is a cell.[42]

Because PCA components consist of contribution from all gene it is possible to identify the genes that give the highest contribution in projecting cells in the space.[82]

The length of an eigenvector represents the largest variance for each gene in the correlation coefficients and the distance between genes in the correlation coefficients was illustrated by the angle formed between the gene eigenvectors.[122] The ordering of the eigenvectors is based on the distance between genes in terms of Pearson's

coefficients. Gene eigenvectors placed close to each other were more similar in the expression patterns and hence were more positively correlated[71,127]. The pattern of samples revealed by PCA can be visualized together with genes having the largest weight for the showed principal components that characterized a specific group of samples.[123]

2.2.3 Silhouette index – clustering score

Silhouette index was used as internal measure to assess quality of clustering (separation of different groups in the PCA PC1/PC2 space). It is a measure of tightness and separation of clusters, which is used to assess the degree of clusters separation. For a given cluster X_j ($j=1,\ldots,c$), this method assigns each sample X_j a quality measure, $s(i)$ ($i=1,\ldots,m$), known as the Silhouette width. The Silhouette width is a confidence indicator of the membership of the *i th* sample in cluster X_j The Silhouette width for the *i th* sample in cluster X_j is defined as:

$$s(i) = \frac{b(i) - \mathrm{a(i)}}{\max\{a(i), b(i)\}}$$

where *a(i)* is the average distance between the *i th* object and all of objects included in *Xj*, and *b(i)* is the minimum average distance between the *i th* sample and all of the sample clustered in X_k (k=1,…,c;k≠j), and this formula follows that $-1 \leq s(i) \geq 1$. $s(i)$ closing to 1 indicates that the *i th* object has been well clustered, i.e. it was assigned to an appropriate cluster. $s(i)$ closing to zero suggests that the *i th* sample could be assigned to the neighboring cluster. If $s(i)$ is close to -1, the object is misclassified.[128]

The extent of separation of cell clusters in the PCA space was quantifies using the clustering silhouette approach. The clustering score provided by this analysis is a measure of how tightly grouped all the data in the cluster are.

2.2.4 Fisher's Exact Test

Fisher's exact test is a statistical significance test for categorical data (Fisher, 1922, 1925). It is based on a 2 × 2 contingency table which measures the association

between two variables. Fisher's exact test can be used when you have two nominal variables, to know whether the proportions for one variable are different among values of the other variable. We used Fisher’s exact test to measure the statistical significance of change in number of cells expressing genes between two groups.

Fisher's exact test is more accurate than the chi-square or G-test of independence when the expected numbers are small. Fisher's exact test is generally used when the total sample size is less than 1000, and chi-square or *G*–test for larger sample sizes.

The null hypothesis is that the relative proportions of one variable are independent of the second variable; in other words, the proportions at one variable are the same for different values of the second variable.

Unlike most statistical tests, Fisher's exact test does not use a mathematical function that estimates the probability of a value of a test statistic; instead, you calculate the probability of getting the observed data, and all data sets with more extreme deviations, under the null hypothesis that the proportions are the same.

Fisher's exact test, like other tests of independence, assumes that the individual observations are independent. Unlike other tests of independence, Fisher's exact test assumes that the row and column totals are fixed, or "conditioned."

An approach used by the *fisher.test* function in R is to compute the p-value by summing the probabilities for all tables with probabilities less than or equal to that of the observed table. In the example here, the 2-sided p-value is twice the 1-sided value; but in general, these can differ substantially for tables with small counts, unlike the case with test statistics that have a symmetric sampling distribution. [129,130]

Chapter 3 – Results

3.1 Overview

In the present study, we implemented an experimental setup for analyzing single-cell transcriptome responses of antigen specific CD8+ T cells following administration of two alternative influenza vaccine formulations. Details of the procedure are presented in Fig 4.1.

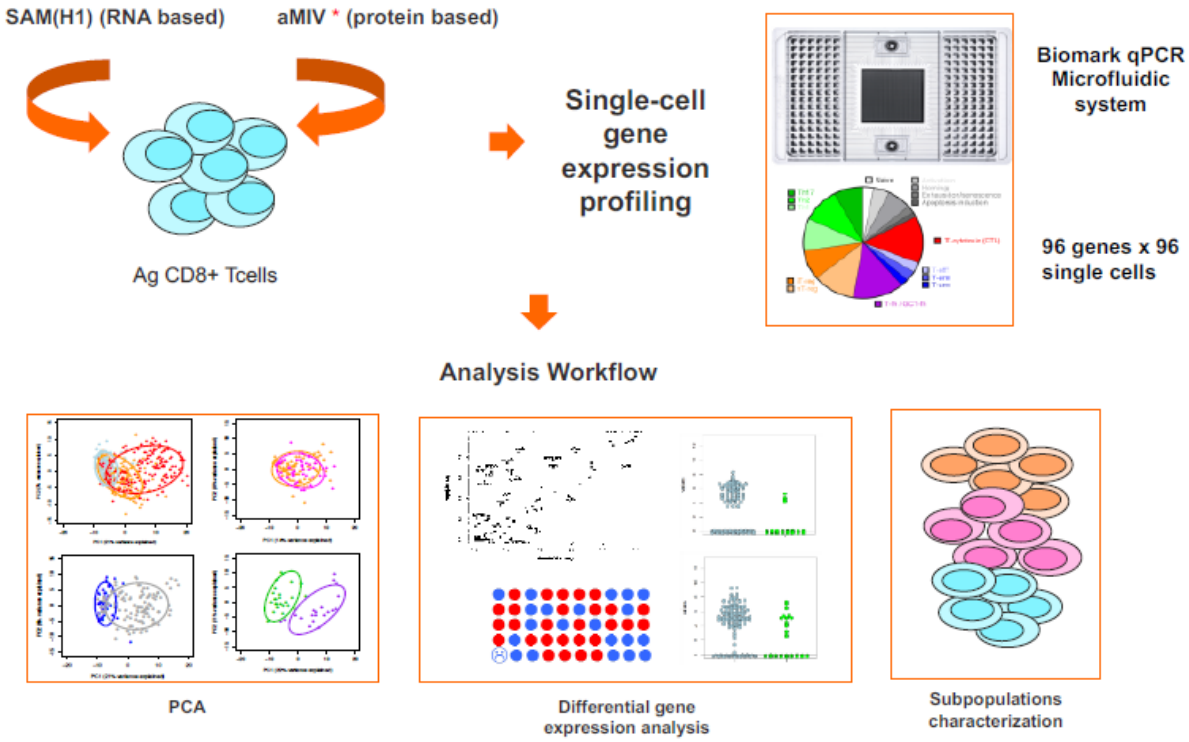


Figure 4.1: Overview on single-cell analysis workflow

Animals were immunized according to the experimental design shown in Fig 4. To understand which cells were activated and their phenotypic properties, a single-cell characterization was performed. To obtain single HA533-541-pentamer+CD8+ T cells, splenocytes from BALB/c mice were stained with H-2Kd-restricted HA533-541-pentamer or a control HIV199-207-pentamer. Cells were single cell sorted based on the binding to and expression of MHC-I monomer. To determine the transcriptional state of $HA_{533-541}$-pentamer+CD8+ T cells, RT-qPCR was applied by

preparation of cDNA. Genes were detected by using TaqMan primer and a fluorescence probing system in a Fluidigm 96.96 dynamic array chip.

96 genes were selected based on known functions in T- cell differentiation, tissue homing, survival, activation, cytotoxicity and regulation of immune response[16,19,70,71,131–133] (Table 1).

After pre-processing steps, genes with no signal detected were filtered out, providing a dataset of 86 genes for further analyses. Overall, expression data from 1,152 single cells were collected and used for subsequent analyses.

Class	**Gene**
Apoptosis	Bim, FasL, PD-1
Cytokine/Chemokine receptors	Ccr4, Ccr5, Ccr6, Ccr7, Ccr10, Cx3cr1, Cxcr3, Cxcr4, Cxcr5, Cxcr6
Interleukin receptor	Il-1R-1, Il-2R-a, Il-2R-g, Il-7R-a, Il-10R-a, Il-12R-b1, Il-21R
Cytokines/Chemokines/Interleukins	Ccl4, Il-2, Il-4, Il-5, Il-9, Il-10, Il-13, Il-21, Il-22, Ifn-γ, Tgf-b1, Tnf, Trail
Killer cell lectins	Klrd1, Klrc1, Klrk1, Klrg1
Secreted proteins/Granzyme	Cd70, Cd40l, Gzma, Gzmb, Gzmk,Mmp2,Mmp9, Prf1
Signaling/Proliferation/Self-renewal	Cd44, Cd8a, Cd69, Cd62l, Spi6, Lfa1, Cd19, Cd27, Cd28, Vla-4b, Cd69, Itgae, Lamp1, Ox40, 41bb, Ctla4, Cd160
Transcription factors	Ahr, Bcl3, Bcl6 ,Bcl11b ,Blimp1, CamkIV, Eomes, Foxo3a, Foxp3, Gata3, Irf4, Mki67, Nfkb1, Relb, Rorc, Stat5a, Tbet

Table 1: 96 selected gene targets grouped by function.

3.2 CD8+ T-cell populations elicited by the two vaccine formulations reveal substantial heterogeneity

Principal component analysis was applied to globally visualize data. Data were projected into the first two principal components (PC1 and PC2), which account for the largest amount of data covariance.

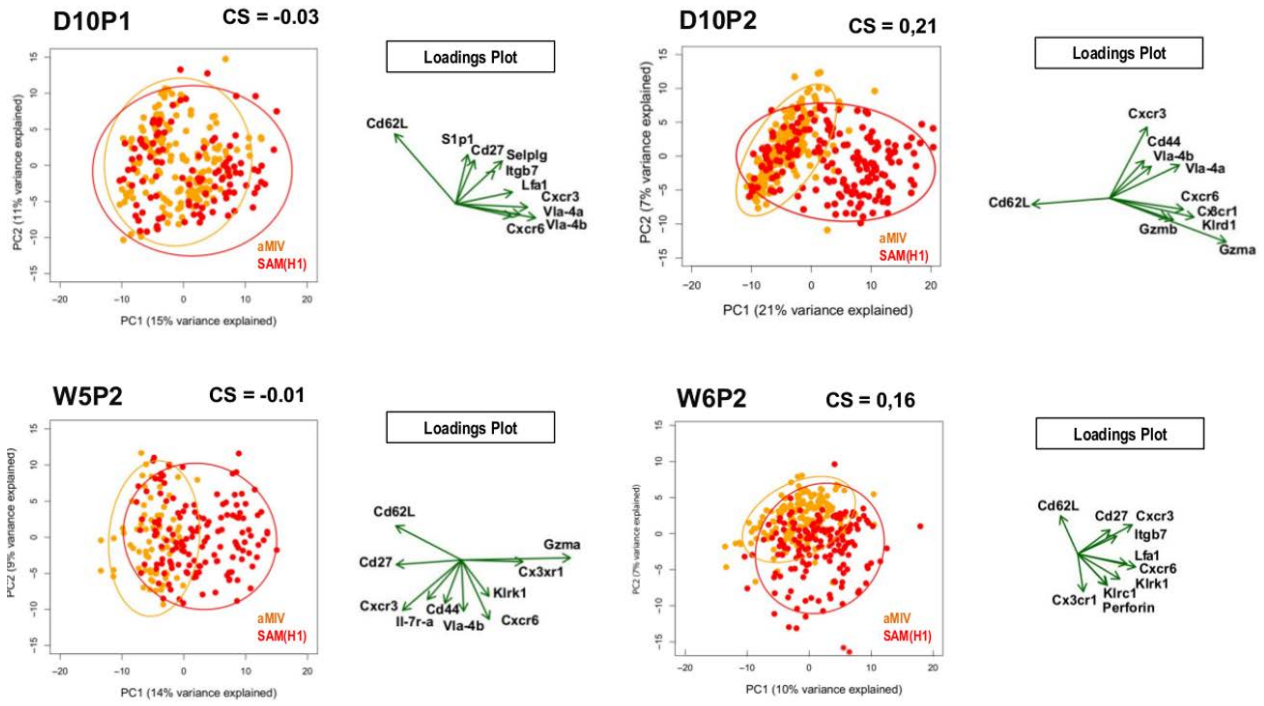


Figure 5: PCA reveals substantial heterogeneity between cells

Data from 1152 single cells were used for subsequent analyses, divided into the two vaccine groups at four time points. PCA performed on transcriptome profiles of antigen-specific CD8+ T cells. aMIV (yellow) and SAM(H1) (red). Clustering score (CS) and loadings plot for the 10 most informative genes are also reported.

PCA reveals extensive overlap between vaccines that could be explained by transcriptional similarity across cells (Fig 5). The first two components captured from 17% to 36% of the variance in our dataset, slightly low compared to similar independent studies.[65] This is probably a reflection of transcriptional pattern that are shared from two CD8+ T cells populations.

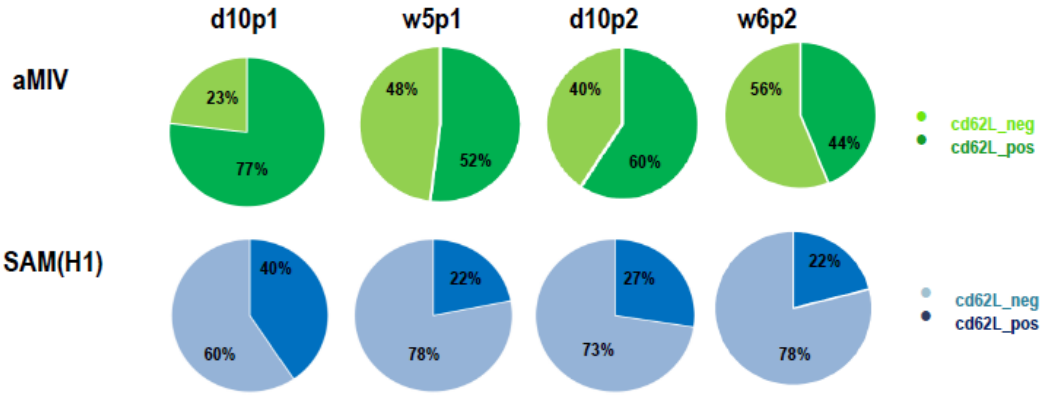
Gene expression space has 86 dimensions because of 86 genes and each data point is a cell. The coordinate in each dimension is the level of expression for a given gene in that cell. Each component has a contribution of all 96 genes since the component cut

across 86D space. A projection of the expression patterns onto PC1 and PC2 revealed cells overlap and weak cluster separation.

A silhouette analysis was used to assess quality of clustering. This method provides both a coefficient representing the tightness and separation of clusters, and a representation of how well the data fits into the clusters. Clustering scores close to 0 indicate that there is no substantial separation between the two clusters. Generally, clusters were not well separated. D10p2 showed a CS of 0,22 (Fig 5), indicating a weak but existing structure. This is confirmed also by the highest percentage of variance explained by the two PC components. It was assumed that best CS of d10p2 was given by well-defined cellular fates and it was reasoned that could be utilized to extrapolate cellular subsets and map them to other time point and looking for differences in molecular insight.

Because PCA components consist of contribution of all 86 genes most information rich genes were identified in classifying cells. Loadings were represented by arrows with gene name. From the selected set of genes, Cd62l contributed clearly to the clustering of aMIV vaccinated cells on the PC1 and interestingly is among most informative genes of PCA of all time points. Cd62l is a key trafficking gene that distinguish naïve from effector cells. Cd62l expression in $CD8^+$ T cells at single-cell level distinguished naïve from short lived effector T cells (T_{SLE}). The two vaccines didn't show any significant difference in level of expression of Cd62l mRNA among time point in each vaccine group. If we consider percentage of cells expressing Cd62l, we observed differences between SAM(H1) and aMIV (Fig 6B). SAM(H1) vaccinated cells were characterized by a higher number of Cd62l_neg cells mostly at later time points. aMIV elicited cells displayed a balance between Cd62l_neg vs Cd62l_pos cells. The majority of SAM(H1) elicited CD8 T cells lack in the expression of Cd62l.

A.



B.

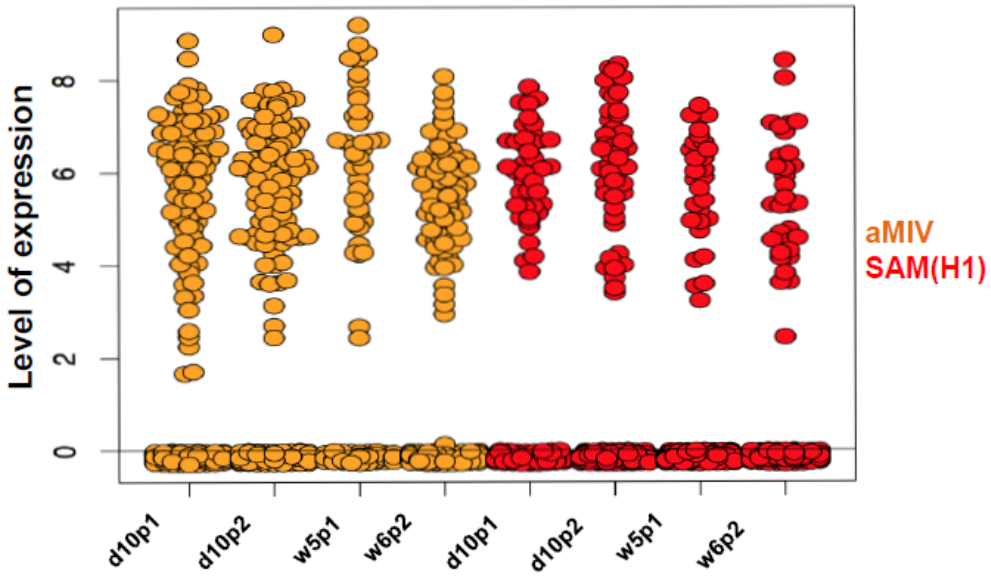


Figure 6: Cd62l mRNA expression in CD8$^+$ T cells

A. Level of expression (Ct_{cutoff}-Ct) of Cd62l at all time points (x axis), aMIV (orange), SAM(H1) (red) **B.** Frequency of cells expressing Cd62l in both vaccine groups, aMIV_Cd62l_neg (light green), aMIV_Cd62l_pos (dark green), SAM(H1)_Cd62l_neg (light blue), SAM(H1)_Cd62l_pos (dark blue).

