UNIVERSITÀ
DI SIENA
1240

Dipartimento di Biotecnologie Mediche

Dottorato in Medical Biotechnologies

38° Ciclo

Coordinatore/Coordinatrice: Prof. Francesco Santoro

Gene expression analysis for the characterization of immune response to vaccines against neglected diseases

Settore scientifico disciplinare: MEDS-03/A

Candidato/a
Chiara Sonnati
Università di Siena

Supervisore
Francesco Santoro

2024/2025

Università degli Studi di Siena
Dottorato in Biotecnologie Mediche
38° Ciclo

Data dell'esame finale
23/02/2026

Commissione giudicatrice
Reiner Siebert
Claudia Sala
Lorenzo Franchi
Flavia Artese
Francesco Santoro
Francesco Iannelli

Supplente
German Ott
Arianna Di Napoli
Teresa Marafioti
Stefano Lazzi

Abstract

Emerging infectious diseases and neglected tropical diseases represent a wide range of diseases caused by a great variety of pathogens, which affect billions of people worldwide. Although pharmaceutical industries have focused for years on eradicating emerging diseases, recently their efforts moved to research and development of drugs to control neglected diseases. In this thesis, we employ transcriptomic analyses to evaluate the efficacy and safety of three vaccine platforms directed against diseases from both categories: the rVSV-ZEBOV-GP vaccine against Ebola virus infection, the ChAd63-KH vaccine for post-kala-azar dermal leishmaniasis, and the novel iNTS-GMMA vaccine targeting invasive non-typhoidal Salmonella infections. The rVSV-ZEBOV- GP vaccine is a live, replication competent VSV vector that expresses the glycoprotein of Zaire ebolavirus, which received a full marketing authorization by EMA in 2021, for adult administration. ChAd63-KH is a replication defective simian adenovirus expressing a novel synthetic gene (KH) encoding two Leishmania proteins KMP-11 and HASPB,whose safety and immunogenicity have been already assessed in a phase 1 study. iNTS-GMMA is a novel platform containing modified S.Typhimurium and S.Enteriditis outern membrane vesicles, cyrrently under evaluation in human studies.

For each vaccine, whole-blood samples were collected at different timepoints and stored at -80°C in PAXGene tubes. RNA was extracted with PAXGene blood RNA extraction QIAcube Connect system and quantified with RNA High Sensitivity Assay kit for Qubit™ 4 Fluorometer. Sequencing libraries were prepared using the Illumina® Ribo-Zero Plus Kit starting from 40 ng of total RNA.

The bioinformatic analysis was conducted with the software R. For each vaccine we investigated gene expression across timepoints, performing differential gene expression (DGE) analysis and enrichment analysis, with DEseq2 and tmod packages respectively. Generally, all vaccine platforms elicited innate immune response soon after the administration with upregulated blood transcriptional modules related interferon responsiveness, dendritic cell and monocyte activation, and antigen presentation, while antibody production was detectable 7 days after injection. The transcriptional data supported the known immunogenicity of the ChAd63-KH vaccine.For iNTS-GMMA, we further explored B-cell

receptor (BCR) repertoires using the MiXCR pipeline, identifying approximately 4,000 distinct clonotypes. In the rVSV-ZEBOV-GP pediatric cohort, correlations between gene expression and antibody titers were investigated to identify early transcriptional determinants of humoral immunity.

Contents

Chapter 1

Introduction

1.1 Emerging and neglected infectious diseases

Emerging infection diseases (EIDs) and neglected tropical diseases (NTDs) represent a group of diseases caused by a variety of pathogens including virus, bacteria, fungi and parasites that affect public health. These diseases affect more than 2 billion people, with a collective daily burden that is equivalent to HIV, tuberculosis and malaria causing devastating health, social and economic consequences [1]. Although they are both infectious diseases, these two categories are essentially socio-political constructs for discriminating those diseases that have been the priority for public surveillance (EIDs), from those for which global public health attention has grown only in the last years (NTDs). It is a matter of fact that global health surveillance and the global pharmaceuticals market have focused their attention and effort in research and development and drug discovery to control and eradicate EIDs (like Ebola virus disease, pandemic influenza, malaria etc.), which today still represent a huge part of pharmaceutical market in the richest countries. On the other hand, NTDs are strongly linked to poverty, they affect vulnerable population in tropical regions, and they are not transmitted rapidly and widely and, as a consequence, they don't represent a big source of profits for biotechnology and pharmaceutical industries as EIDs do [2].

The interest for NTDs materialized only in 2012 in The London Declaration on neglected tropical diseases, which describes the global initiative to eradicate these diseases. The coalition involved in this campaign comprises the WHO, the Bill & Melinda Gates Foundation and some leading pharmaceutical companies. The emergence of new philanthropic donors and the presence of non-governmental organizations (like Médecins sans Frontières) contributed in orienting the attention on poor communities, without threatening to disrupt the world economy.The key role in the battle against NTDs is played

by pharmaceutical industries that recently started taking them in account by engaging public-private cooperation aimed to develop efficient drugs and vaccines for treatment and prevention. Other strategies employed by the WHO for disease management are based on vector control, provision of safe drinking water, basic sanitation and hygiene and veterinary public health. Today WHO recognizes 17 NTDs such as Leishmaniasis, Chagas disease, Dengue, Leprosy and others.

This thesis is focused on the gene expression analysis for the characterization of the immune response elicited by three different vaccines against Ebola virus disease, post-kalazar dermal Leishmaniasis and non-typhoid Salmonella infections respectively. Together these three infectious diseases represent a significant issue for public health worldwide.

1.2 The Ebola Virus

The Ebola virus was discovered in 1976 in the area of Zaire and in Sudan, since this date it has caused multiple outbreaks in African country with high number of deaths. The biggest epidemic occurred between 2013 and 2016 in the central Africa and it resulted in about 11,323 fatalities among the 28 000 cases. The recurrence of outbreaks stimulated the scientific community to develop new experimental drugs and vaccines to treat patients and prevent the infection.

1.2.1 Ebola Virus Biology

The Ebola virus is a member of the *Filoviridae* family, a family of filamentous, enveloped single-stranded, negative-sense RNA viruses, with a genome of approximately 19 kb. In the genus Ebolavirus, five different species are distinguished: *Zaire Ebolavirus (ZEBOV), Sudan Ebolavirus (SUDV), Tai forest Ebolavirus, Bundibugyo Ebolavirus (BDBV) and Reston Ebolavirus.*

The genome encodes for seven distinct genes from which at least nine proteins are expressed: nucleoprotein (NP), polymerase cofactor (VP35), matrix protein (VP40), glycoprotein (GP), soluble GP (sGP), small soluble glycoprotein (ssGP), transcription activator (VP30), minor matrix protein (VP24), and RNA-dependent RNA polymerase (L)[3].

ZEBOV particles have a diameter of 80 nm and their length can vary from 800 nm up to 1400 nm. It binds a huge variety of cellular attachment molecules, including lectins, T-cell immunoglobulin and mucin domain-1 (TIM-1) and Tyrosine kinase receptor Axl. It is internalized mainly by micropinocytosis, or rarely by clathrin-mediated endocytosis;

the viral protein responsible for the attachment and entry is the GP, that also determines the viral tropism. In its native state, GP is a triplet of heterodimers, each of them composed by a receptor binding subunit (GP1) and a fusion subunit (GP2)[4]. After the internalization, the GP protein is degraded by proteolytic cleavage in the endosome, where it binds to the cellular receptor cholesterol transporter Neimann-Pick C1 (NPC1) protein and fuses with the endosomal membrane, causing the release of the nucleocapsid complex in the cytoplasm. NP, VP35 and L proteins are necessary to support genome replication and VP30 is an essential cofactor that stabilizes VP35-L RNA binding. The replication starts with the synthesis of the positive-sense RNA filament, the antigenome, that serves as a template for the generation of new genomes. The 3' region of genome and anti-genome contains the replication promoter for plus- and minus-sense RNA synthesis, ZEBOV promoter is a bipartite structure with two elements separated by a spacer region consisting of the transcription start signal of NP and a downstream located region involved in secondary structure formation and transcription initiation [5]. The RNA genome is transcribed into seven monocistronic mRNA fragments, all of them capped and polyadenylated. The polymerase complex accesses the genes through a single polymerase binding site at 3' end and proceed along the RNA by stopping a reinitiating at each gene junction transcribing each gene individually from 3' to 5' end. NP protein guides the encapsidation together with VP35 and VP24, the first forms helical tubes and the other two interact with these NP-derived helices resulting in the formation of nucleocapsid-like structure. Then nucleocapsids are driven to plasma membrane, here they are packaged together with membrane-inserted GP into progeny viral particles; this process is guided by NP and VP40 [6]. EBOV causes severe hemorragic fever in both humans and NHPs, with fatality rate up to 90%.

1.2.2 Pathogenesis

Ebola virus disease (EVD) is a zoonosis, since the infections in human are mainly linked to exposure to animal reservoirs including bats, squirrels, mice and rats. The primary transmission route from the reservoir to an end host remains unknown, while human-to-human transmission occurs via direct contact with infected tissues, bodily fluids, inoculation by injection of virus into bloodstream. Among these risks factors, the most significant is the contact with infected body fluids, indeed ZEBOV has been recovered from breast milk, saliva, urine, semen, cerebrospinal fluid, and aqueous humor, in addition to blood and blood derivatives, and detected in amniotic fluid, tears, skin swabs and stool[7]. EBOV is able to replicate in a broad range of cell types, including immune cells, endothelial cells, fibroblasts, hepatocyte and adrenal cells. It first replicates in dendritic cells and other cells of the monocytes/macrophages lineage, that travel via lymphatic

vessels to lymph nodes and nodal chains, where virus replication and dissemination initiation occur prior to the onset of symptoms.[6] Incubation period goes from 2 to 21 days and the disease can be divided into three phases; it starts with a few days of fever, headache and myalgia, followed by a gastrointestinal phase with diarrhea and vomiting. The final stage is characterized by multi-focal necrosis of hepatocytes and multi-organ failure, convulsion, shock and death, that occurs within about 2 weeks after the first symptoms appear [3]. Following EBOV exposure, several inflammatory mediators are released due to the presence of GP protein that is shed from the infected cells within the first hour of infection. The quantity of free GP is directly correlated to the degree of DCs and macrophages activation and it is responsible for the abnormal massive release of pro-and anti-inflammatory cytokines. Indeed, another peculiarity of the EVD is the so-called "cytokine storm" originated by an overproduction of IL-1, IL-6, IL-8 and other inflammatory mediators. This effect is the leading cause of coagulopathy and increased endothelial permeability. It is also hypothesized that glycosylation of GP may impede the binding of neutralizing antibodies to the virus, contributing to the immune evasion [6]. The infection of immune cells results in impaired innate and adaptive immune responses characterized by the inhibition of type I interferon (IFN) response, deregulation of the cytokine and chemokine network, a deterioration of DCs and NKs and a massive apoptosis of T cells. The principal viral proteins involved in this activity are VP24 and VP35, two components of the nucleocapsid that block host interferons production and signaling, promote aberrant expression of cytokine and affect the maturation of DCs. The disruption of the immune response mediated by macrophages, dendritic cells and antigen presenting cells contributes to immune evasion and it is correlated with a fatal outcome of Ebola hemorrhagic fever. A fatal outcome is also associated with the lack of specific IgG and a low level of IgM.

Among humans it seems that EVD incidence increase linearly with the age to a peak at 35-44 years [8]. Children usually represent a small number of cases of EVD due to intentional exposure to the infected individual, and differences in susceptibility between different age groups. However they present a shorter incubation period which can vary for 6.9 days in children younger than 1 year of age, to 9.8 day in children 10 to 15 years of age. Younger children also show shorter times from symptom onset to hospitalization and a faster progression to death. Furthermore, children exhibit disease symptoms in an age-dependent manner and at different frequencies compared to adults [9]. The fatality rate is also higher in children compared to adult (about 100% in neonats, 70/80% in children younger than 5 years old and about 50% in children from 10 to 15 years old) [10]. All of these elements highlight the importance of including children in clinical trials to evaluate the efficacy of a possible treatment or vaccine.

1.2.3 Vaccination platforms against Ebola

After the 2013-2016 EBOV epidemic, the necessity for a good prevention strategy became evident and this strongly accelerated the classical human clinical trials on a number of vaccines that had already demonstrated their efficacy in NHPs models. However, due to the sporadic outbreaks of EBOV, is not easy to perform a standard phase III clinical trial. To overcome this limit, the FDA implemented the 'animal rule', under which a countermeasure can be licensed with phase I and phase II clinical trials and animal efficacy data [11]. The major candidates include virus-like-particles, replication incompetent adenovirus serotype 4 vectors (Ad5), replication-competent recombinant human parainfluenza virus 3 (rHPIV3) and recombinant vesicular stomatitis virus (rVSV) [12]. The most promising candidates are rVSV-ZEBOV- GP (Ervebo®) and a rAd26-ZEBOV prime with MVA-BN-Filo boost (Zabdeno®), which received by EMA a full authorization and an authorization under "exceptional circumstances" respectively. The Zabdeno vaccine is able to elicit a strong CD4+ and CD8+ T cells response and the production of neutralizing antibodies, its Ad26 vector encodes the GP of EBOV. Eight weeks later the Zabdeno dose, a booster of a Modified Ankara Bavarian Nordic viral vector (called MvaBea) is administered, it encodes for GP belonging of other species of Ebola virus.

1.2.4 The rVSV-ZEBOV-GP vaccine

The rVSV-ZEBOV-GP vaccine is a live, replication competent VSV (VSV-Indiana serotype) vector that expresses the GP of ZEBOV. Multiple copies of GP are expressed on the viral envelop to induce a protective immunity. The gene encoding the VSV glycoprotein was completely removed and substituted with the one encoding for ZEBOV GP. A co-transfection of cells with a plasmid containing the entire VSV genome deleted for GP, replaced by ZEBOV GP and a helper plasmid is required to produce the final vector. The transcription of the plasmid is controlled by bacteriophage T7 polymerase promoters. The full-length GP protein is anchored to the viral envelope and no soluble GPs are released, this leads to a more efficient immune response and contrasts the natural immune evasion caused by the soluble form. The vector enters the cell by micropinocytosis in a GP-dependent manner and it is then processed by the endosome as for the ZEBOV itself. The complete removal of VSV GP protein leads to a minimal anti-vector immunity.

In 2015 a phase 1 clinical trial showed that the vaccine was immunogenic after a single dose and no serious adverse events were reported. The study involved 158 healthy adults in Europe and Africa. All participants were injected with doses of vaccine ranging from 300 000 to 50 million PFU or placebo. The reactogenicity was evaluated for 7 days after

the injection and at days 14 and 28, the majority of side effects were mild or moderate. Serum antibodies induced by the vaccine were detected and neutralization assays showed a significant increase in neutralizing antibodies in all vaccinees who received similar titers [13]. The most alarming adverse effect was observed in the Swiss cohort, where 11 of 51 (22%) vaccinees developed arthritis. This disease was accompanied by maculopapular rash and vesicular dermatitis in three subjects. Two people presented purpura of the lower legs. The incidence of arthritis was not associated with the dose of the vaccine, but was mainly correlated with the age of the subjects; in fact age was significantly associated with the risk of arthritis in low-dose vaccinees, but not in high-dose ones [14]. In the same year, a phase 3 clinical trial was performed in Guinea where some cases of Ebola appeared. People (contacts of the confirmed case and contacts of the contacts) were treated immediately, or after 21 days, with rVSV-ZEBOV in order to assess the efficacy and the effectiveness of the vaccine. In the immediate vaccination group were no cases of EVD, whereas in the delayed group there were 16 (confirmed before the vaccination). No cases were diagnosed 6 days post vaccination in both groups. These results, together with those provided by additional trials managed in the same area, showed that the efficacy of the vaccine was about 100% after a single dose [15]. The safety and immunogenicity of the vaccine were tested also in children (6-12 years) and adolescents (13-17 years) in a phase 1 trial in Lambaréné (Gabon), who were vaccinated with a single intramuscular dose of $2x10^7$ PFU. Both age groups showed headache, fatigue and subjective fever, but no severe adverse effects occurred, and specific antibodies were detected 28 days after vaccination [16].

1.3 Leishmaniasis and post-Kala-azar dermal Leishmaniasis

1.3.1 Leishmania

The intracellular protozoan parasite genus *Leishmania* comprises more than 54 species, among them around 21 are pathogenic and responsible for cutaneous or visceral leishmaniasis. The vector responsible for the transmission is the female sandfly, an haematophagus, 2-3 mm long arthropod. As a parasite,*Leishmania* needs a mammalian host to reproduce, that are humans for *L.donovani* and mammals for *L.infantum*, that togheter with *L.chagasi*, are the leading cause of visceral leishmaniasis [17]. Despite the huge variability among all the species, the life cycle is common and comprises two major morphological forms, called promastigote (within female sandfly) and amastigote (within mammalian phagocytic cells). Inside the sandfly, the promastigote undergoes a meta-

morphosis from procyclic to metacyclic infective form that moves to the gut of the sandfly. The metacyclic form is inoculated in the host and soon engulfed by phagocytic cells. Here, some parasites succeed in escaping from the killing strategy of neutrophils and spread into macrophages. Inside macrophages, the switch from promastigote to amastigote occurs, in this form parasite are able to massively replicate causing clinical manifestation. During this morphological change, the parasite surface modifies its complexity and it starts exposing hydrophilic acylated surface proteins (HASPs), which are characteristic of infective stage. These proteins are immunogenic, one of them is the hydrophilic acylated surface protein B (HASPB), that is released from the metacyclic parasite surface during phagocytosis by macrophages [18] Genetic variability causes the presence of a lot of different virulence factors which contribute to pathogenesis aspects. The main virulence factors are: proteophosphoglycans, Elongation factor-1α, proteases, acid phosphatases, glycoprotein 63, lipophosphoglycan, heat shock protein and Kinetoplastid membrane protein-11 (KMP-11). In particular KMP-11 is crucial in *L.donovani* infection due to its ability to inhibit IFN-γ production and to induce IL-10 release [19].

1.3.2 Visceral Leishmaniasis and Post-kala-azar dermal Leishmaniasis

Visceral leishmaniasis(VL) is a vector-borne disease caused by intracellular protozoa of the genus *Leishmania*, in particular two species are mainly involved in VL, that are *Leishmania donovani* in southeast Asia and Africa ad *Leishmania infantum* in Mediterranean area, in the Middle east and in Brazil.

The protozoa enters the human body through sandfly bite, then it spreads and replicates in reticuloendothelial system. Parasites also infect macrophages, reach the lymph nodes and lymphoid organs where they evade the immune system of the host. Symptom are various, slowly progressive and in many cases VL can also be asymptomatic or it can behave as latent infection, manifesting itself after years. Most common symptoms include fever, splenomegaly, hepatomegaly,pancytopenia and elevated liver enzyme. In some cases, if untreated, VL can be fatal. The term "kala-azar" usually associated with *Leishmania* infection is due to the greyish appearance of the skin of some patience, probably caused by a cytokine-induced increase in the production of adrenocorticotropic hormone [17]. After infection,the immune response to VL commonly involves T cells, in particular Th1 and Th2 including IL-12, INF-γ, IL-21, IL-27, TGF-β and TNF-α resulting in both pro and anti inflammatory response. Both CD4+ and CD8+ T cells play a crucial role, the first ones are related to high parasite load and production of IL-10, whereas the second ones may differentiate in effector cells, killing target cells or

they can acquire a regulatory role, preventing immunopathology. The humoral response is strong and easily detectable with direct agglutination test and ELISA [20].

After treatment with antileishmania drugs (such as stibogluconate,AmBisome and Fungisome), immune response switches to predominantly Th1 response and macrophages action, whit a slowly decrease of humoral immunity, which can persist for several months [20].

Post-kala-azar dermal leishmaniasis (PKDL) is a complication of VL, whose manifestation can occur months or years after the visceral disease. It is mainly caused by *L.donovani* and it is characterized by the presence of macular, maculopapular and nodular rush. The lesions appear around mouth and then spread around the face, however in most severe cases, mucosal lesions involve also lips and palate [21]. Depending on the lesions pattern, three grade of severity are established:

- Grade one: characterized by scattered maculopapular or nodular rush occurring mainly in face
- Grade two: dense maculopapular or nodular rash covering most of the face and extending to the chest, back, upper arms, and legs.
- Grade three: maculopapular or nodular rash covering most parts of the body, including hands and feet, whit ulceration ans crusting in palate and lips.

In some cases, PKDL occurs concomitantly with VL(para-kala-azar dermal leishmaniasis), resulting in patients manifesting also fever, splenomegaly, hepatomegaly or lymphadenopathy.

The mechanism that determine development of PKDL still remain to be elucidated, but it seems to be strictly bound to the immune response to the parasite that can be found in skin lesions.PKDL is characterized by an immune response mediated by both Th1 and Th2 t cells, with the persistence of IL-10 in skin and IFN-γ sistematically. Also Th17 cells play an important role in induction of inflammatory response,by releasing IL-17 and IL-22, which are inflammatory cytokines. Antigen presenting cells, such as macrophages and dendritic cells are responsible for the initiation, development and maintenance of a protective immunity, however the amastigote is able to modulate macrophage action resulting in the switching into M2 type, that is associated with suppression of cell mediate immunity and disease chronicity. Indeed, the strength of the cell mediate immunity response is higher in acute PKDL, on the other hand, it tends to be weaker in chronic disease. Level of antibodies varies depending of the type of PKDL, it is low in macular PKDL(only elevated IgG1), whereas it is high in the papulo-nodular manifestation(elevated IgG1 and IgG3) [22].

1.3.3 Vaccination platform against Leishmaniasis

Although it is endemic in 88 countries, Leishmaniasis is still considered as a disease of developing countries. This fact, coupled with the difficulties, in the affected areas, in accessing health care and drugs,leads to a poor interest in investing in vaccine research. As a result, Leishmaniasis can be included in the neglected disease[17]. Current measures to address this infection comprises advanced surveillance, production of drug for improved treatment ad identification of animal reservoir. The research of a multi-species *Leishmania* vaccine is still ongoing, at the moment there are numerous candidates that can be grouped in three categories:

- Live attenuated
- Killed parasite
- Defined vaccines like recombinant proteins and DNA vaccines

Although the high numbers of possible candidates, at the moment no vaccine is on the market. The possible issues can be both economic and biological. As a neglected disease, Leishmaniasis is unactractive to the industry,that will not gain enough profits. These candidates still fail in conferring the immunity, in addiction a suitable and stable adjuvant has not been found yet, due to the extreme variability of the infections. *Leishmania* species present essential differences in virulence dynamics and the clinical manifestation varies between the visceral and the cutaneous form, as their virulence factors and the elicited immune response[23]. Furthermore, the genetic of the host seems to be another crucial factor for clinical manifestation [20].

1.3.4 The ChAd63-KH vaccine

Among these candidates, ChAd63-KH demonstrated to be a valid prophylactic and therapeutic solution against VL and PKDL. It is a replication defective simian adenovirus expressing a novel synthetic gene (KH) encoding two Leishmania proteins KMP-11 and HASPB,whose safety and immunogecity have been already assessed in a phase 1 study in 20 healty adults volunteers, through immunogenicity assay and whole blood transcriptomic. Subjects were followed up at days 1, 14, 28, 56 and 90 post-vaccination, none reported severe adverse effects, indicating that the vaccine is safe. Immunogenicity assays highlighted massive activation of CD8+ T cells and the production of relative cytokines, whose activity was validated also through transcriptomic analysis [24]. Indeed, the vaccine is projected to induce CD8+ T cell response, that seems to be one of the most significant correlates of protection in experimental prophylactic vaccination.

In addition several studies in humans affected by VL demostrated that CD8+ T cells are an important contributors to anti-leishmania immunity in asymptomatic patients, and when present at an appropriate frequency, antigen specific CD8+ T cells can thus provide therapeutic benefit [25].

The efficacy of a single dose ($7,5x10^7$) of ChAd63-KH for treatment of PKDL was assessed in a randomized double blind phase 2b trial [25]. Participants were monitored for 7 day after vaccination in hospital and the on day 21,42, 90 and 120. Vaccine-induced immune response was evaluated with whole blood transcriptomic analysis and antibody and IFN-γ ELISPOT [26].

1.4 Invasive non-typhoid Salmonella Infection

1.4.1 Non-typhoid Salmonella serovars and pathogenicity

Invasive non-typhoid Salmonella infection (iNTS) are caused by few serovars of *Salmonella enterica*, that are Gram-negative, facultative, anaerobic, intracellular bacteria belonging to the family of *Enterobactericeae*. They enter human body by consumption of contaminated food or water. The ability to survive acid pH of the stomach allows *Salmonella* to penetrate the intestinal epithelial barrier by invading enterocytes through the expression of a type 3 secretion system (T3SS) encoded by Salmonella pathogenicity island (SPI) . There are two types of T3SS, one is encoded by SPI-1 and it is specific for epithelial cells invasion; the other one is SPI-2 that is expressed once inside the cell and it is required for bacteria survival. Another important factor is the Salmonella virulence plasmid (SPV), that encodes for genes disrupting the immunity [27]. A few portion of bacteria can also penetrate dendritic cells and intestinal macrophages, resulting in the spread of the pathogen through the bloodstream.

The early immune response against *Salmonella* relies on innate immunity in gut mucosa. Macrophages and neutrophils are the first defense line through their phagocytic ability with their production of antimicrobial peptide and IFN-γ. At a later stage of infection, T cell and antibody response become the effective defense, in particular CD4+ T cells play a pivotal role in stimulating antibodies production and their loss is one of the primary correlated of susceptibility to NTS bacteremia. The presence of antibodies anti O-antigens is critical to prevent bacteremia and it is crucial in people with deficiencies in IL-12 and IL-23 production[28]

iNTS infections are always characterized by bacteremia without an evident focus of infection. NTS bacteremia can lead to endocarditis with myocardial abscess and peri-valvular complication with mortality rate of 50%. The presence of endovascular foci of infection

can also cause mycotic aneurysms (more common than endocarditis), especially in late age people. Visceral abscesses, spleen infection can also be possible. NTS gastroentiritis can also lead to damage of joint and bones due to immune inflammatory response, that may result in osteomyelitis (more common in children than in adults). *Salmonella* is able to cross the blood-brain barrier, causing meningitis[27]. The clinical outcome of the infection depends also on the history of the patient. People who are affected by immunodeficiency, or have Malaria coinfection, are characterized by a compromised immunity with a defective T cell activity and complement and antibody deficiency. Both HIV-iNTS and Malaria-iNTS coinfections are characterized by the suppression of gastrointestinal inflammation and an increased production of anti-inflammatory cytokines [29]. The S.Typhimurium spread is also promoted by the consumption of the C3 complement component. Malnutrition is also a key factor especially in children, who have a reduced level of complement components and neutrophils with lower microbicidal capacity [30].

Studies from 1996 to 2016 suggest that the incidence of iNTS is higher in Africa (92%) than in Asia (8%) and the majority of cases are from Eastern Africa [31]. Among all serovars S.Thyphimurium ST313 and S.Enteriditis ST11 are the leading cause of iNTS in Africa (90%). They are characterized by the presence of peculiar SPV with mutations responsible for higher virulence and antibiotic resistance [32]. The genetic differences among NTS serovars is also responsible for peculiar pathogenesis, indeed strains belonging to the same clade can present increase replication rate in macrophages, or specific composition in the O-antigen. Furthermore, the presence of pseudogenes is associated with the adaptation to new host niches [29].S.Thyphimurium ST313 presents four plasmids (pSTL) and five prophage-like elements (BTP),which increase its invasive potential and its acid tolerance.Furthermore a mutation in pgtE gene promoter causes the production of an outern membrane protein that improve the degradation of host antimicrobial proteins, including complement proteins [30]. Both S.Thyphimurium and S.Enteriditis have acquired the resistance to multiple type of antibiotics (like ampicillin and chloramphenicol), actually making fluoroquinolones and third-generation cephalosporin the drugs of choice. However, the resistance is increasing, leading to more frequent treatment failure, especially in a region where antibiotics are commonly use to treat various syndromes.

1.4.2 Vaccination platforms against iNTS

Since humoral response and T-cell mediated immunity result to be strictly necessary to eradicate *Salmonella* infection, vaccine candidates are focused on the induction of this kind of immunity. In addiction, the presence of multidrug resistant serovars and the

consequent failure in antibiotic treatment make really necessary the development of an efficient vaccine against NTS. Several candidates, including live-attenuated, glycoconjugate and recombinant antigen based vaccines are currently under investigation with target S.Typhimurium and S.Enteritidis. Live attenuated vaccines contain strains with attenuating mutations, but with the same immunogenic capacity. These vaccines elicit both cell-mediated and humoral responses. Glycoconjugates are usually based on the presence of the O-antigen (OAg) with carrier proteins able to induce antibodies protection. Another promising approach is based on the use of Outer Membrane Vescicle (OMV), that are plebs which are spontaneously released by Gram Negative bacteria, containing LPS and OAg.Through the presence of different kind of molecules, these vesicles can self-stimulate various types of immunity [29].

Among OMV-based vaccine, one of the most interesting candidate is based the Generalized Modules of Membrane Antigens (GMMA) technology, which consist in the genetic manipulation of the contents of these vesicle in order to modify their reactogenicity and immunogenicity. The novel iNTS-GMMA vaccine platform,developed by GSK, contains S.Typhimurium ans S.Enteriditis outern membrane vesicles[33], that present modifications including the disruption of the links between the outer membrane and the peptidoglycan or inner membrane to greatly increase release of GMMA [34] and modification of the lipid A structure of the LPS, by mutation of msbB and pagP genes, results in decreased reactogenicity. [35]. The long-term immune response was tested in mice through several tests (ELISA, SBA, multiparametric flow cytometry analysis) demonstrated that the vaccine elicits a rapid production of specific serum IgM and IgG persisting for 10 weeks after a single immunization and seven months after the boosting. In addition the presence of memory B cells was assessed by ELispot 28 weeks after the boosting,together with high level of antibodies [36].

1.5 Role of transcriptomic analysis in Vaccine clinical trials and aim of the thesis

In recent years, the development of high-throughput sequencing (also called next generation sequencing NGS) has enormously increased the ability to sequence genome in a short period of time. The advent of NGS technology have redefined the classical workflow of both human and microbial genetic research, and it rapidly became crucial in the study of human infectious disease and vaccine development. NGS allows the sequencing of the entire pathogen genome rapidly, reducing the time needed to investigate the genetic variability that often characterizes human pathogens (like Influenza viruses which present

high mutation rate) and represent a big issue in the development of vaccine. Beside the genome sequencing, another significant application of NGS for vaccine research is the sequencing of host and pathogen transcriptomes(RNA-seq), as well as studies of the host immune response elicited by the vaccination [37]. RNa-seq provides the capacity to investigate differences, between subjects in immune response to a pathogen or a vaccine, are the consequences of alteration in gene expression. Transcriptome analysis is also used to characterize temporal changes in gene expression after an infection, a treatment or a vaccination schedule.

In this work we wanted to focus on the role of whole-blood transcriptomic analysis in identifying possible markers of protection induced by different vaccine platforms. The gene expression was evaluated for each timepoint and compared with the baseline (usually day 0 before vaccination), the number of differentially expressed genes allowed us to determine changes in gene expression, and consequently, which biological pathways were activated by vaccination. We also performed a correlation analysis between gene expression and antibody titers to pinpoint which genes were involved in immunoglobulin production.

Chapter 2

Transcriptomic characterization of efficacy and safety of rVSV-ZEBOV-GP in a pediatric cohort

Authors: $ChiaraSonnati^{1,*}, MariaNovedrati^{1,*}, SimoneLucchesi^{1}, GiorgioMontesi^{1}$, $DonataMedaglini^{1}, FrancescoSantoro^{1,**}$.

On behalf of the VSV EBOPLUS Consortia

Affiliations:

1 Department of Medical Biotechnologies, University of Siena, Siena, Italy

*These authors contributed equally to this work.

**corresponding author

2.1 Abstract

Background: The safety and efficacy of ZEBOV-GP have been already demonstrated in adults [13], [14], however data related to children are still limited. A phase 2 randomized, placebo-controlled trials showed that a single dose of Zebov-GP is sufficient to elicit an antibody response by day 28 after vaccination, that persists for 12 months, and a second dose, administered 56 days after the first one, can boost antibody concentration [38]. Currently, tolerability of the rVSV in children is not clearly understood as the effect of the vaccine on children immune response. The aim of our analysis is to evaluate the ability of the vaccine to induce innate and humoral immune response, through the characterization of a trasncriptomic profile.

Methods: We conducted whole-blood transcriptome analysis on 120 children divided into two age groups depending on the age (1 to 5 years old and 6 to 12 years old). In each group 40 and 20 participants allocated to receive the rVSV-ZEBOV-GP and varicella vaccines VARILRIX, respectively. RNA sequencing was performed at baseline (D0), day 1

(D1), day 2 (D2), day 7 (D7), day 14 (D14) and day 28 (D28) post-vaccination. Differential gene expressions were analyzed using DESeq2, enrichment analysis was performed with blood transcription modules (BTM). Co-expression was assessed with CEMiTool and correlation analysis was performed with psych.