To investigate if Cd62l_neg and Cd62l_pos were characterized by distinct transcriptional patterns, PCA was performed on the Cd62l-defined subpopulations

within and between vaccines at d10p2 (Fig 7). Cd62l_neg and Cd62l_pos cells segregated into distinct clusters both in aMIV and SAM(H1) groups. We hypothesized that this could be explained by different transcriptional patterns. Cd62l_pos cells elicited by the two vaccines were transcriptionally similar whereas Cd62l_neg cells formed distinct clusters. Overall, collected findings suggest that Cd62l_neg populations were transcriptional different (FIG. 7).

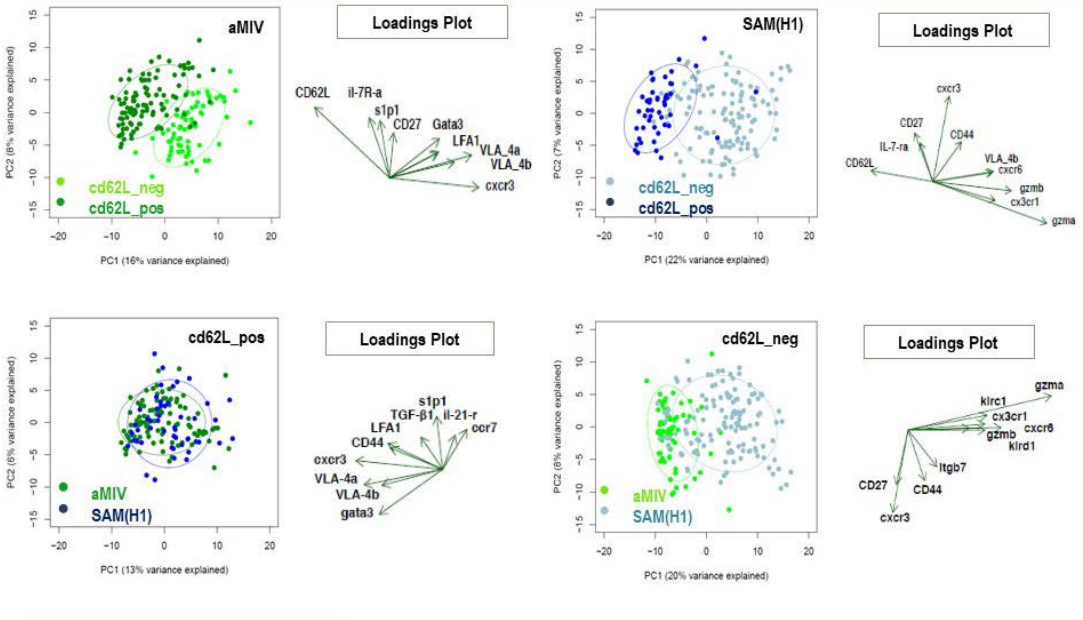


Figure 7: PCA on d10p2 Cd62l_neg and Cd62l_ pos populations highlighted defined clusters

PCA performed at d10p2 on aMIV_*Cd62l*_neg (light green), aMIV_*Cd62l*_pos (dark green), SAM(H1)_*Cd62l*_neg (light blue) and SAM(H1)_*Cd62l*_pos (dark blue) compartments, with indicated 10 most informative genes.

3.3 SAM(H1) and aMIV induce distinct transcriptional programs in Cd62l_neg CD8+ T cells

Loadings from Cd62l_neg populations PCA highlighted a group of six genes that we hypothesized form a distinct group that was prevalent in SAM(H1)_Cd62l_neg cells clustering on PC1. Expression profiles of Cx3cr1, Cxcr6, Gzma, Gzmb, Klrc1 and Klrd1 were examined in both Cd62l_neg and Cd62l_pos populations in both vaccines in all time points (Fig 8). Interestingly, Gzma, Gzmb and Cx3cr1 were uniquely expressed in SAM(H1)_Cd62l_neg populations in all time points. Klrd1 was expressed without any substantial difference in all time points with a peak of expression in SAM(H1)_d10p2_Cd62l_neg group. Klrc1 expression was specific to SAM(H1) in both groups (Cd62l_neg and Cd62l_pos) at every time point, always upregulated in Cd62l_neg populations. Cxcr6 didn’t show appreciable differences.

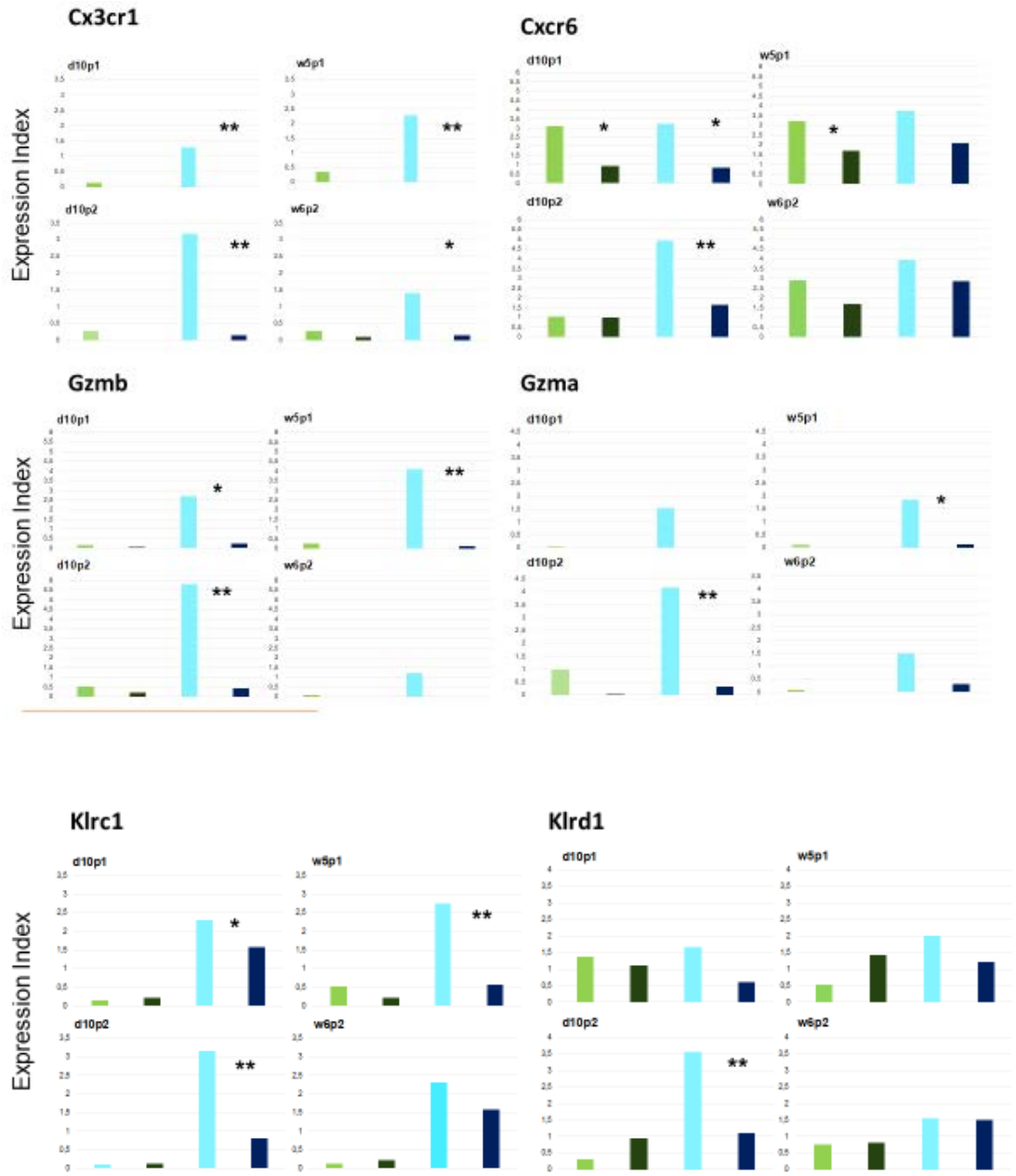


Figure 8: Cytotoxic/inflammation genes characterize SAM(H1)_Cd62l_neg populations

Expression of Cx3cr1, Cxcr6, Gzmb, Gzma, Klrc1, Klrd1 in aMIV_Cd62l_neg (light green), aMIV_Cd62l_pos (dark green), SAM(H1)_Cd62l_neg (light blue) and SAM(H1)_Cd62l_pos (dark blue) compartments. Expression is calculated as expression index (E.I.) = “percentage of positive cells” x “mean of expression of positive cells”. Fisher exact test *=p<0,05, ** p=0,01 (Benjamini-Hochberg threshold 0.05).

3.4 SAM(H1)-induced Cd62l_neg CD8+ T cells are characterized by an effector-cytotoxic phenotype

Gzma, Gzmb, Cx3cr1, Cxcr6, Klrc1 and Klrd1 markers are clustered together on the Cd62l_neg PCA loadings plot, suggesting that they have a similar co-expression pattern (i.e. they are co-expressed in the same cell) (Fig 7 bottom-right panel). 27% of SAM(H1) vaccinated cells co-expressed Gzma, Gzmb, Cx3cr1, Cxcr6, Klrc1 and Klrd1 (Fig 9). Interestingly, these genes were not co-expressed or at very low level (2-7%) in SAM(H1)_Cd62l_neg populations in other time points (d10p1, w5p1, w6p2) (Fig 9B). This finding may underlie the high effector-cytotoxic transcriptional profile at d10p2. Next, we wondered which the most frequent combinations of were 5, 4 and 3 genes. Gzma, Gzmb and Cxcr6 are the 3 genes most co-expressed in d10p2 (57% of cells) (Fig 9B). Frequency of cells co-expressing these genes decreased at other time points. After prime percentage of co-expressed genes varied between 20% at d10 to 27% at w5. Six weeks after the second immunization, most of the co-expression was lost, with only 9% of cells showing Gzma, Gzmb and Cxcr6 co-expression. Levels of expression of these genes (EI) showed a strong decrease of Gzmb and Gzma expression at w6p2, while Cxcr6 abundance remained high at each time point (Fig 9C).

We hypothesized that different expression combination of Cxcr6, Gzma and Gzmb could characterized different SAM(H1)_Cd62l_neg group at different time points. Co-expression of Gzma and Gzmb is higher in d10p2 population (66,14% of cells) and drop at other time points (from 24,1% to 33,6%). However, the percentage of cells that were negative for both markers was 62 % at w6p2 and only 20% of cells were Gzma- and Gzmb-negative at d10p2. The same trend was found for Gzmb and Cxcr6 co-expression (Fig 10).

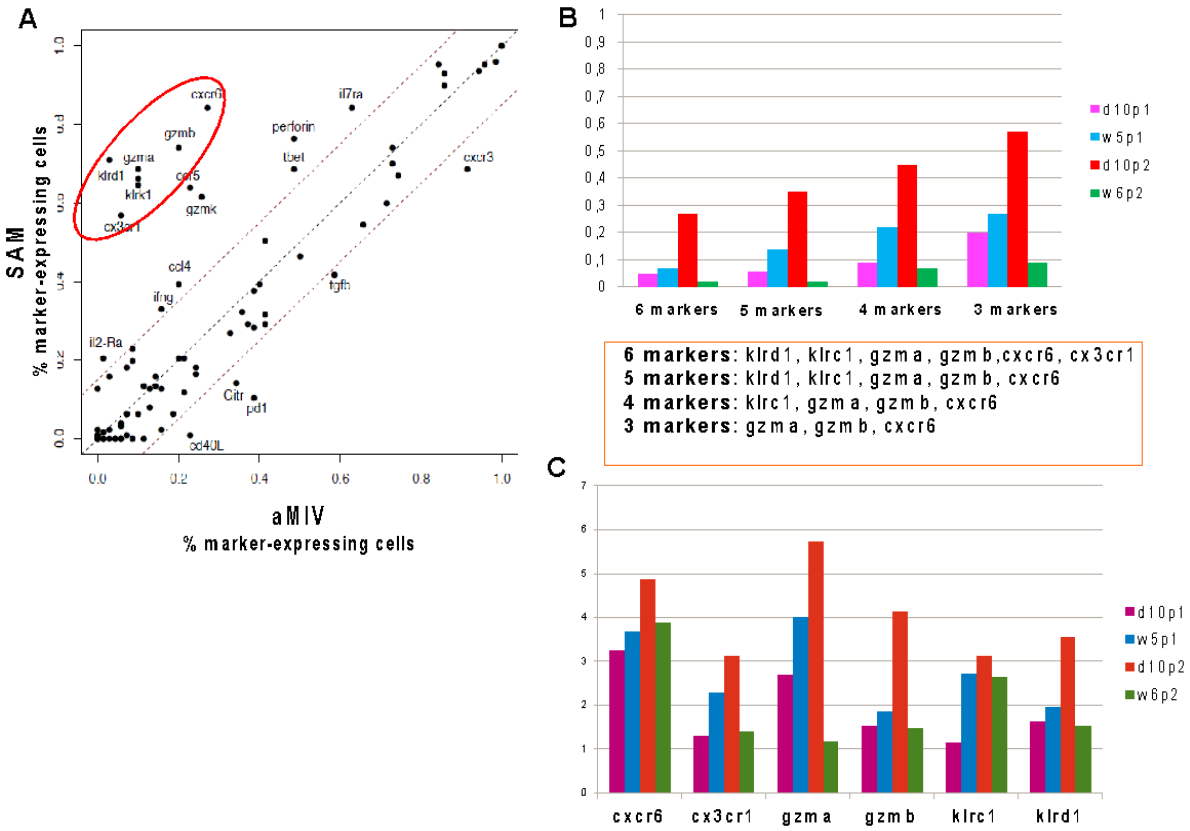


Figure 9: Characterization of Cd62l_neg expression profiles

A. Scatterplots show comparison between aMIV_Cd62l_neg and SAM_Cd62l_neg cells at d10p2 for the frequency of cells expressing the genes analyzed. Genes outside dashed lines are expressed from >15% of cells.**B**. Bar plots show co-expression of the six clustered genes at d10p2 in the SAM(H1)_Cd62l_neg compartment: 6 markers (Klrc1, Gzma, Cx3cr1, Cxcr6, Gzmb and Klrd1), 5 markers (Klrc1, Gzma, Cxcr6, Gzmb and Klrd1), 4 markers (Klrc1, Gzma, Cxcr6 and Gzmb) and 3 markers (Gzma,, Cxcr6 and Gzmb). D10p1 (pink), w5p1 (blue,) d10p2 (red) and w6p2 (green). **C**. Bar plots show expression index of Cxcr6, Cx3cr1, Gzma, Gzmb, Klrc1, Klrd1 in D10p1 (pink), W5p1 (blue,) D10p2 (red) and W6p2 (green).

A

	d10p1	w5p1	d10p2	w6p2
gzmb+/cxcr6+	26,9	33,6	66,14	24,1
gzmb-/cxcr6-	31,7	25,5	8	21,7
gzma+/gzmb+	25,3	30,80	62,9	12,1
gzma-/gzmb-	53,9	31,7	20,4	62,1
gzma+/cxcr6+	31,7	40,2	61,4	14,5
gzma -/cxcr6-	31,7	21,7	8	25,5

B

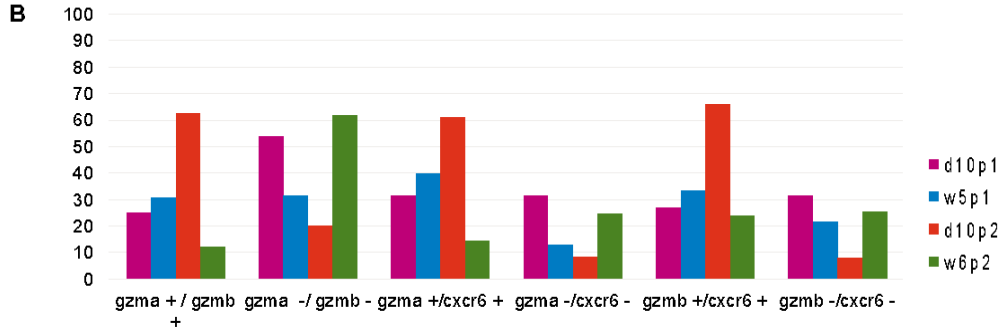


Figure 10: SAM(H1) d10p2_Cd62l_neg populations show highest co-expression of granzymes and Cxcr6

Table (A) and bar plot (B) show frequencies of cells that co-express various combination of couple of genes in SAM(H1) Cd62l_neg populations in d10p1 (pink), w5p1 (blue,) d10p2 (red) and w6p2 (green).