Results: rVSV-ZEBOV-GP induced robust early innate responses with 10.900 differentially expressed genes at D1 (vs VARILRIX), mostly related to innate immune resposnse activation, IFN signature and antigen presentation. Along the entire period of evaluation, rVSV-ZEBOV-GP elicited stronger immune response than VARILRIX, but no module related to immunoglobulins was detectable. A deeper analysis was conducted on the rVSV-ZEBOV-GP group, comparing gene expression of each timepoint with the baseline. Here, gene expression analysis allowed us to appreciate by D7, the upregulation of few genes related to B cells and antibody production. Co-expression analysis revealed significant B-cell module activation (M6 and M9) at D7. According to antibody titers, we classified the entire cohort in two groups, responders and not responders for those a correlation analysis between gene expression and antibody titers was performed. We also investigated the role of age and gender on vaccine efficacy; no differences were found between male and female, but we found that infants (1/2 years old) were characterized by upregulated BTM related to immunoglobulins and B cells since day 0.

Conclusion: rVSV-ZEBOV-GP vaccination induces a rapid, and strong innate immune activation within 24 h, followed by B-cell signatures appreciable by day 7. Vaccine efficacy is not affected by gender, but for infants it was not possible to detect a clear trend for immune activation. These transcriptomic signatures provide insights into rVSV-ZEBOV-GP vaccine immunogenicity in pediatric cohort, though further analyses are needed to clarify the real efficacy in infants.

2.2 Introduction

Ebola is a negative-sense, single-stranded RNA filovirus that causes severe hemorrhagic fever with fatality rates reaching up to 90%. The virus initially infects dendritic cells and macrophages, spreading to lymph nodes and subsequently to major organs, where it triggers a massive inflammatory response ("cytokine storm"), profound immunosuppression, and endothelial dysfunction. Human-to-human transmission occurs through contact with infectious bodily fluids, while the animal reservoir is thought to include bats and small mammals. Incidence and disease severity increase with age, but children experience faster progression and higher mortality rates. The rVSV-ZEBOV-GP vaccine, a recombinant VSV vector expressing the Zaire ebolavirus glycoprotein, has shown strong immunogenicity and approximately 100% efficacy after a single dose in clinical

trials, with an overall favorable safety profile in adults, adolescents and children.

2.3 Materials and Methods

2.3.1 Study design

The study was carried out at the Centre de Recherches Mèdicales de Lambaréné in Lambaréné, Gabon. It is a phase 2, randomized, controlled, open-label pediatric trials aimed to investigate tolerability, safety and immunogenicity of the rVSV-ZEBOV-GP vaccine (at nominal doses of 1 ml 7.8×10^7 plaque-forming units) as compared with a varicella zoster vaccine VARILIX(VZV, GlaxoSmithKline, UK, 0.5 ml dose, $10^{3,3}$ plaque-forming units) in children aged 1 to 12 years [39].

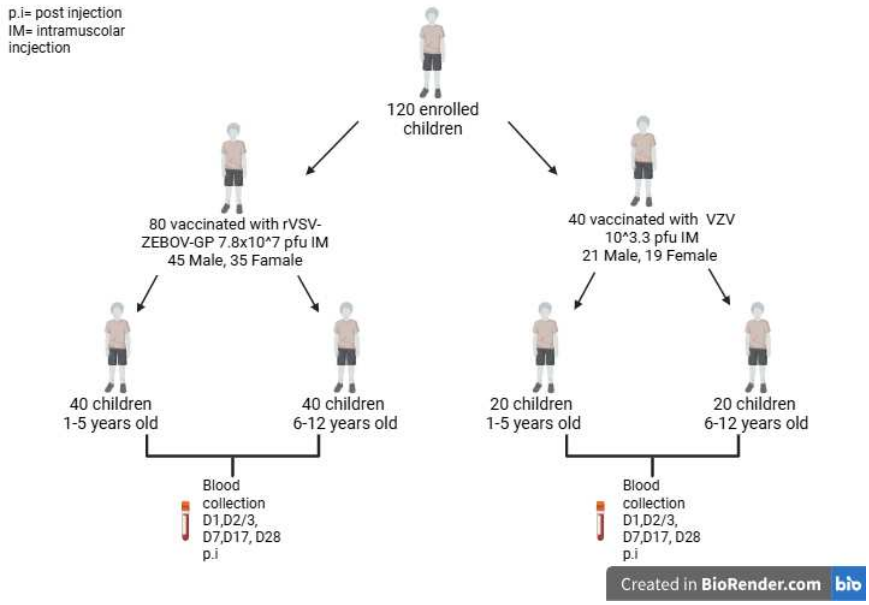


Figure 2.1: *Experimental design.A total of 120 participants were recruited and divided into two age groups (1 to 5 years old and 6 to 12 years old) of 60 children. The sample size had been calculated using a 2:1 ratio, thus 40 and 20 participants per age group were allocated to receive the rVSV-ZEBOV-GP and varicella vaccines VARILRIX, respectively.Blood samples were collected on days 0,1,2/3, 7,14,21,28,56,84,180 and 365 after vaccination in both groups, for occurrence of adverse effects,presence of rVSV ribonucleic acid, concentration and avidity of specific antibodies. For our transcriptomic analysis ,the blood was collected at day 0,1,2,7,14 and 28 post vaccination.*

For our transcriptomic analysis, we received a total of 379 RNA samples (from 116 different subjects belonging from both group of treatment), so we first made a comparison between these two conditions (Figure 2.1). Total number of samples used for each analysis and the criteria for their selection are shown in Figure 2.2.

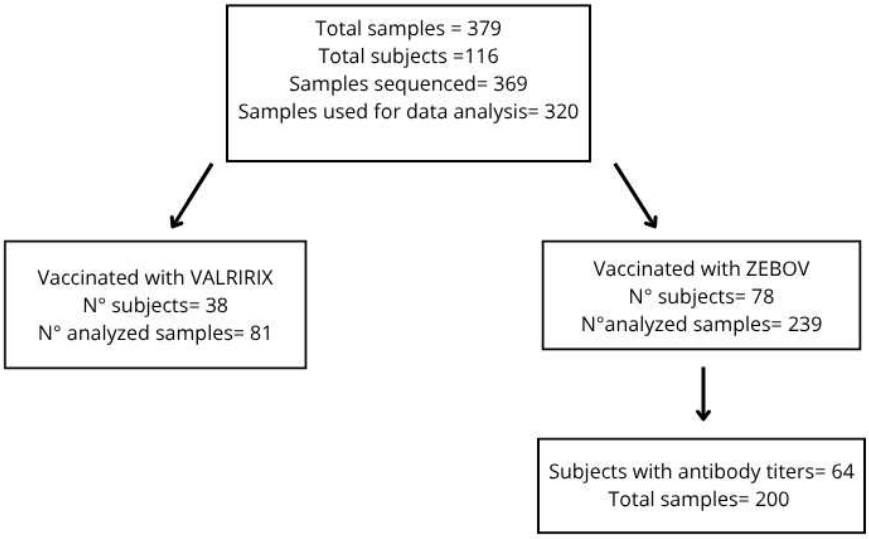


Figure 2.2: *Samples used for data analysis.We received 379 RNA samples, but only 369 were available for the sequencing after RNA quantification. After the sequencing,we selected 320 sample which had at least* $5x10^6$ *total reads, for further analysis. We used all 320 selected samples for the comparisons between the two vaccination groups. For the correlation analysis between antibody titers and gene expression, we received data of 64 subjects for a total number of 200 samples.*

2.3.2 RNA quantification, library preparation and quantification, sequencing

The Invitrogen Qubit™ 4 Fluorometer (ThermoFisher Scientific) was employed for RNA quantification. The assay of choice was the RNA High Sensitivity Assay kit, that allows the quantification of RNA sample with initial concentration from 0.2 to 200 ng/uL.

Sequencing libraries were prepared using the Illumina® Ribo-Zero Plus Kit starting from 40 ng of total RNA. This kit allows the depletion of rRNA from purified total RNA, followed by a reverse transcription and conversion of the remaining RNA into cDNA. The cDNA is then amplified, and adapters are added for clustering and sequencing.

Libraries were quantified with Invitrogen Qubit™ 4 Fluorometer with the DNA Broad Range assay kit as suggested by Illumina protocols. Depending on their concentration, the libraries were grouped into different pools (9 libraries for each pool). The final concentration of each pool was estimated with SS DNA Broad Range kit.

A further quantification and quality check were performed by using the Agilent 2100 Bioanalyzer (Agilent Technologies) and the DNA 1000 kit specific for the analysis of

DNA fragments between 25 and 1000 bp in length.

Sequencing was performed with Illumina NovaSeq6000 Xp 4-lane Kit v1.5 on a S2 flow cell, which can be used for the sequencing of 100 bp both forward and reverse sense a produces an output of about 667-833 Gb.

Before loading, the pools were diluted and denatured following the "Denature and dilute library guide". Then EXAmp Master Mix was added to each pool and finally, all libraries were loaded together into a specific tube.

Reads quality was assessed with FATSQC [40] and Trimmomatic [41] was used to discard low quality reads. Reads were aligned to the reference genome (hg38) by using STAR alignment tool. HTseq [42] was used to count the number of aligned reads. For the analysis, we used samples characterize by at least 5 millions of total reads.

The samples were stored in a descriptive table, showing the respective subject ID, sample number, treatment group, time point, age group, sex, antibody titer on days 0, 28, 180, and 365 and antibody response group(2.1).

Subject ID	sample	Vaccine	Timepoint	age group	Gender	Ab day 0	Ab day 28	Ab day 180	Ab day 365	Response
PEV004	N_4	ZEBOV	0	child	F	-0,204	0,133	0,274	-1,529	Not Responders
PEV004	N_23	ZEBOV	1	child	F	-0,204	0,133	0,274	-1,529	Not Responders
PEV004	N_2	ZEBOV	7	child	F	-0,204	0,133	0,274	-1,529	Not Responders
PEV004	N_16	ZEBOV	14	child	F	-0,204	0,133	0,274	-1,529	Not Responders
PEV004	N_24	ZEBOV	28	child	F	-0,204	0,133	0,274	-1,529	Not Responders
PEV018	N_72	ZEBOV	14	young	M	-0,204	1,227	2,498	1,482	Responders
PEV018	N_71	ZEBOV	7	young	M	-0,204	1,227	2,498	1,482	Responders

Table 2.1: Design of the descriptive table Data of all 379 whole blood samples are stored in this descriptive table

Data obtained from sequencing were stored in a count table whose rows contain gene names and columns contain a sample at a specific timepoint(2.2).

Gene	N_4	N_23	N_2	N_16	N_24	N_72	N_71
ENSG00000000003	10	13	8	2	3	4	56
ENSG00000000419	2284	8902	3894	2250	6262	5612	7909
ENSG00000000938	138	326	172	102	376	174	228

Table 2.2: Design of the count table

2.3.3 Data analysis

Data analysis was performed with Bioconductor and R version 4.1, which is a set of integrated tools including a console, a syntax-highlighting editor that supports direct code execution, and tools for plotting, viewing history, debugging, and managing the workspace.

For RNA-seq data, the strategy taken is to count the number of the reads that fall into annotated genes and to perform st atistical analysis on counts table to discover quantitative changes in expression levels between experimental groups [43]. Two widely used packages were used to perform data analysis, edgeR and DESeq2, which allow both the qualitative and the quantitative analysis of data.

These two packages allow the discrimination of the gene interest from the all the expressed genes and then they are able to highlight the differentially expressed ones between the distinct macaques groups. Although they employ similar strategies to perform the analysis, these packages differ for some important aspects. For the normalization, edgeR uses the trimmed mean of M values, that discards the extreme values and uses the samples whose expression is closest to the mean as reference; whereas DESeq2 uses a relative logarithmic expression approach through which it compares every sample to the others [43]. In addition, they also differ for the dispersion estimation, indeed edgeR is more sensitive to outliers.

edgeR and Principal component analysis

The package edgeR was used to perform a normalization of the row read counts. A list with samples and the respective counts and genes was made using the DGEList function. Genes with < 1 CPM in < 10 samples were filtered out. The remaining counts were normalized with calcNormFactors function, that finds set of scaling factors for the library sizes that minimizes the log-fold changes between the samples for most genes and finally converted in log2(CPM+1) Normalized data were used to perform a principal component analysis (PCA) with prcomp (center=TRUE, scale.=TRUE), to visualize the variation present in the dataset.

DESeq2

DESeq2 allows a quantitative analysis of comparative RNA-seq data using shrinkage estimators for dispersion and fold change [44]. As for edgeR, the starting point of a DESeq2 analysis are the descriptive table and the counts table. First, a DESeq dataset was created by using the function DESeqDataSetFromMatrix, then a differential expression

analysis was performed through the function DESeq. This function performs a default analysis through 3 steps:

- Estimation of size factors
- Estimation of dispersion
- Negative Binomial GLM fitting and Wald statistics

We first performed a differential gene expression analysis ,including all the sequenced samples,to initially compare both vaccination platforms across all timepoints and then rVSV-ZEBOV-GP with baseline (D0).

- rVSV-ZEBOV-GP vs VZV
- rVSV-ZEBOV-GP group vs day 0

Then we focused on the differences between male, female and age groups across all timepoints in rVSV-ZEBOV-GP vaccination group.

- Children (1-5 years old) vs Youngs (6-12 years old)
- Female vs Male

We subsequently stratified vaccinees in two groups (responders and not responders) according to their antibody titers and then we performed a differential gene expression analysis between them.

- Responder vs Not Responders

Each of these analyses produced a file containing a list of genes with the associated *p-value*,the *adj p-value* (also called false discovery rate) and the *log2fold change*.

Enrichment analysis with tmod, ClusterProfiler, KEGG and Gene Ontology

To perform enrichment analysis we use different R packages. *tmod* package provides sets of co-expressed genes. Here we used LI modules which are obtained by the transcriptomic analysis of immune response to different vaccines in human. Three hundred thirty-four Blood Transcription Modules (BTMs) were identified and annotated according to their biological functions and/or tissue -specific expression patterns. [45] For each module significantly up or down-regulated genes are counted, and modules are ranked according to the CERNO test (tmodCERNOtest function), which is an application of Fisher's

method for ranked list in feature set enrichment analysis. The test uses scaled ranks of features, combines them and directly calculate the p-values and it is particularly effective with large sample sizes [46].

clusterProfiler is an R package for comparing biological themes among gene clusters, it also performs enrichment analysis of those clusters starting form differential expression analysis data obtained with DESeq2. This package allows to incorporate the biological knowledge provided by Gene Ontology and KEGG for annotating genes to biological processes [47].

KEGG (Kyoto Encyclopedia of Genes and Genomes) is one of the most important databases for the representation and analysis of biological systems.It contains genomic and functional information about genes and biological pathways in which these genes are involved. KEGG creates pathway maps representing systemic functions of the cell and the organism in terms of molecular interaction and reactions networks [48].

The Gene Ontology (GO) is the most widely used ontology for specifying cellular location, molecular function, and biological process participation of human and model organism genes [49], and it is extremely useful to analyze huge amount of data coming from RNA-seq. GO contains two main blocks, the ontology and the annotation, which are correlated with each other to identify genes and their relationships to each other. Its workflow is similar to tmod, the analysis starts with identifying a list of differentially expressed genes (adjp-value <0.05), that are then investigated to determine those GO terms that are over- or under-expressed within the gene set of interest.

Co-expression analysis with Cemitool

CEMiTool (Co-Expression Modules identification Tool) is another R package that allows to identify and analyze co-expression modules, among normalized levels of expressed genes, in a fully automated manner. This tools also provides additional analysis performing profile plots, module activity analysis, functional enrichment analysis, it is able to generate interaction networks and produces reports HTML web format. For the analysis it only requires a file with genes as rows and samples as columns, but if a sample annotation and gene sets files are provided, it is able to perform a Gene Set Enrichment Analysis (GSEA) and an over representation analysis (ORA), allowing the user to visualize which modules are induced or repressed and to determine the most significant function [50]. The modules table shows the number of modules identifies by CEMiTool, the number and the name of genes for each module.

CEMiTool perform GSEA using the fgsea (Fast Gene Set Enrichment Analysis), that provides an efficient estimation of GSEA p-value in a few minutes. The method consists

of two main procedures: FGSEA-simple and FGSEA-multilevel. FGSEA-simple procedure allows to efficiently estimate P-values with a limited accuracy but simultaneously for the whole collection of gene sets, while FGSEA multilevel procedure allows to accurately estimate arbitrarily low P-values but for individual gene sets [51]. The package first does a normalization based on the Z-score on all genes from co-expression modules and calculates the mean for each sample class, then it applies a pre-ranked GSEA approach. The obtained plot contains circle of different colors and dimensions; red and blue represent higher and lower module activity, respectively, while the dimension and the intensity are proportional to the Normalized Enrichment Score value. In this work, the module activity analysis was also reported as a box plot created with ggplot2.

The over representation analysis is represented as a bar graph in which only the most significantly enriched pathways for each module are represented. Each graph has a specific color that is assigned randomly by the tool itself, but the intensity is proportional to the adjusted p-value.

Mclust and Umap

These two package were used to classify samples in two different classes (Responders and not Responders) depending on their antibody titers. Candidates who were seropositive at D0 (4 children) were excluded from the analysis. UMAP (Uniform Manifold Approximation and Projection) is a technique for dimension reduction such as PCA and t-SNE, but it presents a superior run time performance, guaranteeing the same visualization quality. This algorithm can also be used for working with significantly large datasets [52]. UMAP is classified as a k-neighbor based graph learning algorithm, it can be divided into tow phases, in the first one a weighted k-neighbor graph is constructed and then a low dimensional layout of this graph is computed. Mclust is a R package for clustering, classification and density estimation based of finite Gaussian mixture modeling [52].

Correlation analysis

Correlation between gene expression, BTMs and antibody titers were assessed with the R package *psych*. For each children, a module activity score was calculated to measure BTMs activity and defined as increase or decrease of gene expression compared to baseline level (Day 0). To calculate module activity, gene expression was normalized with vst function of DESeq2 package and subsequently normalized in Z-scores.Then, module activity score at day X was defined for each module in each children by computing the foldchange (DX-D0). We also computed the foldchange of each gene for all children.

Spearman correlation was performed between module activity scores and titers with the corr.test function.

2.4 Results

Comparison between rVSV-ZEBOV-GP and VARILRIX vaccines

We first investigated differences between groups of treatment without taking in account age, gender and antibody titers of patients.

We performed a PCA analysis to visualize the trend of samples at each timepoint, to investigate the presence of eventual outliers and to highlight the different behavior between patients treated with ZEBOV and those treated with VARILRIX. The graph showing PCA is reported in Figure 2.3,

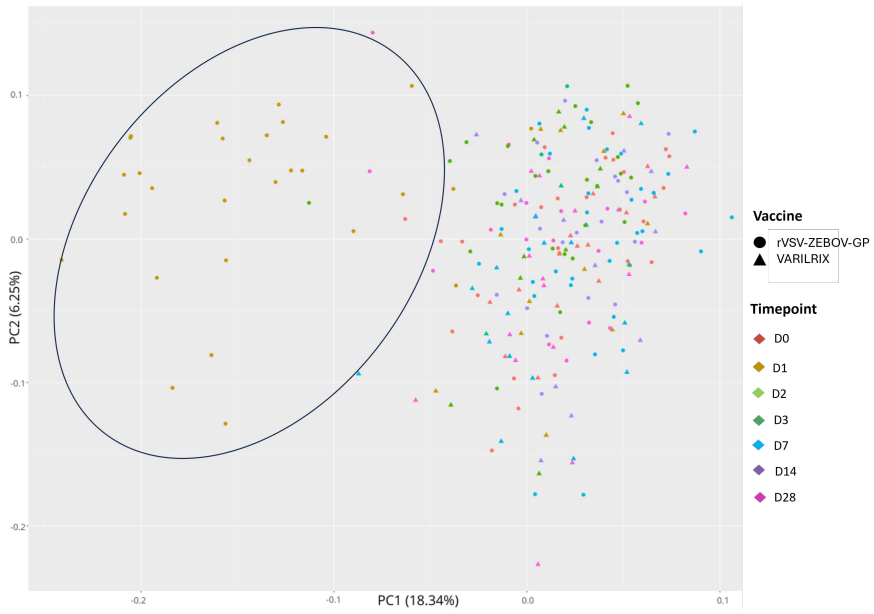


Figure 2.3: *PCA plot.The plot shows a PCA analysis conducted for all 369 sequenced samples. It is possible to see that soon at day 1 after injection, vaccinees receiving ZEBOV cluster together into the ellipse (defining the normal confidence interval of 95% for the group of samples), whereas there are no clear cluster for other timepoints.Principal Component 1(18,3% of total variance) and Principal Component 2 (6,25% of total variance) were chosen for this plot, in order to produce a graph as discriminative as possible*

Differential gene expression analysis was performed with DESeq2 across all timepoints and groups of treatment. The analysis showed that changes in gene expression started soon after the vaccine administration and remained clearly visible until day 7 after vaccination, while only a few changes were still visible at day 14. The following table shows the

amount of differentially expressed genes (DEGs) between the two vaccination schedules for each timepoint (2.3). On day 1 genes related to innate immune response were appreciable, in particular those linked to interferon pathways (IFITM3, IFI6, IFI35), dendritic cells and macrophages activation (RUFY4,CCRL2) and viral immunity (OASL,DDX58). At day 7 the most significant gene was IFI27, which was still upregulated in ZEBOV cohort. The upregulated genes at day 14 were mainly related to antibody production (IGHV3-49,IGLV1-40,IGHV2-5), but also IFI27 was still upregulated.

ZEBOV vs VZV DEGs			
	Upregulated	Downregulated	Unaffected
D0	0	0	19.530
D1	5.306	5.594	7.589
D2	161	80	17.638
D7	112	444	18.479
D14	6	3	19.089
D28	0	0	19.236

Table 2.3: DEGs table.The tables shows the numbers of significant (p.adj <0.005) DEGs between vaccinees with ZEBOV and VZV for each timepoint.

We conducted the same analysis considering only the ZEBOV cohort against the baseline (Day 0). The results confirmed that at day 1 after injection we had the highest number of DEGs (n=16.062,70% of total expressed genes), that drastically decreased after day 7, when they are only 203. Results are shown in table 2.4. Although DEGs were consistent with those found in the previous analysis and genes related to inflammatory response and viral immunity that were upregulated at day 1,2 and 7 (IFI27, IFI44,SIGLEC1), here genes related to antibody production were not appreciable, except for IGHV2-5 and IGSF6 at day 7.

	Upregulated	Downregulated	Unaffected
D1 vs D0	7.439	8.623	7.211
D2 vs D0	1.184	1.138	20.951
D7 vs D0	56	147	23.070
D14 vs D0	2	1	23.270
D28 vs D0	6	2	23.265

Table 2.4: DEGs table.The tables shows the numbers of significant (p.adj <0.005) DEGs in ZEBOV cohort using day 0 as baseline

The enrichment analysis of DEGs was conducted with *tmod* package, that allowed the identification of enriched blood transcription modules (BTMs), which are defined as groups of functionally related and co-regulated genes. Comparisons between ZEBOV-GP and VARILRIX showed enriched BTMs only at D1 after injection, the majority were related to innate immune response (LI.M75,LI.M37.0). We have a massive upregulation of those modules related to the inflammatory response (L.M33, LI.M53), monocytes (LI.M118.0,LI.M11.0, LI.M73), activation of dendritic cells (LI.M67,LI.M64) and the innate antiviral immune response (LI.M150), which have a p-value lower than 10^{-4}. No significant differences were appreciable at the baseline among the two grous. At Day 7 after injection we had a dowregulation in modules related to neutrophils, dendritic cells and T cells, but o modules related to B cells activation was detectable (Figures 2.4, 2.5 and 2.6).

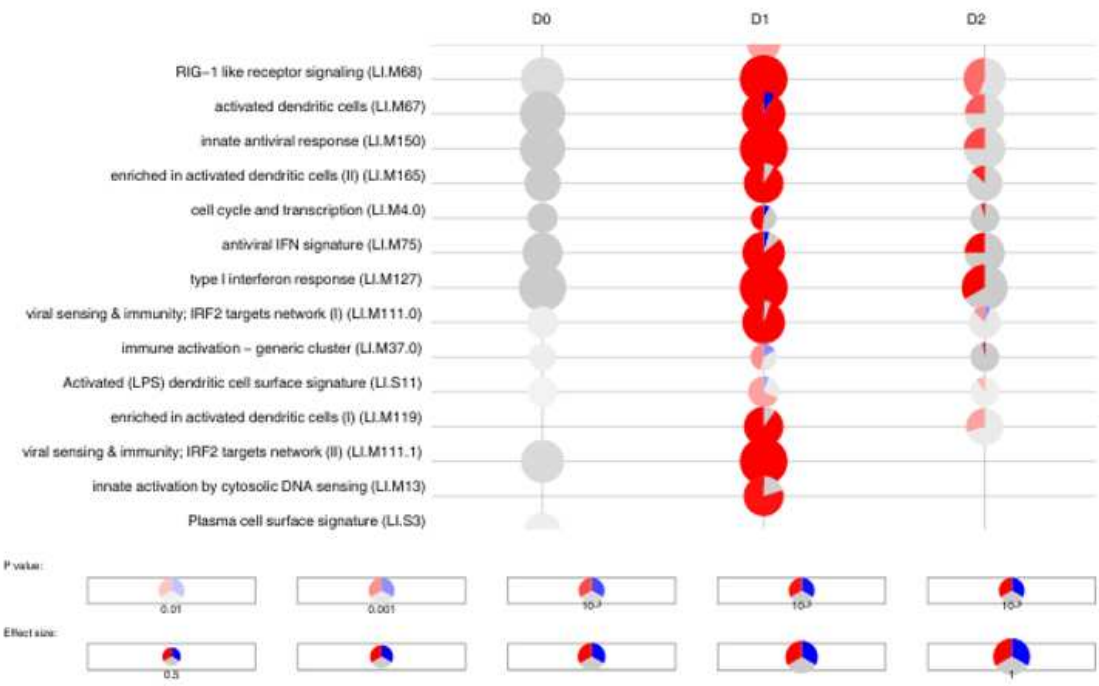


Figure 2.4: *Comparison of activated blood transcription modules (BTMs) by rVSV-ZEBOV- GP and VZV vaccines at day 0,1 and 2 after injection. Activation of modules was tested using the FDR-ranked lists of genes generated by DESeq2 generalized linear model fitting and applying the CERNO test. Rows indicate different BTMs, which were significantly (FDR < 0.05) activated. Each module is represented by a pie in which the proportion of significantly upregulated and downregulated genes is shown in red and blue, respectively. The gray portion of the pie represents genes that are not significantly differentially regulated. The significance of module activation is proportional to the intensity of the pie, while the effect size (area under the curve) is proportional to its size.*

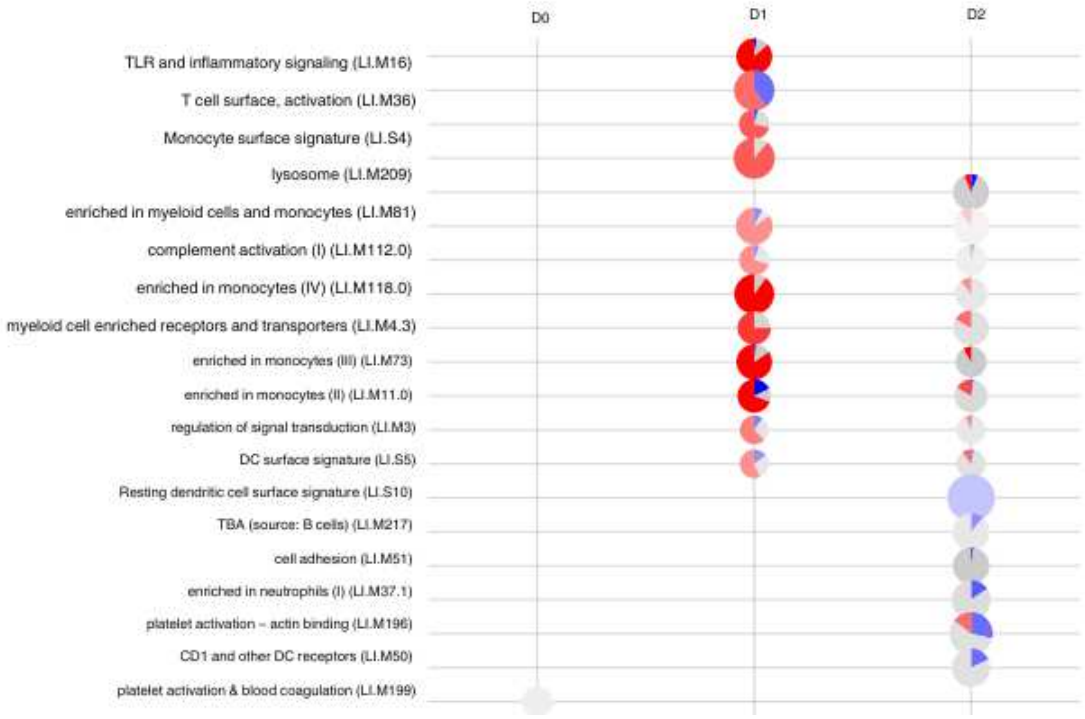


Figure 2.5: *Comparison of activated blood transcription modules (BTMs) by rVSV-ZEBOV- GP and VZV vaccines at day 0,1 and 2 after injection.*

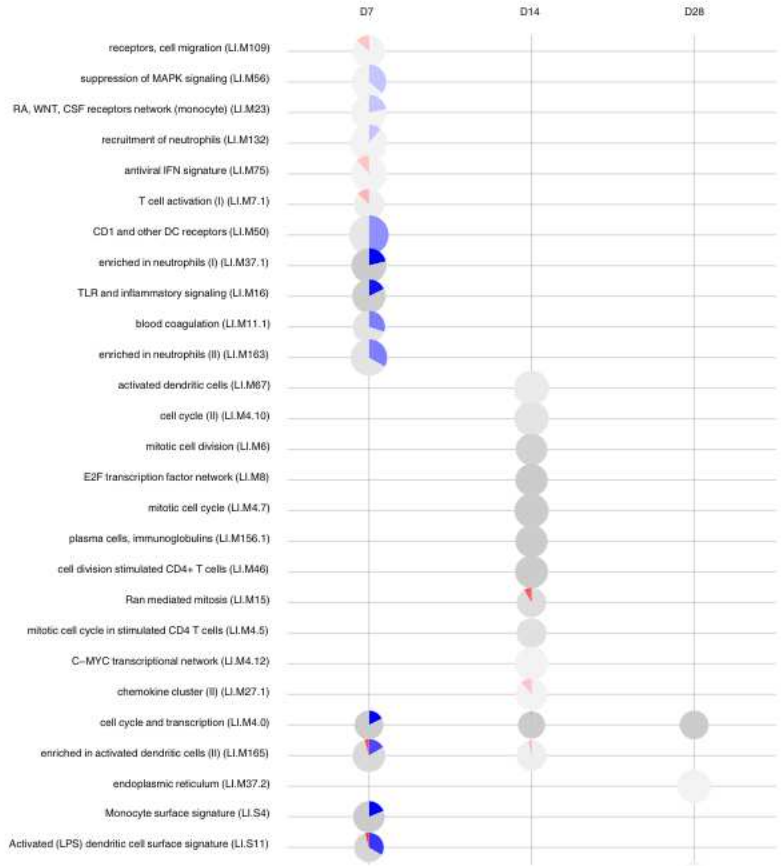


Figure 2.6: *Comparison of activated blood transcription modules (BTMs) by rVSV-ZEBOV- GP and VZV vaccines at day 7,14 and 28 after injection.*

Results of enrichment analysis for ZEBOV-GP against the baseline (D0) were coherent with those found in differential gene expression analysis. At day 1 modules related to innate immune response were strongly upregulated (LI.M33, LI.M53, LI.M4.3, LI.M150), while those regarding T cells activation and differentiation (LI.M7.3,LI.M52,LI.M18) were downregulated. B cells related BTMs were downregulated at day 2 after vaccination, but unfortunatly they were not detectable at later timepoints.

A gene set enrichment analysis (GSEA) and an over-representation analysis were performed with *CEMiTool* for ZEBOV vs VZV and for ZEBOV vs baseline. The over representation analysis identified eight new modules of co-expressed genes, among them the most interesting were related to innate antiviral immune response (M1), NK cells and T cells differentiation (M4), and B cells (M5) for ZEBOV-GP vs VZV analysis and NK cells and T cells differentiation (M4), B Cells (M6 and M9) for ZEBOV vs baseline analysis. The box plots generated for the GSEA provided an overview of each co-expression module across the entire course of the experiment, confirming the trend of immune response for both comparisons (Figure 2.7), with additional information about enrichment

of B cells (Figure 2.8), that were not previously detected.