3.5 SAM(H1)-induced Cd62l_neg CD8+ T cells are characterized by a terminal effector profile.

The balance between T-bet and Eomes has been shown to determine effector and memory cell fate in CD8+ T cells[20,134]. T-bet and Eomes have a crucial role in the formation and function of effector and memory T cells SAM(H1)_Cd62l_neg populations expressed higher level of Tbet than Cd62l_pos cells(Fig 11A). In contrast, Eomes is preferentially expressed in Cd62l_pos populations except for d10p1 time point in which Eomes is expressed in both populations (Fig 11B). It was observed that SAM_Cd62l_neg populations exhibited a positive Tbet/Eomes ratio (Fig 11C). Interestingly, after prime, the Tbet/Eomes ratio increased at 5w, whereas after boost, the Tbet/Eomes ratio performed an opposite trend, with higher ratio at d10 and decrease at 6w.

In both SAM(H1)- and aMIV-elicited cells, Tbet was expressed preferentially in Cd62l_neg cells. Eomes was preferentially expressed in Cd62l_pos cells. aMIV vaccinated cells presented a lower expression index compared to SAM(H1) vaccinated cells (Fig 11D). It was hypothesized that in SAM(H1)-indiced cells, boost could induce a strong cytotoxic response at earlier time point which would disappear after 6 weeks, caused by possible exhaustion of cells. After prime at d10, transcriptional cytotoxic response is moderate and increased in late time point showing that cells seemed to continue transcriptional cytotoxic activity also after weeks post prime.

It was reported that expression of Eomes and Cd62l promotes the expression of Tbet and Cx3cr1. Additionally, blimp and Tbet expression induces short-lived effector T cells (T_{sle}) differentiation of CD8+ T cells, while Klrg1 is a marker of T_{sle} phenotype. SAM(H1)_Cd62l_neg cells were characterized by a transcriptional profile consistent with that of T_{sle} differentiation (Fig 12). T_{sle} cells have a significantly shorter lifespan and reduced proliferative capacity. Interestingly, d10p2 exhibited expression of Il2ra. Il2ra is a marker for terminal T_{eff}. It was shown that at single cell level CD8+ T cells exhibit a pronounced asymmetric distribution in cells that were preparing for division. Cells that receive a prolonged il2 signals acquire terminal Teff characteristic (increased capacity for INF and Gzmb production).

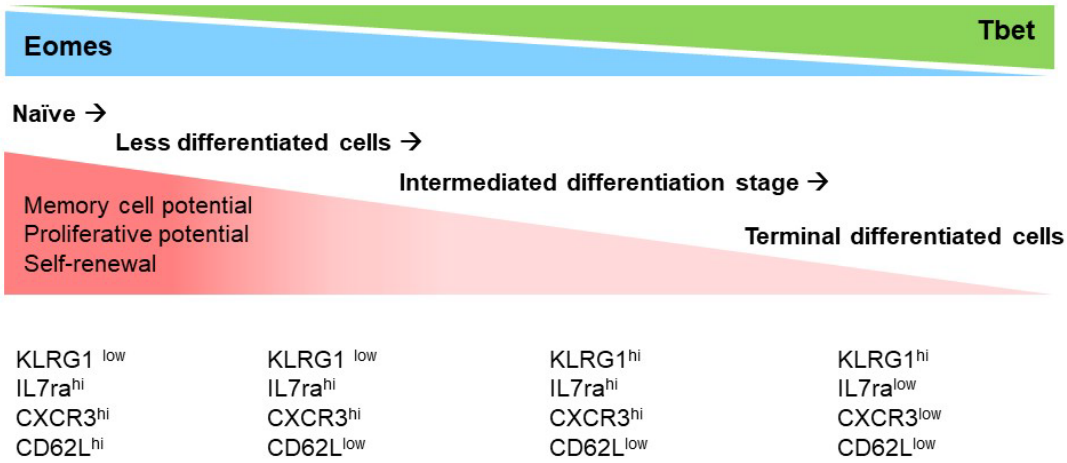
Expression index of Klrg1, Blimp1 and Il2ra (Fig 12) confirmed the effector pattern in Cd62l_neg compartments and showed the presence of cells with terminal effector

phenotype in SAM(H1)-induced cells at d10p2; Klrg1 was found to be uniquely expressed in SAM(H1)_Cd62l_neg cell population.

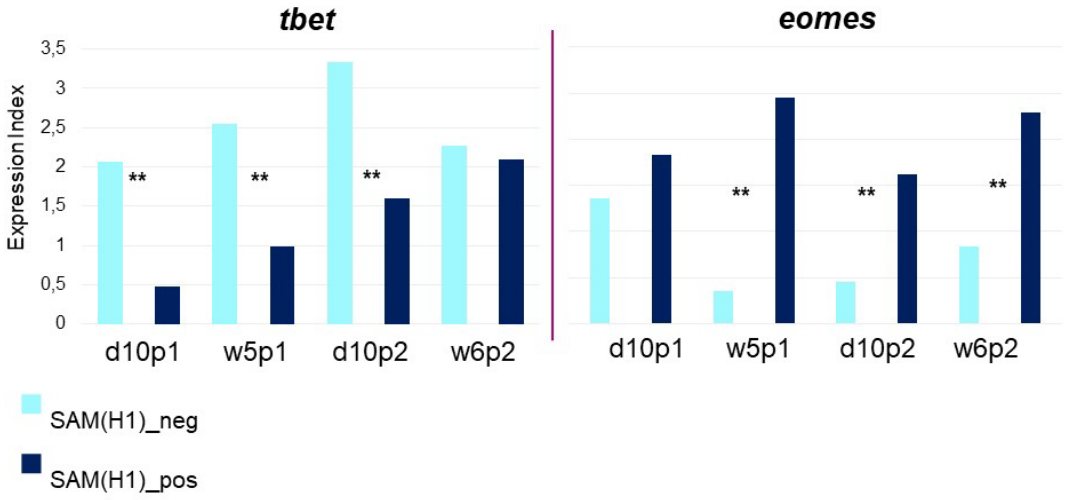
In early activated CD8+ T cells, Blimp1, Klrg1 and Il2ra cooperate through partially redundant activity to induce cytotoxic T cells inducing INF, GZMB, Perforin and CXCR3. [16,19,20,36]

Taken together, these results demonstrate that gene expression pattern in SAM(H1)_Cd62l_neg cells is characterized by a strong cytotoxic profile supported by the high ratio Tbet/Eomes. Changes in the percentage of co-expressed genes follow the trend of Tbet/Eomes. SAM(H1)_Cd62l_neg population expressed Klrg1 and blimp at all time points (Fig 12). Blimp is part of transcriptional program that increase the formation of $Klrg1^{hi}$ $Il7ra^{low}$ terminal effector cells and enhance functions, such as migration to site of inflammation and the expression of INF and GZMB.[16,21]

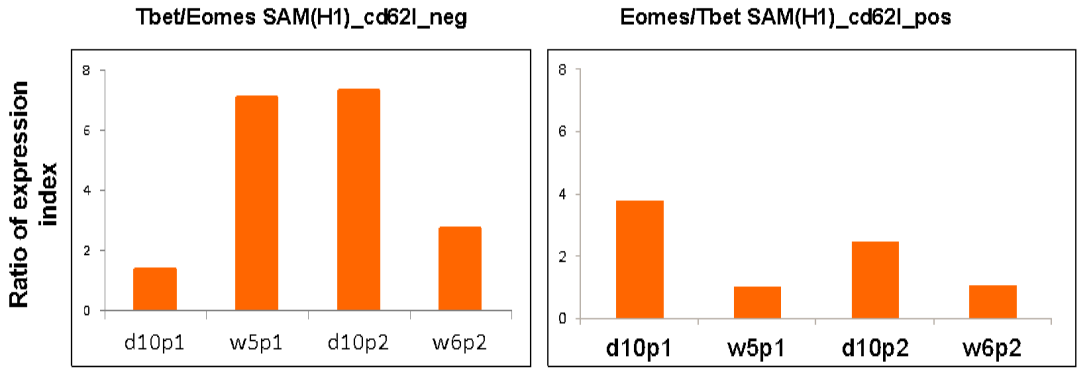
A.


Eomes
Tbet
Naïve →
Less differentiated cells →
Intermediated differentiation stage →
Terminal differentiated cells
Memory cell potential
Proliferative potential
Self-renewal
KLRG1 low
IL7ra hi
CXCR3 hi
CD62L hi
KLRG1 low
IL7ra hi
CXCR3 hi
CD62L low
KLRG1 hi
IL7ra hi
CXCR3 hi
CD62L low
KLRG1 hi
IL7ra low
CXCR3 low
CD62L low

B.


tbet
eomes
Expression Index
3,5
3
2,5
2
1,5
1
0,5
0
**
d10p1
w5p1
d10p2
w6p2
SAM(H1)_neg
SAM(H1)_pos

C.


Tbet/Eomes SAM(H1)_cd62l_neg
Eomes/Tbet SAM(H1)_cd62l_pos
Ratio of expression index
8
6
4
2
0
d10p1
w5p1
d10p2
w6p2

D.

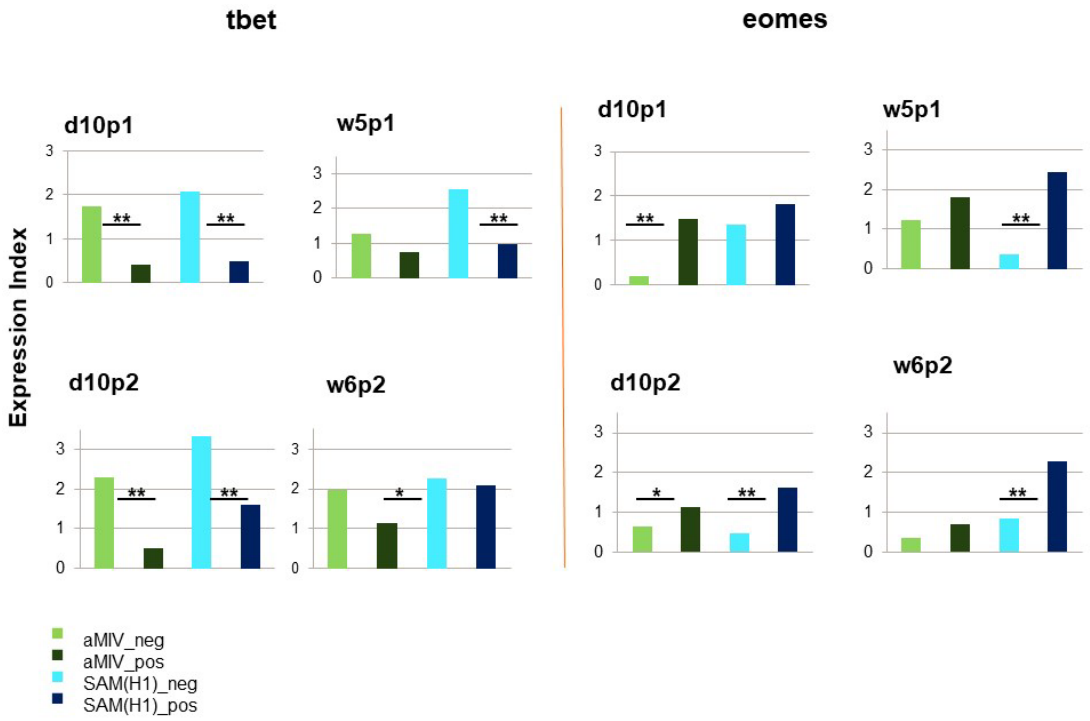


Figure 11: Tbet and Eomes expression in different compartments

A. CD8 + T cell differentiation model. Adapted from "Transcriptional control of effector and memory $CD8^+$ T cell differentiation" Kaech & Cui, Nature Review, 2012 **B**. Expression of Tbet and Eomes expression in SAM(H1)_Cd62l_neg (light blue) and SAM(H1)_Cd62l_pos (dark blue) compartments in different time points. **C**. Ratio between expression of Tbet/Eomes in SAM(H1)_Cd62l_neg cells and Eomes/Tbet in SAM(H1)_Cd62l_pos cells. **D.** Expression of Tbet and Eomes expression in aMIV_Cd62l_neg (light green), aMIV_Cd62l_pos (dark green), SAM(H1)_Cd62l_neg (light blue) and SAM(H1)_Cd62l_pos (dark blue) compartments.

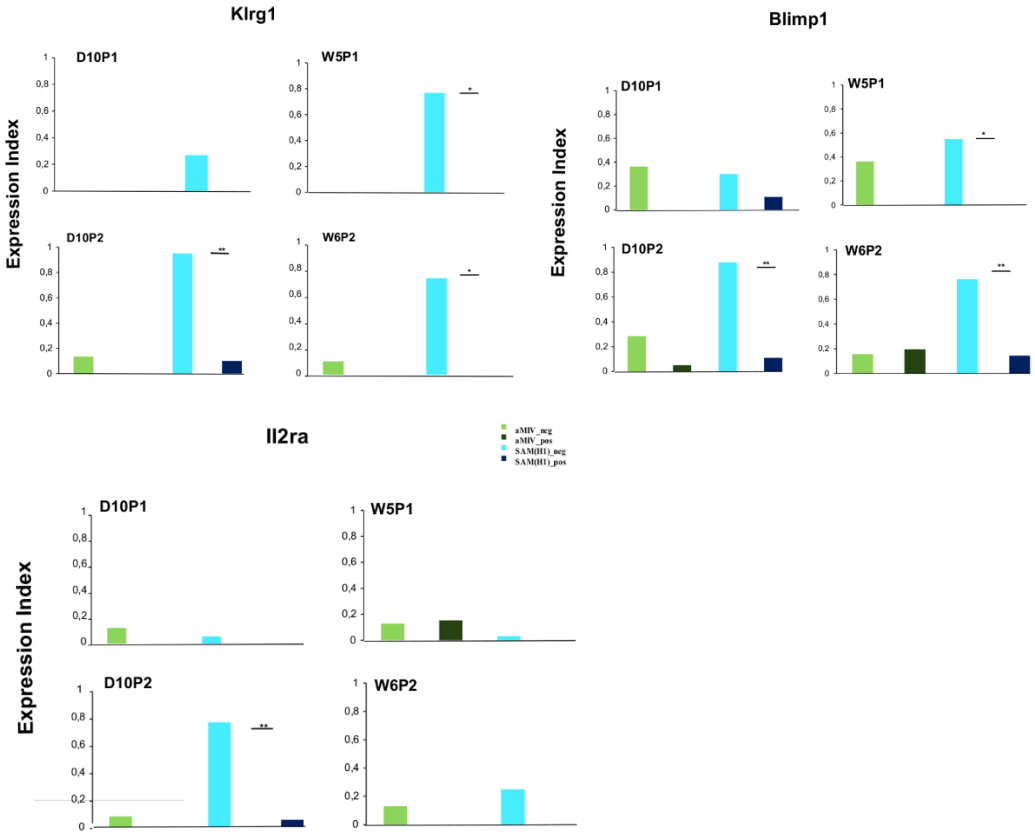


Figure 12: SAM(H1)_*Cd62l*_neg population at d10p2 is characterized by a terminal effector profile

Expression of *Klrg1*, *Blimp1* or *Il2ra* in aMIV_*Cd62l*_neg (light green), aMIV_*Cd62l*_pos (dark green), SAM(H1)_*Cd62l*_neg (light blue) and SAM(H1)_*Cd62l*_pos (dark blue) compartments. Expression is calculated as expression index (E.I.) = "percentage of positive cells" x "mean of expression of positive cells". Fisher exact test *=p<0,05, ** p=0,01 (Benjamini-Hochberg threshold 0.05).

3.6 Cd62l_neg CD8+ T cells shows consistent transcriptional differences between SAM(H1) and aMIV

Cd62l_neg CD8 T cell subsets induced by the two vaccine formulations were characterized by distinct transcriptional patterns at all analyzed time points. SAM(H1) Cd62l_neg populations were represented by a subset of cells characterized by the expression of multiple cytotoxic genes (Fig 13). Interestingly none of the aMIV-induced cells were transcriptionally similar to these. aMIV Cd62l_neg cells were transcriptionally less active than their SAM(H1) counterparts. Separation between clusters was driven by the expression of Il7ra, Cd27, Cxcr3 and Gata3 genes (Fig 13).

Scatter plots exhibited comparison in frequency of cells expressing genes in Cd62l_neg populations at all time points (Fig 14). The most relevant difference is given by the presence of higher number of cells in SAM(H1) groups that express cytotoxic/inflammation genes. A greater frequency of aMIV_Cd62l_neg cells from all time points than SAM(H1)_Cd62l_neg cells expressed genes encoding molecules that have inhibitory (Pd1), regulatory (Cd40l) or activation of CD8 + T cells through TNFSF pathway (gitr). Mainly gitr can be found higher expressed in aMIV_Cd62l_neg populations in all time points.

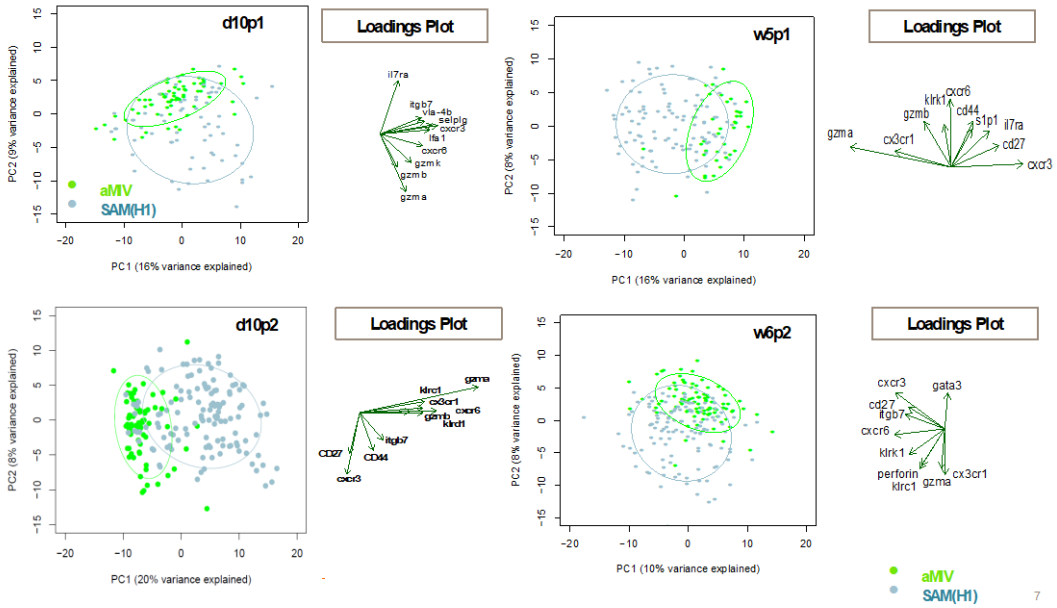


Figure 13: Cd62l_neg groups were transcriptionally different at all time points. PCA performed at d10p2 on aMIV_*Cd62l*_neg (light green), SAM(H1)_*Cd62l*_neg (light blue).

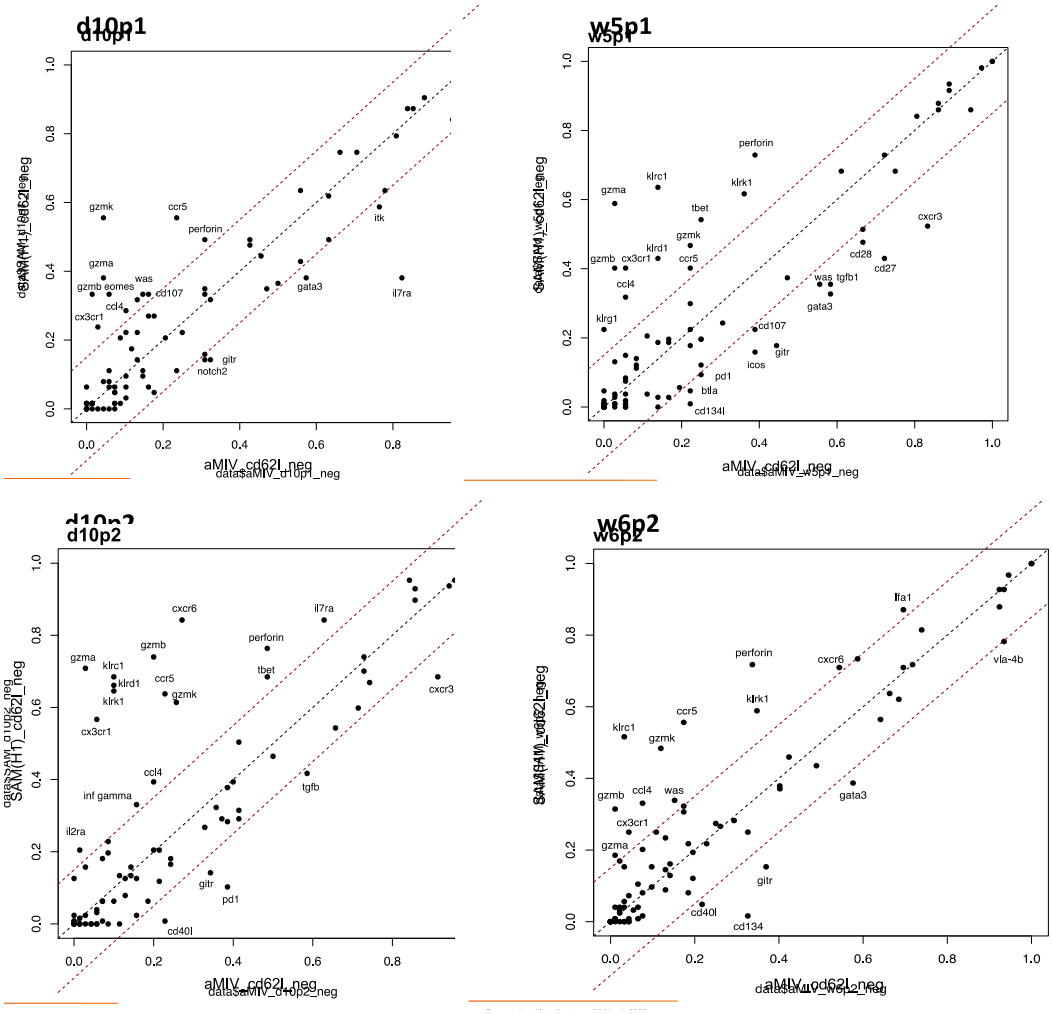


Figure 14: Cd62l_neg groups were transcriptional different in all time points Scatter plots show comparisons between A. aMIV_Cd62l_neg and aMIV_Cd62l_pos cells or B. SAM(H1)_Cd62l_neg and SAM(H1)_Cd62l_pos cells for the frequency of cells expressing the genes analyzed. Genes outside dashed lines are expressed from >15% of cells

3.7 Cd62l_pos cells exhibit transcriptional similarity between vaccines

A comparison between Cd62l_pos population and Cd62l_neg population were performed for both vaccines, to gain further insight into transcriptional patterns in the Cd62l_pos populations (Fig 5A+B). Cd62l_pos populations were generally characterized by a decreased percentage of transcriptionally active cells.

aMIV_Cd62l_pos population is characterized by an increased number of cells expressing Ccr7, Cxcr4, Il7ra compared to Cd62l_neg population (Fig 15A). Similarly, in SAM(H1)-induced cells, an increase in Ccr7, Btla, Eomes, Cxcr4 and Cd27 expression was more evident in Cd62l_pos cells than Cd62l_neg (Fig 15B). In both vaccine groups Cd62l_pos populations revealed fewer transcriptional differences compared to related Cd62l_neg populations. Specifically, Ccr7 and Cxcr4 were shared by both vaccine groups (Fig 15A+B), where Ccr7 was mainly expressed by Cd62l_pos populations in both vaccines at all time points (Fig 4C). Differently, Cxcr4 was observed in Cd62l_pos population in both vaccines at early time points but not at later time points, where differences were not evident (Fig 4D). Expression of both CCR7 with CD62L has been shown to characterize murine memory cells. This is well characterized at protein level30, where CXCR4 has been shown to have a critical role on CD8+ T cells homing to the bone marrow and in maintaining CD8 + T-cell memory pool31. Results suggest that presence/absence of Cd62l mRNA identified two cellular subsets: a Cd62l_pos memory like subpopulation and a Cd62l_neg effector like subpopulation.

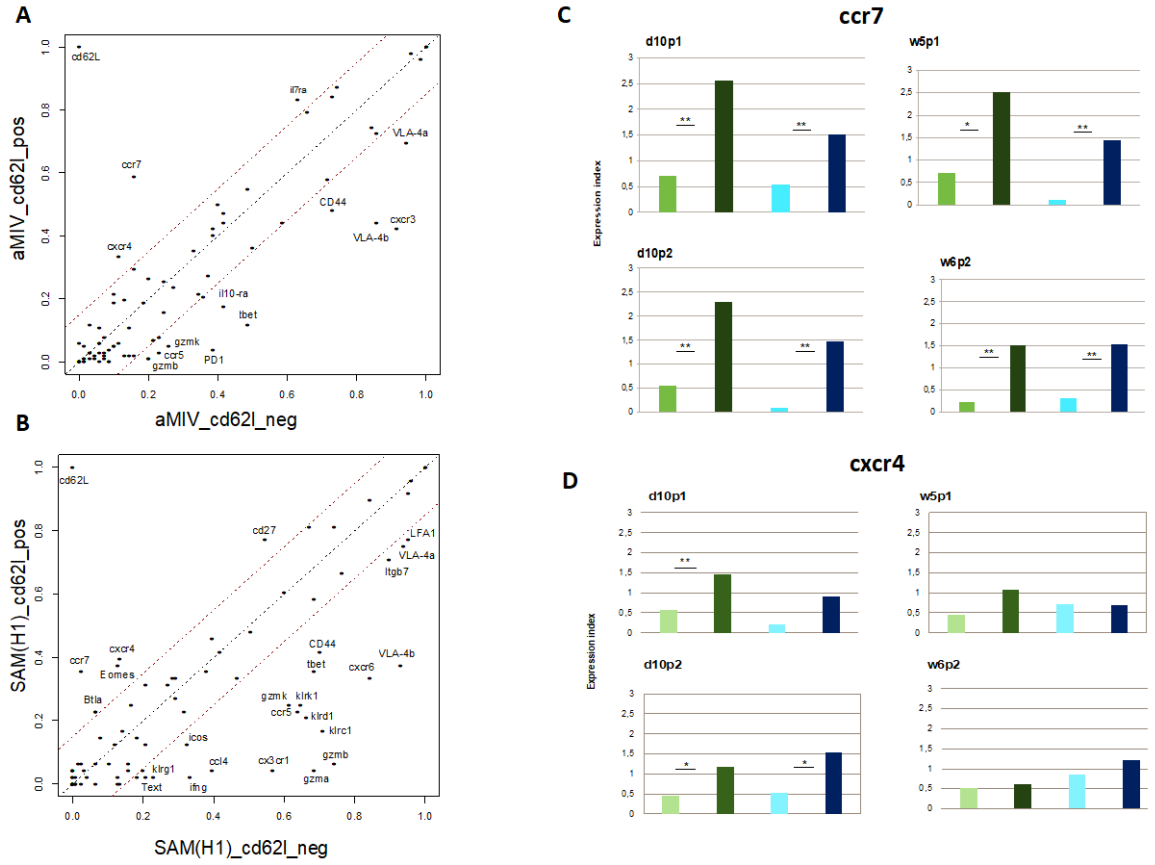


Figure 15: Cd62l_pos cells exhibit transcriptional similarity between vaccines

Scatterplots show comparisons between A. aMIV_Cd62l_neg and aMIV_Cd62l_pos cells or B. SAM(H1)_Cd62l_neg and SAM(H1)_Cd62l_pos cells for the frequency of cells expressing the genes analyzed. Genes outside dashed lines are expressed from >15% of cells. Bar plots show expression of C. Ccr7 gene and D. Cxcr4 gene across aMIV_Cd62l_neg (light green), aMIV_Cd62l_pos (dark green), SAM(H1)_Cd62l_neg (light blue) and SAM(H1)_Cd62l_pos (dark blue) compartments. Fisher exact test *=p<0,05, ** p=0,01 (Benjamini-Hochberg threshold 0.05).

3.8 Klf2 may act as a master regulator of Cd62l_pos CD8+ T cells

Ccr7 in preferentially expressed in Cd62l_pos populations. (Fig. 15 C). We hypothesized that this could be explained with the common transcription factor Klf2 that regulate the expression of both. To investigate this hypothesis, we look the behavior of another gene co-regulated by the same transcription factor S1p1. Klf2 regulates expression of Cd62l, S1p1, Ccr7 and Cxcr3. [83,135] S1p1 is expressed in all compartments (slightly upregulated in Cd62l_neg groups). Cxcr3 is downregulated in Cd62l_neg in aMIV but we didn't find appreciable differences in SAM cd62_neg versus cd62_pos compartments. Klf2 is highly expressed in naïve and memory T cells but only expressed at low levels in effector T cells such as CTL.

Klf2 expression in T cells is dynamic and determined by both the strength and duration of antigen receptor or cytokine signaling. Klf2 expression is controlled by FoxO.[136,137] Klf2 in naïve T cells control expression of Cd62l, an adhesion receptor essential for T cell transmigration from the blood into secondary lymphoid tissue.

Transcriptional profiling reveals that the loss of Klf2 is essential for T cells to acquire the full effector profile (Klf expressed upregulation of IFN gamma and perforin). Klf can upregulate the expression of Spi6. Klf2 expression also controls T cells trafficking by maintaining expression of Cd62l and S1p1; key molecules that control T cell entry and positioning in secondary lymphoid tissue.

CD8 T cells that maintain Klf2 failed to express the inflammatory chemokine receptor Cxcr3 and did not acquire the ability to migrate to his ligand CXCL10. Reacquisition of Cd62l and Ccr7 (after rapamycin treatment that induce expression of Klf2) and loss of Cxcr3.[135,136]

Ccr7-S1p1 tend to be co-expressed in Cd62l_pos populations mainly in aMIV at early time points (Fig 16). SAM(H1) was characterized by this trend but with lower percentage of cells that co-expressed this genes in Cd62l_pos populations. The transcriptomic profiling at single-cell level seemed to confirm that cells expressing Cd62l, also expressed Ccr7 and S1p1, genes regulated by Klf2.

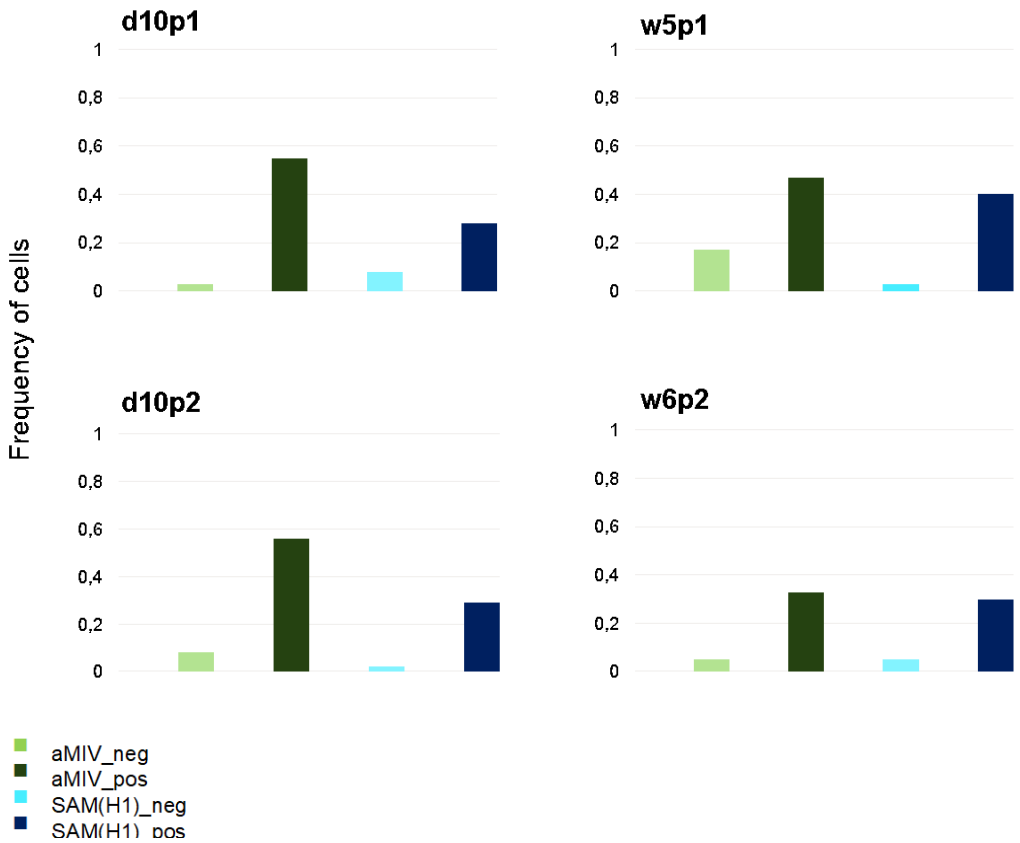


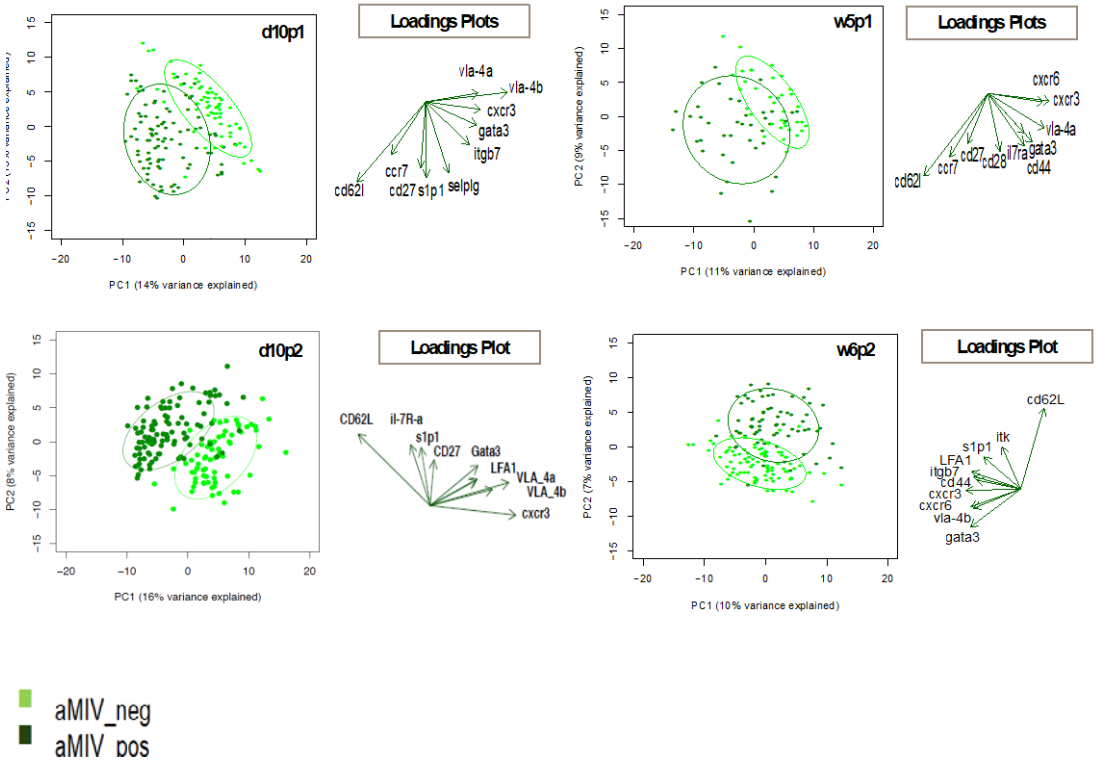
Figure 16: Ccr7-S1p1 co-expression

S1p1-Ccr7 co-expression across aMIV_Cd62l_neg (light green), aMIV_Cd62l_pos (dark green), SAM(H1)_Cd62l_neg (light blue) and SAM(H1)_Cd62l_pos (dark blue) compartments at all time points (d10p1, w5p1, d10p2, w6p2)