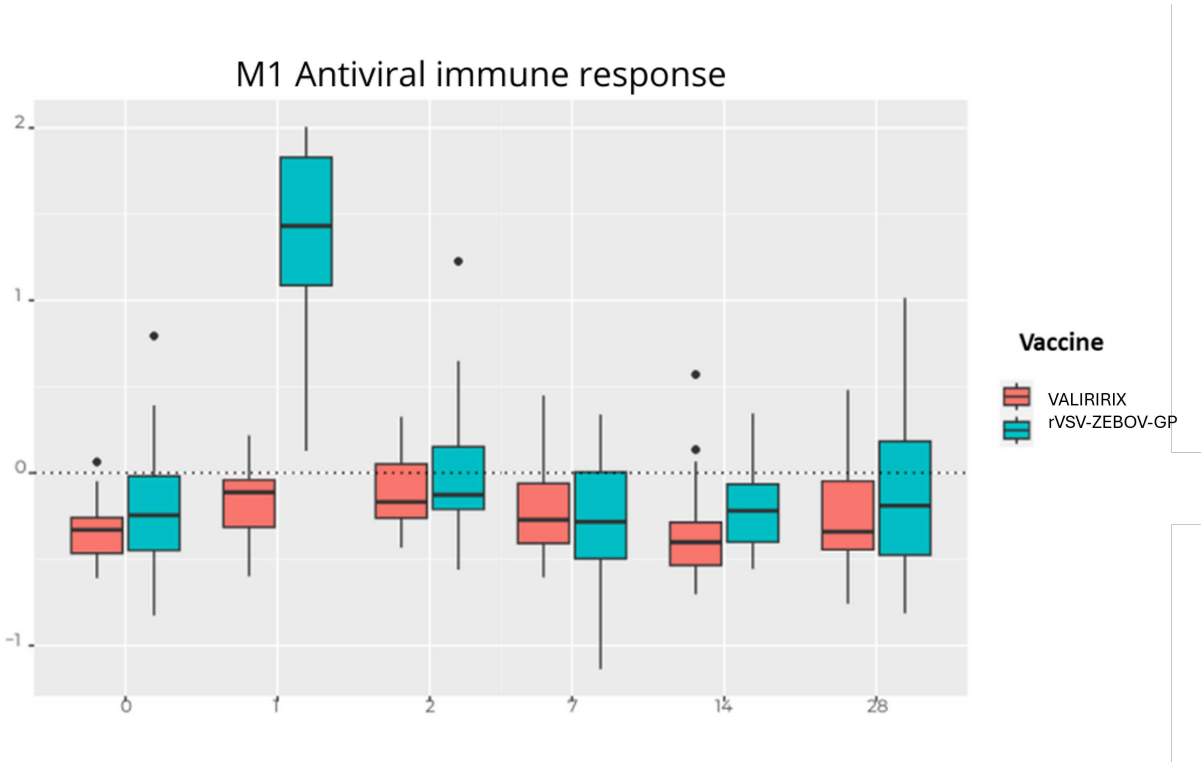


Figure 2.7: *GSEA of innate immune response module M1 with all meaningful timepoints. On X axis timepoints were represented , while on Y axis the Z-score of the module respect to each timepoint. The innate immune response is soon activated after the administration of the vaccine, with a peak at day 1 after the dose. It slowly decreases until it comes back to the baseline at day 14 post dose.It is visible that ZEBOV elicited the strongest response at D1*

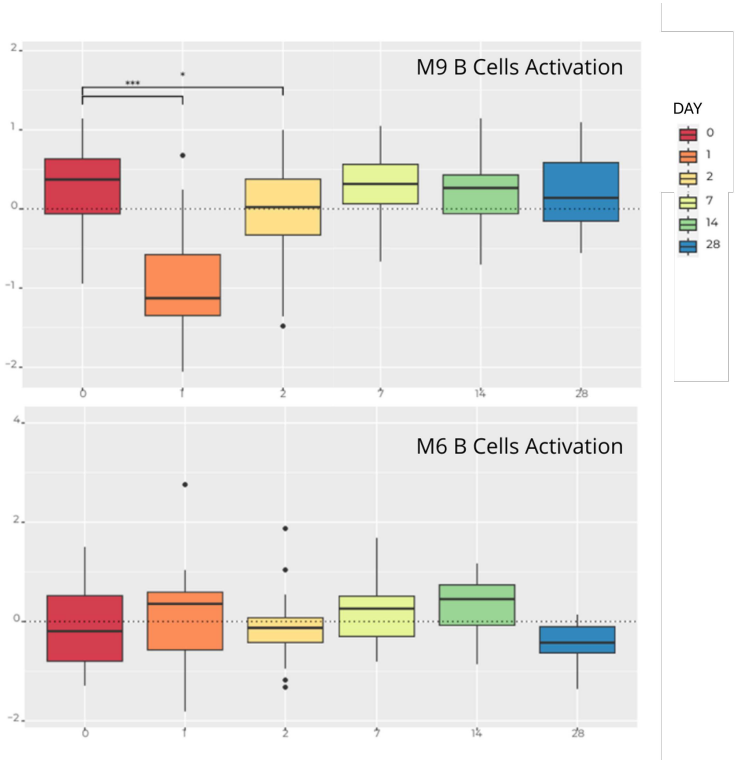


Figure 2.8: *GSEA of innate immune response module M6 and M9 with all meaningful timepoints. Both modules suggest that there is not a strong activation and proliferation of B cells after vaccination, but values remain stable. The significant differences are visible only in M9 and in the first two day after vaccination, when we can appreciate a decrease in the proliferation value at day 1, that reach the baseline at day 2.*

Clustering

Once we received data on antibody titers, we performed a clustering analysis to identify the eventual groups of patients depending on the result of the serological tests. Our unsupervised classification provided two final groups of vaccinees (Not responders N=88, Responders N=112) visible in Figure2.9. To better visualize the kinetics of antibody titers, another plot was created as reported in Figure2.10.

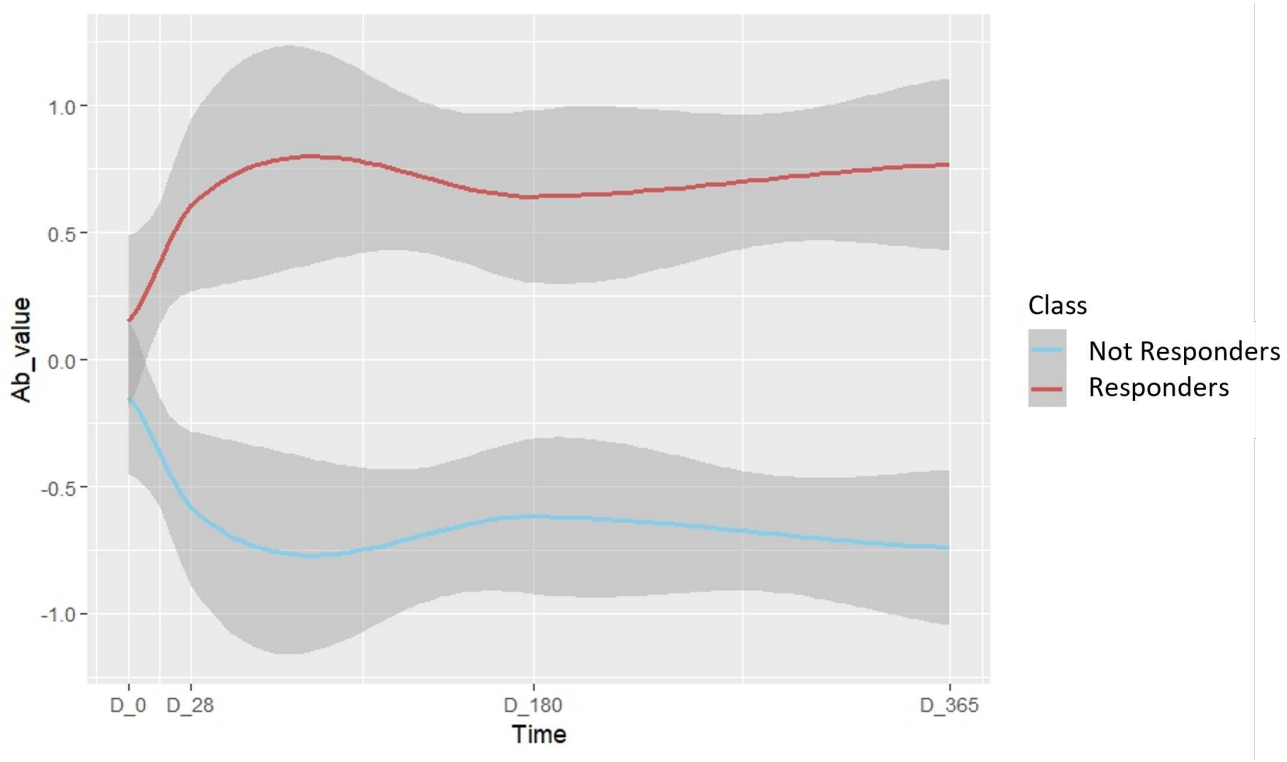


Figure 2.9: *MClust clustering analysis result. The picture shows the two final clusters created by MClust algorithm. Not responders are represented in blu, while Responders in red.*

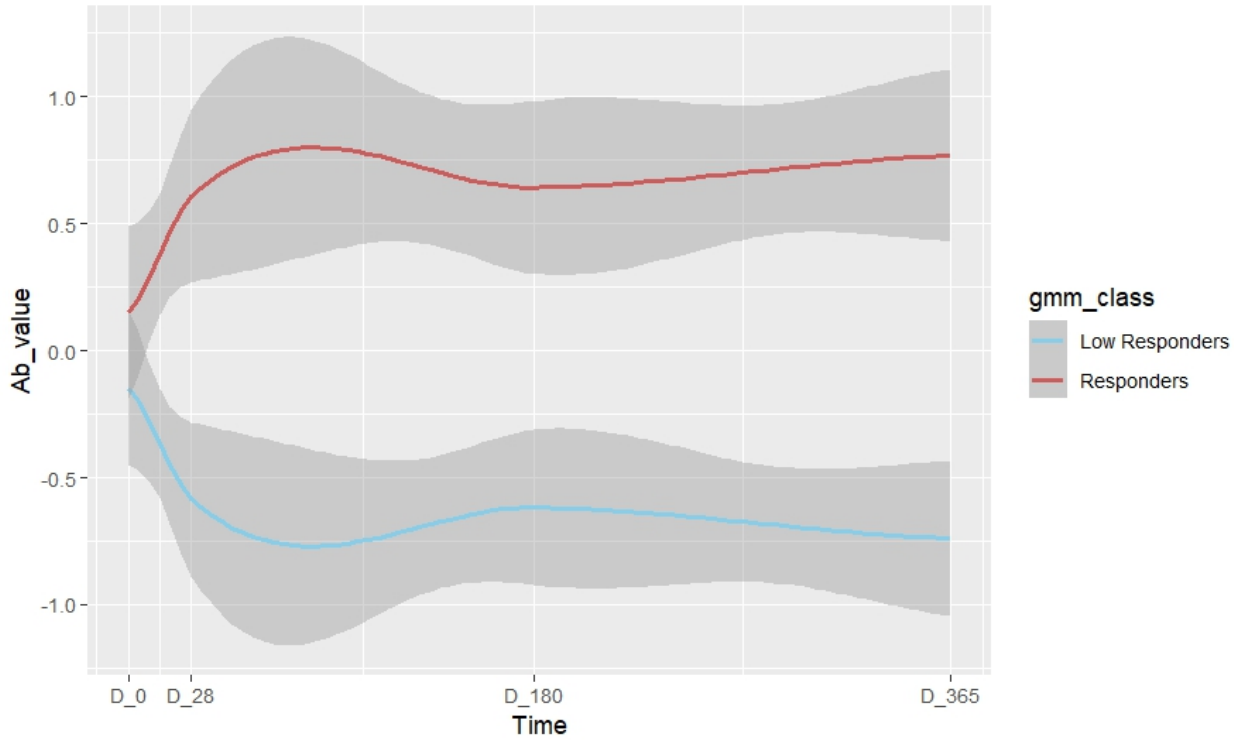


Figure 2.10: *Antibody titers kinetics.*

We performed again PCA, DEGs and enrichment analyses to identify those genes responsible for the classification. We created a PCA plot for each timepoint to visualize the two classes and we created a list of the first ten genes in PCA components which cover the 90% of the total variance. We performed an enrichment analysis for all the lists with DAVID and STRINGR annotation tools, but no significant result was obtained.

Differential gene expression analysis with DESeq2 was performed using Not responders as control group. It revealed that only few genes presented changes in the expression between the two classes, as visible in table2.5

Responders vs Not Responders			
	Upregulated	Downregulated	Unaffected
D0	0	5	23.032
D1	0	1	23.036
D7	10	4	23.023
D14	2	11	23.024
D28	8	0	23.029

Table 2.5: DEGs table.The tables shows the numbers of significant (p.adj <0.005) DEGs Responder class using Not Responders as control group.

The highest number of DEGs were found at day 7 and 14 after injection. At day 7 in Responders group we had an upregulation of two important genes the are CXCL10 (related to stimulation of monocytes, natural killer and T-cell migration, and modulation of adhesion molecule expression) and IGKV3D-20 that is part of immunoglubulin complex. Among the downregulated genes, we had TUSC2 that is involved in inflammatory response,natural killer cell differentiation and regulation of cytokine production. At day 14 changes in expression involved genes not related to immune response, such as ZNF235, HMGXB4, except for FPR3 (downregulated) that is part of several processes including complement receptor mediated signaling pathway and neutrophils activation. DEGs at day 28 were the most relevant because we had a meaningful upregulation of genes C1QC,C1QA,C1QB related to classical complement pathway activation.

Results of DESeq2 analysis were confirmed by tmod enrichment pie plot, visible in Figure 2.11.

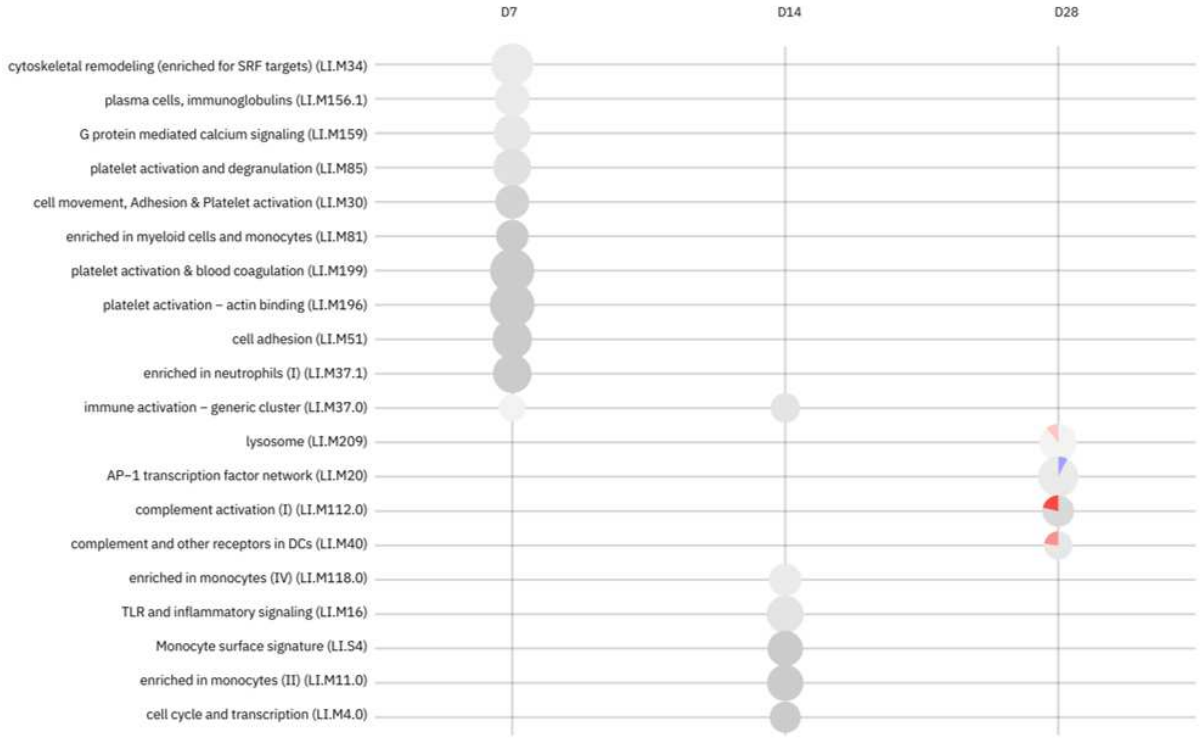


Figure 2.11: *Activation of blood transcription modules (BTMs) by rVSV-ZEBOV- GP in Responders vs Not responders comparison. The picture shows that no modules are significantly enriched at day 7 and 14, but at day 28 the upregulated module related to complement activation is clearly visible.*

Gender and Age analysis

Generally,the immune response elicited by a vaccine could vary among people of distinct age and gender, so we assessed these differences to better clarify the immunogenicity

of our vaccine. In table 2.6 are reported information about gender and age for all 200 samples.

Age and Gender		
	Female	Male
Children (1 to 5)	66	36
Youngs (6 to 12)	32	66

Table 2.6: *Age and Gender.The tables reports age and gender of 200 samples. We used only those 200 samples, whose antibody titers was known.*

As for the previous analyses, we performed a PCA, a differential gene expression and an enrichment analysis at first using age information, then using gender.

The PCA plots soon highlighted that the two age groups had a peculiar response to the vaccine, suggesting that gene expression presented relevant differences among them, as shown in Figure 2.12.

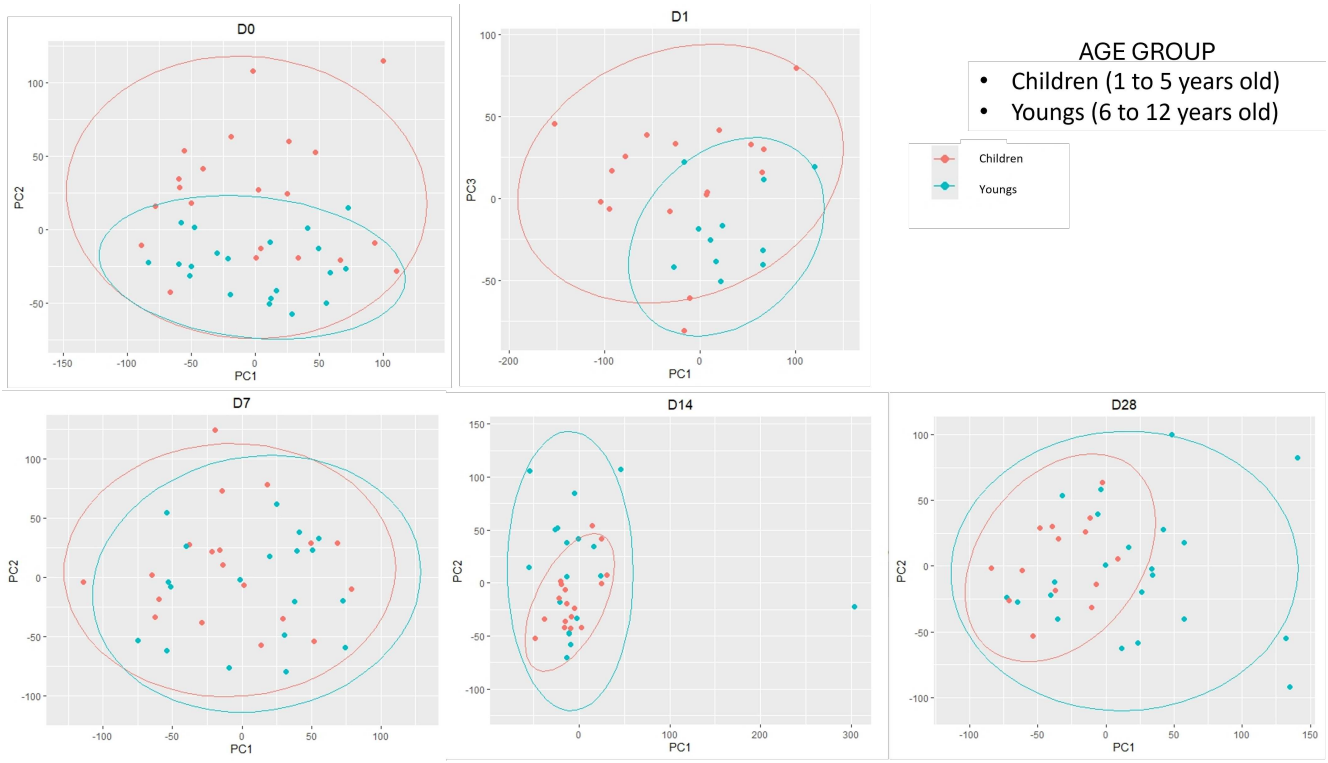


Figure 2.12: *Age Group PCA plot. The figure shows PCA plots for each timepoint for child and young age groups. The groups had a diverse genetic profile before and after vaccination, indeed two different clusters are clearly visible at all timepoints except for day 7. Plot of day 0 suggested that these two groups present a distinct gene expression also before the vaccination.*

The differential gene expression for age groups (Children vs Youngs) showed generally an higher numbers of DEGs than the one performed on gender groups, meaning that changes in gene expression were mostly related to the age than to the gender and confirming the trend suggested by PCA plots. Unusually, for age groups the highest number of DEGs was found at day 0 (n=756) and day 28 (n=817). At day 0 we had an unexpected upregulation of genes related to immunoglobulin production (like IGHV3-43,IGLV4-69,IGKV3-11,NEO1), T cells activation (CD84,CCR3) and general immune response activation (TLR7,IRF4,CXCL10). Other genes were related to cell cycles, transcription and other biological processes (CDC6,KIF21A,EPHB2). The upregulation of immunoglobulin related genes still persisted at other timepoints. At day 14 we found an upregulation in genes related complement activation (C1QA,CR2). At day 28 complement receptor gene CR2 was still upregulated, but the majority of DEGs were involved in cell cycle. DEGs are visible in table2.7

Children vs Youngs			
	Upregulated	Downregulated	Unaffected
D0	456	300	20.004
D1	261	224	17.846
D7	21	23	19.957
D14	23	35	20.269
D28	362	455	19.400

Table 2.7: DEGs table.The tables shows the numbers of significant (p.adj <0.005) DEGs in the comparison between the two age groups

Volcano plots were used to better visualize DEGs (Figure 2.13.

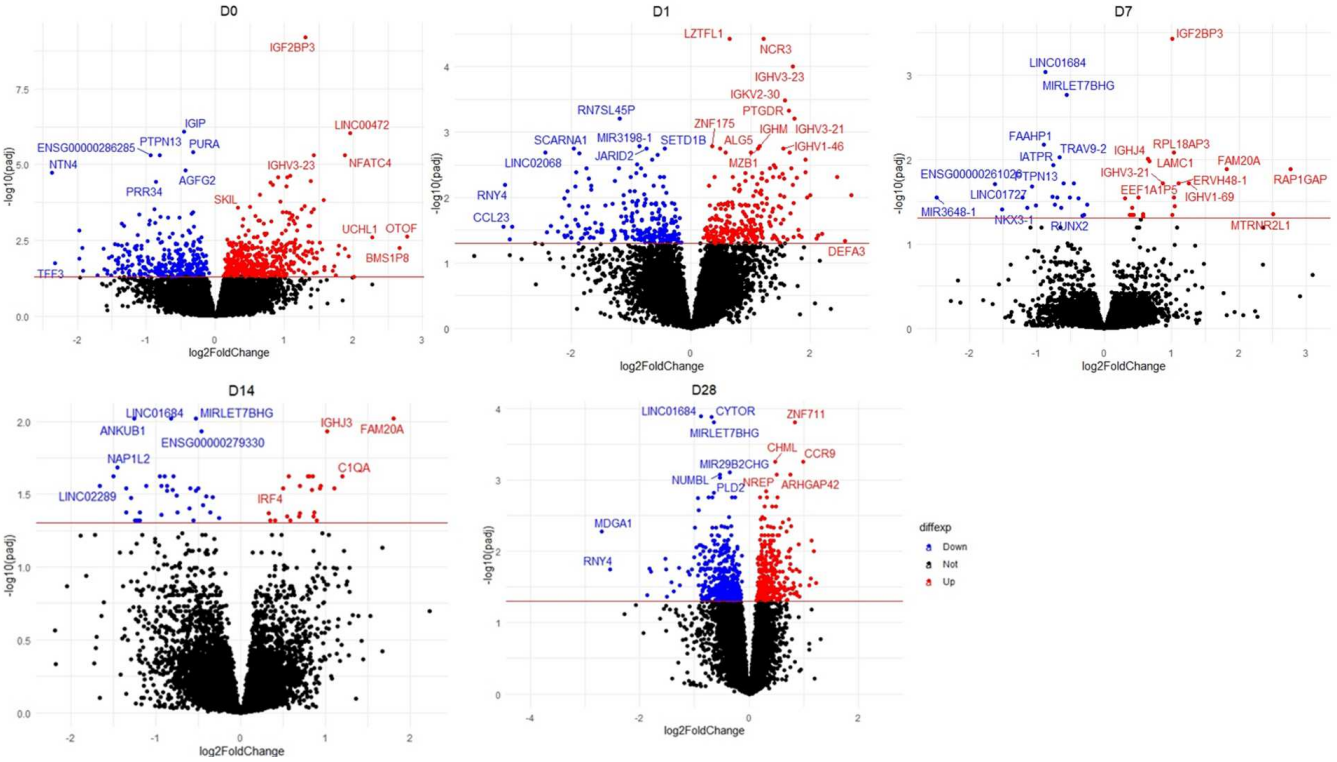


Figure 2.13: *DEGs volcano plot. These volcano plots show upregulated, downregulated and unaffected genes in red, blue and black respectively. The value on X axis is the log2FoldChange, while on Y axis there is the* $-log10(padj)$.

The enrichment analysis with tmod confirmed the activation of blood transcription modules related to B cells and immunoglobulin (LI.M47.0, LI.M47.1, LI.M47.2, LI.M47.3, LI.M156.0) at all timepoints. At day 1 for younger children we found also an activation of modules of NK cells activation (LI.M6.0,LI.M7.2,LI.S1) and T cells division (LI.4.5,LI.4.6, LI.4.11). Modules related to complement pathways (LI.40.0, LI.M112.0) are enriched at day 14. At day 28 there is a downregulation in monocytes associated modules (LI.M11, LI.S4), suggesting a decrease in innate immune response. Enrichment analysis results are visible in Figure 2.14

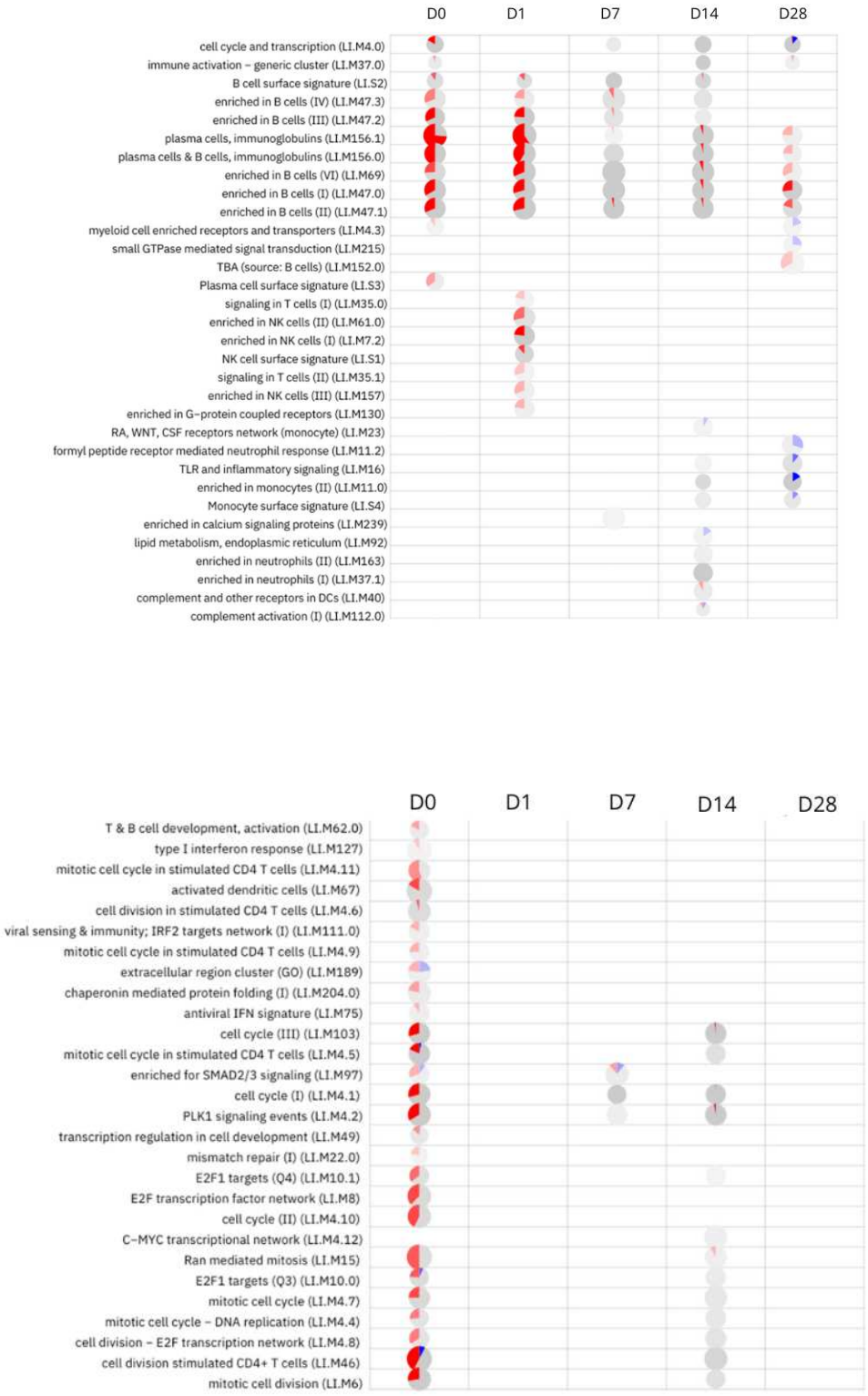


Figure 2.14: *Activation of blood transcription modules (BTMs) in child and young group*

The Cemitool GSEA and Over-expression analysis confirmed that the immune response is similar for the two groups, that presented similar activation for modules related to antivi-

ral response and interferon (M1), neutrophils activation (M2) and T cells (M4),whereas the module involved in B cells activation (M6) was the only one with significant differences in expression (Figure2.15.

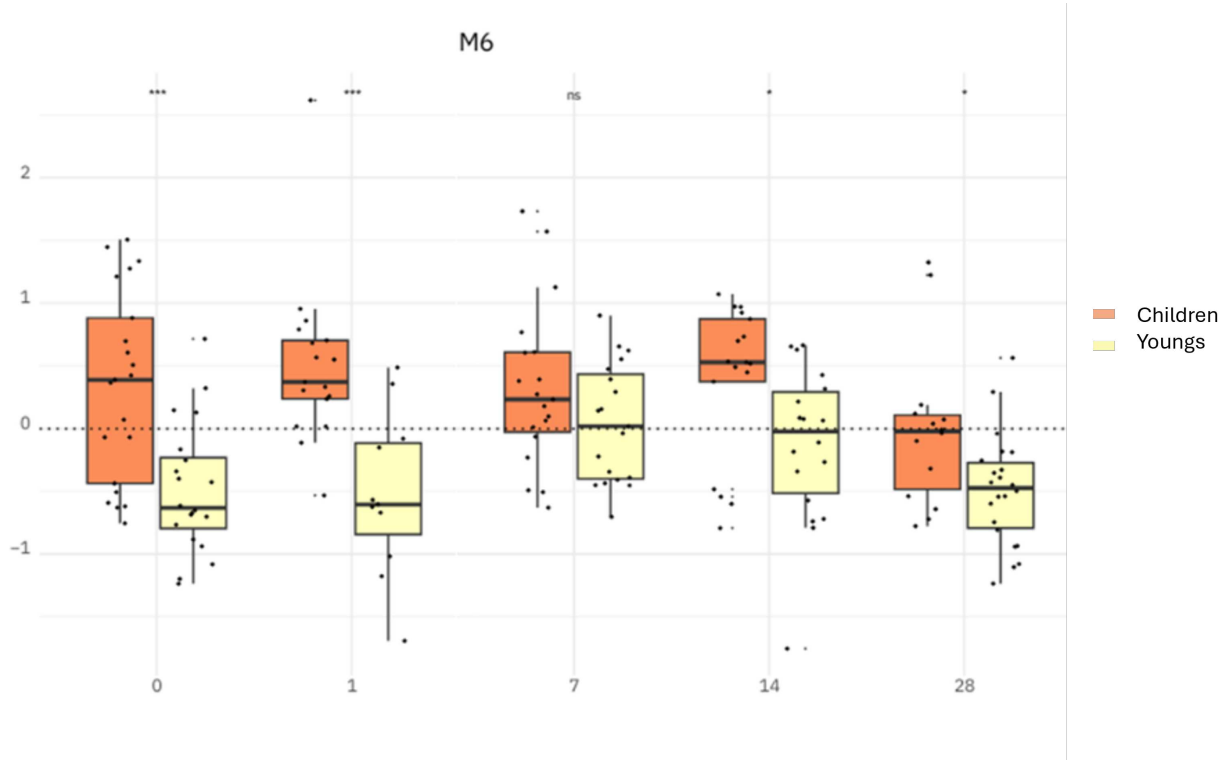


Figure 2.15: *Activation of M6.The picture shows the behavior of B cells related module during the entire time of our study. The comparisons between the two groups are significant at all timepoints except for day 7, they show that children present higher level of plasma cells and B cell enrichment since day 0.Children group is represented in orange, while youngs in yellow.*

A deeper analysis was conducted among younger children, who was further divided into two groups according to their ages; we had infants of 1/2 years and children from 3 to 5 years.DGE, enrichment and co-expression analyses were performed in order to define DEGs and related biological pathways. We compared gene expression among the two groups at each timepoint, using children as reference. The highest number of DEGs was detectable at day 0 (n=957), while about 40 genes were appreciable at day 7 and 28. On day 1 we had only two DEGs that are TRBV28 (upregulated) and MYO7A(downregulated), while on day 14 only a downregulation of NHS gene was visible (table 2.8).

Infants vs Children			
	Upregulated	Downregulated	Unaffected
D0	350	607	18.281
D1	1	1	16.775
D7	21	17	18.850
D14	0	1	18.372
D28	14	24	17.689

Table 2.8: DEGs table.The tables shows the numbers of significant (p.adj <0.005) DEGs in the comparison between infants and children

Tmod enrichment analysis showed up regulated BTMs related to immunoglobulins and plasma cells (LI.M156.1,LI.M156.0) at day 0

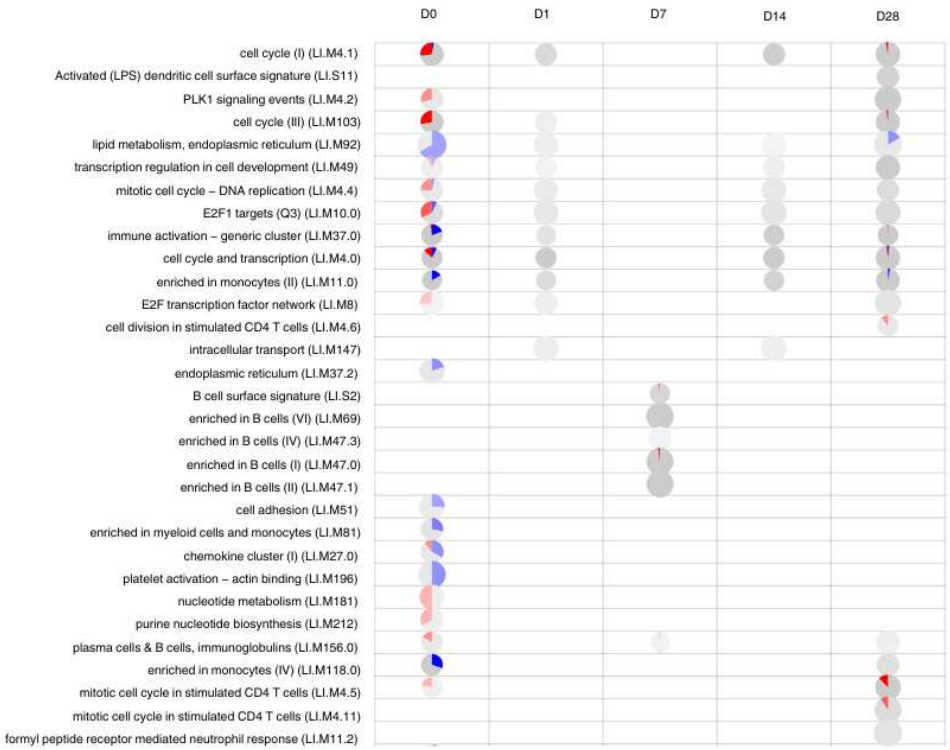


Figure 2.16: *Activation of blood transcription modules (BTMs) by rVSV-ZEBOV- GP in infants vs children comparison.*

We also investigated the impact of gender of vaccine efficacy. No important variations emerged for the PCA for the gender (Figure 2.17 of the two groups, suggesting it did not affect the efficacy of the vaccine.

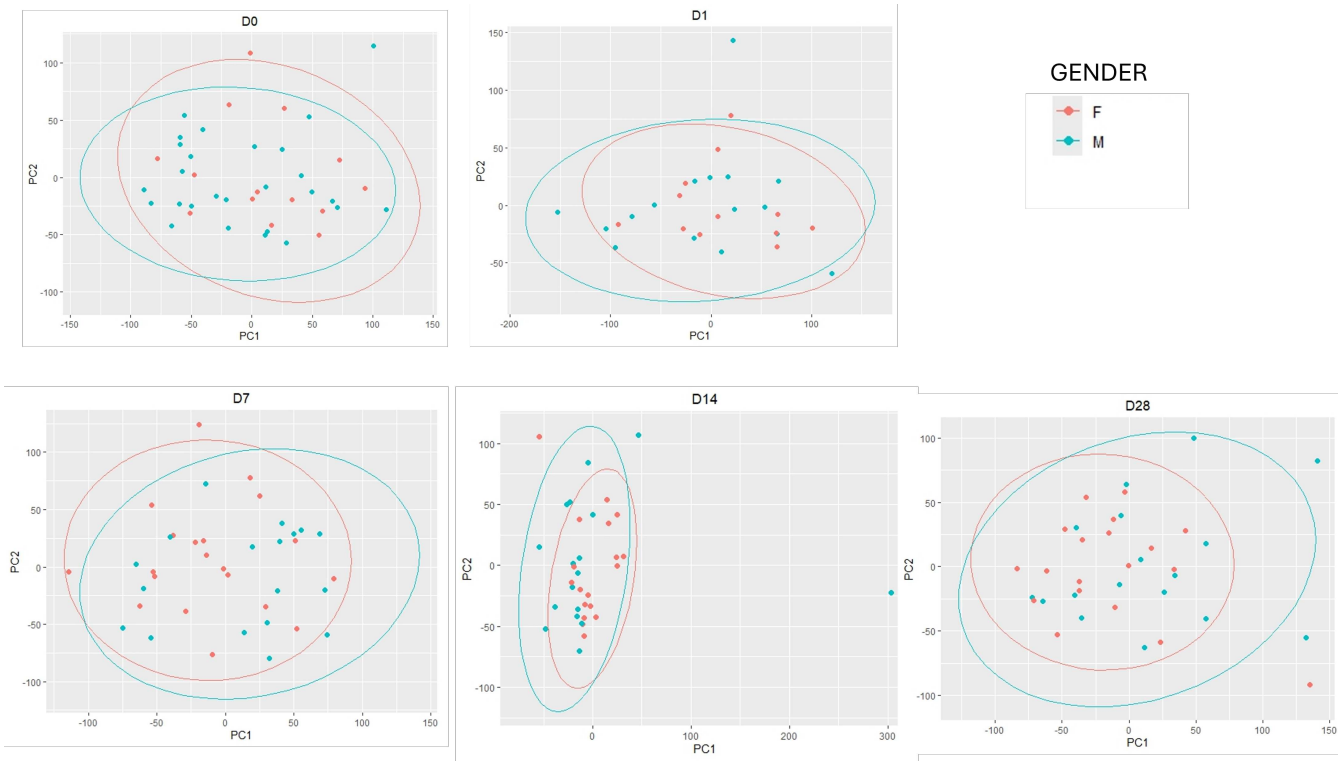


Figure 2.17: *PCA plot for gender. The figure shows PCA plots for all timepoints of our study. Female are represented in red, while male in blu. Samples are not clearly separated depending on their gender, meaning that it does not affect gene expression.*

The differential gene expression analysis returned a lower number of DEGs than the other analyses. DEGs for each timepoint are reported in table 2.9

Female vs Male			
	Upregulated	Downregulated	Unaffected
D0	35	27	20.698
D1	13	21	18.297
D7	32	29	17.697
D14	22	22	20.283
D28	52	34	20.131

Table 2.9: DEGs table.The tables shows the numbers of significant (p.adj <0.005) DEGs in the comparison between Female and Male

At all timepoints obviously there were differences in expression of gender related genes. No changes in DEGs involved in immune response were discovered at day 1. At day 7

some genes linked to immunoglobulin production were upregulated in female (IGHJ4,IGHV1-69,IGHV3-21), but they were not present in significant DEGs at day 14 and 28. The enrichment analysis with tmod surprisingly did not show modules related to immunoglobulin and B cells activation at day 7, but they appeared at day 28 (LI.M47.0, LI.M47.1, LI.M47.2, LI.M47.3, LI.M156.0), as visible in Figure 2.18.

Figure 2.18: *Activation of blood transcription modules (BTMs) female and male groups*

CEMITool GSEA and over-representation analysis assessed the presence of significant differences in B cells related module (M9, Figure) at day 14 and 28.

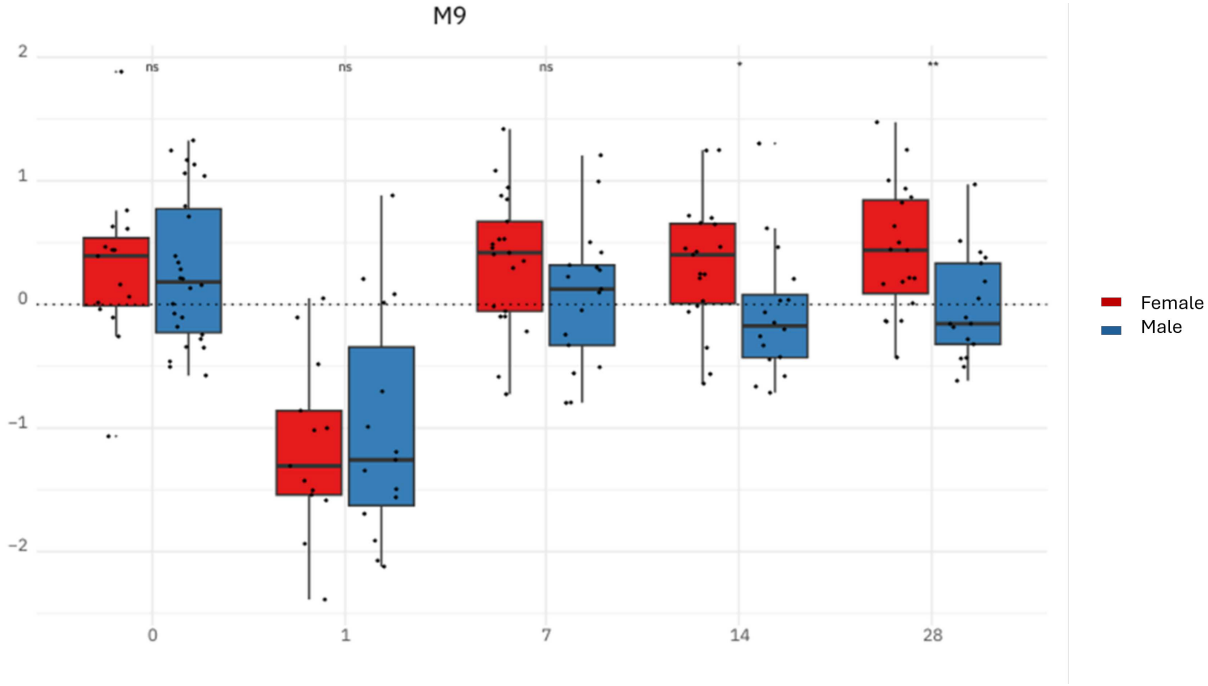


Figure 2.19: *Activation of M9.The picture shows the behavior of B cells related module during the entire time of our study. The comparisons between the two groups are significant at day 14 and 28. Female group is represented in red, while male in blu.*

2.4.1 Correlation analysis

Once data were collected about gene expression, gender, age and antibody titers, we performed a final complete analysis including correlation between gene expression and antibody titers and between tmod blood enrichment module activity scores and antibody titers. For each timepoint, we selected samples whose data were present for the respective timepoint and the baseline, we performed again differential gene expression analysis with DESeq2, enrichment analysis with tmod and ClusterProfiler and a correlation analysis with the psych package.

D1 vs D0

Here we had a total of 17 subjects (9 Not responders and 8 responders, 2.10).

	Upregulated	Downregulated	Unaffected
D1	6.670	7.322	5.413

Table 2.10: **DEGs table**.The tables shows the numbers of significant (p.adj <0.005) DEGs in the comparison between D1 and the baseline

The DGE analysis revealed 13.530 significant genes, among them, we selected 3261 genes with |log2 fold change| >=1 and padj <0.05 including 350 immunity related genes identified through the ImmPort database (2.20

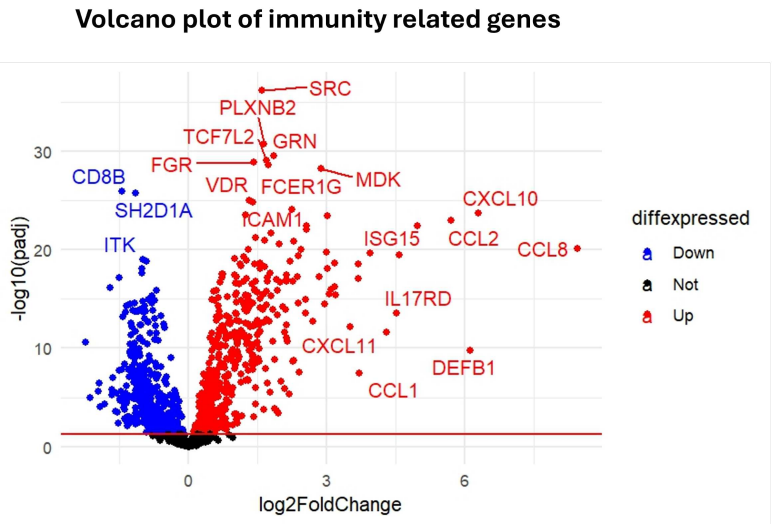


Figure 2.20: *Volcano plot representing genes related to immunity*

These 3261 genes were used for the enrichment analysis with Cluster Profiler, that allowed us to perform an over-representation and GSEA with both KEGG and Gene Ontology database (2.21. Results confirmed the activation of an innate immune response mediated by cytokine production.

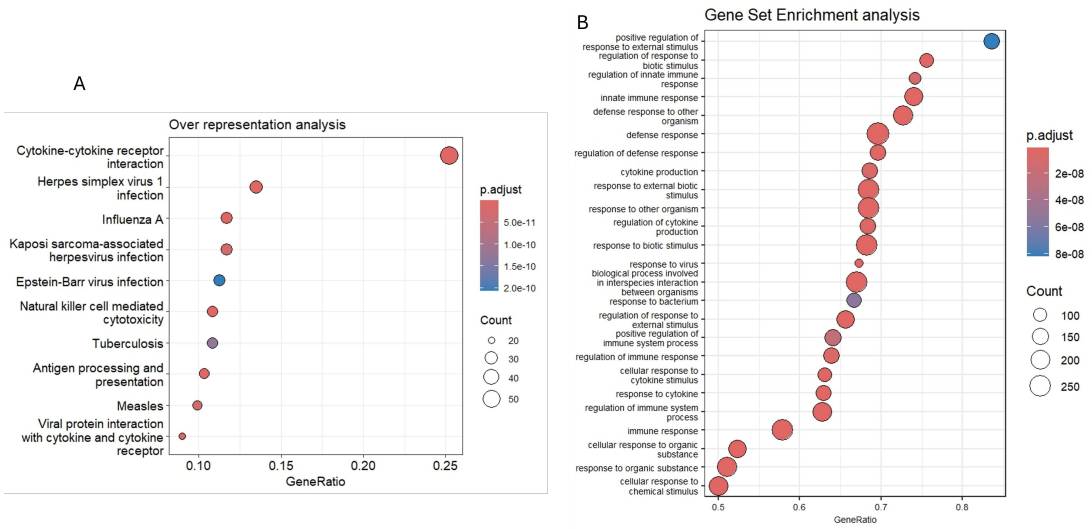


Figure 2.21: *GSEA analysis.(A) Over-representation analysis performed with KEGG database showed the enrichment of cytokine-cytokine receptor interaction (59 genes),natural killer cells mediated toxicity (23 genes), antigen processing and presentation (23 genes) and viral protein interaction with cytokine and cytokine receptors (21 genes). (B) GSEA performed with GO showed the enrichment of biological pathways related to innate immune response activation and defense regulation.*

Genes involved in biological pathways pinpointed by KEGG were correlated with antibody titers at day 28, 180 and 365 after vaccination, among them only CCL1 showed a significant (padj<0.05) positive correlation, while KURC3 and IL10 showed a negative correlation (2.22.

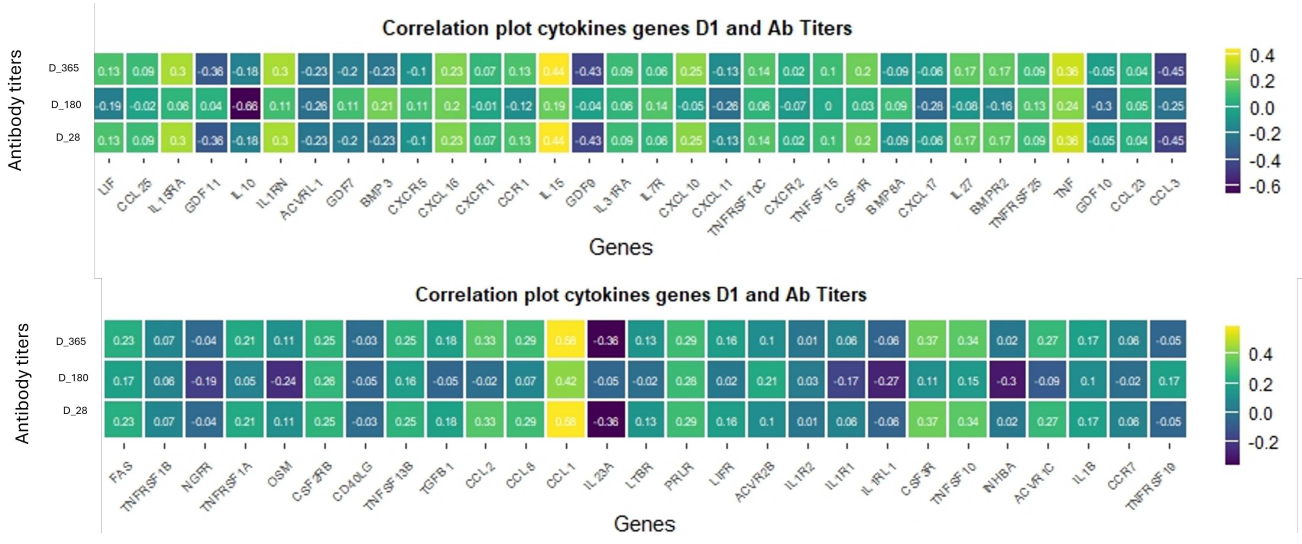


Figure 2.22: *Correlation plot.(A)Cytokine-cytokine receptor interaction biological pathway genes, here CCL1 gene shows a significant positive correlation with all antibody titers.(B) Antigen processing and presentation pathway genes, here KURC3 gene shows a negative and significant correlation with antibody titer at day 28 and 365.*

We then correlated BTMs activity for each subject with antibody titers. Unfortunately,no correlation was found with antibody titers at 180 and 365 days after vaccination, while correlations were identified with titers measured at day 28. Only one module was identified, related to NK cells surface signature (LI.S1) with a negative correlation coefficient. We also investigate the foldchange and the absolute expression of each single gene inside this module through the creation of heatmap plot (2.23).

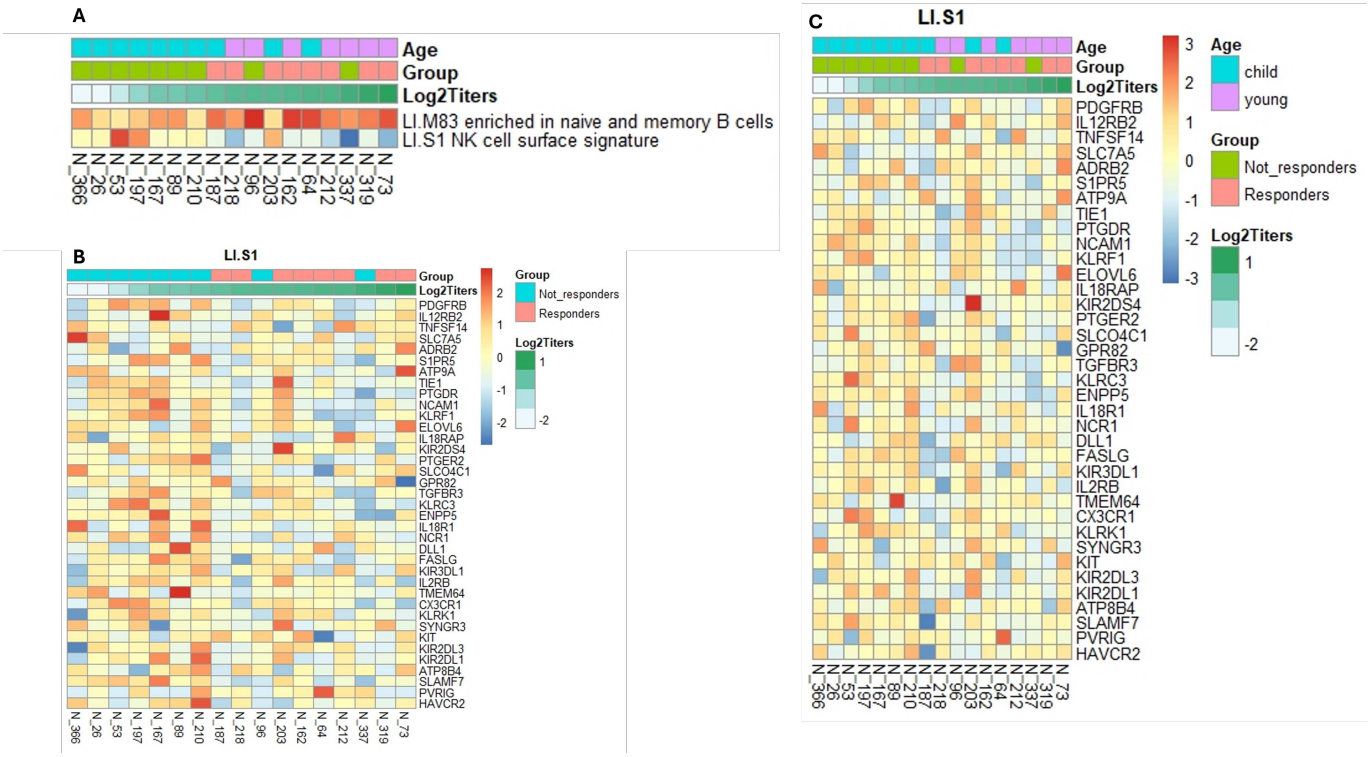


Figure 2.23: *Heatmap plot.(A)This heatmap shows the foldchange of the LI.S1 module for all 17 subjects. The negative correlation between the module activation and the increase of antibody titers is clearly detectable. The plot also highlight how the majority of responders subject belong to the older group of patients.(B) This heatmap shows the value of expression of genes inside LI.S1 module, it is slightly visible how the expression decrease with the increasing of antibody titers.(C) This heatmap shows the foldchange of all genes inside the module.*

D7 vs D0

Here we had a total of 21 subjects (7 Not responders and 14 responders, 2.11).

	Upregulated	Downregulated	Unaffected
D7	9	9	2.160

Table 2.11: DEGs table.The tables shows the numbers of significant (p.adj <0.005) DEGs in the comparison between D7 and the baseline

The DGE analysis revealed 20.178 significant genes, among them, we had IFI27, IGHV3-48,IGHV3-21, IGHM, IGLV2-8 and SIGLEC1 that are upregulated (Figure 2.24).

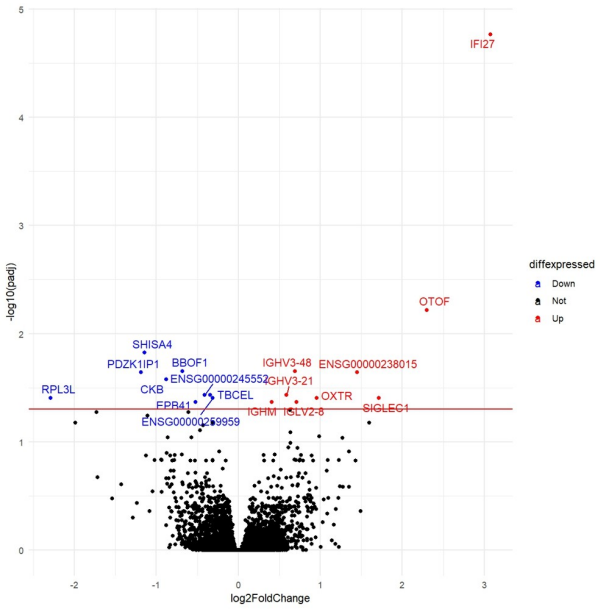


Figure 2.24: *Volcano plot representing DEGs at D7*

As we did for D1, we computed the z-score of gene expression and the respective fold-change and we used it to make correlation of all 18 DEGs with antibody titers.We obtained significant positive correlation between SIGLEC1 and OTOF and antibody titers at day 180 and a negative one between IGLV2-8 with titer at day 365(Figure 2.25).

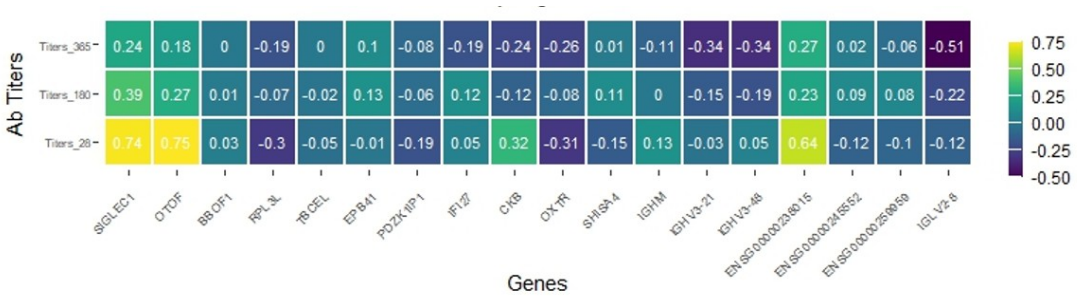


Figure 2.25: *Correlation plot between DEGs at day 7 and antibody titers*

The correlation between enriched blood transcription module and antibody titers did not show any significant results.

Day 14 vs Day 0 and Day 28 vs Day 0

DGE analysis revealed only one significant gene,IFI27, that is still upregulated on day 14. No DEGs were detected on day 28.

2.5 Conclusion

Whole-blood transcriptomic data showed rVSV-ZEBOV-GP elicits an early innate immune response within 24 hours and a following adaptive signature. This temporal model is coherent with the common kinetics of vaccine responses, which is usually characterized by a first inflammatory signaling and a subsequent emergence of B-cell and immunoglobulin. Among the activated transcript at day 1, there were genes primarily related to activation of interferon pathways, dendritic cells and macrophages activation and viral sensing (e.g IFITM3,IFI6, RUFY4,OASL,,CCRL,DDX58). TMOD enrichment highlighted up-regulation of modules linked to dendritic cells activation and antiviral interferon signature alongside down-regulation of T-cell–associated modules, a pattern frequently observed in whole-blood systems vaccinology [53], [54]. Among the most significantly upregulated genes at day 1, IFI27 and SIGLEC1 showed one of the wider fold-change.The upregulation of these two genes has been reported in several systems vaccinology studies in which a predominant IFN-I induced immunity was appreciable soon after vaccination. [55]. IFI27, togheter with IFI44, are able to prevent an excessive immune response, since exacerbated innate immune responses can be deleterious to the host, playing a regulatory role [56].It encodes a 122-amino acid hydrophobic protein that interact with nucleic acids and RIG-I, leading to impaired RIG-I activation. SIGLEC1 is a type I transmembrane protein expressed only by a subpopulation of macrophages and is involved in mediating cell-cell interactions. Cotrary to IFI27, it has been reported to be a marker for the early response to IFNs and it is also involved in the circulation of activated dendritic cells[57]. At day 7, the most prominently upregulated genes were related to immunoglobulin production (IGHV3-49,IGHV2-5, IGHV2-5, IGSF6). These finding are consistent with those assessed in other clinical trials which had previously tested vaccine efficacy and safety in adults and adolescents cohorts [38], [39], [13].

The comparison with VARILRIX vaccine platoform pinpointed that rVSV-ZEBOV-GP is responsible for a stronger immune response activation, as showed in the TMOD enrichment and CeMiTool over-expression analyses. The majority of activated BTMs resulted

to be upregulated in rVSV-ZEBOV-GP cohort, in particular those related to innate antiviral response, dendritic cells and monocytes activation and IFN signaling pathways (LI.M67, LI.M150, LI.M75, LI.M111.1).

The unsupervised classification with MClust allowed the arrangment of rVSV-ZEBOV-GP vaccinees into two groups (responders and not responders) according to their antibody titers. The comparison between these two classes asserted the activation of classical complement pathway through C1Q subcomplex ,which specifically binds IgG or IgM immunoglobulins complexed with antigens, forming antigen-antibody complexes on the pathogens surface. The relationships between the complement pathway and the innate immunity has been known from years, indeed B cells express complement receptor 2 (CR2; CD21), which specifically binds C3dg, a degradation product of C3 acting as an opsonin. When antigens are coated with C3dg, their simultaneous recognition by antigen-specific B cell receptors and CR2 facilitates their internalization and lowers the activation threshold of B cells, thereby enhancing antibody production [58].

We also investigated the role of gender and age on vaccine efficacy; confirming that gender didn't affect it. Conversely, the two age groups showed relevant dissimilarities, especially at day 0, 1 and day 28 after vaccination. Children between 1 and 5 years old presented enriched BMTs linked to innate immune response and B cells activation at day 0. Modules correlated with immunoglobulins and plasma cells were still detectable at day 28 after vaccination.

The foldchange analysis of both genes and module highlighted the presence of some correlation between changes in gene expression and antibody titers, in particular we detected on day 1 positive correlation with CCL1 and negative with IL10 and KURC3 genes, whereas blood transcription module LI.S1 (NK cells signature) presented a negative correlation with antibody titers at day 28. On day 7, SIGLEC1 and OTOF presented positive correlation with antibody titers at day 180, whereas IGLV2-8 was negatively correlated with titer at day 365. The correlation between high expression of SIGLEC1 molecule seven day after vaccination and the production of neutralizing antibody was already assessed for YF17D vaccination, where this molecule helped the recruitment and consequent activation of antigen presenting cells [57].No enriched module showed significant correlation with antibody. At day 14 only IFI27 was still upregulated in responder class, while at day 28 DEGs were not appreciable.

Overall, the data presented indicate that rVSV-ZEBOV-GP can induce a rapid and strong innate immune response, followed by the activation of humoral immunity. Our classification, based on antibody titers, detects transcriptional differences between responders and non-responders, among them the key one is the activation of the classical complement pathway via C1Q in the responders group. The vaccine efficacy seems to

be affected by the age of vaccinees, since younger children are characterized by B-cell activation and immunoglobulins production soon after the vaccination. However this transcriptome profile is detectable also at the baseline and it may influence in a profound way the real impact of vaccination. Further analysis are needed to verify the role of those gene expression profile on vaccine efficacy.

Chapter 3

Evaluation of iNTS-GMMA vaccine against invasive non-typhoidal Salmonella infection immune response through transcriptome analysis

Authors: $MariaNovedrati^{1,*}, ChiaraSonnati^{1,*}, SimoneLucchesi^{1}$
$HendersonZhu^{2,3}, LeilaGodfrey^{5}, BramaHanumunthadu^{2,3}, MaheshiN.Ramasamy^{2,4}$
$, DonataMedaglini^{1}, DanielO'Connor^{2,3}, FrancescoSantoro^{1,**}$
On behalf of Vacc-iNTS Consortium. Affiliations:

- 1 Department of Medical Biotechnologies, University of Siena, Siena, Italy
- 2 Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom
- 3 NIHR Oxford Biomedical Research Centre and Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom.
- 4 Oxford University Hospitals NHS Foundation Trust, Oxford, United Kingdom
- 5 Nuffield Department of Medicine, The Jenner Institute, University of Oxford, Oxford, United Kingdom
- 6 The members of VacciNTS Consortium are listed at the end of the paper.

*These authors contributed equally to this work. **corresponding author

Key words: iNTS; Salmonella infections; Bioinformatics; BCR, RNA sequencing, GMMA Vaccines.

3.1 Abstract

Background: Invasive non-typhoid Salmonella (iNTS) is a leading cause of morbidity and mortality in sub-Saharan Africa. The iNTS-GMMA vaccine, based on genetically engineered outer-membrane vesicles from Salmonella Typhimurium and Salmonella Enteritidis, has shown promising safety and immunogenicity, but the early transcriptional signatures associated with vaccine-induced immunity remain to be elucidated.

Methods: We conducted whole-blood transcriptional analysis on 31 healthy adults receiving either full-dose (40 µg O-antigen, n=15), low-dose (10.6 µg, n=4), or placebo (n=12). RNA sequencing was performed at baseline (D0), day 1 (D1), day 7 (D7) post-vaccination. Differential expression analysis (DESeq2), modular enrichment (Blood Transcription Modules), co-expression networks (CEMiTool), and BCR profiling (MiXCR) were employed to characterize immune responses

Results: Full-dose vaccination triggered rapid, dose-dependent innate immune activation within 24 hours, characterized by transcriptional changes (>3,600 differentially expressed genes) and upregulation of TLR2/4 signaling pathways and neutrophil markers including CD177. This early innate response moved to adaptive immunity by day 7, marked by B-cell module activation and selective upregulation of immunoglobulin genes. BCR repertoire analysis revealed over 4,000 unique clonotypes with significant clonal expansion (Gini index difference D0 vs. D7, $p<0.05$) and preferential usage of IGHV3-49. Notably, we identified four convergent BCR clusters shared across multiple vaccinees, including a dominant IGHV3-49-expressing cluster present in both vaccine groups but absent in placebo recipients

Conclusion: The GMMA-based iNTS vaccination triggers an early innate immune response followed by a dose-dependent adaptive immune reaction. Our study identifies early TLR activation as a key marker of immunogenicity and highlights specific BCR clonotypes, particularly IGHV3-49, as potential public clonotypes for further investigation. These findings contribute to a deeper understanding of vaccine-induced immunity and to identification of molecular signatures associated with vaccine response

3.2 Introduction

Invasive non-typhoidal Salmonella infections (iNTS) are caused by specific serovars of Salmonella enterica like S.Typhimurium, S.Enteriditis, S. Choleraesuis, S. Dublin. Transmission occurs through ingestion of contaminated food or water. Upon reaching the intestine, Salmonella invades enterocytes by utilizing a type III secretion system (T3SS)

encoded by Salmonella Pathogenicity Islands (SPIs). A subset of bacteria can also penetrate dendritic cells and intestinal macrophages, resulting in the spread of the pathogen through the bloodstream.