3.9 Transcriptional differences between aMIV Cd62l_pos vs aMIV Cd62l_neg at all time points

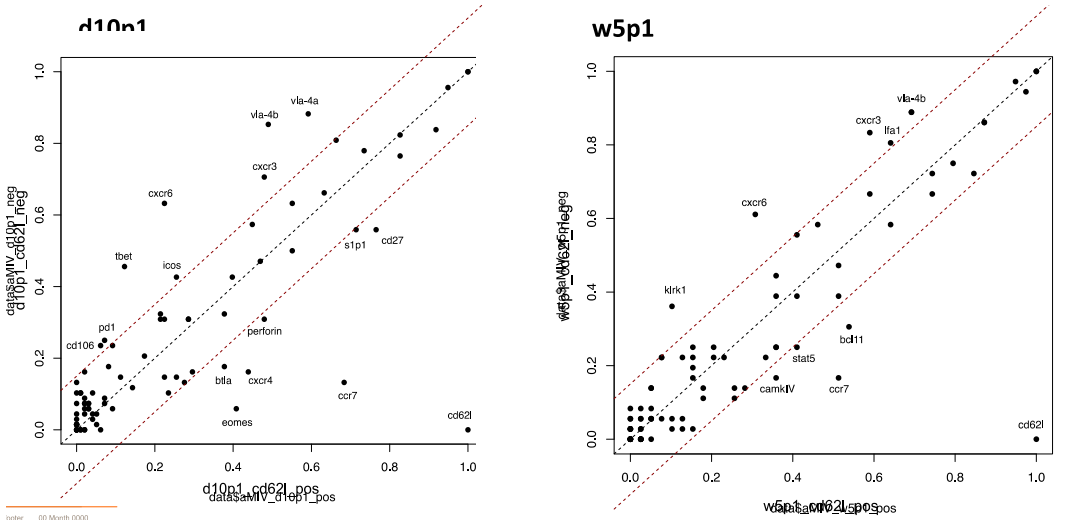
Cd62l_pos population showed different transcriptional patterns compared with Cd62l_neg populations in all time points. This confirm that the status of Cd62l expression led to two different transcriptional patterns. aMIV neg populations exhibited an increased number of cells expressing Vla-4a, Vla-4b in all time points, compared to Cd62l_neg population (Fig 17B). In aMIV_pos populations higher frequencies of cells express Ccr7 in all time points, Cxcr4 only in early time points. (Fig 17B). Interestingly Cxcr3 was among most informative gene in all time points, showing a contribute of this gene in clustering aMIV_Cd62l_neg populations in PCA plot. (Fig 17A).

A.



aMIV_neg
aMIV_pos

B.



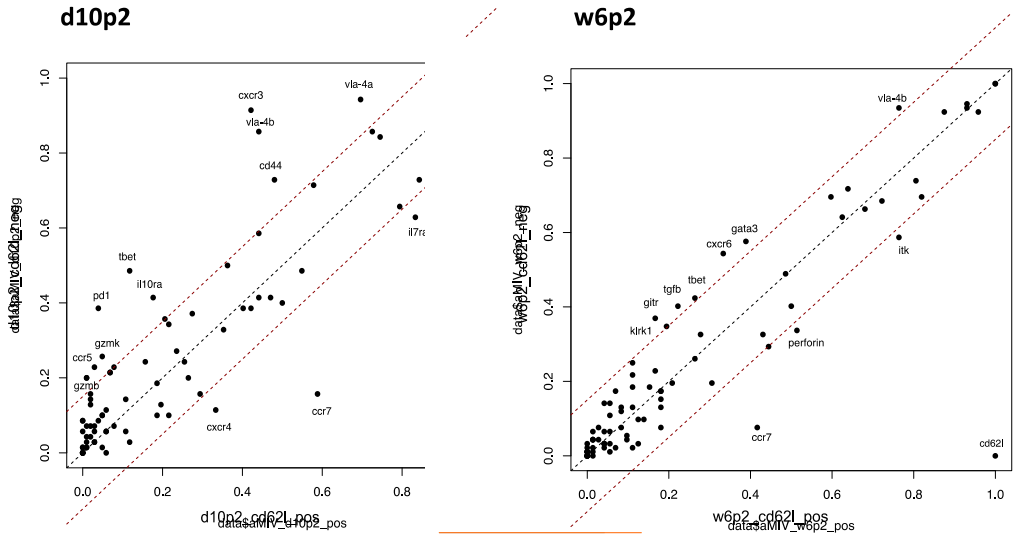


Figure 17: aMIV Cd62l_neg vs Cd62l_pos groups were transcriptional different in all time point

A. PCA performed on aMIV_Cd62l_neg (light green), aMIV_Cd62l_pos (darkgreen) compartments, in all time points, with indicated 10 most informative genes. **B.** Scatter plots show comparisons between aMIV_Cd62l_neg and aMIV_Cd62l_pos cells in all time points for the frequency of cells expressing the genes analyzed. Genes outside dashed lines are expressed from >15% of cells.

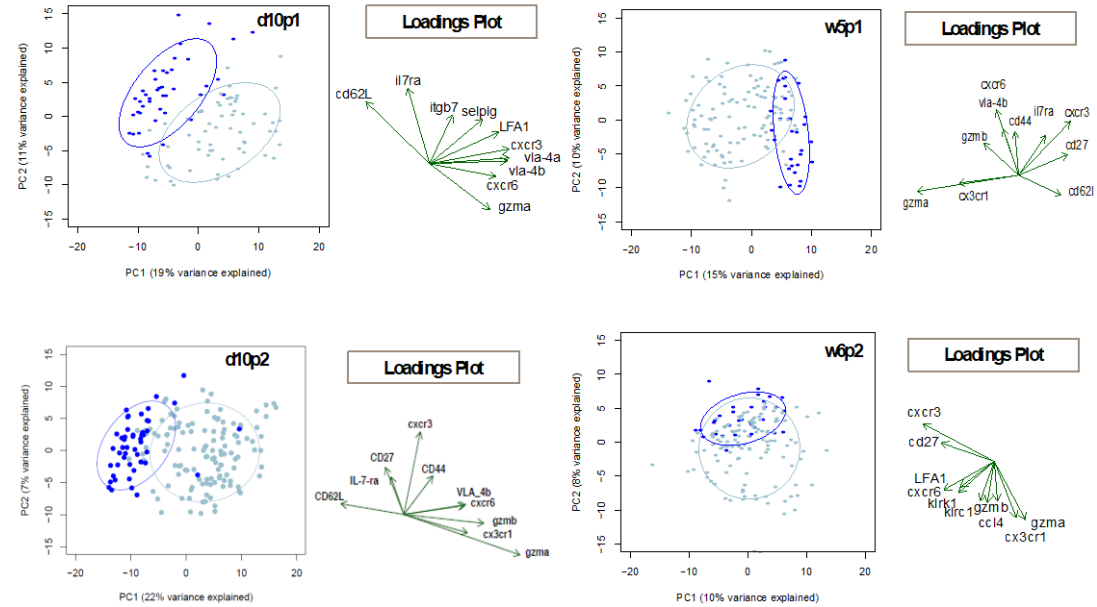
3.10 Transcriptional differences between SAM(H1) Cd62l_pos vs Cd62l_neg at all time points

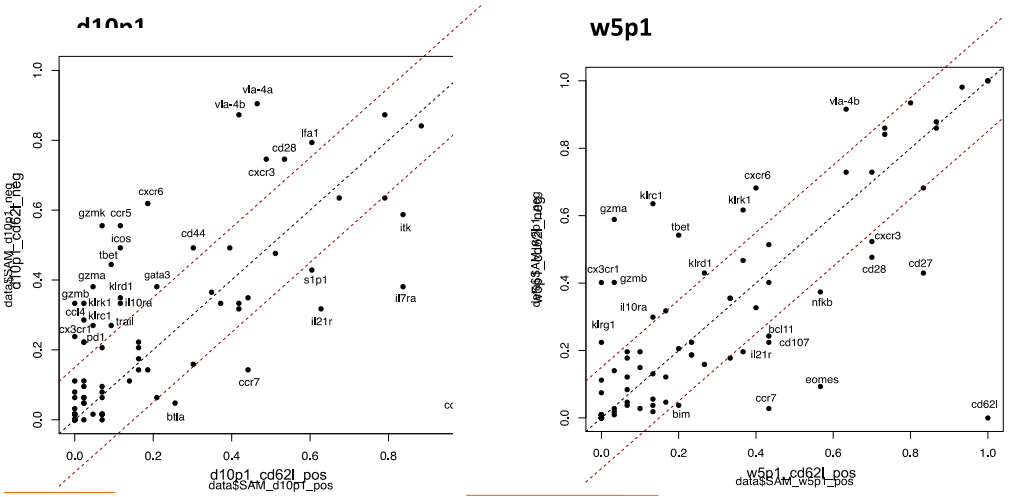
Cd62l_pos population exhibited different transcriptional patterns compared with Cd62l_neg populations in all time points. This confirms that the status of Cd62l expression led to two different transcriptional patterns. Most informative genes (Fig 18A) that distinguished SAM(H1)_Cd62l_neg populations highlighted higher frequencies of cells expressing cytolytic, cytotoxic, inflammatory genes, confirming the effector cytotoxic pattern found at d10p2 that is strongly diminished at w6p2. W6p2 PCA revealed a weak structure, due to the low number of positive cells.

Interestingly, loading plots highlighted that Cd62l is opposite to Gzma, confirming the found that Gzma was not expressed in Cd62l_pos population.

SAM(H1) neg populations showed an increased number of cells expressing granzymes. Cxcr6, Klr genes, compared to Cd62l_neg population (Fig 18B). In aMIV_pos populations higher frequencies of cells express Ccr7 in all time points. (Fig 18B).

A.


d10p1
Loadings Plot
w5p1
Loadings Plot
d10p2
Loadings Plot
w6p2
Loadings Plot
PC1 (19% variance explained)
PC2 (11% variance explained)
PC1 (15% variance explained)
PC2 (10% variance explained)
PC1 (22% variance explained)
PC2 (7% variance explained)
PC1 (10% variance explained)
PC2 (8% variance explained)
cd62L
il7ra
itgb7
selplg
LFA1
cxcr3
vla-4a
vla-4b
cxcr6
gzma
cd27
cd44
gzmb
cx3cr1
CD62L
IL-7-ra
CD27
CD44
VLA_4b
ccl4
klrc1
klrk1
SAM(H1)_neg
SAM(H1)_pos


d10p1
w5p1
d10p1_cd62l_pos
w5p1_cd62l_pos
data$SAM_d10p1_neg
data$SAM_d10p1_pos
data$SAM_w5p1_pos
vla-4a
vla-4b
lfa1
cd28
cxcr3
cxcr6
gzmk
ccr5
icos
cd44
tbet
itk
gzma
gata3
klrd1
s1p1
gzmb
klrk1
il10ra
il7ra
cd4
klrc1
traf1
il21r
cx3cr1
pd1
ccr7
btla
klrk1
klrc1
gzma
tbet
cxcr3
cd28
cd27
klrd1
cx3cr1
gzmb
il10ra
nfkb
klrg1
pcl11
cd107
il21r
eomes
ccr7
bim
cd62l

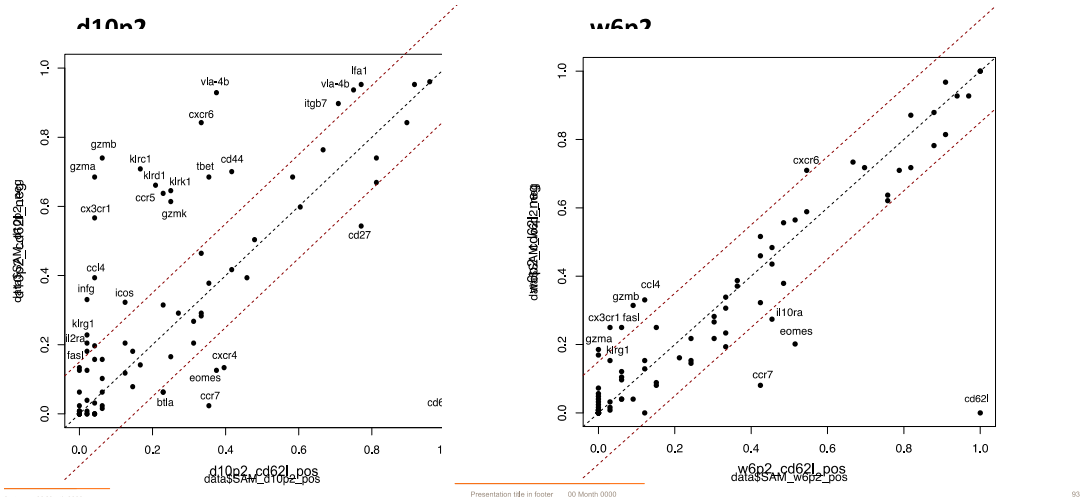


Figure 18: SAM(H1)_Cd62l_neg vs Cd62l_pos groups were transcriptional different in all time points

A. PCA performed on SAM(H1)_Cd62l_neg (light blue), SAM(H1)_Cd62l_pos (dark blue) compartments, in all time points, with indicated 10 most informative genes. **B.** Scatter plots show comparisons between SAM(H1)_Cd62l_neg and SAM(H1)_Cd62l_pos cells in all time points for the frequency of cells expressing the genes analyzed. Genes outside dashed lines are expressed from >15% of cells.

3.11 Cd62l and Il7ra combinations define distinct subpopulations

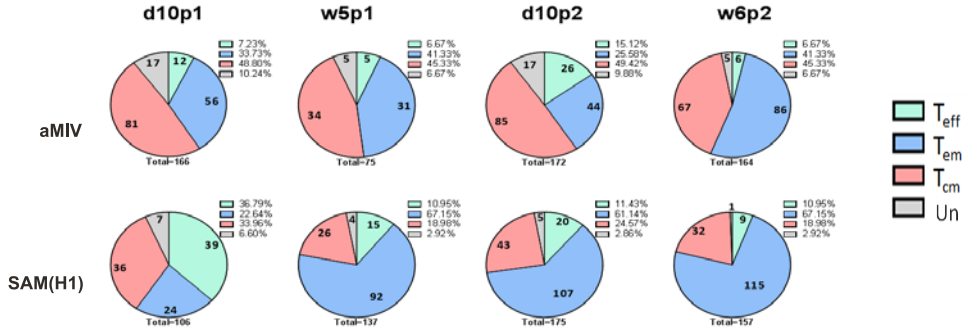
Mouse analysis of surface proteins such as IL7R and CD62L allowed the definition of three differentiation states: effector T_{eff} (IL7R-, CD62L-), effector memory T_{em}(IL7R+, CD62L-) and central memory T_{cm} (IL7R+, CD62L+), characterized by a gradient of proliferative and cytotoxic potential. Previous studies on single CD8+ T-cell transcriptomic profiles have shown different set of genes expressed in T_{cm} and T_{em} elicited by different vaccine regimes.[113] It was hypothesized that sub-sampling cells in T_{eff}, T_{em} and T_{cm} cells based on the presence or absence of mRNA for Cd62l and il7r might provide useful insights to capture differences between vaccines PCA revealed that T_{eff}, T_{em} and T_{cm} cells clustered distinctly and a similar trajectory in both groups were found, which supports the gradient of cytotoxic and proliferative potential (T_{cm}< T_{em} < T_{eff}) (Fig 19B).