Macrophages and neutrophils represent the first line of defense through phagocytosis production of antimicrobial peptides and release of IFN-gamma. At later stages of infection, T cells and antibody response become the effective defense, in particular CD4+ T cells play a pivotal role in stimulating antibodies production and their loss is one of the primary correlates of susceptibility to NTS bacteremia. The presence of antibodies anti O-antigens is critical to prevent bacteremia and it is crucial in people with deficiencies in IL-12 and IL-23 production [28].

iNTS infections are always characterized by bacteremia without an obvious focus of infection. NTS bacteremia can lead to endocarditis with myocardial abscess and peri-valvular complication with a mortality rate of 50%. The presence of endovascular foci of infection can also cause mycotic aneurysms (more common than endocarditis), especially in elderly people. Visceral abscesses and spleen infection are also possible. NTS gastroenteritis can lead to joints and bones damage due to immune inflammatory response, that may result in osteomyelitis (more common in children than in adults). Salmonella can cross the blood-brain barrier, causing meningitis [27]. The spread of S.Typhimurium is promoted by the consumption of the C3 complement component.

Among all serovars, S.Typhimurium ST313 and S.Enteritidis ST11 are the leading cause of iNTS in Africa (90%). They are characterized by the presence of distinctive plasmids containing genes responsible for higher virulence and antibiotic resistance [32]. Genetic differences among NTS serovars are also responsible for peculiar pathogenesis, indeed strains belonging to the same clade can present increased replication rate in macrophages, or specific composition in the O-antigen.

Since the humoral response and T cell-mediated immunity are necessary to eradicate Salmonella infection, vaccine candidates are focused on the induction of this kind of immunity. Moreover, the presence of multidrug resistant serovars and the consequent failure in antibiotic treatment make necessary the development of an efficient vaccine against NTS. Several candidates, including live-attenuated, glycoconjugate and recombinant antigen-based vaccines are currently under investigation with target S. Typhimurium and S. Enteritidis. A promising strategy involves the use of Outer Membrane Vesicles (OMVs), which are blebs spontaneously released by Gram-negative bacteria and contain lipopolysaccharide (LPS) and O-antigen (OAg). Through the presence of different kind of molecules, these vesicles can self-stimulate various types of immunity [29].

Among OMV-based vaccines, one of the most interesting candidates is based on the Generalized Modules of Membrane Antigens (GMMA) technology, which consist in the

genetic manipulation of the contents of these vesicles to modify their reactogenicity and immunogenicity. The novel iNTS-GMMA vaccine platform, developed by GSK, contains S.Typhimurium and S.Enteriditis outer membrane vesicles [33], that present modifications including the disruption of the links between the outer membrane and either the peptidoglycan layer or the inner membrane to greatly increase release of GMMA [34] and modification of the lipid A structure of the LPS, by mutation of msbB and pagP genes, results in decreased reactogenicity [35]. The long-term immune response, tested in mice through several assays (ELISA, SBA, multiparametric flow cytometry analysis), demonstrated that the vaccine elicits a rapid production of specific serum IgM and IgG persisting for 10 weeks after a single immunization and seven months after boost. In addition, memory B cells, detected by ELISpot, were present 28 weeks after the booster dose [36].

A first clinical trial started in Oxford was launched to evaluate safety and immunogenicity in healthy adults, who received two different doses of the vaccine [33]. Our objective was to investigate the immune response induced by the GMMA-based vaccine. To this end, we conducted a transcriptomic analysis of whole blood to identify changes in gene expression associated with protective immunity.

3.3 Material and Methods

3.3.1 RNA extraction, library preparation and sequencing

Whole-blood RNA was collected in PAXgene® Blood RNA tubes (Qiagen) and extracted according to the manufacturer's instructions. Total RNA was quantified with Qubit™ (hsRNA kit, Thermo Fisher Scientific) and 50 ng per sample were used to prepare dual-indexed stranded cDNA libraries with the Illumina Stranded Total RNA Prep Ligation with Ribo-Zero Plus (Illumina) following manufacturer's instructions (Illumina).

Dual-indexed libraries were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter), quantified on a Qubit Fluorometer (Thermo Fisher) with the HS DNA kit, and assessed for size distribution on a BioAnalyzer DNA 1000 chip (Agilent Technologies, Germany), with an expected fragment size of 300–400 bp. Libraries were then diluted to 0.5 nM, pooled in a final volume of 100 µL, and loaded into an Illumina SP flow cell according to the manufacturer's protocol (Illumina, v1.0).. Sequencing reactions were performed for 200 cycles on the Novaseq 6000 platform (Illumina) to obtain 100 bases paired-end reads. FASTQ files from the two technical runs were processed independently and later batch-corrected with ComBat-Seq (batch = run) [59]. Quality assessment of sequencing reads was performed using FASTQC v 0.11.9 [40]. Adapter sequences and

low-quality bases were trimmed using Trimmomatic v0.39[41], with parameters SLIDINGWINDOW:4:5 and MINLEN:36. Cleaned reads were aligned to the human reference genome (GRCh38/hg38) using the STAR aligner v2.7.3a [60]. Gene-level quantification was conducted using HTSeqv2.0.2 [42].

Batch correction and count merging

All analyses were conducted in R v4.4.1..To correct for technical batch effects, raw gene-level counts were adjusted using ComBat-Seq (sva v3.50.0) with batch=run and group as the biological covariate (Timepoint and vaccine dose or placebo).

After batch correction with ComBat-Seq and merging of technical replicates, we proceeded with sample filtering based on sequencing depth. Samples with fewer than 3×10^6 total reads were excluded. A total of 75 high-quality samples were retained for downstream analysis.

3.3.2 Exploratory normalization and visualization

Raw counts were loaded into edgeR (v 4.4.1) as a DGEList. Genes with < 1 CPM in $<$ 10 samples were filtered out. Library-size bias was corrected with TMM (calcNormFactors), and counts were converted to Log2(CPM + 1). Principal component analysis was performed with prcomp (center=TRUE, scale.=TRUE), and UMAP embeddings were computed with the umap package (umap::umap(); default settings, seed = 123)[61].

Differential gene expression

The DESeq2 package [44]was used to perform differential expression analysis and multiple test correction, returning values of LogFC, and adjusted P values, using the Benjamini-Hochberg procedure. Genes exhibiting $|log2FC| \geq 0$ and an adjusted p-value < 0.05 were classified as differentially expressed.

Enrichment analysis

Enrichment analyses were performed using the Blood Transcription Modules (BTM) database[62] , significance was assessed by the CERNO test [63] from tmod package [46].

Module and network analysis

Modular gene co-expression network analysis was carried out with CEMiTool v1.20.0[64]. A single dataset comprising all the samples was transformed by variance-stabilizing trans-

formation (vst) in DESeq2 and used as input [65]. CEMiTool automatically determines the optimal soft-thresholding power (*beta*) and identifies co-expression modules in an unsupervised manner. For each module, enrichment in BTMs was evaluated using hypergeometric tests, module activity scores (mean z-scores of member genes) were calculated, and group comparisons of module activity were performed using the non-parametric Mann–Whitney test.

Immunome data extraction and IGHV-gene usage

Immunome (BCR) repertoires were extracted from bulk RNA-Seq FASTQ files using MiXCR (v3.0.13) with the functions Align, Assemble partial, Assemble, Export clones min, and Export clones under default settings ([66], [67]), [68]. The resulting clonotype tables were reformatted in R v4.4.1 to include, for each clone, CDR-H3 amino-acid sequence, V-gene, participant ID, time-point, and read count. Clonotypes were then aggregated by vaccine group and time point to compute the proportion of each IGHV gene in the total BCR repertoire per sample. IGHV-usage plots were generated with ggplot2 (v3.5.1)([69]).

BCR CDR-H3 single-linkage clustering

Clonotype tables prepared as above were imported into R and stratified by CDR-H3 length. For each length group, a pair-wise Hamming-distance matrix was computed using stringdistmatrix (method = "hamming"). Single-linkage clustering was then performed with a distance threshold of 1: the first unassigned sequence in each group was taken as an initial seed and assigned to cluster 1, and all sequences at Hamming distance < 1 from that seed were added to the same cluster. Each newly added sequence subsequently acted as an additional seed, and the procedure was iterated until no further neighbors were found. Clustering then proceeded to the next unassigned sequence (incrementing the cluster label each time) until all sequences were assigned. This approach yields non-overlapping clusters of highly similar CDR-H3s.

Statistical analysis

DGE analysis was performed following the DESeq2 function pipeline ([44]), p-values were adjusted for multiple testing with the Benjamini-Hochberg method, and significant genes were selected as p.adjust$<$0.05. The statistical significance of CEMiTool scores was assessed with Mann-Whitney U test. All tests were two-tailed.

3.4 Results

In this study, we conducted a whole-blood transcriptional analysis (WBTA) of 31 healthy adult volunteers: 15 received the invasive non-typhoidal Salmonella GMMA vaccine at full dose (FD; 40 µg O-antigen), 4 received a low dose (LD; 10.6 µg O-antigen), and 12 received placebo (Alhydrogel alone). Gene expression analysis was performed at four time points: day 0 (prevaccination), day 1, day 7, and day 14 after vaccination. The experimental design is summarized in Figure 3.1.

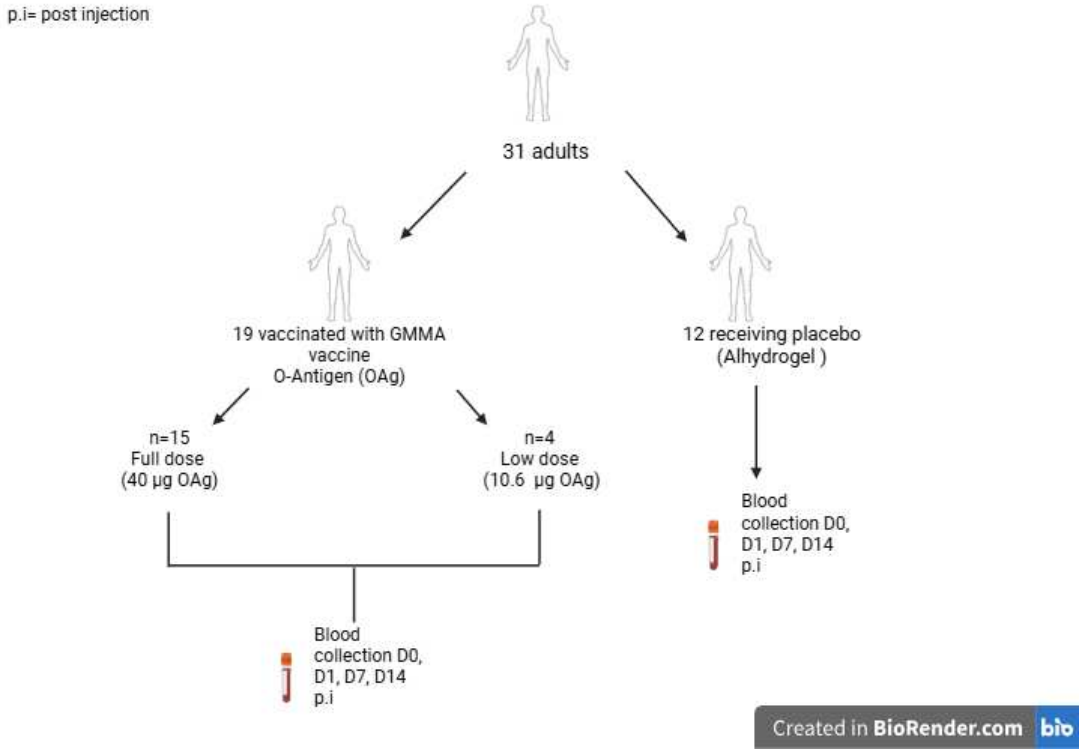


Figure 3.1: *Study design.*

3.4.1 Batch effect correction with ComBat-Seq

Each library was sequenced in duplicate on separate NovaSeq SP flow cells. The initial run on a NovaSeq 6000 SP flow-cell produced a median of about 2.4×10^6 paired-end reads per sample. Because this read depth was considered inadequate for reliable transcriptomic analysis, the same libraries were re-sequenced on a second SP flow-cell under identical run parameters, yielding a further 3.0×10^6 reads per sample. To assess batch effects, we performed PCA on log-transformed raw counts, revealing separation by sequencing run 3.2A. Batch correction was applied using ComBat-Seq (sva v3.50.0), treating run as batch and group (timepoint × treatment) as biological condition (3.2A,B). Technical replicates were summed post-correction to yield a final count matrix for anal-

ysis.Across both runs, this corresponded to a combined median sequencing depth of 5.4×10^6 paired-end reads per sample.

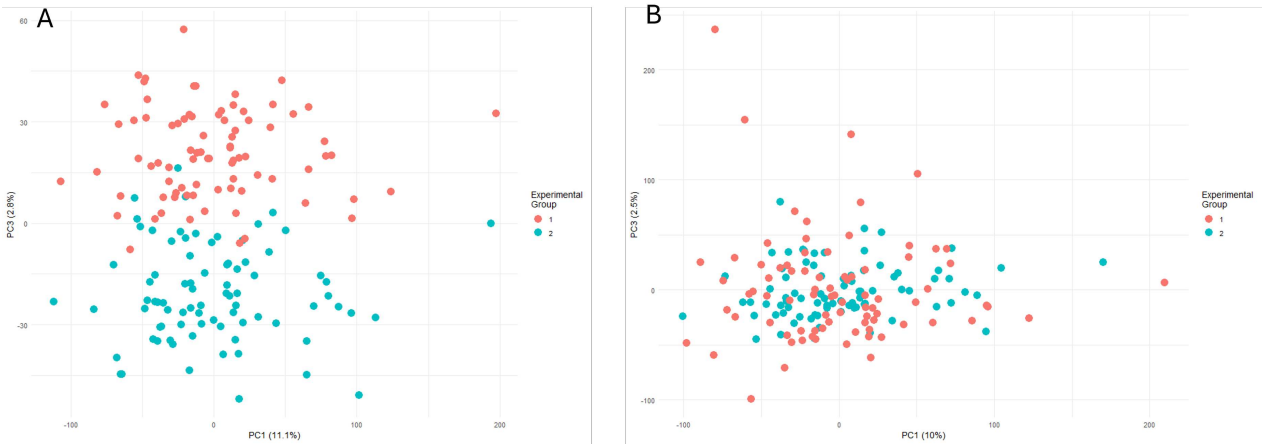


Figure 3.2: *Principal component analysis (PCA) of log-CPM normalized Salmonella samples before and after batch correction with ComBat-Seq. (A) Before batch correction: PC1 explains 11.0% of the variance and PC2 explains 6.5%. Samples, colored by experimental group, show partial clustering driven by batch. (B) After batch correction: Samples are mixed along PC1 and PC2, indicating effective removal of the batch effect while preserving the biological grouping.*

3.4.2 Principal component and Uniform Manifold Approximation

Principal component analysis (PCA) of all the volunteers across all timepoints revealed that samples cluster primarily along the first principal component (PC1), which accounts for 14.8% of the total variance; PC2 explains an additional 7.8% (3.3A). A subtle shift along PC1 is apparent at Day 1 post-vaccination.

To further explore sample relationships at each timepoint, we applied Uniform Manifold Approximation and Projection (UMAP). Three-dimensional UMAP projection at Day 1 demonstrates clear separation between placebo, low-dose, and high-dose vaccine groups (3.3B), suggesting an early, dose-dependent transcriptional response.

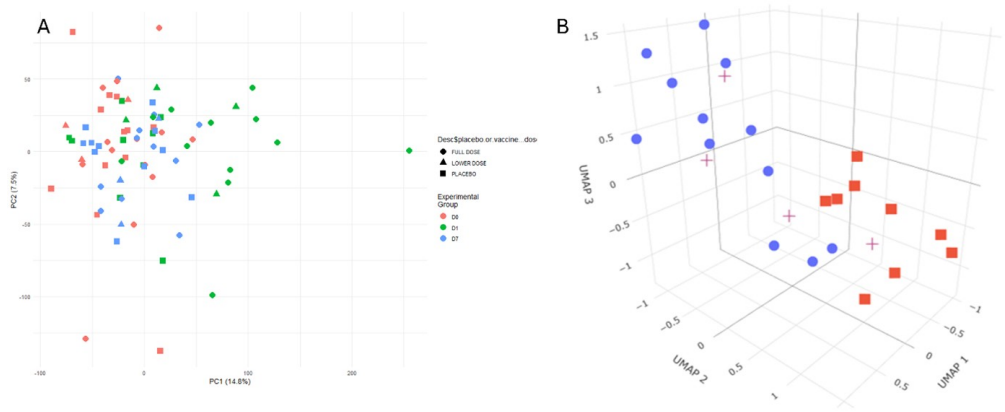


Figure 3.3: *Principal component analysis (PCA) and Uniform Manifold Approximation and Projection(Umap) of log-CPM normalized Salmonella samples (A) 2D PCA of all samples. Shapes denote timepoints, colors denote dose group PC1 (14.8% variance) and PC2 capture the main axes of transcriptional variation, although no clear dose-dependent clusters emerge when all timepoints are considered. (B) 3D UMAP of Day 1 samples only (UMAP-1, UMAP-2, UMAP-3), showing improved visual separation of dose groups at this time point.*

Differential gene expression analysis

Differential gene expression (DGE) analysis was conducted using Day 0 as baseline. In the full-dose group, the highest number of differentially expressed genes (DEGs) was observed at Day 1 (3,628), predominantly related to innate immunity (e.g., TLR1, TLR2, TLR5, IFNGR1, IL1B, IFITM2). Day 7, only 7 genes remained significantly modulated, six of which were upregulated immunoglobulin genes IGHV3-49, IGHV3-15, IGHV2-24, IGHV3-74, IGLV2-11, IGKV-30 3.4. In contrast, the low-dose group showed a markedly reduced transcriptional response, with 246 DEGs at Day 1 and only one gene (XIST) at Day 7. The number of DEGs is shown in table 3.1.

Group	Day Contrast	Upregulated	Downregulated	Unaffected
Full dose				
	D1 vs D0	2,326	1,302	13,277
	D7 vs D0	6	1	16.898
	D7 vs D1	628	1,743	14,345
Low Dose				
	D1 vs D0	246	0	16,559
	D7 vs D0	0	1	16,904
	D7 vs D1	0	1	16,904

Table 3.1: Number of Differentially Expressed genes using using both D0 and D1 as baseline for low dose group

DEGs were visible in the volcano plots following (Figure 3.4).

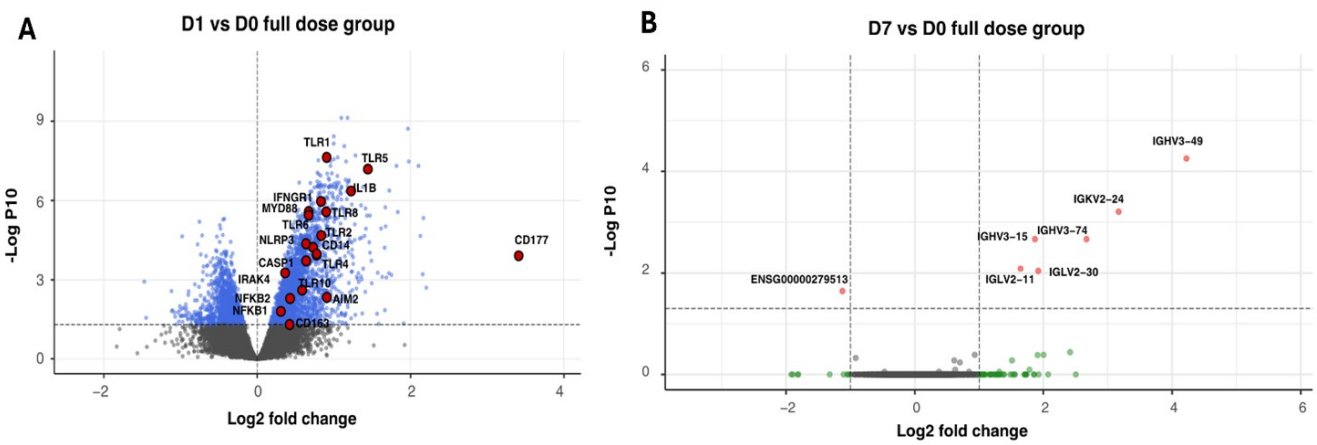


Figure 3.4: *(A) Volcano plot showing genes differentially expressed between day 1 (D1) and baseline (D0) in the full-dose group. Red dots indicate significantly upregulated genes, including several related to innate immune signaling. (B) Volcano plot showing differential expression between day 7 (D7) and baseline (D0 A few immunoglobulin genes were significantly upregulated*

TMOD enrichment analysis

Enrichment analysis was conducted starting from DESeq results obtained from the comparisons with the baseline for both cohorts (full-dose and low-dose) and for pooled samples. Modules enriched for pooled samples resembled those found for the full dose cohort, indeed at day 1 both were characterized by upregulated modules regarding neutrophils activation (LI.M163,LI.M37.1,LI.M11.2), antigen presentation (LI.M95.1), monocyte activation (LI.S4, LI.M11), TLR and inflammatory signaling (LI.M16), myeloid cells (LI.M81,

LI.M4.3) and downregulated modules related to T cells (LI.M52,LI.M7.1). Other modules were linked to cell migration (LI.M109), blood coagulation (LI.M11.1) cell adhesion (LI.M117) and other biological functions. In the low dose cohort, only a few modules associated with monocyte,neutrophils and TLR were enriched (LI.S4, LI.M37.1 and LI.M16). On day 7 a mild upregulation of modules linked to plasma cells and immunoglobulin was observed in the full dose group. Results of tmod enrichment analysis were visible in Figures 3.5 and 3.6.

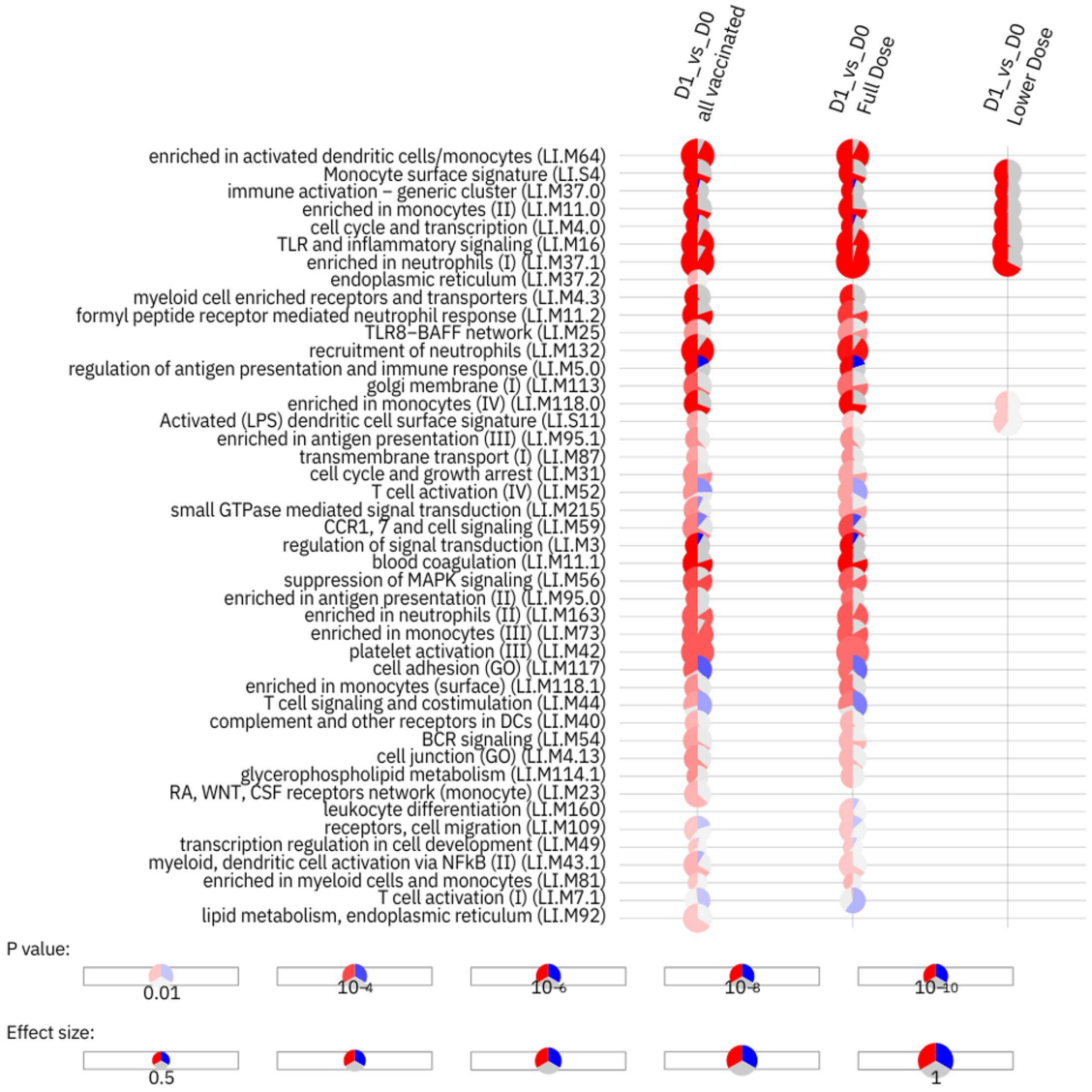


Figure 3.5: *Activation of blood transcription modules (BTMs) by GMMA vaccine*

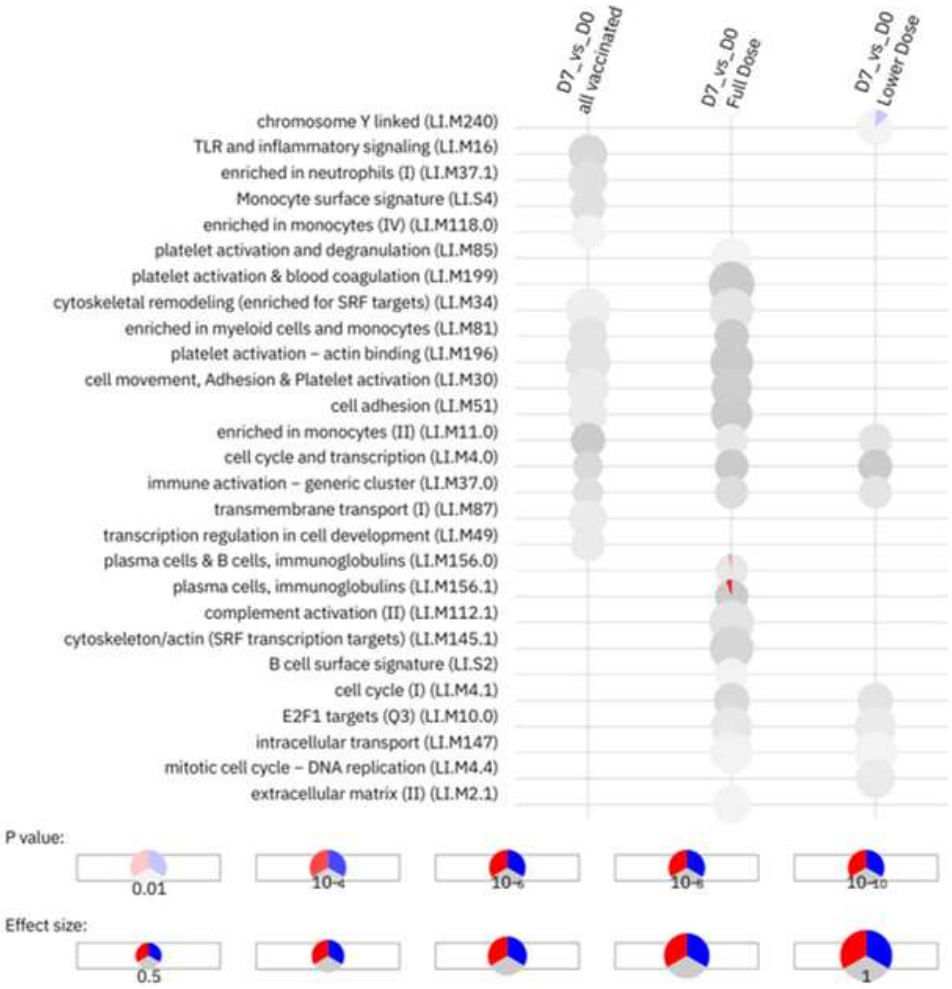


Figure 3.6: *Activation of blood transcription modules (BTMs) by GMMA vaccine.*

Figure 3.5 and 3.6 .Each column represents a comparison between a day post vaccination versus the baseline. Activation of modules was tested using the FDR-ranked lists of genes generated by DESeq2 generalized linear model fitting and applying the CERNO test. Rows indicate different BTMs, which were significantly (FDR < 0.05) activated. Each module is represented by a pie in which the proportion of significantly upregulated and downregulated genes is shown in red and blue, respectively. The gray portion of the pie represents genes that are not significantly differentially regulated. The significance of module activation is proportional to the intensity of the pie, while the effect size (area under the curve) is proportional to its size.

Given the consistent enrichment of module LI.M16, we further examined key TLR genes (TLR4, TLR2). Expression levels were significantly elevated in vaccinated groups compared to placebo at Day 1, indicating early innate immune activation. These TLRs recognize pathogen-associated molecular patterns (PAMPs), initiating inflammatory cascades and promoting cytokine and interferon production (Figure 3.7 , Figure 3.8

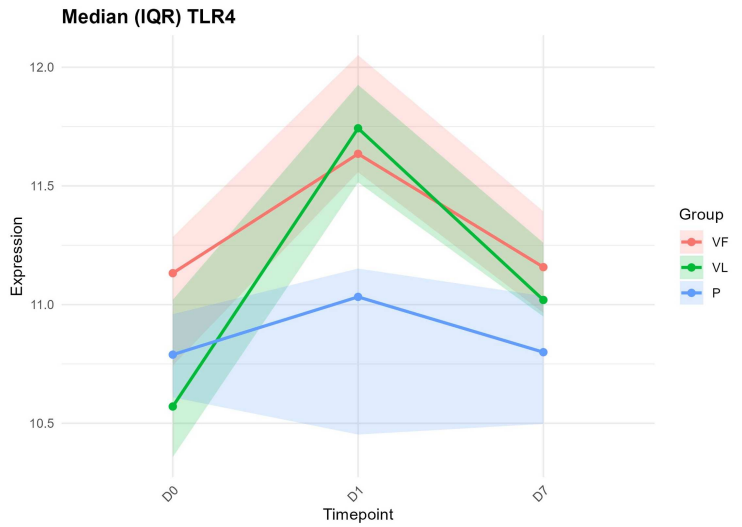


Figure 3.7: *Expression of TLR4 gene*

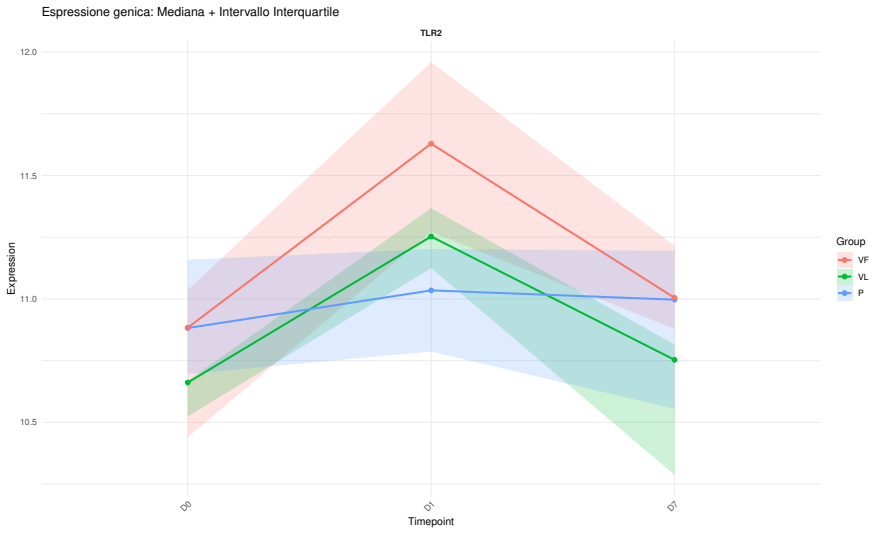


Figure 3.8: *Expression ofand TLR2 gene.*

Figure 3.7 and 3.8. The median expression was calculated for full dose, low dose and placebo groups. Vaccinated with full dose (VF) are represented in red, low dose (VL) in green and placebo (P) in blue.