Our analysis was focused on the comparison of each subpopulation between vaccines. PCA highlighted a great overlap in T_{cm} populations, indicating transcriptional similarity between vaccines (Fig 19C). Differently, PCA on T_{eff} and T_{em} revealed separation between cell clusters, highlighting different transcriptional patterns activated by the two vaccine formulations (Fig 19B). Higher proportions of cells within SAM(H1)_T_{em} compartment expressed cytotoxic and pro-inflammatory genes; confirming the cytotoxic profile of CD8+ T cells elicited by SAM(H1) (Fig 5C). Interestingly, aMIV_T_{em} population showed higher proportion of cells expressing Pd1, Cd40l and Gitr compared to SAM(H1)_T_{em} populations (Fig 20A). Proportion of cells expressing these genes increased in aMIV_T_{eff} population (Fig 20B). Moreover, a trend of higher number of cells expressing genes that belong to TNFR (Gitr, Cd27, Trail) and CD28 family (Pd1, Icos) was observed. Members of the CD28 family provide co-stimulation to CD8+ T cells whereas, members of TNFR family have being implicated in the survival and maintenance of activated CD8 T cells.[138–140]

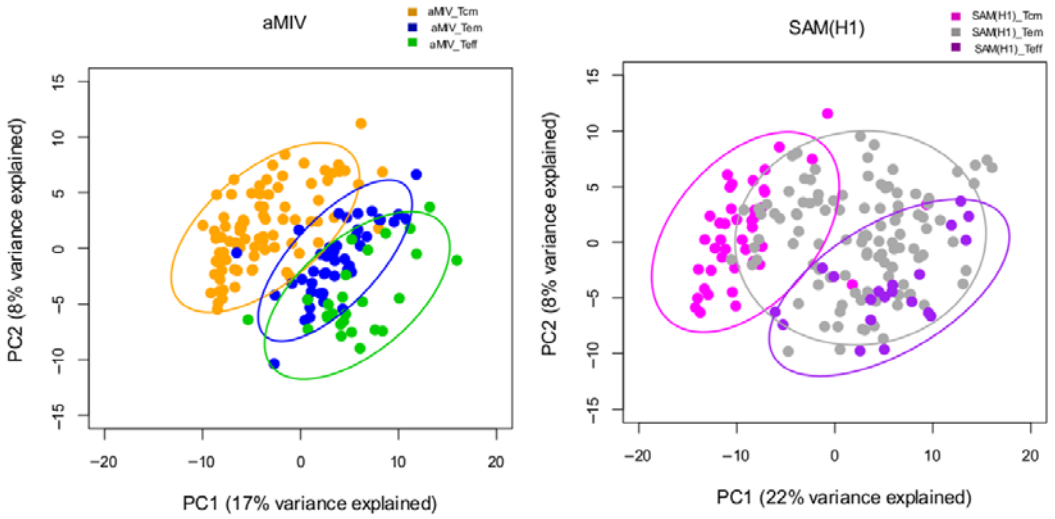
These results suggest that CD8+ T cells elicited by the aMIV vaccine contain a subpopulation of activated cells characterized by a regulatory and exhausted profile. It was observed that Pd1 is expressed mainly in aMIV_Cd62l_neg populations (Fig 20C). It was hypothesized that the expression Pd1 at d10p2 inhibited the effector function in aMIV CD8+ T cells and they switched through a regulatory/helper

profile. It is possible that Gitr expression may have some critical immune-regulatory role that could compensate for the missing cytotoxic activity. Finally, a small subset of aMIV elicited cells expressed Cd40l (Fig.20). We hypothesized that aMIV induces a particular CD8+CD40L+ T cell sub-population, which have been characterized in various immune responses as a subset of CD8+ memory/effector T cells.[141,142]

A.



B.



C.

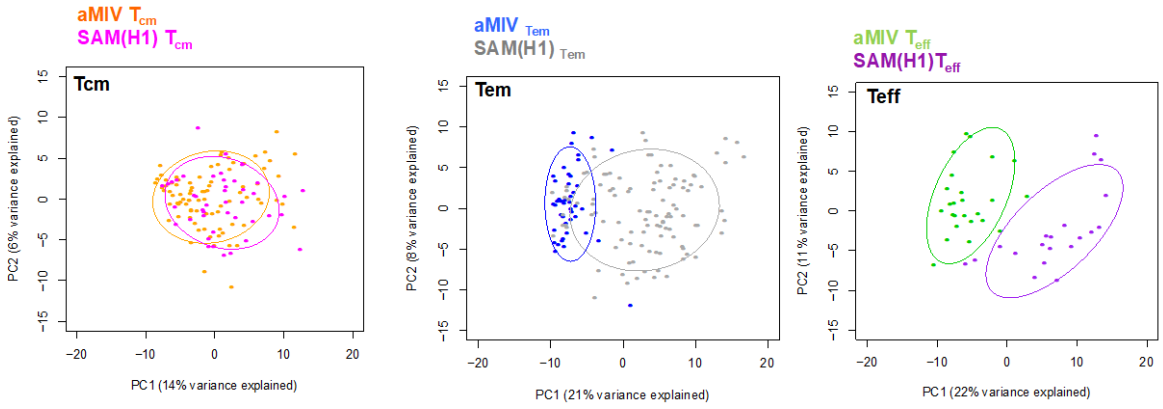
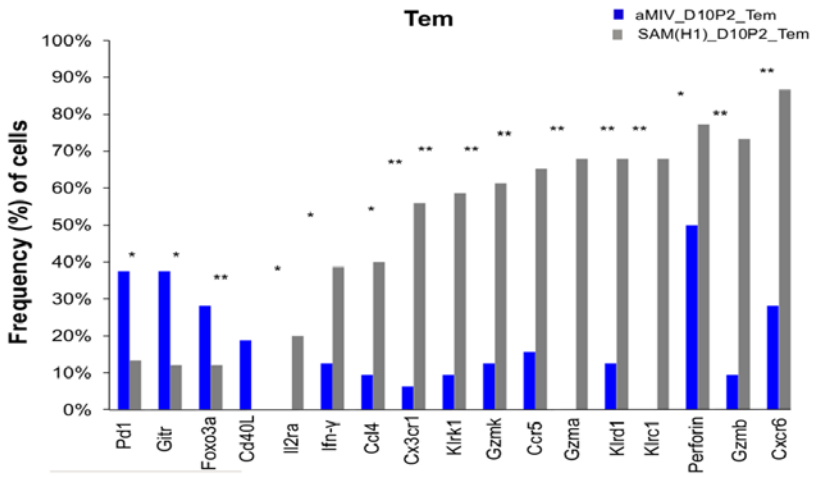


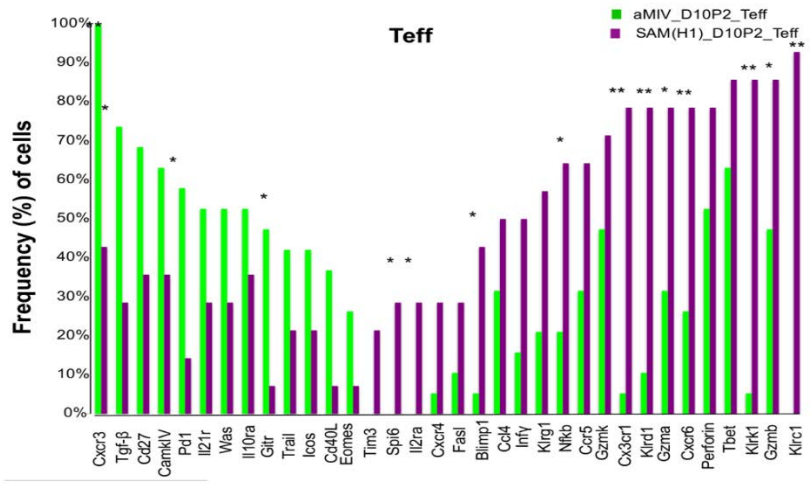
Figure 19: Cd62l and Il7ra combinations highlight distinct subpopulations

A. Pie charts show frequencies of Teff, Tem, Tcm cells in different vaccines in all time points. **B.** PCA of aMIV effector T_{EFF} ($IL7R^-$, $CD62L^-$) (orange), effector memory T_{EM} ($IL7R^+$, $CD62L^-$) (blue) and central memory T_{CM} ($IL7R^+$, $CD62L^+$) (light orange) and SAM(H1) T_{EFF} (pink), T_{EM} (grey) and T_{CM} (purple) cells at d10p2. **C.** PCA comparison between vaccines in the T_{EFF}, T_{EM}, and T_{CM} compartments.

A.



B.



C.

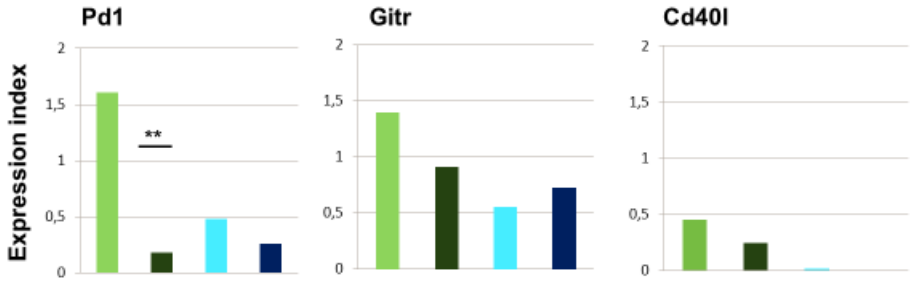


Figure 20: Tcm and Teff subpopulations showed distinct transcriptional patterns.

Bar plots show frequencies of cells expressing genes in T_{EM} (**A**) and T_{EFF} (**B**) cells at d10p2. **C.** Bar plots showing expression of Pd1, gitr or Cd40l across

aMIV_Cd62l_neg (light green), aMIV_Cd62l_pos (dark green), SAM(H1)_Cd62l_neg (light blue) and SAM(H1)_Cd62l_pos (dark blue) compartments. Expression is calculated as expression index (E.I.) = “percentage of positive cells” x “mean of expression of positive cells”. Fisher exact test * = p < 0,05; ** p = 0,01 (Benjamini-Hochberg threshold 0.05).

4 – Discussion and Conclusions

4.1 Discussion

In recent years single-cell transcriptomic analysis has been applied to dissect heterogeneity and understand molecular mechanism in immune responses[70–72]. It has been shown that information on heterogeneity of single-cell gene expression cannot be appreciated using conventional bulk approaches[71]. Gene expression analysis at single-cell level is a powerful method that could help gaining a deeper understanding of the biological processes involved in the response to vaccination. Single-cell approach can refine our cellular classification schemes and enable the direct assessment of molecular mechanisms that have been obscured in bulk samples where cells of a family (*eg* CD8+ T lymphocytes) were considered identical. Understanding new mechanisms of actions following vaccination and revealing new and functionally distinct subsets of immune cells will provide the opportunity to improve and develop better cross-protective vaccines.

Protein-based influenza vaccines have been commercially available for a long time and continue to be the most successful strategy to reduce disease burden and economical costs associated with this particular virus infection.[27,29,30] Currently, the efficacy of licensed vaccines depends on the match between the annual vaccines and circulating strains and better cross-reactive vaccines need to be developed. In order to overcome this problem, the nature of the immune response needs to be investigated and exactly understood to direct the vaccine to target both humoral and cellular immunity.

In this study, we focused on the characterization of transcriptome responses of $HA_{533\text{-}541}$-pentamer+ CD8 T cells at single cell level. For this purpose, a mouse model of influenza was used in which mice were immunized with RNA_based SAM(H1) or MF59-adjuvanted subunit H1N1 aMIV. These two vaccine formulations were previously shown to induce similar protection rates in mice after infection.[25,26,40] Moreover, both vaccines stimulated antigen-specific CD4 T cells, but only SAM(H1) induced a robust CD4 T_H1 cell response, detected in splenocytes after in vitro stimulation with HA peptides. Such antigen specific responses were characterized by using a combination of five cytokines (IL2, IL4/Il3, TNF-α, IFN-γ). In addition, SAM(H1) but not aMIV, induced cytokine-producing CD8 T Th1 cells[25].

For a deeper characterization of antigen-specific CD8+ T cells after vaccination, BALB/c restricted MHC-I $HA_{533-541}$-pentamers were used ex vivo to identify the population of interest[143,144]. Both aMIV and SAM(H1) immunization induced pentamer+ CD8 T cells, despite the common belief that protein vaccines, like aMIV, only induce humoral and cellular CD4+ T cells responses, and not CD8+ T cells.[25]

In vitro stimulation is currently the most common way for measuring antigen-specific T cells. While this type of analysis fit for identification of already known cellular populations, it cannot properly describe the complex and heterogeneous immune response landscape induced after vaccination and new methods for characterization of CD8 T cells need to be developed.

In this study, single-cell transcriptomic analysis was applied to antigen-specific $CD8^+$ T cells elicited by two different influenza vaccine formulations to assess differences within cellular immunity. This provided a new analytical approach to characterize and study $CD8^+$ T-cell responses to vaccination. This new approach could help CD8 T cells characterization given that the magnitude of $CD8^+$ T cells alone is usually not a reliable correlate of protection[70].

In the past, several studies of influenza vaccine-induced transcriptome responses have been focused on samples derived from peripheral whole-blood of vaccinated individuals.[8,11,140,145] These studies allowed for the identification of several different metabolic pathways related to T cell activity and differentiation. Single-cell transcriptomic studies, on the other end, were able to go deeper revealing consistent heterogeneity in the transcriptional response of activated T lymphocytes and allowed the characterization distinct subpopulations that were precursor of specific T cells lineages. [15,71]

Using single-cell high throughput RT-qPCR analysis, it was possible to capture differences in the transcriptome response of $CD8^+$ T cells elicited by the two vaccine formulations tested in this study. After immunization, CD8 T cells go through a transition state from quiescent, poor effector cells to metabolically active, proliferating cells with cytolytic functions[16,17,27,146,147].

Cd62l has been identified as a key gene in distinguish transcriptionally different CD8+ T cell subpopulations. Specifically, SAM(H1)-induced Cd62l_neg cells were characterized by an effector/cytotoxic expression profile, in agreement with previous findings[71]. Differently, aMIV-induced Cd62l_neg cells were transcriptionally less active than their SAM(H1) counterparts. Instead, Cd62l_pos populations showed

fewer transcriptional differences between vaccines and were generally characterized by a memory-like phenotype, characterized by the expression of Ccr7, Cxcr4 and Il7ra markers. (Fig. 15)

We next focused on the identification of gene co-expression patterns and the characterization of vaccine-induced transcriptional states. Single-cell gene expression analysis allowed us to perform this kind of analysis given the higher power of analysis resolution. In particular, 27% of the assessed SAM(H1)_Cd62l_neg T cells co-expressed Cx3cr1, Cxcr6, Gzma, Gzmb, Klrc1 and Klrd1 (data relative to the d10p2 time point), indicating a cytotoxic effector phenotype (Fig. 9 B). This same co-expression pattern was not found at other time points. It was also investigated which were the most co-expressed 6, 5, 4 or 3 genes and a similar trend was observed (Fig 9 B). The three most co-expressed genes were Cxcr6, Gzma and Gzmb, occurring in 57% of cells in d10p2. This percentage dropped at w6p2. Interestingly, after prime, the percentage ranged between 20% at d10 and 27% at w5. In SAM(H1)_Cd62l_neg cells an upregulation of Blimp, Klrg1 and a group of cells expressing Il2ra was observed. Klrg1 has been previously characterized as a marker of terminal differentiation in $CD8^+$ T cells[21]. In addition, Blimp and Il2ra cooperate for short-lived effector cells (T_{SLE}) formation[131,148]. We hypothesized that after boost, $CD8^+$ T cells elicited by SAM(H1) received very high transcriptional activation with a strong co-expression of cytotoxic and inflammation genes to enhance a strong effector profile.

It was previously reported that CD8+ T cell effector molecules share several regulatory elements and it was proposed that once an individual cell would acquire some of these components de novo expressed gene would be expressed preferentially in that cell.[16,149] This could explain the trend of co-expression of cytotoxic-effector genes that we described (Fig. 9 and Fig 10).

The trend of co-expression that we have shown it is supported by a previous study in which two putatively different CD8 memory T cells populations were characterized in response Listeria monocytogenes OVA infection[150]. The authors isolated individual cells at the different points of immune reaction and in each cell, they evaluated the simultaneously expression of 14 T-cell effector genes. The authors highlighted that different effector genes were induced at different time points of the response and transcribed during different time periods. Moreover, CD8 T cells tend

to co-express Gzms, Infg, Fasl and Prf1 and they are characterized by the loss of co-expression of killer molecules in the effector–memory transition.[150]

It has been previously suggested that the balance between Tbet and Eomes determine effector cell fate in $CD8^+$ T cells[21,36,148]. Given that our results highlighted a strong effector/cytotoxic transcription profile related to SAM(H1)_Cd62l_neg populations we investigated the level of expression of Tbet and Eomes in the different CD8_ Cd62l negative and positive populations. Transcript abundance analysis of the SAM(H1) induced cells revealed that Tbet is expressed preferentially in Cd62l_neg cells, whereas Eomes was mostly expressed in Cd62l_pos population (Fig 11B). Ratio between Tbet and Eomes in Cd62l_neg populations at all time points exhibited an opposite trend between prime and boost. Ratio increased from d10 to w5 after prime. Conversely, after boost the highest ratio was found at d10 followed by a decrease at w6 (Fig 11C). Positive Tbet/Eomes ratio would support the hypothesis that SAM(H1)_Cd62l_neg subpopulation activated a transcriptional effector profile that follows two opposite trends after prime and boost.

Cd62l_pos cells exhibited great overlap between cellular response elicited by two vaccines, explained by similar transcriptional pattern. Increased number of cells in the Cd62l_pos population, for both vaccines, expressed Ccr7 and Cxcr4. Ccr7 and Cd62l have been characterized in murine $CD8^+$ T cells as marker of T_{cm} population[151]. This finding supports our hypothesis that Cd62l mRNA presence and absence define two transcriptional states that can be defined as memory and effector like.