3.4.3 Co-expression analysis

Cemitool GSEA and over-expression analyses highlighted significant differences between placebo and all vaccinated samples in the innate immune response (M1) activation and B cells (M6). M1 expression peaked at Day 1 and returned to baseline by Day 7. Although only a few DEGs were identified at day 7, CEMITool revealed significant changes in gene

co-expression, indicating B-cell activation at the network level.The trends of M1 and M6 are shown in Figure 3.9

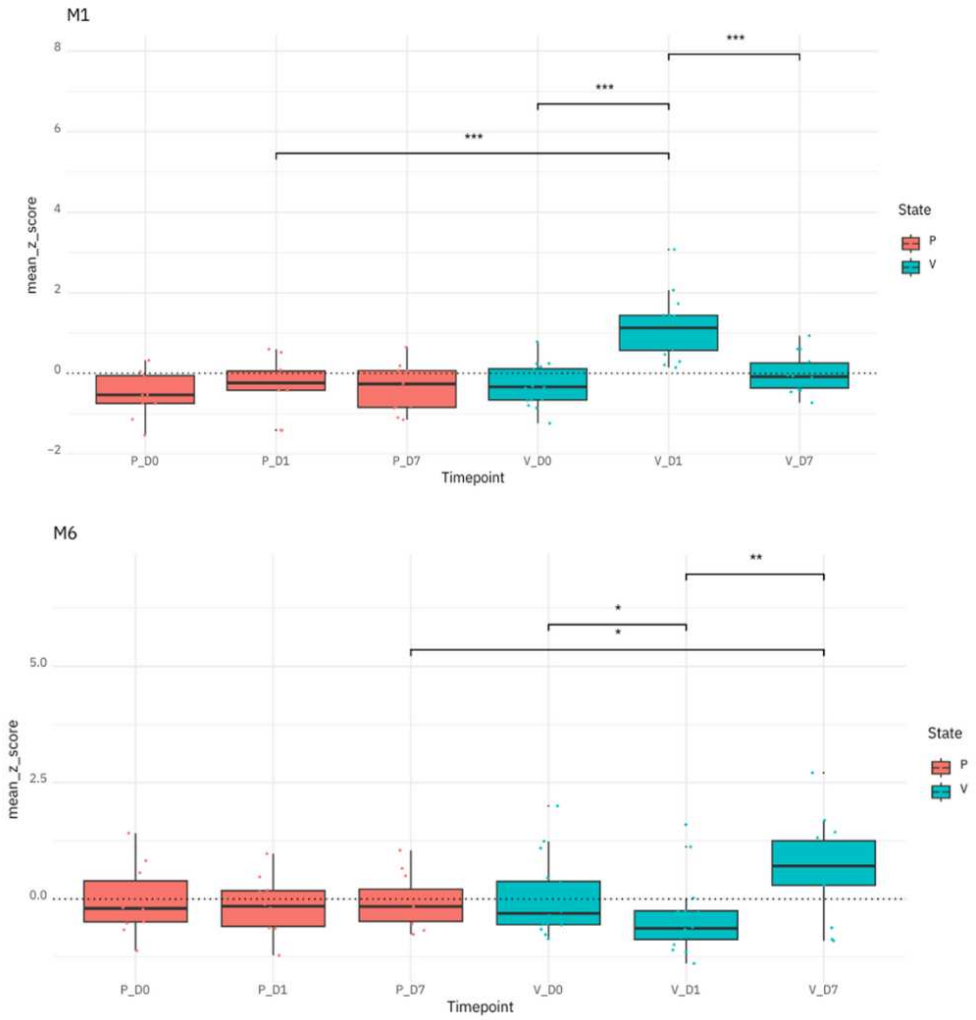


Figure 3.9: *CEMITool boxplot. Mean z-score of module expression is described here for each timepoint and for the two groups. Placebo (P) are represented in orange, while vaccines (V) in blue. The plot assesses the significant variance of expression for M1 (general innate immune response) and M6 (plasmacells activation) between the two groups.*

3.4.4 BCR clonal expansion and gene usage after vaccination

Differential gene expression analysis revealed the upregulation of immunoglobulin-related genes in vaccinated compared with non-vaccinated participants at day 7 (D7) relative to baseline (D0). To further characterize the B-cell receptor (BCR) repertoires, we retrieved and filtered CDR3 heavy chain (CDR H3) sequences from MiXCR-processed data. Across all time points, 4,214 unique BCR clonotypes were identified in full-dose

vaccine recipients and 390 in low-dose recipients. Among full-dose participants at day 7, 16.1% of assigned reads mapped to clonotypes using IGHV3-49, followed by IGHV3-21 (9.7%) and IGHV3-74 (5.6%). In contrast, low-dose participants at day 7 predominantly used IGHV3-23 (13.4%), IGHV1-18 (4.9%), and IGHV3-74 (5.0%) (Figure 3.10 A). Based on the upregulation of IGHV genes and changes in gene usage, we next assessed BCR repertoire diversity by calculating the Gini index for each repertoire in full- and low-dose vaccine recipients, using clonotype counts derived from MiXCR-processed repertoires. The mean Gini index did not differ significantly between the two groups (full-dose: 0.35 $\pm$ 0.14; low-dose: 0.25 $\pm$ 0.02; $p = 0.14$, two-sided Wilcoxon test) (Figure3.10 B). We then asked whether repertoire inequality changed from baseline (D0) to day 7 (D7) within each group. In the full-dose group, the Gini index increased from D0 to D7 in 9/11 participants, reaching statistical significance (D0: 0.24 $\pm$ 0.05; D7: 0.35 $\pm$ 0.14; $p = 0.045$, two-sided paired Wilcoxon test) with one participant showing a pronounced decrease that influenced the paired comparison. No significant change was observed in the low-dose group ($p = 0.5$) or placebo recipients ($p = 0.79$) (Figure3.10 C)

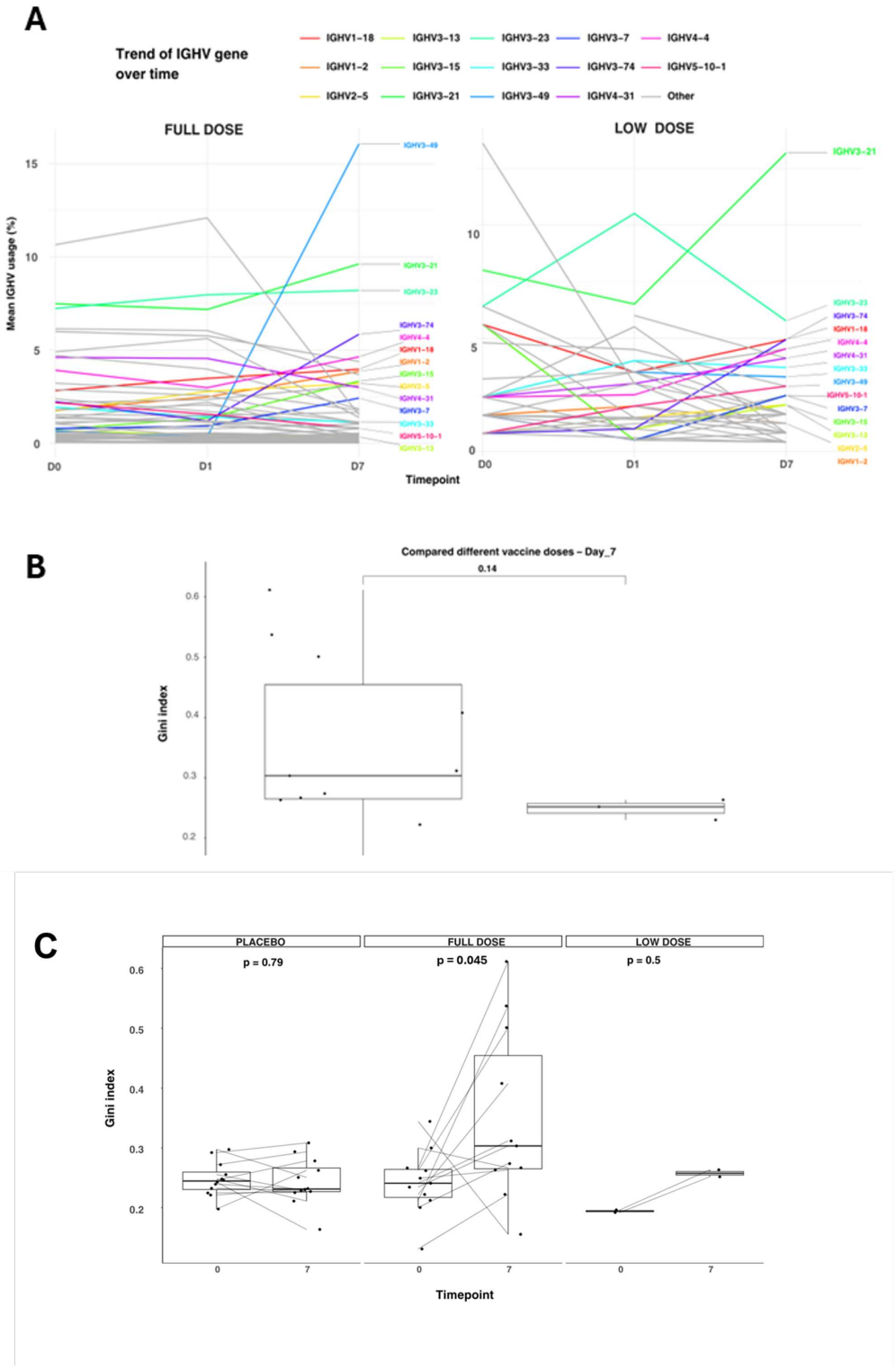


Figure 3.10: Line plots showing the dynamics of IGHV gene usage (mean percentage) across three timepoints (D0, D1, D7) for Full Dose (left panel) and Low Dose (right panel) vaccine recipients.

Each line represents a specific IGHV gene, with genes showing >5 % usage at D7 or >1% increase from D0 to D7 highlighted in different colors, while remaining genes are shown

in grey. The most prominent changes are observed for IGHV3-49 and IGHV3-21 in the Full Dose group, and IGHV3-21 and IGHV3-23 in the Low Dose group. (B)Gini index comparison between vaccine doses at Day 7. Box plots comparing the Gini index between Full Dose and Low Dose vaccine recipients at Day 7 post-vaccination. No significant difference was observed between the two groups (p = 0.14, Mann-Whitney U test (C) Gini index at D0 and D7 in placebo (n=12), full-dose (n=11), and low-dose (n=4) groups. Points represent individuals; lines connect paired samples. Paired Wilcoxon tests indicate a significantly increased repertoire inequality at D7 in the full-dose group (p=0.045), with no significant change in placebo (p=0.79) or low-dose (p=0.5).

CDR H3 clustering identifies potentially public clonotypes

BCR clonotypes with similar CDR H3 sequences and V, D, and J gene usages may confer similar binding capabilities [70], [71].In order to identify if there are shared BCR clonotypes among vaccinated individuals derived from IGHV upregulation, BCR sequences using the previously selected IGHV genes by amino acid similarities were clustered using a Hamming distance-based clustering approach. After assigning all unique clonotypes to clusters, we filtered for those (i) shared by at least two vaccinated individuals and (ii) detected exclusively at Day 7, with confirmed absence at baseline (Day 0) and Day 1.

These criteria aimed to isolate putative vaccine-induced, convergent clonotypes. This yielded four clusters potentially representing convergent responses: cluster 20 (IGHV3-49), cluster 36 (IGHV3-7), and clusters 107 and 320 (IGHV3-23). None of these clusters were observed in placebo recipients or at any pre-vaccination time point, supporting their specificity to the post-vaccine response. Cluster 20 used IGHV3-49, an IGHV gene that showed strong upregulation in RNA-seq analysis of full-dose recipients. This cluster was found in one full-dose and one low-dose participant (Figure 3.11). The remaining clusters involved only full-dose participants: cluster 36 (IGHV3-7), and clusters 107 and 320 (IGHV3-23).

To evaluate whether our clusters corresponded to previously characterized antibodies, we aligned their defining CDRH3 sequences against public BCR repositories (e.g., the Observed Antibody Space database and ClonoMatch libraries of virus-binding antibodies (26,27)) but found no significant similarities to any known antibody sequences. No significant matches were found between our CDRH3 sequences and public repositories,

In addition to the shared clonotypes described above, full-dose vaccinees exhibited a greater number of unique CDR-H3 clusters that were not shared across individuals.These unshared clusters may reflect private, individual-level B-cell responses to vaccination. Overall, the number of clusters identified was higher in the full-dose group compared to

the low-dose group (Figure 3.12), consistent with a broader clonal activation at higher antigen doses.

CloneId	Count	Sample	Timepoint	Dose	V	D	J	C	CDRH3_aa	CDRH3_lenght
178	1	N_1647_D7.IGH_clones	Day_7	VL	IGHV3-49	IGHD3-10	IGHJ5	NA	CTRDEYGSGSPFDPW	15
179	1	N_1647_D7.IGH_clones	Day_7	VL	IGHV3-49	IGHD3-10	IGHJ5	IGHM	CTRDEYGSGSPFDPW	15
180	1	N_1647_D7.IGH_clones	Day_7	VL	IGHV3-49	IGHD3-10	IGHJ5	IGHM	CTRDEYGTGSPFDPW	15
194	2	N_1676_D7.IGH_clones	Day_7	VF	IGHV3-49	IGHD3-10	IGHJ5	IGHA2,IGHA1	CTRDEYGSGSPFDPW	15
195	2	N_1676_D7.IGH_clones	Day_7	VF	IGHV3-49	IGHD3-10	IGHJ5	IGHM	CTRDEYGTGSPFDPW	15
196	2	N_1676_D7.IGH_clones	Day_7	VF	IGHV3-49	IGHD3-10	IGHJ5	IGHM	CTRDEYGTGSPFDPW	15
432	1	N_1676_D7.IGH_clones	Day_7	VF	IGHV3-49	IGHD3-10	IGHJ5	IGHM	CTRDEYGSGSPFDPW	15

Figure 3.11: *Characteristics of convergent BCR clonotype cluster 20 (IGHV3-49). Clonotypes in cluster 20 were detected only at Day 7 in two vaccinees (one full-dose, one low-dose), absent at Day 0/1 and in placebo. The table lists each clone's ID, count, sample, timepoint, dose, V/D/J gene usage, constant region, CDRH3 amino-acid sequence, and CDRH3 length.*

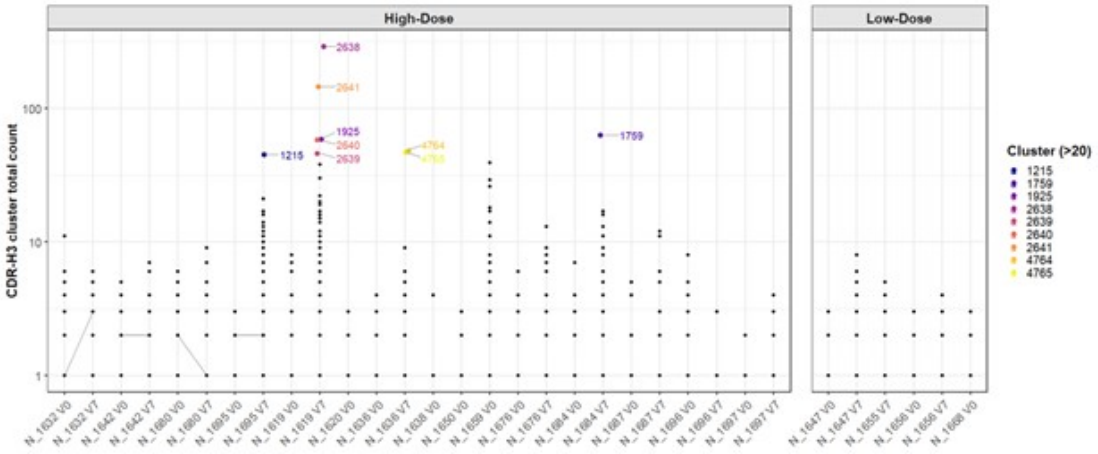


Figure 3.12: *BCR clusters in participants receiving the GMMA vaccine. Clusters with the highest total clones are shown in individual colors. The same clusters between V0 and V7 are connected by gray lines.*

3.5 Conclusions

Overall, data showed that GMMA-based iNTS vaccination elicits an early innate transcriptional program within 24 hours and a subsequent adaptive signature by day 7, with clearer effects in full-dose recipients. This temporal pattern is consistent with the established kinetics of vaccine responses, in which innate sensing and inflammatory signaling precede the emergence of B-cell and antibody-related programs. GMMA

vaccines are engineered outer-membrane vesicles that retain the architecture of Gram-negative bacteria while incorporating genetic modifications designed to reduce endotoxicity. GMMA can be considered multi-valent antigens, as they may present multiple polysaccharide molecules and proteins in natural conformation[72]. This combination supports a self-adjuvanting mode of action in which early innate sensing by pattern-recognition receptors—particularly TLR pathways—precedes the emergence of antibody and B-cell programs [73]. Among the most induced Day 1 transcripts, there were genes mapping to canonical TLR signaling and its downstream effectors (e.g., TLR1/2/4/5, CD14, LY96/MD2, MYD88, IRAK4, NFKB1/2) as well as inflammatory mediators such as IL1B. These findings are biologically plausible for GMMA/OMV platforms: even when lipid A is genetically modified to reduce reactogenicity, vesicles retain LPS and abundant outer-membrane lipoproteins and proteins, which provide ligands for TLR4 and TLR2 pathways, and can act in combination to amplify early innate activation [74]. Beyond TLR signalling, pathway-level analyses showed on a predominance of myeloid/neutrophil-associated programs at Day 1. TMOD enrichment highlighted up-regulation of modules linked to neutrophil activation and myeloid cells alongside down-regulation of T-cell–associated modules, a pattern frequently observed in whole-blood systems vaccinology [53],[54].

Among the most prominently upregulated genes at day 1 in full-dose recipients, CD177 showed one of the largest fold-changes. CD177 encodes a neutrophil-specific glycoprotein implicated in neutrophil activation. Similar CD177 upregulation has been reported in systems vaccinology studies, including the OMV-containing 4CMenB meningococcal vaccine, where CD177 contributed strongly to the early post-vaccination transcriptional stratification and tracked with neutrophil abundance [75]. While not specific to OMV-based vaccines, early enrichment of neutrophil-related genes has been described after multiple immunizations. This early gene signature probably reflects the rapid recruitment of neutrophils to the site of vaccination [76]. By day 7 post-vaccination, a clear shift from innate to adaptive immune signatures emerged, particularly in the full-dose group. We observed specific upregulation of immunoglobulin genes at this timepoint. This pattern is characteristic of plasmablast expansion and early B-cell differentiation into antibody-secreting cells. The emergence of immunoglobulin gene expression by one week post-vaccination aligns with the well-established kinetics of the adaptive response as documented across multiple vaccine platforms several Gram-negative bacteria [77],[78]. For instance, analyses of typhoid conjugate vaccine responses in a controlled infection model showed that while interferon and innate genes predominated at day 1, by day 7 the most upregulated genes were Ig heavy-chain genes [79]. Integrating B-cell receptor (BCR) repertoire profiling provided deeper insights into the vaccine-induced adaptive

response. On day 7, the full-dose group exhibited preferential usage of IGHV3-49 followed by IGHV3-21 and IGHV3-23, whereas the low-dose group predominantly used IGHV3-23, IGHV3-21 and IGHV1-18. This dose-dependent shift in IGHV gene usage suggests that antigen dose may influence which B-cell clones expand and become measurable in peripheral blood. Similar context-dependent differences in IGHV usage have been reported in Vi-based typhoid vaccination studies [79], supporting the concept that vaccine dose and type could shape the clonal composition of the plasmablast response. Although the mean Gini index at day 7 did not differ significantly between full-dose and low-dose groups, within-group comparisons showed that 9 of 11 full-dose recipients had increased Gini indices from baseline to day 7 (paired Wilcoxon p=0.045). This significant trend suggests that full-dose vaccination may promote clonal expansion of responding B cells, resulting in a more unequal repertoire dominated by a subset of expanded clonotypes. By clustering CDRH3 sequences, we identified four convergent clonotype groups (clusters20, 36, 107, 320) present only at Day 7 and shared by >2 vaccinees. Cluster20 (IGHV3-49) appeared in both vaccine doses, marking it as a possible candidate for further investigation. The phenomenon of convergent antibody responses, wherein individuals independently generate antibodies with similar genetic features in response to the same antigen, has been extensively documented. Public clonotypes have been identified following infection or vaccination against influenza [80],SARS-CoV-2 [81] and have also been reported for natural infection with viral pathogens, including hepatitis C virus (HCV), dengue virus, and Zika virus [82]. Similar convergent/public responses have been described against bacterial antigens, including Haemophilus influenzae type b (Hib) and after Hib–Meningococcus C polysaccharide–protein conjugate vaccination [82] as well as in responses to Salmonella Typhi Vi [83]. In this context, our identification of four convergent clusters in a relatively small cohort (n=15 full-dose; n=4 low-dose) may reflect the possibility that iNTS-GMMA presents immunodominant epitopes capable of eliciting stereotyped antibody responses. None of the convergent sequences we identified matched entries in public BCR databases (e.g., Observed Antibody Space, ClonoMatch). This likely reflects limited representation of iNTS- and bacterial polysaccharide–specific repertoires in current repositories, consistent with the absence of a licensed iNTS vaccine. More broadly, public datasets are dominated by well-studied viral pathogens. Thus, the lack of matches may indicate novel vaccine-induced clonotypes but could also be influenced by study the small cohort size (n=15 full-dose; n=4 low-dose). In addition to the shared public clusters, we also observed more "private" (individual-specific) clusters in the full-dose group, implying that higher antigen doses may affect the B-cell response. This finding suggests that higher antigen dose may promote broader clonal diversity, allowing expansion of B-cell clones targeting individual-specific epitopes. This study has some limitations. First, our BCR repertoire analysis was based on bulk RNA-seq rather

than single-cell sequencing, limiting our ability to definitively pair heavy and light chains or to quantify the absolute frequencies of antigen-specific clonotypes. Second, the small sample size, particularly in the low-dose cohort (n=4), reduced statistical power to detect dose-dependent differences. Third, we extracted BCR sequences from whole blood rather than sorting plasmablasts, meaning that our repertoire data include contributions from all circulating B-cell subsets, not just vaccine-induced plasmablasts. Future studies could employ single-cell RNA-seq combined with BCR sequencing to precisely define vaccine-induced clonotypes and their transcriptional states. Functional characterization of the putative public antibodies identified here will be critical to assess their protective potential. Finally, validation in larger cohorts and diverse populations will be necessary to confirm whether clusters 20, 36, 107, and 320 represent public clonotypes with potential as biomarkers or therapeutic targets. Taken together, these findings demonstrate that combining whole-blood transcriptomics with BCR profiling provides a powerful systems vaccinology approach to characterize GMMA-induced immunogenicity and to identify molecular signatures associated with vaccine response.

3.5.1 Acknowledgments

This work was supported by the EU Framework Programme for Research and Innovation, Horizon 2020 (grant agreement number 815439, Vacc-iNTS consortium).

3.5.2 Members of the iNTS Vaccine Consortium

Francis Agyapong, Gianluca Breghi, Annalisa Ciabattini, John A. Crump, Melita A Gordon, Liselotte Hardy, Samuel Kariuki, Stefano Malvolti, Carsten Mantel, Christian S. Marchello, Florian Marks, Donata Medaglini, Tonney S. Nyirenda, Mercy Ngetich, Ellis Owusu-Dabo, Francesco Santoro, J. Anthony G. Scott, Bassiahi Abdramane Soura, Tiziana Spadafina, Bieke Tack.

Chapter 4

A randomized, double-blind phase 2b trial to evaluate efficacy of ChAd63-KH for treatment of post kala-azar dermal leishmaniasis

$BrimaM.Younis$,[1] $RebeccaWiggins$,[2] $EltahirA.G.Khalil$,[1] $MohamedOsman$,[2] $FrancescoSantoro$,[3] , $ChiaraSonnati$,[3] , $AdaKeding$,[4] $MariaNovedrati$,[3] $GiorgioMontesi$,[3] $AliNoureldein$,[1] , $ElmukashfiT.A.$,[1] , $AlaEldinMustafa$,[1] $MohammedAlamin$,[1] $MohammedSaeed$,[1] , $KhalidSalman$,[1] $AhmedJ.Suliman$,[1] $AminE.A.Musa$,[1] $AlisonM.Layton$,[2] , $CharlesJ.N.Lacey$,[2,5] $PaulM.Kaye$,[2,5] $andAhmedM.Musa$[1,5]

- 1 Department of Clinical Pathology & Immunology, Institute of Endemic Diseases, University of Khartoum, Khartoum 11111, Sudan
- 2 York Biomedical Research Institute Hull York Medical School, University of York, Heslington, York YO10 5DD, UK
- 3 Department of Medical Biotechnologies, University of Siena, 53100 Siena, Italy
- 4 Department of Health Sciences, University of York, Heslington, York YO10 5DD, UK

4.1 Abstract

A recent phase 2a clinical trial (LEISH2a) assessed the safety and the immunogenicity of ChAd63-KH as treatment against PKDL in Sudanese patients with persistent PKDL. The vaccine showed minimal adverse effects, while it induced a strong innate and cell-mediated immune response with a clinical improvement higher than 90% in patients

who completed the follow up [84]. However, its efficacy as a stand-alone therapeutic was still unknown.We conducted a randomized, double-blind, placebo-controlled phase 2b clinical trial (LEISH 2b) to investigate the therapeutic efficacy of ChAd63-KH. In this trial the primary outcomes were safety and efficacy, while secondary outcomes were change in severity grade and vaccine-induced immune response. 86 participants with PKDL for at least 6 months were enrolled to receive $7,5x10^10$ vaccine dose or a placebo and then checked at day 90 post-vaccination for clinical improvement. No serious adverse effects were observed and at the end of the follow-up 15% and 11% of participants in the vaccine and placebo groups, respectively showed clinical improvement higher than 90%.Whole-blood transcriptomic analysis identified transcriptional modules associated with interferon responses and monocyte and dendritic cell activation Therefore, a single ChAd63-KH vaccination did not show therapeutic efficacy in this subset of Sudanese patients with PKDL [26].

4.2 Introduction

Post kala-azar dermal leishmaniasis (PKDL) is a chronic dermatological sequela associated with treatment for visceral leishmaniasis (VL; kalaazar),but it may occur without a history of VL or even during treatment of VL (when it is known as para kala-azar dermal leishmaniasis). PKDL typically involves the face and later spreads to the extremities and trunk.[85],[86],[87] PKDL has often been confused with leprosy, and because of its chronic but not debilitating nature, it is often tolerated by infected people, who fail to seek treatment. Case rates for PKDL are intimately linked to the waxing and waning of VL incidence. [88]The VL elimination campaign in South Asia has served to focus attention on PKDL, as those with the disease have been shown to harbor parasites in their skin and be infectious to the sand fly vector. As such, people with PKDL may serve to maintain infection in inter-epidemic periods, thus posing a threat to elimination.[89] Epidemiologicalmodeling has demonstrated the value of a PKDL vaccine for mitigating this risk in a post elimination era.[90] Recent calls to action for VL elimination on the African continent also recognize the need to have effective means to control PKDL[91] Drug regimens for PKDL are generally arduous and invasive and have multiple potential side effects.[92],[93] PKDL can also be exacerbated byHIV, and such cases respond poorly to treatment.[94] New combination therapies have been evaluated and show promise both in reducing the burden of PKDL[95] and in treatment,[96] but there has been a strong and persistent argument for developing additional means of control, including therapeutic vaccines.[97][98],[99]

PKDL has several intriguing features epidemiologically, clinically, and immunologically.

VL is caused by infection with two species of the protozoan parasite Leishmania, namely L. donovani and L. infantum, yet PKDL is restricted to infection with L. donovani and found only in the Old World.[86] PKDL is also restricted within the geographical range of L. donovani, being found in Sudan and South Asia but rarely in other countries in East Africa. PKDL typically occurs after treatment, whether this be with pentavalent antimonials,amphotericin B, miltefosine, or paromomycin/antimonial combinations.[100] However, the onset of disease varies geographically, being rapid(weeks to months) in Sudan compared to delayed (often years) in South Asia. Clinical presentation also varies geographically and at different body sites, with various hypotheses, including ultraviolet light (UVB) exposure, being put forward to explain differences in clinical presentation. [101][102] In Sudan, disease presents mainly as nodules and papules that,in 80% of cases, self-resolve over several months.[92] [103] In the remainder, lesions may persist, often for years. In contrast, in South Asia, cases can show hypopigmented macular lesions or be polymorphic, with both macular and nodular lesions. Immunological and histopathological differences have also been noted.[101] [102] [104] [105] [106] [107] Collectively, these features suggest that PKDL is a heterogeneous disease, with implications for the development of new therapeutics.

Therapeutic vaccination against PKDL in Sudan has its historical roots in the use of first-generation vaccines for VL.Khalil and colleagues conducted phase 1/2 randomized trials in volunteers with no history of VL and showed that a two intradermal doses of autoclaved L. major (ALM) administered with Bacille Calmette-Guérin (BCG) was safe, well tolerated, and immunogenic, as measured by skin test conversion.[108] [109] [110] However, no efficacy with regards to protection against VL was observed. A follow-up study of alum-adjuvanted ALM + BCG also demonstrated safety and immunogenicity,[111][112] and this formulation was evaluated in combination with sodium stibogluconate (SSG) as a potential therapeutic in patients with PKDL with persistent disease, reporting a cure rate at 60 days of 87% in the combined therapy group vs. 53% in the SSG alone group (SSG + vaccine efficacy = 71%, 95% confidence interval [CI] for risk ratio [RR], 0.7–1.16).[113] Based on these encouraging results, we first conducted an open-label phase 2a trial (LEISH2a; ClinicalTrials.gov: NCT02894008) in patients with persistent PKDL of ChAd63-KH, a new adenoviral vaccine incorporating two well-documented candidate antigens.[114] This study confirmed that ChAd63-KH was both safe and immunogenic in this patient group,[115] opening the way for the randomized, controlled efficacy trial reported here.

4.3 Results

4.3.1 Site initiation and recruitment

The LEISH2b study was conducted over approximately 4 years, punctuated by periods of significant challenge that resulted in temporary trial suspensions. The site initiation visit, attended by all field site personnel, trial monitors, and trial staff from the Universities of Khartoum and York, took place in August 2019 in Addis Ababa due to a popular uprising and deteriorating security situation in Sudan. Recruitment began on April 4, 2020, and continued through June 2020, with 17 participants recruited in this period. Recruitment was then paused due to the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. The second round of recruitment took place between December 2020 and February 2021, enrolling 28 participants. A third round of recruitment took place in April and May 2021, adding 23 participants. Delayed by a military coup, the final round of recruitment (May to June 2022) added 18 more, bringing the total to 86 participants. Following discussion with the LEISH2b data safety and monitoring board (DSMB), a meeting of the trial steering group with the funders was convened. All attendees agreed that recruitment should halt at 86 participants due to logistical difficulties in completing the trial prior to expiry of the investigational medicinal product (IMP) and the onset of further military clashes. 75/86 (87%) of enrolled participants completed the study to day 90 post-vaccination, with the last patient last visit on September 20, 2022.Missed visits and losses to follow-upwere due to occupational priorities. A further day 120 visit was arranged to allow follow-up of anydrug treatment provided at the completion of the study. A summary of study recruitment is provided in the CONSORT diagram (Figure 4.1).

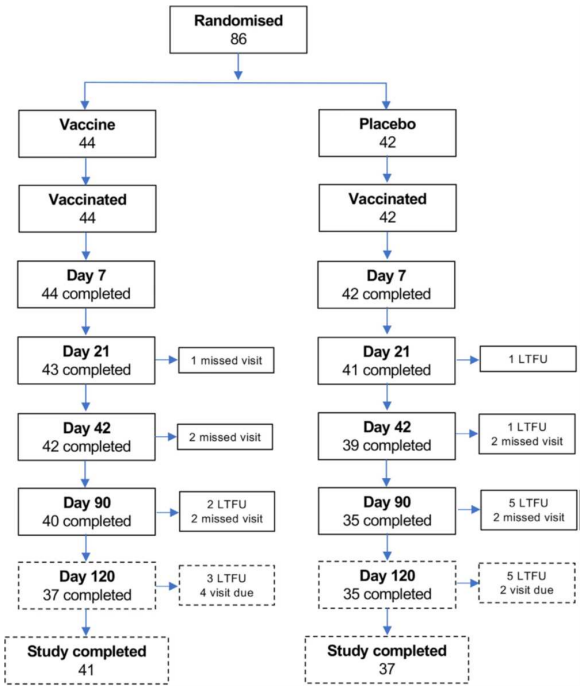


Figure 4.1: *Study CONSORT diagram Solid boxes indicate participants to primary outcome at day 90 post-vaccination (Vx). Treatment was offered per protocol from day 90 (see main text). Dotted boxes were outside the trial window for primary outcome and representscheduled follow-up to provide or monitor standard of care. LTFU, lost to follow up*

4.3.2 Study population

Demographics and PKDL grading of the 86 participants enrolled in this study are shown in Table 1. All participants had PKDL for a 6 month duration or longer, with the majority (88%) graded as either PKDL grade 1 or 2 Table (4.1). Day 90 visits were completed by 75 (87%) participants (n = 11 missed this time point or had been lost to follow-up; figure 4.1). As per protocol, participants were offered standard of care (AmBisome) at day 90 if clinical improvement was less than 90%. Overall, 26/86 (30%) elected to receive treatment (8/20 adults; 18/66 adolescents), and these participants remain in long-term follow-up to confirm drug effectiveness.