Surface expression of IL7R and CD62L markers can be used to define three T cell differentiation states, generally referred to as effector cells ($IL7R^{LOW}$, $CD62L^{LOW}$), effector memory ($IL7R^{HIGH}$, $CD62L^{LOW}$), and central memory ($IL7R^{HIGH}$, $CD62L^{HIGH}$) [24,70,113]. The memory compartment consists of T cells that can rapidly acquire effector functions to kill infected cells and/or secrete inflammatory cytokines that inhibit replication of the pathogens. The memory compartment is important for long-lived immunological protection and cells express a pattern of surface proteins that are involved in cells adhesion and chemotaxis. [16,149]

The question was if it was possible to combine Cd62l and il7r also as mRNA expression? In a previous work, a global transcriptional profile of distinct CD8 T cells subsets (T_{naive}, T_{eff}, T_{em}, T_{cm}) study was performed[113]. CD8 T cells were induced by three distinct prime-boost vaccine regimens and microarray profiling was

performed. Authors have shown that transcriptional profiles were similar between the same population from distinct vaccines but when subsets were compared in the order T_{naive}> T_{cm}> T_{em}> T_{eff} high number of genes up- and down- regulated raised. Microarray profiling showed downregulation of Sell (encoding CD62L) specifically in T_{EM} and T_{EFF} and downregulation of Il7r expression in T_{EFF} cells alone, consistent with sorting strategy. Moreover, changes in the expression of individual genes between distinct CD8+ T-cell subsets were similar between all three vaccination protocols.[113]

Starting from these results, T_{eff}, T_{em} and T_{cm} were selected based on presence/absence of Cd62l and Il7ra mRNA. PCA on selected sub-population revealed a structure that highlight a continuum from T_{cm} > T_{em}> T_{eff} transition that confirmed previous findings[113]. By comparing cells in distinct subsets between vaccines, it was observed that T_{cm} cells appeared to be transcriptionally similar. However, T_{em} and T_{eff} are characterized by major differences (Fig 19C). SAM(H1)_T_{em} and SAM(H1)_T_{eff} confirmed the cytotoxic and inflammatory transcription pattern related to Cd62l_neg population.(Fig 20 A and B). Interestingly, higher percentage of aMIV_T_{em} cells expressed Pd1, Gitr and Cd40l (Fig 20C). PD1 biomarker is a regulatory receptor in the CD28 superfamily and has been found to serve as an important regulator of T-cell function. Exhausted $CD8^+$ T cells express high levels of PD1, and some studies have found a high PD1 expression to inhibit $CD8^+$ T-cell function[139,152]. Exhausted T cells progressively lose their ability to kill other cells and produce cytokines[139]. GITR belongs to TNFR superfamily member and positively regulate survival. It is expressed on T-regulatory cells and the expression of GITR is maintained in effector cells. It regulates functional balance between regulatory and effector T cells, while enhances IL2R, IL-2 and INF-gamma expression[138]. CD40L contribute to the activation of $CD8^+$ T cells. Key helper molecule $CD8^+CD40L^+$ T cells were previously characterized, and these cells were present in various immune response and they have a cytokine expression signature resembling conventional $CD4^+$ helper T cells rather than cytotoxic T cells[141].

A previous study, focused on how multiple antigen encounters impact memory $CD8^+$ T cell, demonstrated that every additional antigen stimulation (primary to quaternary) leads to an increase in the number of differentially regulated genes and thus to further differentiation of memory $CD8^+$ T cells and repeated antigen stimulation results in memory CD8+ T cell populations that possess a unique repertoire of

regulated genes and phenotypic peculiarities[153]. According to these results, we have shown that CD8 T cells elicited by SAM(H1), increased effector cytotoxic profile after boost at early time point (Fig 9 and Fig 13). This transcriptional specialization was associated with no expression of Cd62l. Conversely, specialization into a memory-like phenotype was associated with expression of Cd62l (Fig 15). Overall, these findings confirm previously reported data, according to which Cd62l expression within cells may influence the transcriptional profile of CD8 T cells.[71]

In conclusion, we have established an analysis procedure for the isolation and transcriptional profiling of vaccine-induced individual $CD8^+$ T-cell responses. Analysis of the collected transcriptome information allowed getting new insights into cellular and molecular basis of CD8+ T cell responses to protein- and mRNA-based vaccines. Our findings revealed that antigen-specific $CD8^+$ T cells elicited by SAM(H1) exhibited a cytotoxic-effector phenotype, which was confirmed by a cytotoxicity killing assay (data not shown) as well as independent studies[25,26]. For the first time ever, we have detected and characterized aMIV-elicited $CD8^+$ T cells, which exhibited an inhibitory/regulatory profile, defined by the presence of Pd1 and Cd40l mRNAs molecules (Fig. 20C).

Overall, we proofed that single-cell transcriptome profiling is a valuable analytical tool, which allows for a high resolution analysis of the cellular events triggered by vaccination and holds the potential to unravel biological dynamics that are not captured by conventional bulk-approaches.

Disclaimer about sponsorship, financial support and conflict of interest:

This study was sponsored by Novartis Vaccines, now acquired by the GSK group of companies. All authors have declared the following interests: FB was PhD student at University of Turin at the time of the study. Following the acquisition of Novartis Vaccines by the GSK group of companies in March, 2015

References

1 Rappuoli R, Pizza M, Del Giudice G, De Gregorio E. Vaccines, new opportunities for a new society. *Proc Natl Acad Sci U S A* 2014; **111**: 12288–12293.

2 Plotkin SA. Vaccines: past, present and future. *Nat Med* 2005; **11**: S5-.

3 Plotkin SA, Plotkin SA. Correlates of Vaccine-Induced Immunity. *Clin Infect Dis* 2008; **47**: 401–409.

4 Plotkin SA. Correlates of Protection Induced by Vaccination. *Clin Vaccine Immunol* 2010; **17**: 1055–1065.

5 Pantaleo G, Koup RA. Correlates of immune protection in HIV-1 infection: what we know, what we don't know, what we should know. *Nat Med* 2004; **10**: 806.

6 Valmaseda A, Macete E, Nhabomba A, Guinovart C, Aide P, Bardají A *et al.* Identifying Immune Correlates of Protection Against Plasmodium falciparum Through a Novel Approach to Account for Heterogeneity in Malaria Exposure. *Clin Infect Dis* 2018; **66**: 586–593.

7 Six A, Bellier B, Thomas-Vaslin V, Klatzmann D. Systems biology in vaccine design. *Microb Biotechnol* 2012. doi:10.1111/j.1751-7915.2011.00321.x.

8 Nakaya HI, Wrammert J, Lee EK, Racioppi L, Marie-Kunze S, Haining WN *et al.* Systems biology of vaccination for seasonal influenza in humans. *Nat Immunol* 2011. doi:10.1038/ni.2067.

9 Pulendran B. Systems vaccinology: Probing humanity's diverse immune systems with vaccines. *Proc Natl Acad Sci* 2014. doi:10.1073/pnas.1400476111.

10 Furman D, Davis MM. New approaches to understanding the immune response to vaccination and infection. *Vaccine* 2015. doi:10.1016/j.vaccine.2015.06.117.

11 Gomez Lorenzo MM, Fenton MJ. Immunobiology of influenza vaccines. *Chest* 2013. doi:10.1378/chest.12-1711.

12 Buonaguro L, Pulendran B. Immunogenomics and systems biology of vaccines. *Immunol Rev* 2011; **239**: 197–208.

13 Kondo M. Lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors. *Immunol Rev* 2010. doi:10.1111/j.1600-065X.2010.00963.x.

14 Sridhar S, Begom S, Bermingham A, Hoschler K, Adamson W, Carman W *et al.* Cellular immune correlates of protection against symptomatic pandemic influenza. *Nat Med* 2013. doi:10.1038/nm.3350.

15 Kakaradov B, Arsenio J, Widjaja CE, He Z, Aigner S, Metz PJ *et al.* Early transcriptional and epigenetic regulation of CD8+ T cell differentiation revealed by single-cell RNA sequencing. *Nat Immunol* 2017. doi:10.1038/ni.3688.

16 Kaech SM, Hemby S, Kersh E, Ahmed R. Molecular and Functional Profiling of Memory CD8 T Cell Differentiation. *Cell* 2002; **111**: 837–851.

17 Appay V, Douek DC, Price DA. CD8+ T cell efficacy in vaccination and disease.

Nat Med 2008. doi:10.1038/nm.f.1774.

18 Lugli E. T-Cell Dii erentiation Methods and Protocols Methods in Molecular Biology 1514. .

19 Rutishauser RL, Kaech SM. Generating diversity: Transcriptional regulation of effector and memory CD8+ T-cell differentiation. *Immunol Rev* 2010; **235**: 219–233.

20 Kaech SM, Cui W. Transcriptional control of effector and memory CD8+ T cell differentiation. *Nat Rev Immunol* 2012; **12**: 749–761.

21 Rutishauser RL, Martins GA, Kalachikov S, Chandele A, Parish IA, Meffre E *et al.* Transcriptional Repressor Blimp-1 Promotes CD8+ T Cell Terminal Differentiation and Represses the Acquisition of Central Memory T Cell Properties. *Immunity* 2009. doi:10.1016/j.immuni.2009.05.014.

22 Palendira U, Chinn R, Raza W, Piper K, Pratt G, Machado L *et al.* Selective accumulation of virus-specific CD8+ T cells with unique homing phenotype within the human bone marrow. *Blood* 2008. doi:10.1182/blood-2008-02-138040.

23 Samji T, Khanna KM. Understanding memory CD8 + T cells. *Immunol Lett* 2017; **185**: 32–39.

24 Sallusto F, Lenig D, Forster R, Lipp M LA. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* 1999; **401**: 708–712.

25 Brazzoli M, Magini D, Bonci A, Buccato S, Giovani C, Kratzer R *et al.* Induction of Broad-Based Immunity and Protective Efficacy by Self-amplifying mRNA Vaccines Encoding Influenza Virus Hemagglutinin. *J Virol* 2016. doi:10.1128/JVI.01786-15.

26 Magini D, Giovani C, Mangiavacchi S, Maccari S, Cecchi R, Ulmer JB *et al.* Self-Amplifying mRNA Vaccines Expressing Multiple Conserved Influenza Antigens Confer Protection against Homologous and Heterosubtypic Viral Challenge. 2016. doi:10.1371/.

27 Grant EJ, Quiñones-Parra SM, Clemens EB, Kedzierska K. Human influenza viruses and CD8+ T cell responses. *Curr Opin Virol* 2016. doi:10.1016/j.coviro.2016.01.016.

28 O'Gorman WE, Huang H, Wei Y-L, Davis KL, Leipold MD, Bendall SC *et al.* The Split Virus Influenza Vaccine rapidly activates immune cells through Fcγ receptors. *Vaccine* 2014; **32**: 5989–5997.

29 Wong S-S, Webby RJ. Traditional and New Influenza Vaccines. doi:10.1128/CMR.00097-12.

30 Van Buynder PG, Konrad S, Van Buynder JL, Brodkin E, Krajden M, Ramler G *et al.* The comparative effectiveness of adjuvanted and unadjuvanted trivalent inactivated influenza vaccine (TIV) in the elderly. *Vaccine* 2013. doi:10.1016/j.vaccine.2013.07.059.

31 O' Hagan DT, Ott GS, Nest GV, Rappuoli R DGG. The history of MF59® adjuvant: a phoenix that arose from the ashes. *Expert Rev Vaccines* 2013; **12**: 13–30.

32 Vesikari T, Pellegrini M, Karvonen A, Groth N, Borkowski A, O 'hagan DT *et al.* Enhanced Immunogenicity of Seasonal Influenza Vaccines in Young Children Using MF59 Adjuvant. *Pediatr Infect Dis J* 2009; **28**: 563–571.

33 O'Hagan DT, Ott GS, De Gregorio E, Seubert A. The mechanism of action of MF59 - An innately attractive adjuvant formulation. Vaccine. 2012. doi:10.1016/j.vaccine.2011.09.061.
34 Nicole L. La Gruta and Stephen J. Turner. T cell mediated immunity to influenza: mechanisms of viral control. *Trends Immunol* 2014; **35**: 396–402.
35 Seder RA, Darrah PA, Roederer M. T-cell quality in memory and protection: implications for vaccine design. *Nat Rev Immunol* 2008. doi:10.1038/nri2274.
36 Kaech SM, Wherry EJ AR. Effector and memory T-cell differentiation: implications for vaccine development. *Nat Rev Immunol* 2002; **2**: 251–262.
37 Della Cioppa G, Vesikari T, Sokal E, Lindert K, Nicolay U. Trivalent and quadrivalent MF59 ®-adjuvanted influenza vaccine in young children: A dose- and schedule-finding study. *Vaccine* 2011. doi:10.1016/j.vaccine.2011.08.111.
38 Altenburg AF, Rimmelzwaan GF, de Vries RD. Virus-specific T cells as correlate of (cross-)protective immunity against influenza. Vaccine. 2015. doi:10.1016/j.vaccine.2014.11.054.
39 Iavarone C, O' Hagan DT, Yu D, Delahaye NF UJ. Mechanism of action of mRNA-based vaccines. *Expert Rev Vaccines* 2017.
40 Hekele A, Bertholet S, Archer J, Gibson DG, Palladino G, Brito LA *et al.* Rapidly produced SAM® vaccine against H7N9 influenza is immunogenic in mice. *Emerg Microbes Infect* 2013. doi:10.1038/emi.2013.54.
41 Geall AJ, Verma A, Otten GR, Shaw CA, Hekele A, Banerjee K *et al.* Nonviral delivery of self-amplifying RNA vaccines. doi:10.1073/pnas.1209367109.
42 Trapnell C. Defining cell types and states with single-cell genomics. Genome Res. 2015; **25**: 1491–1498.
43 Bendall SC, Davis KL, Amir EAD, Tadmor MD, Simonds EF, Chen TJ *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human b cell development. *Cell* 2014; **157**: 714–725.
44 Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014. doi:10.1038/nbt.2859.
45 Mahata B, Zhang X, Kolodziejczyk AA, Proserpio V, Haim-Vilmovsky L, Taylor AE *et al.* Single-cell RNA sequencing reveals T helper cells synthesizing steroids De Novo to contribute to immune homeostasis. *Cell Rep* 2014; **7**: 1130–1142.
46 Shalek A, Satija R,Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, Schwartz S, Yosef N, Malboeuf C, Lu D, Trombetta JT, Gennert D, Gnirke A, Goren A, Hacohen N, Levin JZ, Park H RA. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 2013; **498**: 236–240.
47 Newell EW, Sigal N, Bendall SC, Nolan GP, Davis MM. Cytometry by Time-of-Flight Shows Combinatorial Cytokine Expression and Virus-Specific Cell Niches within a Continuum of CD8 + T Cell Phenotypes. *Immunity* 2012. doi:10.1016/j.immuni.2012.01.002.
48 Bengtsson M, Ståhlberg A, Rorsman P, Kubista M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal

distribution of mRNA levels. 2005; : 1388–1392.
49 Chubb JR, Liverpool TB. Bursts and pulses: Insights from single cell studies into transcriptional mechanisms. Curr. Opin. Genet. Dev. 2010. doi:10.1016/j.gde.2010.06.009.
50 Kouno T, de Hoon M, Mar JC, Tomaru Y, Kawano M, Carninci P *et al.* Temporal dynamics and transcriptional control using single-cell gene expression analysis. *Genome Biol* 2013; **14**: R118.
51 Li G-W, Sunney Xie X. Central dogma at the single-molecule level in living cells. doi:10.1038/nature10315.
52 Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. *Proc Natl Acad Sci* 2008. doi:10.1073/pnas.0803850105.
53 Dobrzy M, Bruggeman FJ. Elongation dynamics shape bursty transcription and translation. .
54 Kumar N, Singh A, Kulkarni R V. Transcriptional Bursting in Gene Expression: Analytical Results for General Stochastic Models. *PLoS Comput Biol* 2015. doi:10.1371/journal.pcbi.1004292.
55 Kharchenko P V, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. *Nat Methods* 2014. doi:10.1038/nmeth.2967.
56 McDavid A, Dennis L, Danaher P, Finak G, Krouse M, Wang A *et al.* Modeling Bi-modality Improves Characterization of Cell Cycle on Gene Expression in Single Cells. *PLoS Comput Biol* 2014; **10**. doi:10.1371/journal.pcbi.1003696.
57 Proserpio V. Single-cell technologies to study the immune system. 2015; : 133–140.
58 Chattopadhyay PK, Gierahn TM, Roederer M, Love JC. Single-cell technologies for monitoring immune systems. doi:10.1038/ni.2796.
59 Ning L, Liu G, Li G, Hou Y, Tong Y, He J. Current Challenges in the Bioinformatics of Single Cell Genomics. *Front Oncol* 2014. doi:10.3389/fonc.2014.00007.
60 Liu S, Trapnell C. Single-cell transcriptome sequencing: recent advances and remaining challenges [version 1; referees: 2 approved]. *F1000 Fac Rev* 2016; **72231**: 182182–25.
61 Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* 2011. doi:10.1038/nature09807.
62 Li Y, Xu X, Song L, Hou Y, Li Z, Tsang S *et al.* Single-cell sequencing analysis characterizes common and cell-lineage-specific mutations in a muscle-invasive bladder cancer. *Gigascience* 2012. doi:10.1186/2047-217X-1-12.
63 Xu X, Hou Y, Yin X, Bao L, Tang A, Song L *et al.* Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. *Cell* 2012. doi:10.1016/j.cell.2012.02.025.
64 Hou Y, Song L, Zhu P, Zhang B, Tao Y, Xu X *et al.* Single-cell exome sequencing and monoclonal evolution of a JAK2-negative myeloproliferative neoplasm. *Cell* 2012. doi:10.1016/j.cell.2012.02.028.
65 Voet T, Kumar P, Van Loo P, Cooke SL, Marshall J, Lin ML *et al.* Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res* 2013. doi:10.1093/nar/gkt345.
66 Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K *et al.*

Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 2012. doi:10.1016/j.cell.2012.08.023.

67 Ye F, Huang W, Guo G. Studying hematopoiesis using single-cell technologies. *J Hematol Oncol* 2017; **10**: 27.

68 Proserpio V, Piccolo A, Haim-Vilmovsky L, Kar G, Lönnberg T, Svensson V *et al.* Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. *Genome Biol* 2016; **17**: 103.