Table 1. Characteristics of participants in LEISH 2b trials		
Characteristics	Vaccine, N=44	Placebo,N= 41
Adults (18 to 50)	n=11	A n=9
Sex		
Male	n=8(73%)	n=6(67%)
Female	n=3(27%)	n=3 (33%)
PKDL grade		
1	n=7(64%)	n=2(22%)
2	n=3(27%)	n=5(56%)
3	n=3(27%)	n=2(22%)
Adolescents (12 to 17)	n=33	A n=33
Sex		
Male	n=17(52%)	n=15(45%)
Female	n=16(48%)	n=18 (55%)
PKDL grade		
1	n=14(42%)	n=18(55%)
2	n=15(45%)	n=12(36%)
3	n=4(12%)	n=3(9%)

Table 4.1: Table showing characteristic of LEISH 2b participants

4.3.3 Safety Outcomes

There were 70 (25 local and 45 systemic) adverse events (AEs) reported by 43 participants during the study, split 1.6:1 between the vaccine and placebo groups . The average numbers of local AEs per participant were 0.30 in the vaccine vs. 0.29 in the placebo arms(median of 0 in both arms), and systemic AEs per participant were 0.68 in the vaccine vs. 0.36 in the placebo arms (median of 0 in both arms). Event numbers did not significantly differ between arms (Mann-Whitney U p = 0.921 and p = 0.501 for local

and systemic events, respectively). AEs were limited to grade 1 and 2, with no grade 3 AEs, serious AEs (SAEs), or suspected unexpected serious adverse reactions (SUSARs) reported. All local and systemic AEs were deemed to be not serious and recovered. 44 AEs (25 local and 19 systemic) were considered possibly, probably, or definitely related to vaccination, again showing a bias toward vaccine recipients (Figure 4.2). These included itch, pain or soft swelling at the injection site, headache, vomiting, fever, and general muscle pain.Medication was not required for any of the local AEs, but paracetamol (for headache, chills, and fever) and chlorphenamine (for whole-body itch) were prescribed for systemic AEs. No clinically relevant changes in blood biochemistry or hematology were observed. The most common systemic AE unrelated to the study was malaria (16 clinical episodes, 15 patients). One case of thrombocytopenia was recorded but deemed unrelated to vaccination (occurring in the placebo group). Since the completion of the study, there have been no formal follow-up visits, but participants remain able to contact the study team should they have any future health issues. None have been reported to date.

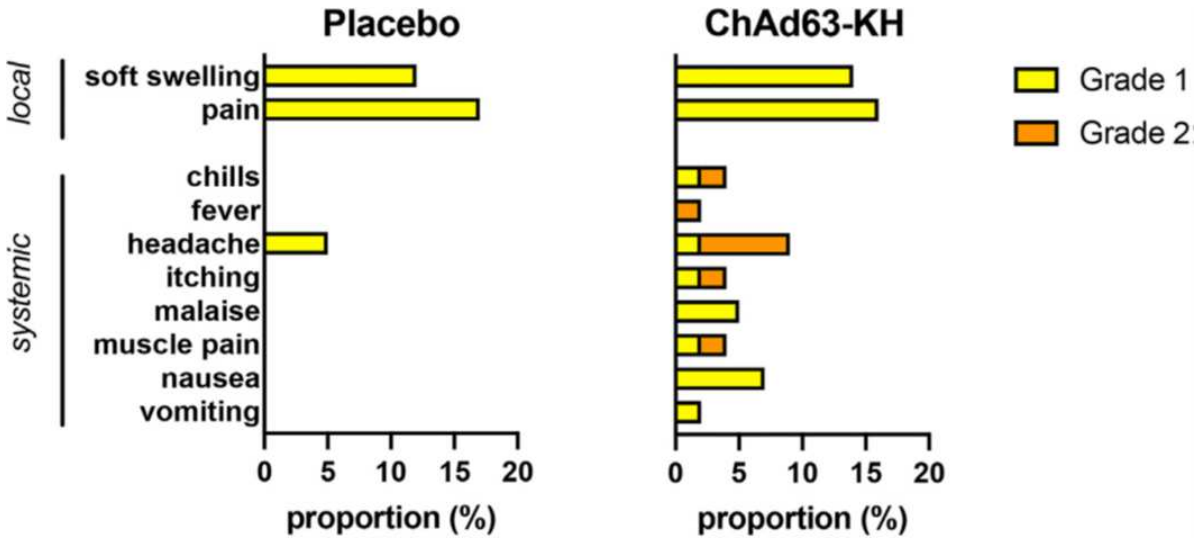


Figure 4.2: *Adverse events associated with the LEISH2b trial.Data are shown for local and systemic adverse events judged to be possibly, probably, likely, or definitely associated with Vx. Data are shown as proportions of participants in the study (n = 86). Only grade 1 AEs (mild, yellow; 26 vs. 17 events overall in vaccine vs. placebo groups) and grade 2 AEs (moderate; orange; 17 vs. 10 events overall in vaccine vs. placebo groups) were observed.*

4.3.4 Clinical Outcomes

The per-protocol primary outcome measure was 90% or greater improvement in clinical PKDL as determined by two clinical (39 female [F], 36male [M]), 6/40 (15%; 2 F, 4M) and 4/35(11%; 3 F, 1 M) in the vaccine and placebo arms, respectively,reached this

threshold (RR 1.31 [95% CI, 0.40–4.28], p = 0.742). To explore whether there was any vaccine effect at lower levels of improvement, we calculated the proportion of participants that attained 25%, 50%, 75%, and 90% of improvement over time of follow-up, and the trajectories of recovery appeared to be very similar between arms (Figure 4.3(A)). Of note, all participants that reached the primary outcome did so between days 42 and 90 of follow-up. The secondary outcome of PKDL grading (grade 1 to grade 4) was evaluated, as PKDL severity is often used to monitor disease status.[116] Grade distributions improved only marginally over time and were similar between arms (p = 0.36, p = 0.53, and p = 0.38 for days 21, 42, and 90, respectively; Fisher's exact test; Figure4.3(B)).

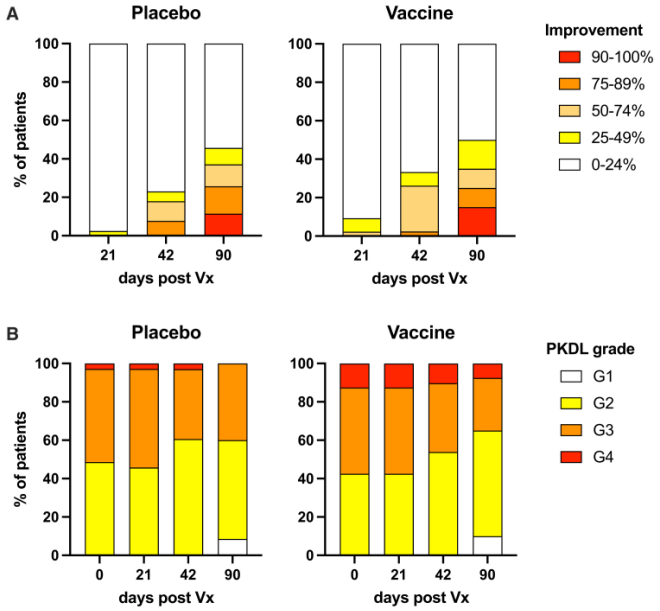


Figure 4.3: *Clinical improvement in LEISH2b trial. for vaccine and placebo groups, respectively). Participants were categorized based on degree of improvement at different times post-Vx. Color key indicates improvement category. No significant differences were detected at any time point or over time (Fisher's exact test; see main text). (B) Distribution of overall PKDL grades by study arm (n = 40 and n = 35 for vaccine and placebo groups, respectively). Participants were scored for PKDL grade at Vx (day 0) and at indicated days post-Vx. Color key indicates grades. No significant differences were detected at any time point or over time (Fisher's exact test; see main text).*

As duration of PKDL has been suggested to influence cure rate, we additionally obtained prior duration in months data for a subgroup of 57/75 participants completing day 90 follow-up (45 adolescents, 12 adults; 33 F, 24 M). We stratified this subgroup into two classes based on duration of PKDL (<18 months: median: 11 months, range: 6–11 months, n = 26; R18 months, median: 36 months, range: 18– 120 months, n = 31) and evaluated whether there was any difference in the proportion of patients reaching a

conservative 50% improvement. In the <18 month duration class, 6/11 receiving vaccine and 6/15 receiving placebo reached this threshold compared to 5/19 and 5/12 in the >18 month duration class. Hence, the duration of PKDL in this limited cohort did not appear to be associated with vaccine response or overall improvement. Collectively, these analyses indicate that under the trial conditions employed, ChAd63-KH lacked efficacy as a single treatment in patients with persistent PKDL in Sudan.

4.3.5 Whole-blood transcriptome prior to and after vaccination

We used whole-blood transcriptional analysis (WBTA) to confirm vaccine reactogenicity and capacity to induce immune responses in a subset of patients (n = 23 placebo and n = 27 vaccinated) for whom PAXGene-collected blood samples were available for analysis. Compared to prevaccination, no differentially expressed genes (DEGs) were identified at day 1 post-vaccination in patients receiving placebo (adjusted $p < 0.05$). In contrast, in patients receiving ChAd63-KH, we identified 318 DEGs using this threshold (311 UP, 7 DOWN). Using g:Profiler, we identified enriched terms associated with the response to ChAd63-KH, demonstrating activation of antiviral, inflammatory, and immune response genes (Figure 4.4(A)). To provide a comparison with data from our previous phase 2a study and responses to a range of other vaccines,[117] we next identified transcriptional modules associated with vaccination. In keeping with the above analysis and our previous analysis of ChAd63-KH responsiveness in patients with PKDL,[115] highly upregulated transcriptional modules included those associated with interferon responsiveness, dendritic cell and monocyte activation, and antigen presentation (Figure 4.4(B)). Out of 57 significantly enriched transcriptional modules, 31 were also enriched in the adolescent cohort of the LEISH2a study.[115] Given that only 2/50 of the patients for whom transcriptomic data were available achieved the primary outcome of 90% clinical response, it was not possible to conduct a robust analysis to identify DEGs or modules associated with clinical cure.

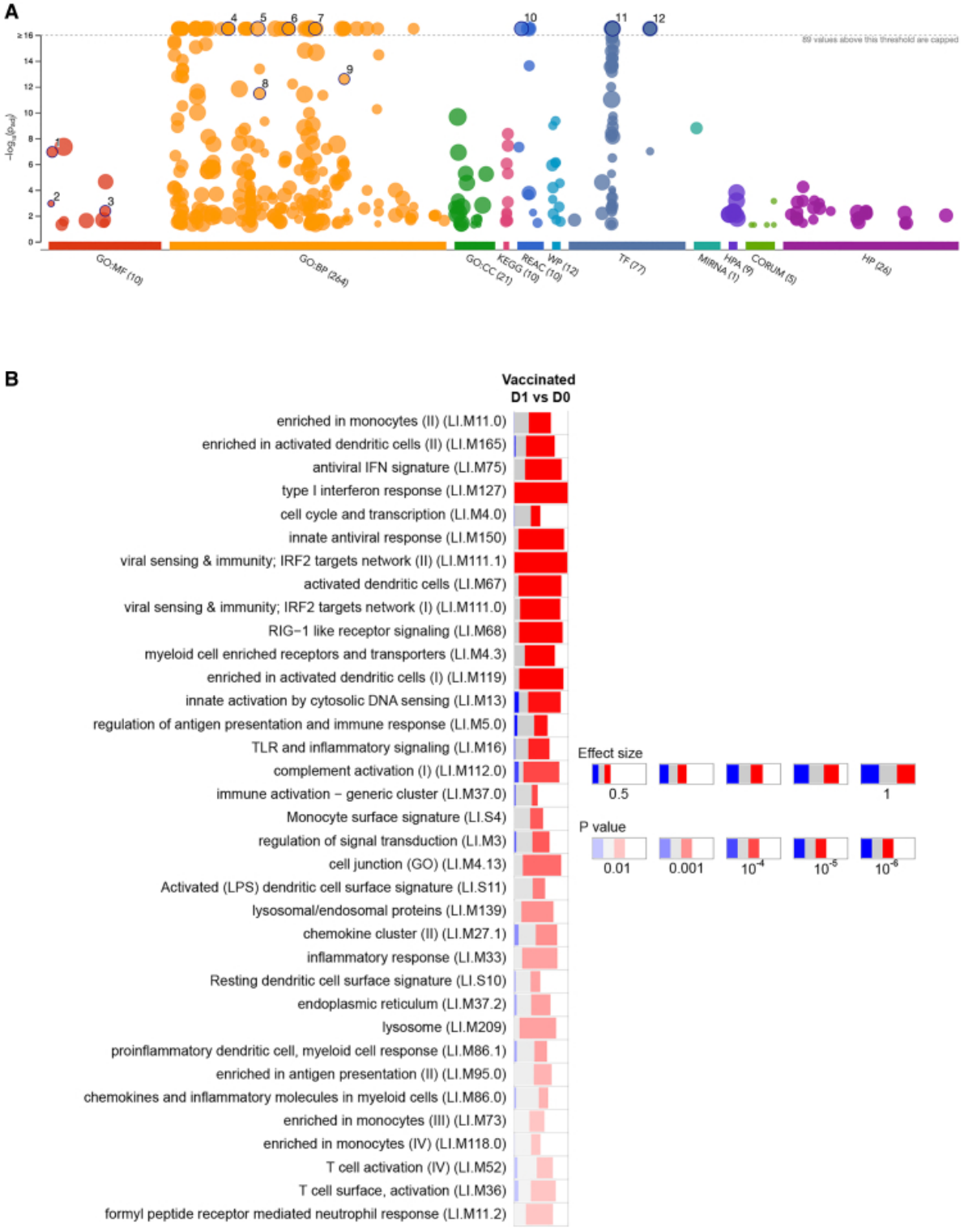


Figure 4.4: *WBTA of vaccine response. WBTA was conducted on blood drawn prior to and 1 day post-Vx. Data are shown for 50 vaccinated participants (19 F, 31 M; 9 adult, 41 adolescent).(A) g:Profiler analysis of 311 upregulated genes. Representative enriched pathways are numbered: (1) GO0003725, double-stranded RNA binding; (2) GO0001730, 20-50 oligoadenylate synthetase activity;(3) GO0042379, chemokine receptor binding;*

(4) GO0019221, cytokine mediated signaling pathway; (5) GO0034097, response to cytokine; (6) GO0045089, positive regulation of innate immune response; (7) GO0051607, defense response to virus; (8) GO0034341, response to type II interferon; (9) GO0071357, cellular response to type I interferon; (10) REAC:R-HAS-1, cytokine signaling in immune system; (11) TF:M00772, IRF motif; and (12) TF:M10080, STAT2 motif. (B) Significantly enriched immune-related modules were identified applying the CERNO test on the adjusted p value-ranked lists of genes generated by DeSeq2. Bars represent the proportions of significantly upregulated (red), downregulated (blue), or unchanged (gray) genes. The significance of module activation is proportional to the intensity of the bar, while the effect size is proportional to its width. Only the top 35 modules (of 57; see Table S6) are shown for clarity. No DEGs, and hence no modules, were identified in patients receiving placebo.

4.4 Discussion

ChAd63-KH is a third-generation vaccine based on adenoviral delivery of a gene construct encoding two well-characterized Leishmania antigens. In a phase 2b stand-alone therapeutic trial in Sudanese patients with persistent PKDL, we have confirmed the safety of this vaccine. However, this study failed to demonstrate clinical efficacy when measured by clearance of PKDL lesions over a 90 day follow-up period.

The study was conducted in Gedaref state, South East Sudan, against a backdrop of significant adversity. Over the study period, Sudan sequentially experienced a revolution, a pandemic, a coup, and military conflict that remains ongoing to date. Nevertheless, throughout this period, the trial team was able to maintain most aspects of trial management and governance, although considerable delays in recruitment and the ability to undertake sample analysis occurred. Following consultation with the DSMB and trial sponsor, the decision was taken in November 2022 to curtail recruitment at 86 participants, short of the 100 originally proposed. Full data analysis for the primary clinical outcome is reported here, with additional subgroup analysis being reported where patient metadata or biological samples were incomplete (either due to sample loss or data inaccessibility).

The primary clinical outcome of the study was to determine whether a single administration of ChAd63-KH vaccine was able to induce R90% improvement of PKDL in patients with persistent disease of R6 month duration. Our analysis of 75 patients that that completed follow-up to day 90 indicated that this was not the case. Similarly, PKDL grade was not improved over this time in vaccinated patients compared to those receiving placebo. Relaxing the criterion for clinical response also did not reveal any significant

vaccine effect or effect of prior PKDL duration. Hence, we conclude that this phase 2b clinical trial has not demonstrated efficacy of ChAd63-KH under the conditions tested.

It is therefore important to consider the premise on which this study was based and potential reasons why efficacy was not observed. The immunopathogenic mechanisms that result in the chronic dermal presentation of PKDL are not fully understood. Immunohistological analysis of biopsies from Sudanese patients with PKDL have noted the presence of CD4+ and CD8+ T cells, but their functional state has not been characterized. Granuloma formation is seen in some cases, but the histopathological picture may not correspond with clinical outcome.[107] Keratinocyte interleukin (IL)-10 expression during VL was observed to be associated with the later development of PKDL,[118] but the specific role of IL-10 in PKDL per se is not understood. In Indian PKDL, more extensive studies have characterized CCR4+ CD8+ T cells with a phenotype associated with exhaustion and immune regulation (PD-1hi IL-10+).[104] Collectively, these findings lend support to the hypothesis underpinning this trial, i.e., that enhancing CD8+ T cell activation through vaccination may augment the ability of the host to clear parasites/antigens and reduce local pathology. Although we were unable to study antigen-specific T cell responses in this study due to the loss of all frozen cells, our previous studies have indicated that KMP-11- and HASPB-specific CD8+ interferon (IFN)$gamma$+ responses are induced in patients with PKDL [115] and healthy volunteers [114] following ChAd63-KH vaccination. Given the comparability of the transcriptional response to vaccination seen in this and our other studies, we have no reason to suspect that similar T cell responses were not induced in this patient group. However, in the absence of efficacy, it is difficult to ascertain whether this CD8+ T cell response was merely insufficient in magnitude to elicit a clinical response or whether a broader response (in terms of antigen specificity or immune response quality) might be needed to achieve a clinical response. Broader antigen exposure along with concurrent inflammation are features of both naturally acquired immunity and that afforded by "leishmanization" and provide a rationale for developing live attenuated vaccines[119] With their known ability to rapidly induce antigen-specific CD8+ T cells, adenovirus-vectored vaccines have been extensively studied as potential therapeutic vaccines in cancer and chronic hepatitis B virus (HBV) infection. However, studies in these other diseases have, to date, also met with limited success. While a poor choice of immunogens was previously mooted as the explanation for therapeutic failure in cancer therapeutic trials, a common attribute of cancer and chronic viral disease is the development of a microenvironment that favors immune regulation or suppression. These changes often reflect the outcome of signaling through inhibitory checkpoint pathways (e.g., PD-1) or the induction of metabolic checkpoints such as IDO1, and it has been proposed that these disease-associated regulatory constraints on T cell function hinder

vaccine efficacy.

Experimental evidence to support this notion has been provided by Maini and colleagues, who identified both natural killer cell and PD-1/PD-L1 signaling as negative regulators of HBV vaccine efficacy in animals.[120] While evidence for such an immunoregulatory environment is less well established in PKDL and other forms of dermal leishmaniasis, some similarities to these other chronic diseases exist. For example, both IDO1 and PD-L1 are abundant in chronic cutaneous leishmaniasis (CL) lesions resulting from L. donovani infection in Sri Lanka, with a reduction in PD-L1 expression predicting the rate of cure after antimonial chemotherapy.[121] More recently, we have identified expression of these checkpoint molecules in CL patients infected with L. (V.) braziliensis and also in patients with PKDL infected with L. donovani in India.[122] PKDL lesions are also enriched in M2 macrophages,[105] contributing to a lack of leishmanicidal capacity and an immunoregulatory environment. Hence, it is possible that the local inhibitory environment in the skin of patients with persistent PKDL provides an obstacle to effector CD8+ T cell function at this site. If this is indeed the case, it is likely that therapeutic vaccination in PKDL might also benefit from interventions that target regulatory pathways, such as those proposed in HBV infection [120]. While studies using existing biologics or other clinically approved drugs might provide proof-of concept data in PKDL, it is recognized that such high-cost adjunct therapies are unlikely to drive vaccine implementation in a lower- and middle-income country (LMIC) setting. Alternatively, vaccination in conjunction with antileishmanial chemotherapy should also be considered, given previous successes with immunochemotherapy in PKDL [113] and more recent evidence suggesting that such chemotherapy may also diminish the immunoregulatory environment in CL patients [121].

We confirm in this study the potent ability of ChAd63-KH to induce innate responses characterized by antiviral gene signatures and dendritic cell and monocyte activation, with weaker representation of modules associated with neutrophil response and complement activation. The module signatures observed are also in keeping with those observed in a multi-vaccine analysis conducted by Hagan et al.[117],which included an adenoviral vaccine candidate (MRKAd5-HIV), and with our previous analysis of patients with PKDL.[115] In the latter study, we observed that adolescents, but not adults, differentially expressed modules associated with B cell activation (M47.0 and M47.1). However, we did not observe differential expression of these modules in the current study, despite the patient subgroup studied for WBTA being mainly composed of adolescents. The reasons for this difference are currently unknown and subject to ongoing investigation. We had also previously identified 11 whole-blood transcriptional modules predictive of 90% clinical cure, two of which were significantly differentially expressed between cure and

non-cure participants (M139: lysosomal/endosomal proteins and M118.0: enriched in monocytes).Unfortunately, as only 2/50 of the samples analyzed here achieved this level of clinical cure, we are unable to confirm or not the predictive value of these modules.

In addition to the loss of samples/data due to the ongoing situation in Sudan and other issues discussed above, a limitation in study design was the evaluation of only one vaccination schedule. ChAd63-KH was administered as a stand-alone therapy using a single dose given intramuscularly. This approach was chosen for both scientific and pragmatic reasons. Extensive data derived from experimental studies and early-phase human trials [123] and late-phase trials and real-world evidence obtained during the SARS-CoV-2 pandemic [124] have indicated that repeat dosing with homologous adenovirus vaccines fails to significantly augment the CD8+ T cell response. Though direct evidence is lacking, it seems likely that a similar outcome would occur with vaccination in patients with PKDL. Pragmatically, single-dose schedules have multiple benefits in an LMIC setting, including reducing logistical complexities and a reduction in cost, not least due to the savings from not requiring manufacture of two clinical grade vaccines. Similarly, intramuscular dosing is preferred in many clinical situations due to the ease of administration and its standardization and tolerability but may not be the most appropriate route for eliciting skin-homing CD8+ T cells[125]. Thus, we cannot rule out that an efficacy signal may have been observed using a repeat-dosing schedule, by extending the period of follow-up, using a heterologous prime-boost strategy, [123] and/or varying the route of administration. Given the heterogeneity of clinical presentation and histopathology observed in PKDL, our data also do not rule out the possibility of a therapeutic benefit from ChAd63-KH vaccination in either Sudanese patients with less persistent disease or in patients with PKDL in South Asia. In addition, the prophylactic efficacy of this vaccine against different types of leishmaniasis remains to be evaluated. Such studies would need to be cognizant of the rare AEs reported for Vaxzevria, the ChAdOx1-based SARS-CoV-2 vaccine[126].In conclusion, however, this phase 2b study did not provide evidence to support progressing ChAd63-KH as a standalone therapeutic in Sudanese patients with persistent PKDL.

4.5 Materials and methods

4.5.1 Ethics statement

The LEISH2b study (ClinicalTrials.gov: NCT03969134) was approved by the Sudan National Medicines and Poisons Board and the ethical review committees of the Institute of Endemic Diseases, University of Khartoum, and the Department of Biology, University

of York. LEISH2b was sponsored by the University of York. The study was conducted according to the principles of the current revision of the Declaration of Helsinki 2008 and ICH guidelines for GCP (CPMP/ICH/135/95). All participants provided written informed consent before enrollment. Consent forms are available as extended data accompanying the published protocol. [127]

4.5.2 Study design and participants

LEISH2b was a randomized, double-blind, placebo controlled therapeutic trial designed to evaluate the safety and efficacy of therapeutic vaccination with the investigational vaccine ChAd63-KH. The initial study design allowed for the recruitment of 100 participants diagnosed with PKDL aged between 18 and 50 (adults) or 12 and 17 (adolescents) and with persistent PKDL of greater than 6 month duration. The full details of the inclusion and exclusion criteria are provided in the published protocol.[127] Participants were recruited from an endemic area in Gedaref state, Sudan, and all study procedures were conducted at the Professor El-Hassan's Center for Tropical Medicine, Dooka, Sudan. Clinical monitoring of the study was performed under contract by ClinServ (http://www.clinserv.net). An independent DSMB was established and met throughout the study period. The DSMB charter is available as extended data accompanying the published protocol [127].

4.5.3 Eligibility criteria

Full inclusion and exclusion criteria are provided in the protocol [127] Key inclusion criteria included the following: age 12–50 years on the day of screening; females must be unmarried, single, or widowed willing and able to give written informed consent/assent; uncomplicated PKDL ofR6 month duration; otherwise good health; negative for malaria on blood smear; Leishmania PCR positive on the screening skin biopsy; and willing to undergo urinary pregnancy tests (females only). Key exclusion criteria included the following: has mucosal or conjunctival PKDL; treatment for PKDL within 21 days; negative rK39 strip test; receipt of a live attenuated vaccine within 60 days or other vaccine within 14 days of screening; history of allergic disease or reactions to vaccines or their components; history of severe local or general reaction to vaccination; fever $>$39.5°within 48 h; anaphylaxis, bronchospasm, laryngeal edema,collapse, convulsions, or encephalopathy within 48 h; pregnancy, less than 12 weeks post-partum, lactating, or willingness/intention to become pregnant during the study and for 3 months following vaccination (females only); seropositive for hepatitis B surface antigen or hepatitis C (antibodies to HCV); and tuberculosis, leprosy, or malnutrition (malnutrition in adults

defined as a BMI $<$ 18.5 and in children and adolescents [8–17 years] as a Z score cutoff value of $<$-2 SD.

4.5.4 Vaccine and study procedure

As described previously,[115] ChAd63-KH encodes two leishmanial proteins (KMP-11 and HASPB1) and was manufactured to cGMP by Advent Srl. (Pomezia, Italy; lot B0004; $7.5x10^10$ viral particles per mL). ChAd63-KH or placebo (saline) was administered as a single dose in 1 mL volume intramuscularly into the deltoid muscle. Participants were monitored in a hospital for 7 days post-vaccination and thereafter as out patients on days 21, 42, 90, and 120 post-vaccination. Recruitment occurred over four rounds, punctuated by a revolution, a pandemic, and a coup. Meetings of the independent DSMB were held at the end of each round of recruitment. Participants were evaluated for clinical response at days 42 and 90. Those with less than 75% improvement at day 90 were offered standard treatment (AmBisome;2.5 mg/kg/day for 20 days). If improvement was between 75% and 90%, they were offered conservative treatment or AmBisome, and those with greater than 90% clinical improvement were deemed to not require further treatment (and scored as clinical cure). Some participants defaulted from scheduled visits and were evaluated and treated at unscheduled visits based on their availability. Decisions to treat and evaluation of PKDL were performed by two clinicians. The clinical grade of PKDL was also recorded before and after vaccination using the 4 point grading system described previously.[116]

4.5.5 Randomizing and blinding

A computer-generated (Stata 16) randomization list was prepared by the trial statistician, allocating participants 1:1 to either the active or placebo injection, using randomly permuted blocks, stratified by age group (adult or adolescent). The randomization allocation was communicated to pharmacy staff at the study site through prepared, sealed envelopes with only external sequential participant numbers. During the trial, adults were found to be more difficult to recruit, and the initial aim of an equal number of adolescents and adults in the trial was not feasible. However, the number of available randomization envelopes in Sudan was limited for each age group. Therefore, once allocations for the adolescent group had been used up, allocations for the adult stratum were used in order of presentation irrespective of age group, thus creating a single, unstratified randomized sample.

The vaccine and placebo injections were prepared in blacked-out syringes labeled only with the participant identification number. The participants and clinical investigators

(who administered the IMP and conducted the study follow-up) were blinded as to which injection participants received. The trial statistician (who generated the randomization list), trial coordinator (who prepared sealed envelopes according to the randomization list), assistant pharmacist (who prepared the IMP), and study nurse (who delivered the IMP to the clinical investigators) were not blind to the treatment allocation.

4.5.6 Outcomes

Primary outcome measures were safety, as recorded as AEs from clinical examination and evaluation of blood biochemistry and hematology, and efficacy, as determined by the proportion of participants reaching a 90% improvement in clinical disease. Secondary outcomes included trajectory of lesion improvement and measures of vaccine-induced immune response. Planned immune analyses included whole-blood transcriptomics and analysis of peripheral lymphocyte responses (antibody and IFNg ELISpot). Peripheral lymphocyte and antibody responses could not be measured due to sample loss resulting from the military conflict in Sudan.

4.5.7 Statistical analysis

Total sample of size of 100 (randomly assigned 1:1 to be vaccinated with ChAd63-KH or placebo) was determined to be sufficient to detect an increase of >25% in the proportion of participants achieving clinical cure, assuming 90% power, 5% statistical significance, a spontaneous clearance rate of <2%, and loss to follow-up of <5% (Fisher's exact test). Baseline characteristics were summarized descriptively. Continuous measures are reported as mean, SD, median, and range, while categorical data are reported as counts and percentages. The numbers of local and systemic AEs per participant are presented as median, minimum, maximum, and interquartile range. The median numbers of AEs per participant (separately for local and systemic events) were compared between groups using the Mann-Whitney U test. A primary efficacy outcome of >90% clinical improvement between the two arms was evaluated using a relative RR with statistical significance determined by Fisher's exact test. Event numbers were too low to allow for an adjusted regression analysis. Comparison of categorical data across groups was analyzed using Fisher's exact test. Analyses were performed using Stata v.18, Rv.4.3.2, or Prism 10 for macOS (v.10.1.1; GraphPad).

4.5.8 Whole-blood transcriptomic analysis

Whole-blood samples (2.5 mL) were collected into PAXGene tubes immediately prior to vaccination and at 1 day post-vaccination. All reagents and equipment for these analyses were supplied by Thermo Fisher Scientific and processes carried out per the manufacturer's protocols unless otherwise stated. Total RNA was extracted using the PAXgene Blood RNA kit (PreAnalytiX, QIAGEN). RNA was quantified using the Qubit 2.0 Fluorometer with the RNA HS Assay Kit. *tilde*50 ng of total RNA was used to construct sequencing libraries with the Ion AmpliSeq Transcriptome Human Gene Expression Kit. Libraries were barcoded, purified with 2.5x Agencourt AMPure XP Magnetic Beads (Beckman Coulter), and then quantified using the Ion Library TaqMan Quantitation Kit on a QuantStudio 5. Libraries were diluted to a concentration of about 50 pMol and pooled in groups of 8 for sequencing on Ion PI Chips. Chips were loaded using the Ion Chef System and the IonPI Hi-Q Chef Kit. Sequencing was performed on an Ion Proton Sequencer using an Ion PI Hi-Q Sequencing 200 Kit.

Differential gene expression analysis was performed using DeSeq2. After count data normalization, differential gene expression analysis was performed using pooled day 0 data from the two study cohorts as the baseline for all contrasts. Enrichment of blood transcription modules at each time point in the different groups was assessed with the tmod R package using as an input the lists of DEGs ranked by the p value after multiple test correction, as computed by DeSeq2. Significance of module enrichment was assessed using the CERNO statistical test (a modification of Fisher's combined probability test) and corrected for multiple testing using the Benjamini-Hochberg correction.

4.6 Role of funders

The funders played no part in study design, data collection, analysis, interpretation, writing, or decision to publish. All authors had full access to study data and final responsibility for the decision to submit for publication.

4.7 DATA AND CODE AVAILABILITY

Processed gene expression data are available as supplemental information.Raw gene expression data are available from GEO:GSE266275.

4.8 SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtm.2024.101310.

4.9 ACKNOWLEDGMENTS

The clinical trial was funded by a Wellcome Trust Translation Award(WT108518MA; https://wellcome.org). Additional support for transcriptomicstranscriptomics studies was provided by a Wellcome Trust Senior Investigator Award (to P.M.K.; WT104726) and by the TRANSVAC2 program supported by the European Union's Horizon 2020 Research and Innovation programme under grant agreement no. 730964 (TNA1802-02; https://www.transvac.org). GSK is allowing University of York to use the cell line ProCell92 (proprietary of GSK) for their adeno program. GSK had the opportunity to review the publication, but the content is solely the responsibility of the authors. The authors also express their gratitude to the DSMB (Dr. Rob Davidson, Dr. Koert Ritmeijer, Dr. Wolfgang Stohr, and Prof. Alaadin Ahmed) for their advice, continual support, and encouragement and to all the study participants and their families.