69 Brennecke P, Reyes A, Pinto S, Rattay K, Nguyen M, Küchler R *et al.* Single-cell transcriptome analysis reveals coordinated ectopic gene-expression patterns in medullary thymic epithelial cells. *Nat Immunol* 2015; : 1–67.

70 Flatz L, Roychoudhuri R, Honda M, Filali-Mouhim A, Goulet J-P, Kettaf N *et al.* Single-cell gene-expression profiling reveals qualitatively distinct CD8 T cells elicited by different gene-based vaccines. *Proc Natl Acad Sci U S A* 2011; **108**: 5724–5729.

71 Janilyn Arsenio, Boyko Kakaradov, Patrick J Metz, Stephanie H Kim GWY& JTC. Early specification of CD8+ T lymphocyte fates during adaptive immunity revealed by single-cell gene-expression analyses. *Nat Immunol* 2014; **15**: 365–372.

72 Mcheyzer-williams LJ, Milpied PJ, Okitsu SL, Mcheyzer-williams MG. Class-switched memory B cells remodel BCRs within secondary germinal centers. *Nat Immunol* 2015; **16**: 1–12.

73 Neu KE, Tang Q, Wilson PC, Khan AA. Single-Cell Genomics: Approaches and Utility in Immunology. Trends Immunol. 2017. doi:10.1016/j.it.2016.12.001.

74 Buettner F, Natarajan KN, Casale FP, Proserpio V, Scialdone A, Theis FJ *et al.* Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat Biotechnol* 2015; **33**. doi:10.1038/nbt.3102.

75 Villani A-C, Shekhar K. Single-Cell RNA Sequencing of Human T Cells. 2017 doi:10.1007/978-1-4939-6548-9_16.

76 Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat Rev Immunol* 2017. doi:10.1038/nri.2017.76.

77 Kippner LE, Kim J, Gibson G, Kemp ML. Single cell transcriptional analysis reveals novel innate immune cell types. *PeerJ* 2014; **2**: e452.

78 Kim JK, Marioni JC. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. .

79 Min JW, Kim WJ, Han JA, Jung YJ, Kim KT, Park WY *et al.* Identification of distinct tumor subpopulations in lung adenocarcinoma via single-cell RNA-seq. *PLoS One* 2015. doi:10.1371/journal.pone.0135817.

80 Moignard V, Macaulay IC, Swiers G, Buettner F, Schütte J, Calero-Nieto FJ *et al.* Characterization of transcriptional networks in blood stem and progenitor cells using high-throughput single-cell gene expression analysis. *Nat Cell Biol* 2013; **15**: 363–372.

81 Corporation F. Real-Time PCR Analysis. .

82 Guo G, Huss M, Tong GQ, Wang C, Li Sun L, Clarke ND *et al.* Resolution of Cell Fate Decisions Revealed by Single-Cell Gene Expression Analysis from Zygote

to Blastocyst. *Dev Cell* 2010. doi:10.1016/j.devcel.2010.02.012.

83 Buganim Y, Faddah D a., Cheng AW, Itskovich E, Markoulaki S, Ganz K *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* 2012; **150**: 1209–1222.

84 CyTOF ® 2 Mass Cytometer User Manual. .

85 Su Y, Shi Q, Wei W. Single cell proteomics in biomedicine: High-dimensional data acquisition, visualization, and analysis. Proteomics. 2017. doi:10.1002/pmic.201600267.

86 Qiu P, Simonds EF, Bendall SC., Gibbs KD Jr., Bruggner RV, Linderman MD, Sachs K, Nolan GP PS. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nat Biotechnol* 2011; : 886–891.

87 Bjornson ZB, Nolan GP, Fantl WJ. Single-cell mass cytometry for analysis of immune system functional states. Curr. Opin. Immunol. 2013. doi:10.1016/j.coi.2013.07.004.

88 Amir ED, Davis KL, Tadmor MD, Simonds EF, Levine JH, Bendall SC *et al.* viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol* 2013; **31**: 545–52.

89 Diggins KE, Ferrell PB, Irish JM. Methods for discovery and characterization of cell subsets in high dimensional mass cytometry data. *Methods* 2015; **82**: 55–63.

90 Ye F, Huang W, Guo G. Studying hematopoiesis using single-cell technologies. *J Hematol Oncol* 2017; **10**: 27.

91 Cheow LF, Courtois ET, Tan Y, Viswanathan R, Xing Q, Tan RZ *et al.* Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat Methods* 2016; **13**: 833.

92 Schwartzman O, Tanay A. Epigenetics can be defined as the study of the mecha- nisms that allow cells to translate the nearly constant genome content of a multicellular organism into multi- ple functional and stable cellular conditions. *Nat Publ Gr* 2015; **16**. doi:10.1038/nrg3980.

93 Dulken BW, Leeman DS, Boutet SC, Hebestreit K, Brunet A. Single-Cell Transcriptomic Analysis Defines Heterogeneity and Transcriptional Dynamics in the Adult Neural Stem Cell Lineage. *Cell Rep* 2017. doi:10.1016/j.celrep.2016.12.060.

94 Dalerba P, Kalisky T, Sahoo D, Rajendran PS, Rothenberg ME, Leyrat A a *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nat Biotechnol* 2011; **29**: 1120–1127.

95 Finak G, Mcdavid A, Chattopadhyay P, Dominguez M, De Rosa S, Roederer M *et al.* Mixture models for single-cell assays with applications to vaccine studies. *Biostatistics* 2014; **15**: 87–101.

96 Lin L, Finak G, Ushey K, Seshadri C, Hawn T, Frahm N *et al.* Combinatorial polyfunctionality analysis of antigen-specific T-cell subsets identifies novel cellular subsets correlated with clinical outcomes. *Nat Biotechnol* 2015; : Under review.

97 Mcdavid A, Gottardo R, Simon N, Drton M. GRAPHICAL MODELS FOR ZERO-INFLATED SINGLE CELL GENE EXPRESSION. .

98 McDavid A, Finak G, Chattopadyay PK, Dominguez M, Lamoreaux L, Ma SS *et al.* Data exploration, quality control and testing in single-cell qPCR-based

gene expression experiments. *Bioinformatics* 2013; **29**: 461–467.

99 Mahata B, Zhang X, Kolodziejczyk AA, Proserpio V, Haim-Vilmovsky L, Taylor AE *et al.* Single-cell RNA sequencing reveals T helper cells synthesizing steroids De Novo to contribute to immune homeostasis. *Cell Rep* 2014. doi:10.1016/j.celrep.2014.04.011.

100 Gaublomme JT, Yosef N, Lee Y, Gertner RS, Yang LV, Wu C, Pandolfi PP, Mak T, Satija R, Shalek AK, Kuchroo VK, Park H RA. Single-cell Genomics Unveils Critical Regulators of Th17 cell Pathogenicity. *Cell* 2015; **163**: 1400–1412.

101 Proserpio V, Piccolo A, Haim-Vilmovsky L, Kar G, Lönnberg T, Svensson V *et al.* Single-cell analysis of CD4+ T-cell differentiation reveals three major cell states and progressive acceleration of proliferation. doi:10.1186/s13059-016-0957-5.

102 Bargaje R, Trachana K, Shelton MN, Mcginnis CS, Zhou JX, Chadick C *et al.* Cell population structure prior to bifurcation predicts efficiency of directed differentiation in human induced pluripotent cells. doi:10.1073/pnas.1621412114.

103 Daigle BJ, Soltani M, Petzold LR, Singh A. Inferring single-cell gene expression mechanisms using stochastic simulation. *Bioinformatics* 2015. doi:10.1093/bioinformatics/btv007.

104 Padovan-Merhar O, Raj A. Using variability in gene expression as a tool for studying gene regulation. *Wiley Interdiscip Rev Syst Biol Med* 2013. doi:10.1002/wsbm.1243.

105 Wang J, Xia S, Arand B, Zhu H, Machiraju R, Huang K *et al.* Single-Cell Co-expression Analysis Reveals Distinct Functional Modules, Co-regulation Mechanisms and Clinical Outcomes. *PLOS Comput Biol PLoS Comput Biol* 2016; **21**.

106 Satija R, Shalek AK. Heterogeneity in immune responses: from populations to single cells. *Trends Immunol* 2014; **35**: 219–29.

107 Gabrielle T. Belz, Dominik Wodarz, Gabriela Diaz,Martin A. Nowak and PCD. Compromised Influenza Virus-Specific CD8 -T-Cell Memory in CD4 -T-Cell-Deficient Mice. *J -virology* 2002; **Vol.76**, **No. 23**: 12388–12393.

108 Saadatpour A, Lai S, Guo G, Yuan G-C. Single-cell analysis in cancer genomics. doi:10.1016/j.tig.2015.07.003.

109 Saadatpour A, Guo G, Orkin SH, Yuan G. Characterizing heterogeneity in leukemic cells using single-cell gene expression analysis. 2014; : 1–13.

110 Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. doi:10.1126/science.1254257.

111 Natarajan KN, Teichmann SA, Kolodziejczyk AA. Single cell transcriptomics of pluripotent stem cells: reprogramming and differentiation. *Curr Opin Genet Dev* 2017. doi:10.1016/j.gde.2017.06.003.

112 Gawad C., Koh W. QS. Single-cell genome sequencing: current state of the science. *Nat Rev Genet* 2016; **17**: 175–188.

113 Roychoudhuri R, Lefebvre F, Honda M, Pan L, Ji Y, Klebanoff C a *et al.* Transcriptional profiles reveal a stepwise developmental program of memory CD8 + T cell differentiation. *Vaccine* 2015; **33**: 914–923.

114 Fluidigm. Real-Time PCR Analysis. 2017.

115 Ståhlberg A, Bengtsson M. Single-cell gene expression profiling using reverse transcription quantitative real-time PCR. *Methods* 2010; **50**: 282–288.
116 Ståhlberg A, Rusnakova V, Forootan A, Anderova M, Kubista M. RT-qPCR work-flow for single-cell data analysis. *Methods* 2013; **59**: 80–88.
117 Livak KJ, Wills QF, Tipping AJ, Datta K, Mittal R, Goldson AJ *et al.* Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. *Methods* 2013; **59**: 71–79.
118 Ståhlberg A, Rusnakova V, Kubista M. The added value of single-cell gene expression profiling. *Brief Funct Genomics* 2013. doi:10.1093/bfgp/elt001.
119 Karlen Y, McNair A, Perseguers S, Mazza C, Mermod N. Statistical significance of quantitative PCR. *BMC Bioinformatics* 2007; **8**: 131.
120 Bergkvist A, Rusnakova V, Sindelka R, Garda JMA, Sjögreen B, Lindh D *et al.* Gene expression profiling – Clusters of possibilities. *Methods* 2010; **50**: 323–335.
121 Abdi H, Williams LJ. Principal component analysis. *Wiley Interdiscip Rev Comput Stat* 2010; **2**: 433–459.
122 Todorov H, Fournier D, Gerber S. GENOMICS AND COMPUTATIONAL BIOLOGY Principal Components Analysis: Theory and Application to Gene Expression Data Analysis. 2018; **4**.
123 Ringnér M. What is principal component analysis? *Nat Biotechnol* 2008; **26**.
124 Springer ITJ. Principal Component Analysis, Second Edition. .
125 Lever J, Krzywinski M, Altman N. Points of Significance: Principal component analysis. Nat. Methods. 2017. doi:10.1038/nmeth.4346.
126 Scholz M, Selbig J. Visualization and Analysis of Molecular Data. In: Weckwerth W (ed). *Metabolomics: Methods and Protocols*. Humana Press: Totowa, NJ, 2007, pp 87–104.
127 Vincent Barra. Analysis of gene expression data using functional principal components. *Comput Methods Programs Biomed* 2004; **75**: 1–9.
128 Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. .
129 Fisher RA. On the Interpretation of X2 from contingency tables, and the calculation of p. *J R Stat Soc* 1922; **85**: 87–94.
130 Barnard GA. Significance Tests for 2 x 2 Tables. *Biometrika* 1947; **34**: 123–138.
131 Sarkar S, Kalia V, Haining WN, Konieczny BT, Subramaniam S, Ahmed R. Functional and genomic profiling of effector CD8 T cell subsets with distinct memory fates. *J Exp Med* 2008. doi:10.1084/jem.20071641.
132 Ichii H, Sakamoto A, Hatano M, Okada S, Toyama H, Taki S *et al.* Role for Bcl-6 in the generation and maintenance of memory CD8. 2002. doi:10.1038/ni802.
133 Olson J, McDonald-Hyman C, Jameson S, Hamilton S. Effector-like CD8+ T Cells in the Memory Population Mediate Potent Protective Immunity. *Immunity* 2013; **38**: 1250–1260.
134 Joshi NS, Kaech SM. Effector CD8 T cell development: a balancing act between memory cell potential and terminal differentiation. *J Immunol* 2008; **180**: 1309–1315.
135 Preston GC, Feijoo-Carnero C, Schurch N, Cowling VH, Cantrell DA. The

Impact of KLF2 Modulation on the Transcriptional Program and Function of CD8 T Cells. *PLoS One* 2013; **8**: e77537.

136 Takada K, Wang X, Hart G, Odumade OA, Weinreich MA, Hogquist KA *et al.* KLF2 is required for trafficking but not quiescence in post-activated T cells. *J Immunol* 2011; **186**: 775–783.

137 Pabbisetty SK, Rabacal W, Maseda D, Cendron D, Collins PL, Hoek KL *et al.* KLF2 is a rate-limiting transcription factor that can be targeted to enhance regulatory T-cell production. *Proc Natl Acad Sci U S A* 2014; **111**: 9579–9584.

138 Duttagupta PA, Boesteanu AC, Katsikis PD. COSTIMULATION SIGNALS FOR MEMORY CD8 + T CELLS DURING VIRAL INFECTIONS. .

139 Bennett F, Luxenberg D, Ling V, Wang I-M, Marquette K, Lowe D *et al.* Program Death-1 Engagement Upon TCR Activation Has Distinct Effects on Costimulation and Cytokine-Driven Proliferation: Attenuation of ICOS, IL-4, and IL-21, But Not CD28, IL-7, and IL-15 Responses. *J Immunol* 2003. doi:10.4049/jimmunol.170.2.711.

140 Kennedy RB, Ovsyannikova IG, Haralambieva IH, Oberg AL, Zimmermann MT, Grill DE *et al.* Immunosenescence-related transcriptomic and immunologic changes in older individuals following influenza vaccination. *Front Immunol* 2016. doi:10.3389/fimmu.2016.00450.

141 Frentsch M, Stark R, Matzmohr N, Meier S, Durlanik S, Schulz AR *et al.* CD40L expression permits CD8+ T cells to execute immunologic helper functions. *Blood* 2013. doi:10.1182/blood-2013-02-483586.

142 Burocchi A, Pittoni P, Gorzanelli A, Colombo MP, Piconese S. Intratumor OX40 stimulation inhibits IRF1 expression and IL-10 production by Treg cells while enhancing CD40L expression by effector memory T cells. *Eur J Immunol* 2011; **41**: 3615–3626.

143 Vemula S V, Pandey A, Singh N, Katz JM, Donis R, Sambhara S *et al.* Adenoviral Vector Expressing Murine β-Defensin 2 Enhances Immunogenicity of an Adenoviral Vector based H5N1 Influenza Vaccine in Aged Mice. *Virus Res* 2013; **177**: 10.1016/j.virusres.2013.07.008.

144 Toapanta FR, Ross TM. Impaired immune responses in the lungs of aged mice following influenza infection. *Respir Res* 2009; **10**: 112.

145 Ovsyannikova IG, Oberg AL, Kennedy RB, Zimmermann MT, Haralambieva IH, Goergen KM *et al.* Gene signatures related to HAI response following influenza A/H1N1 vaccine in older individuals. *Heliyon* 2016. doi:10.1016/j.heliyon.2016.e00098.

146 Brinza L, Djebali S, Tomkowiak M, Mafille J, Loiseau C, Jouve PE *et al.* Immune signatures of protective spleen memory CD8 T cells. *Sci Rep* 2016. doi:10.1038/srep37651.

147 Weng N, Araki Y, Subedi K. The molecular basis of the memory T cell response: differential gene expression and its epigenetic regulation. *Nat Rev Immunol* 2012. doi:10.1038/nri3173.

148 Xin A, Masson F, Liao Y, Preston S, Guan T, Gloury R *et al.* A molecular threshold for effector CD8+ T cell differentiation controlled by transcription factors Blimp-1 and T-bet. *Nat Immunol* 2016. doi:10.1038/ni.3410.

149 Jung YW, Rutishauser RL, Joshi NS, Haberman AM, Kaech SM. Differential Localization of Effector and Memory CD8 T Cell Subsets in Lymphoid Organs

during Acute Viral Infection. doi:10.4049/jimmunol.1001948.

150 Peixoto A, Evaristo C, Munitic I, Monteiro M, Charbit A, Rocha B *et al.* CD8 single-cell gene coexpression reveals three different effector types present at distinct phases of the immune response. *J Exp Med* 2007; **204**: 1193–205.

151 Bjorkdahl O, Barber KA, Brett SJ, Daly MG, Plumpton C, Elshourbagy NA *et al.* Characterization of CC-chemokine receptor 7 expression on murine T cells in lymphoid tissues. .

152 Wei F, Zhong S, Ma Z, Kong H, Medvec A, Ahmed R *et al.* Strength of PD-1 signaling differentially affects T-cell effector functions. doi:10.1073/pnas.1305394110.

153 Wirth TC, Xue HH, Rai D, Sabel JT, Bair T, Harty JT *et al.* Repetitive antigen stimulation induces stepwise transcriptome diversification but preserves a core signature of memory CD8+ T cell differentiation. *Immunity* 2010. doi:10.1016/j.immuni.2010.06.014.