4.10 AUTHOR CONTRIBUTIONS

E.A.G.K., P.M.K., A.M.L., C.J.N.L., A.M.M., and B.M.Y. conceptualized the work. Clinical studies were conducted by B.M.Y., E.A.G.K., M.O., A.N., E.T.A.E., A.E.M., M.A., M.S., K.S., A.J.S., A.E.A.M., and A.M.M. Trial data analysis was conducted by A.K., P.M.K., A.M.L., C.J.N.L., A.M.M., and R.W. Transcriptomic analysis was conducted by P.M.K., G.M., M.N., C.S., and F.S. Statistical analysis was conducted by A.K. and P.M.K. The original draft was written by P.M.K. The draft was reviewed and edited by C.J.N.L., A.M.L., A.M.M., B.M.Y., R.W., and A.K. The final draft was written by P.M.K. All authors read and approved the final version of the manuscript.

4.11 DECLARATION OF INTERESTS

P.M.K. and C.J.N.L. are co-inventors of a patent that covers the gene insert used in ChAd63-KH.

Chapter 5

Final Conclusions

This work provides an integrated analysis of the immune responses elicited by three distinct vaccine platforms recombinant vesicular stomatitis virus based (rVSV–ZEBOV-GP), GMMA-derived (iNTS-GMMA), and adenoviral-based (ChAd63-KH) using transcriptomic, immunological, and bioinformatic approaches. Despite their different biological properties, all vaccines tested induced a rapid activation of innate immunity, followed by adaptive responses whose characteristics varied according to the vaccine modality.

In the first part of the work, whole-blood transcriptomics revealed that vaccination triggers an early and pronounced innate immune response, with the peak of differentially expressed genes occurring at Day 1 and involving canonical innate pathways (dendritic cells, interferon pathways and antiviral signature) . By Day 7, transcriptional profiles shifted toward immunoglobulin-associated signatures, indicating the initiation of the adaptive phase. Comparative analyses demonstrated that ZEBOV-GP induces a stronger activation than VARILRIX, confirming its high immunogenic potential. Classification of vaccinees based on antibody titers identified transcriptional differences between responders and non-responders, including the activation of the classical complement pathway via C1Q. Correlations between specific genes or transcriptional modules and antibody titers suggest that early gene expression changes contribute to both the magnitude and persistence of the humoral response. Age-related differences, but not gender-related differences, were observed, with younger children exhibiting enhanced B-cell–related activity at before vaccine administration.

The second part of the thesis focused on a GMMA-based vaccine targeting invasive Salmonella. Both vaccine doses activated innate immune pathways, particularly those involving TLR4 and TLR2, indicating that the modified LPS component retains substantial immunostimulatory capacity. At Day 7, analysis of the B-cell receptor repertoire identified upregulation of IGHV genes, modest clonal expansions, and the emergence of convergent clonotypes shared across multiple vaccinees. These findings indicate early

shaping of the antibody repertoire following a single immunization, although confirmation of clonotype specificity and functional relevance will require larger cohorts and antigen-specific assays.

In the final part, the adenoviral vaccine ChAd63-KH was shown to induce transcriptional profile dominated by antiviral signatures and activation of monocytes and dendritic cells. The module patterns observed were consistent with previous analyses of adenoviral vectors and with earlier transcriptomic studies in PKDL, although modules associated with B-cell activation—previously detected in adolescents—were not observed here. Due to the limited number of participants achieving clinical cure, it was not possible to validate previously identified predictive transcriptional modules. Nevertheless, the data confirm the capacity of ChAd63-KH to generate robust innate immune activation relevant for protection against intracellular pathogens.

Together, these findings highlight the utility of integrative immunogenomic approaches for characterizing vaccine-induced immunity. While shared innate signatures were observed across platforms, each vaccine exhibited distinct adaptive features and B-cell repertoire profiles. The results contribute to a deeper understanding of early immune events following vaccination and provide a framework for the identification of biomarkers and mechanistic correlates that may support the rational design and evaluation of next-generation vaccines.

Appendix A

Quantification of micro-RNA with digital PCR

A.1 Introduction

Psoriasis is an immune-mediated, genetic, chronic inflammatory skin disease involving the skin or joint or both in adults. Psoriasis affects about 2/3% of worldwide population and it dramatically impairs the quality of life of affected people . Five types of psoriasis have been reported:plaque psoriasis (also known as psoriasis vulgaris), guttate or eruptive psoriasis, inverse psoriasis, pustular psoriasis, and erythrodermic psoriasis, among which psoriasis vulgaris is the most common one and accounts for about 90% of cases [128]. There are not specific trigger factors, but the presence of mild trauma like scratching or piercing and tattoos, or the usage of systemic drugs as non-steroidal anti-inflammatory agents can exacerbate the disease. Psoriasis skin lesion originate from a dysregulated interaction of innate and adaptive components of immune system with cutaneous cells. A central role is played by TNFα , interleukin-23 and T helper cell 17 (Th17), which have a pro-inflammatory action leading to the release of other pro-inflammatory peptides, which induce keratinocyte proliferation and an over production of inflammatory cytokines.

Although the pathogenesis of this disease is not yet fully understood, due to its complex multifactorial nature; it seems that aberrations in microRNA (miRNA) expression are involved in psoriasis onset. MiRNAs are natural mediators of post-transcriptional gene silencing, they can be involved in regulation of differentiation, proliferation and cytokine response of keratinocytes and T cells [129]. In this study, we focus on four specif miRNA, which have already been demonstrated to be involved in psorias development, which are miR-223,mir-146a,miR-21-5p, and miR-155-5p. MiR-223 is one of the most upregulated miRNA in psoriatic skin, it is expressed in in dermal inflammatory infiltrates of psoriatic skin and in Th17. Several studies of expression of miRNA in peripheral blood mononuclear cells (PBMCs) in affected people showed an upregulation of both miR-223 and miR-146a, t positively correlated with IL-17 expression [129].Mir-21-5p is frequently

upregulated in inflammatory disease and it promotes inflammation by silencing anti-inflammatory genes, while mir-155-5p is regulator of immune response and it is involved in T-cell activation. All of these miRNAs were determined to be markers for psoriasis, detectable in blood and in extracellular vesicles inside plasma [130], in our study we demonstrated their presence also in saliva.

A.2 Material and Methods

A.2.1 Saliva collection and micro-RNA extraction

Saliva samples were collected from 184 volunteers at tree timepoints (baseline, T10 and T20). Saliva was collected from patients who stopped eating,smoking and drinking at least one hour before the collection. Patients were asked to wash their mouth for a minute with distillate water and then spit about 5ml of saliva in a sterile tube.The expectorate was centrifuged at 2000xg for 20 second,the supernatant was then collected in 2ml eppendorf and stored at -80°C. miRNA was extracted by using Maxwell RSC miRNA Plasma and Serum kit by Promega, a final elution volume was 50 μl or 40μl .

A.2.2 Retrotranscription and quantification with digital-droplets PCR

miRNA was retrotranscribed according to the miRCURY LNA RT Kit instruction by QIAGEN. For digital-droplets PCR we tested different condition for the amplification step, such as samples pre-dilution, amount of primer into the final reaction mix and annealing temperature (see A.1).

Primer	Volume	T°	Dilution
hsa-miR-155-5p	1μl	55°C	1:10
hsa-miR-223-3p	0.5μl	55°C	1:300
hsa-miR-146a-5p	1μl	58°C	1:10
hsa-miR-21-5p	1μl	52°C	1:20

Table A.1: **PCR condition for each primers**

Droplets generation and reading were performed with QX200 Droplet Generator and QX200 Droplets Reader according to Biorad procedures for EVA Green die. The QX200 Droplets reader was used to detect the assorbance on FAM channel of each samples. Data collected by the reader were analyzed using the QX Manager software (v 2.1)provided by BIORAD and then stored in a database.

A.3 Discussion and results

Each miRNA resulted to be expressed in all samples, confirming their presence also in saliva of patents. MiR-223 was the most expressed, with a medium number of copies of about $22x10^6$, followed by miR-21-5p with $14x10^6$ medium copies copies. A lower expression characterized miR-155-5p, for which we found about 175.000 copies, while for miR-146a we extimated about $1x10^6$ copies.

Appendix B

Supervised Ensemble Learning Identifies Minimal Consensus Gene Signatures for Crohn's Disease Classification

Title: Supervised Ensemble Learning Identifies Minimal Consensus Gene Signatures for Crohn's Disease Classification

Authors: Giorgio Montesi[1], Gabriel Dos Santos Mouta[1], Maria Novedrati[1], André F. Cunha[2,3], Alessandro Fuschi[4], Simone Lucchesi[1], Chiara Sonnati[1], Annalisa Ciabattini[1], Francesco Santoro[1], Donata Medaglini[1,*] and Helder I. Nakaya[2,3,5,*].

Affiliations:

[1] Laboratory of Molecular Microbiology and Biotechnology, Department of Medical Biotechnologies, University of Siena; Siena, Italy

[2] Department of Clinical and Toxicological Analyses, School of Pharmaceutical Sciences, University of São Paulo; São Paulo, Brazil

[3] Institut Pasteur de São Paulo; São Paulo, Brazil

[4] Department of Physics and Astronomy, University of Bologna; Bologna, Italy

[5] Hospital Israelita Albert Einstein; São Paulo, Brazil

*To whom correspondence should be addressed:

Prof. Donata Medaglini: Tel.: +39 0577 233307; email: donata.medaglini@unisi.it

Prof. Helder I. Nakaya: Tel: +55(11)2648-1130; email: helder.nakaya@einstein.br

One Sentence Summary: TENTACLES, a machine learning consensus framework, identifies minimal, reproducible gene signatures for Crohn's disease stratification across cohorts.

Abstract:

Background: Crohn's disease (CD) is a complex chronic inflammatory bowel disease driven by heterogeneous immune-mediated mechanisms. Although transcriptomic studies have revealed disease-associated gene expression patterns, identifying reproducible minimal gene signatures across independent cohorts remains challenging due to methodological variability, dataset heterogeneity, and reliance on single-algorithm approaches that may not generalize effectively.

Materials and Methods: We developed a supervised ensemble framework, TENTACLES, that prioritizes robust transcriptomic features through consensus across diverse ML algorithms. The framework was validated using bulk RNA-Seq data spanning three independent CD cohorts (n=515 samples). During training on the first dataset, consensus genes were selected based on agreement across ten ML algorithms. The resulting 28-gene panel was evaluated on a second independent cohort through univariate ranking, dimensionality reduction, and non-linear predictive modeling. A refined 14-gene subset, showing consistent cross-cohort importance, was further validated in a third independent cohort using six unsupervised clustering approaches.

Results: The 28-gene consensus panel achieved robust discrimination between CD and healthy controls across training and testing datasets, demonstrating superior generalizability compared to conventional differential expression approaches. The refined 14-gene subset maintained high discriminatory power in unsupervised analyses. A minimal 5-gene signature (*CSF3R*, *OSM*, *BACE2*, *HK2*, and *IL15RA*) achieved optimal performances across multiple clustering algorithms.

Conclusion: TENTACLES overcomes limitations of single-algorithm approaches providing a generalizable consensus framework for identifying minimal, reproducible gene signatures in heterogeneous transcriptomic datasets. The validated gene panel offers biologically interpretable biomarkers with potential applications in CD diagnosis, therapeutic monitoring, and mechanistic investigation. This framework represents a broadly applicable approach for robust feature discovery in complex immune-mediated diseases.

Main Text:

INTRODUCTION

Crohn's disease (CD) is a chronic and relapsing form of inflammatory bowel disease (IBD) characterized by transmural segmental inflammation that can affect any portion of the gastrointestinal tract*(1)*. CD presents with distinctive discontinuous lesions ("skip lesions") that may involve the entire digestive tract from mouth to anus. Clinical manifestations are highly variable, ranging from abdominal pain and diarrhea to severe extra-intestinal complications, often resulting in diagnostic delays*(2)* and the need for invasive diagnostic procedures*(3, 4)*. These delays carry significant clinical consequences, as progressive tissue damage frequently leads to stricturing, penetrating disease, and the eventual need for surgical intervention in up to 70% of patients*(5)*.

High-throughput RNA sequencing has emerged as a powerful approach for elucidating the molecular mechanisms underlying complex diseases like CD. Transcriptomic analyses of intestinal biopsies and blood samples have successfully identified disease-associated gene expression patterns capable of distinguishing affected individuals from healthy controls. When combined with machine learning (ML) approaches, these datasets have yielded promising multigene signatures with diagnostic and prognostic potential*(4, 6–8)*.

However, current computational pipelines predominantly employ a conventional two-step approach: initial identification of differentially expressed genes (DEGs) followed by training of a single classification model*(7–11)*. This methodology presents several critical limitations. First, it ignores DEG interactions and subtle but consistent multivariate patterns that may be biologically meaningful. Second, it relies on single-algorithm classifiers that are vulnerable to model-specific biases and may yield signatures that perform well on training data but generalize poorly to independent cohorts with different experimental conditions, patient populations, or technical platforms. Third, the arbitrary statistical thresholds used in DEG filtering may inadvertently exclude genes with modest but consistent discriminatory power across multiple analytical approaches. These limitations underscore the need for more robust, generalizable approaches to biomarker discovery in transcriptomic datasets. Ensemble

methods, which aggregate predictions or feature importance across multiple algorithms, have demonstrated superior performance and generalizability in numerous machine learning applications but remain underutilized in transcriptomic biomarker discovery. By leveraging the complementary strengths of diverse algorithms—from linear models sensitive to additive effects to tree-based methods capable of capturing complex interactions—ensemble approaches can identify more stable and reproducible feature sets.

To address these methodological gaps, we developed TENTACLES (Transcriptomic Exploration Tool through Aggregation of Classifiers), a supervised ensemble framework specifically designed for robust feature prioritization in heterogeneous transcriptomic datasets. Unlike conventional approaches that rely on single algorithms, TENTACLES systematically aggregates feature importance rankings across multiple diverse ML classifiers to identify consensus gene signatures that are both predictive and reproducible across independent cohorts. In this study, we applied TENTACLES to three independent CD cohorts encompassing 535 samples from four bulk RNA-Seq data of intestinal biopsies. Our analytical strategy employed a rigorous three-phase validation approach: (1) consensus gene identification through multi-algorithm feature importance aggregation in a training cohort, (2) comprehensive evaluation of generalizability in an independent testing cohort using complementary analytical strategies, and (3) final validation through unsupervised clustering in a third independent cohort. This framework enabled us to systematically refine our initial 28-gene consensus panel to a minimal 5-gene signature while maintaining robust discriminatory performance across all validation phases.

RESULTS

Multi-Algorithm Ensemble Performance Reveals Robust Classification Patterns

Our ensemble framework evaluated ten diverse machine learning algorithms on the training cohort (*PRJNA248469*: 218 CD patients, 42 healthy controls, **Table 1**, **Fig. 1**), encompassing complementary methodological approaches including tree-based methods (Random Forest, C5-rules, Decision Tree), neural networks (MLP, Bagged-MLP), kernel methods (SVM-linear), gradient boosting (LightGBM, XGBoost), regression splines (MARS), and dimensionality reduction (PLS). Performance assessment during cross-validated hyperparameter optimization and subsequent full-dataset refitting revealed distinct algorithmic behaviors and convergent classification patterns (**Fig. 2A** and **Table S1**).

Nine of ten algorithms demonstrated robust and consistent performance, with median accuracy of 0.87 (IQR: 0.85-0.88) and F1-scores of 0.70 (IQR: 0.68-0.73). SVM-linear achieved superior performance (Accuracy=0.91, F1=0.78, Precision=0.64, Recall=1.0), followed closely by tree-based methods—C5-rules, Random Forest, and Decision Tree—all exceeding 0.88 accuracy with F1-scores above 0.72. The consistent improvement from cross-validation to full-dataset refitting across algorithms indicated stable learning dynamics with minimal overfitting risk.

Notably, XGBoost exhibited markedly poor performance (Accuracy=0.43, F1=0.30, Precision=0.19, Recall=0.78), likely reflecting suboptimal hyperparameter configuration or sensitivity to the class imbalance ratio (5.2:1 CD:HC). Sample-level error analysis (**Fig. 2B**) confirmed that the majority of misclassifications originated from XGBoost, while 85% of samples were consistently and correctly classified by all high-performing algorithms. The small subset of samples misclassified by multiple algorithms (approximately 8% of the cohort) likely represents cases with intermediate transcriptomic profiles, possibly reflecting disease heterogeneity, early-stage disease, or technical variability inherent to biological samples.

Consensus Feature Selection Identifies Core Transcriptomic Signatures

Feature importance analysis across the nine well-performing classifiers revealed substantial but structured overlap in gene prioritization (**Fig. 3**). Neural network methods (MLP, Bagged-MLP) contributed the largest feature sets with 265 genes exclusively shared between them, while tree-based methods showed strong convergence with neural networks through multiple intermediate intersections ranging from 8-42 genes. The intersection architecture demonstrated clear algorithmic families, with the largest single intersection comprising 265 genes exclusively shared by the two MLP-based models, followed by a secondary intersection of 42 genes common to MLPs and Random Forest. Intermediate-size intersections of 21-8 genes connected these core methods with PLS, SVM-linear, and C5-rules in various combinations, while smaller intersections involving fewer than four genes displayed high model connectivity, often bridging classifiers from distinct methodological families. Notably, MARS contributed only marginally to the shared feature space, and XGBoost selections appeared largely unique with weak intersections with other classifiers, supporting their exclusion from consensus formation.

Based on these convergence patterns, a 28-gene consensus panel was defined by selecting features deemed important by at least six of the nine high-performing classifiers. The resulting signature encompassed genes with established roles across multiple biological domains critical to CD pathogenesis. The panel included key innate and adaptive immune regulators such as *CTSS*, *CXCL8*, *OSM*, *CSF3R*, *FCGR1A*, *FCGR3A*, and *FCER1A*, which mediate inflammatory cell recruitment, activation, and antibody-dependent cellular responses. Metabolic reprogramming and cellular stress response were represented by *ACSL1*, *HK2*, *HK3*, *CYP2E1*, *CHAC1*, *XBP1*, and *TXNDC5*, reflecting the altered energy metabolism and endoplasmic reticulum stress characteristic of inflamed intestinal tissue.

Signal transduction and cellular communication pathways were captured through *IL15RA*, *SEMA3E*, *WNT5A*, *HCAR2*, and *ADRA1B*, which regulate immune cell survival, tissue remodeling, and neurotransmitter signaling. Transcriptional and structural cellular regulators including *ELL2*, *RPA1*, and *LIMK2* represented the altered gene expression programs and cytoskeletal dynamics associated with tissue inflammation and repair. The consensus panel was completed by additional genes with established or emerging roles in immune and metabolic contexts, including *BACE2*, *C6*, *BATF2*, *TNFAIP6*, *ADCYAP1*, and *AQP9*, many of which have been implicated in recent IBD-focused studies but had not previously been identified through traditional differential expression approaches.

Consensus Signatures Outperform Conventional DEG Approaches

Differential expression analysis of the training cohort identified 582 significantly altered genes (FDR<0.05, $|\log_2 FC|>1$), with only a subset of the 28 consensus genes reaching conventional statistical thresholds (**Fig. 4A**, **Data file S1**). This divergence indicated that ensemble feature selection captured discriminatory signals overlooked by univariate testing.

Principal component analysis revealed clear class separation for both gene sets, with distinct structural characteristics (**Fig. 4B-C**). The consensus panel showed tighter clustering of healthy controls with greater dispersion among CD patients, suggesting capture of disease heterogeneity. Key metabolic genes (*HK2*, *BACE2*, *TXNDC5*) and inflammatory mediators (*AQP9*, *TNFAIP6*, *CSF3R*, *CXCL8*, *OSM*) exhibited high-magnitude PC1 loadings, driving the discriminatory pattern. The DEG-based analysis showed comparable separation but with more evenly distributed gene contributions (**Fig. S1**).

Univariate AUROC analysis identified eight consensus genes (*CXCL8*, *OSM*, *AQP9*, *FCGR3A*, *FCGR1A*, *TNFAIP6*, *HK2*, *ELL2*) achieving near-perfect discrimination (AUROC>0.9), while five genes (*C6*, *SEMA3E*, *RPA1*, *ADRA1B*, *FCER1A*) showed protective associations (AUROC<0.1) (**Fig. 4E**). Notably, discriminatory power was not strictly correlated with fold-change magnitude, emphasizing the value of multivariate selection approaches. Similar patterns were observed in the DEG-based analysis (**Data file S1**).

Multivariate classification confirmed superior efficiency of the consensus approach. The 28-gene MLP classifier achieved excellent performance (Accuracy=0.962, AUROC=0.99, Brier=0.111) with clear feature importance hierarchy led by *CXCL8*, *ELL2*, *FCGR3A*, *CYP2E1*, and *TXNDC5* (**Fig. 4F**). The full DEG-based classifier reached similar but slightly lower performance (Accuracy=0.899, AUROC=0.947, Brier=0.124) with more diffuse variable importance patterns (**Data file S1**).

Coordinated expression analysis of the top-ranking genes *CXCL8* and *OSM* revealed strong correlation in CD samples (R^2=0.87, $p<0.001$) versus moderate correlation in controls (R^2=0.55, $p<0.001$), with CD patients displaying broader expression ranges consistent with inflammatory amplification (**Fig. 4D**). These results demonstrate that the consensus panel achieved comparable discriminatory performance to the full DEG set using 95% fewer features, while maintaining superior biological coherence and interpretability across multiple analytical approaches.

Cross-Cohort Generalization Validates Consensus Signature Robustness

To assess preservation of predictive structure across independent cohorts, the 28-gene panel was evaluated on the merged *PRJNA565216* and *PRJNA985602* datasets. PCA revealed partial separation between CD and healthy controls, consistent in direction with training data (**Fig. S2A**).

Univariate AUROC analysis showed most genes retained their classification trends, with *BACE2*, *HK2*, *WNT5A*, and *ELL2* remaining top performers (**Fig. S2B**). Conversely, *ADRA1B*, *C6*, and *RPA1* continued to associate with healthy controls. AUROC values were moderately reduced compared to training but remained consistently robust for an independent dataset. The full 582-DEG panel showed weaker PCA separation and more diffuse AUROC distributions, supporting superior generalizability of the consensus approach (**Fig. S2C**, **Data file S2**).

Multivariate classification confirmed these trends. The 28-gene MLP classifier achieved AUROC=0.821, accuracy=0.8, and Brier score=0.18 (**Fig. 5**), demonstrating effective generalization. Variable importance analysis identified 14 genes with consistent directional contributions across

datasets and model architectures. In contrast, the full DEG-based classifier showed reduced performance (AUROC=0.723, accuracy=0.6, Brier score=0.199; **Data file S2**).

Independent DEG analysis on the merged test cohort identified 3,687 differentially expressed genes (FDR<0.05, $|\log_2 FC|>1$; **Fig. S3**, **Data file S2**). Intersection with training DEGs revealed a reproducible core of 385 genes. Notably, 17 of the 28 consensus genes belonged to this validated core, including *FCGR3A*, *CSF3R*, *HK2*, *TNFAIP6*, *WNT5A*, *CXCL8*, *ACSL1*, *HK3*, *ADRA1B*, *C6*, *TXNDC5*, *SEMA3E*, *BATF2*, *HCAR2*, *AQP9*, *OSM*, and *BACE2*. This enrichment demonstrates that consensus-based selection captured both predictive and biologically validated disease signatures, including features overlooked by traditional differential expression approaches.

Unsupervised Validation Confirms Minimal Gene Signature Robustness

To evaluate consensus gene panel robustness and identify minimal feature sets, 14 genes with concordant variable importance across training and testing phases were selected: *BACE2*, *HK2*, *HCAR2*, *OSM*, *FCGR1A*, *IL15RA*, *HK3*, *ELL2*, *BATF2*, *RPA1*, *C6*, *FCER1A*, *ADRA1B*, and *CSF3R*. Unsupervised classification was performed on the third independent dataset (*PRJNA702434*) using six clustering algorithms: GMM, hierarchical clustering, k-means, and k-means applied to PCA, t-SNE, and UMAP embeddings. All possible gene combinations were systematically evaluated and ranked by average F1-score.

The top ten gene combinations (C1-C10) demonstrated robust classification performance (**Fig. 6**). Most combinations achieved accuracy, F1-score, and precision above 75%, confirming strong discriminatory capacity under unsupervised conditions. Hierarchical clustering showed lower precision (mean=0.55, IQR=0.44-0.78), while k-means on t-SNE and UMAP embeddings achieved the most consistent results across metrics and combinations.

The best-performing combination (C1) comprised five genes: *CSF3R*, *OSM*, *BACE2*, *HK2*, and *IL15RA*, achieving accuracy=77%, F1=72%, precision=70%, and recall=79%. The second-ranked combination (C2) expanded to nine genes by adding *ELL2*, *ADRA1B*, *RPA1*, and *FCER1A*, maintaining comparably

high performance. Additional high-ranking combinations alternated inclusion of these four supplementary genes alongside the five-gene core. These results independently validate the robustness and biological coherence of consensus-selected transcriptional signatures, demonstrating that a compact 5-9 gene panel is sufficient for accurate CD stratification across both supervised and unsupervised analytical frameworks.

DISCUSSION

This study presents TENTACLES, a consensus-based machine learning framework designed to prioritize robust and reproducible transcriptional features across heterogeneous RNA-seq datasets. By aggregating outputs from multiple ML classifiers and applying rigorous validation across three independent CD cohorts, TENTACLES enables extraction of compact, generalizable gene signatures that are resilient to dataset-specific biases and modeling assumptions.

Unlike traditional univariate differential expression analyses that evaluate genes in isolation, TENTACLES integrates multiple supervised learning algorithms to identify features carrying predictive power across models and datasets. This multivariate consensus strategy effectively captures non-linear, weakly expressed, or combinatorially active features that conventional statistical thresholds often miss*(12)*. The framework also incorporates mechanisms to handle imbalanced class distributions, a frequent challenge in clinical datasets. Notably, none of the 28 consensus-prioritized genes appeared in the KEGG IBD pathway annotation*(13)*, and only 17 overlapped with genes consistently differentially expressed across cohorts. These findings highlight the complementary value of data-driven prioritization frameworks, particularly for diseases with complex, multi-axis pathophysiology such as CD.

A central strength of TENTACLES lies in its ability to consistently extract stable transcriptional signals across fully independent cohorts, encompassing diverse experimental conditions and patient populations. The 28-gene signature prioritized during training maintained substantial discriminative power in external testing datasets, with 14 genes preserving concordant variable importance and directionality. This refined subset was subsequently validated in a third independent cohort through unsupervised clustering, confirming stratification capability without prior disease knowledge. Among the most stable features, a minimal 5-gene panel comprising *BACE2*, *HK2*, *OSM*, *IL15RA*, and *CSF3R* emerged as sufficient for robust CD classification across analytical frameworks.

Each gene in the minimal signature has established roles in immune regulation and metabolic dysfunction relevant to CD pathogenesis. *OSM*, a cytokine regulator, drives epithelial barrier disruption

and anti-TNF therapy resistance*(14–16)*. *HK2* gene, encoding a key glycolytic enzyme, correlates with inflammatory severity in IBD patients*(17)*. *IL15RA* promotes immune cell survival and proliferation, with documented upregulation in IBD*(18, 19)*. *BACE2*, a membrane protease, represents an emerging IBD biomarker*(6, 20)*. Finally, *CSF3R*, which regulates neutrophil homeostasis, shows altered expression patterns suggesting disrupted immune cell trafficking in CD*(21, 22)*. This biological coherence, combined with robust statistical validation, strengthens confidence in the clinical relevance of these signatures.

The identified gene panels offer multiple translational applications. The 5-gene minimal signature could serve as a diagnostic classifier requiring fewer resources than comprehensive transcriptomic profiling. The 14-gene refined panel provides additional granularity for molecular subtyping and therapeutic stratification. Both signatures demonstrated consistent performance across supervised and unsupervised analytical frameworks, suggesting utility for diverse clinical applications including disease monitoring, treatment response prediction, and patient stratification for clinical trials.

Some limitations merit consideration. The analysis relied on publicly available bulk RNA-seq datasets with inherent experimental and clinical heterogeneity. Despite rigorous batch correction and quality control measures, residual confounding cannot be entirely excluded. Future studies should validate these signatures in prospectively collected cohorts with standardized protocols. Additionally, single-cell RNA sequencing could provide insights into cell-type-specific expression patterns underlying the bulk tissue signatures. Integration with clinical metadata, including disease phenotypes, treatment responses, and longitudinal outcomes, would further enhance the clinical utility of these biomarkers.

Beyond Crohn's disease, TENTACLES represents a broadly applicable approach for robust biomarker discovery in complex diseases. The framework's modular design enables adaptation to different data types, disease contexts, and clinical objectives. Its emphasis on consensus-based feature selection and multi-cohort validation addresses persistent reproducibility challenges in computational biology and precision medicine. As transcriptomic datasets continue expanding, such ensemble approaches will

become increasingly valuable for extracting reliable, clinically actionable insights from heterogeneous biological data.

This study demonstrates that TENTACLES effectively identifies minimal, reproducible gene signatures for Crohn's disease classification through consensus-based machine learning. The resulting biomarker panels offer biologically interpretable features with potential applications in diagnosis, molecular stratification, and therapeutic monitoring. This framework represents a scalable, generalizable approach for robust feature discovery that could advance precision medicine efforts across immune-mediated diseases and beyond.

MATERIALS AND METHODS

Acquisition and preprocessing of public RNA-Seq data from CD intestinal samples

RNA sequencing datasets related to CD were obtained from the Sequence Read Archive (SRA). Specifically, intestinal biopsy samples from the following BioProjects were downloaded: *PRJNA248469(23, 24)* (composing Dataset 1), *PRJNA565216(25)* and *PRJNA985602(26)* (composing Dataset 2) and *PRJNA702434(27)* (composing Dataset 3). Inclusion criteria were restricted to datasets originating from healthy controls and patients diagnosed with CD. All 535 resulting samples were uniformly processed using a standardized custom pipeline to ensure consistency across datasets. Quality control was performed with **fastp***(28)* (v1.0.1), followed by alignment to the human reference genome GRCh38.p14 (NCBI RefSeq assembly GCF_000001405.40) using **STAR***(29)* (v2.7.10) with default parameters. Read strandedness was assessed with **RSeQC***(30)* (v5.0.1), and transcript quantification was carried out using **Rsubread***(31)* (v2.22.1). To avoid non-interpretable results and reduce computation time, genes not annotated in GO or KEGG databases were filtered out before running the classification pipeline.

Machine Learning Analysis

A three-phase strategy involving separate training, testing and validation was applied to three different dataset combinations independently pre-processed following the same workflow (**Fig. 1**). Data preprocessing was performed through normalization with the **edgeR***(32)* (v4.4.2) package and batch-effect correction with the **sva***(33)* (v3.54.0) package. The quality of the pre-processed data was examined using Principal Component Analysis (PCA) and Principal Variance Component Analysis (PVCA; **pvca***(34)* v1.46.0) plots. Genes with less than 1 CPM in 1/10 of the samples were filtered out.

Training Phase

The training phase, applied to the study *PRJNA248469*, comprised sequential steps that included feature selection and hyperparameter tuning, model fitting with ten machine learning classifiers, and consensus-based gene selection. Feature selection began with feature scaling and the removal of near-zero variance

features using the **recipes***(35)* (v1.3.1) package, followed by class balancing through downsampling with the **themis***(36)* (v.1.0.3) package. This step, performed within each fold of the cross-validation, was then refined using three alternative approaches: correlation filtering with a threshold of 0.95, ROC-based ranking with a ROC threshold of 0.95 implemented in the **colino***(37)* (v0.0.1) package, and the Boruta algorithm (**Boruta***(38)* v8.0.0) with a maximum of 100 iterations.

The classification stage involved Random Forest (**ranger***(39)* v0.17.0), Multi-Layer Perceptron (MLP) and Bagged-MLP (**nnet***(40)* v7.3-20), Support Vector Machine with a linear kernel (SVM; **kernlab***(41)* v0.9-33), C5-rules (**C50***(42)* v0.2.0), LightGBM (**lightgbm***(43)* v4.6.0), Decision Tree (**rpart***(44)* v4.1.24), Multivariate Adaptive Regression Splines (MARS; **earth***(45)* v5.3.4), Extreme Gradient Boosting (XGBoost; **xgboost***(46)* v1.7.11.1), and Partial Least Squares (PLS; **mixOmics***(47)* v6.30.0). Each classifier was paired with a specific feature selection strategy: correlation-based filtering was applied to Random Forest; ROC-based selection was used for MLP, Bagged-MLP, and LightGBM; Boruta was applied to SVM, C5-rules, Decision Tree, MARS, and PLS; XGBoost was run without additional feature selection.

Model parameters were optimized independently for each classifier through a grid-search procedure with 5-fold cross-validation implemented in the **workflowsets***(48)* (v1.1.1) package. The following hyperparameters were tuned: Random Forest (number of variables randomly sampled at each split, *mtry*; minimum node size, *min_n*), MLP and Bagged-MLP (number of hidden units, weight decay penalty, number of training epochs), linear SVM (regularization cost, margin parameter), C5-rules (number of boosting iterations, *min_n*), LightGBM (*mtry*, number of trees, *min_n*, maximum tree depth, learning rate, minimum loss reduction required for further partitioning), Decision Tree (maximum tree depth, *min_n*), MARS (number of terms, maximum interaction degree), XGBoost (*mtry*, number of trees, *min_n*, maximum tree depth, learning rate, minimum loss reduction, subsample size, early stopping rounds), and PLS (number of components, proportion of predictors used). The optimal configuration for each model was selected based on the highest F1-score across folds.

Model performance was assessed using Accuracy, F1-score, Precision, and Recall, calculated against the true class labels. Variable importance was estimated with the **vip***(49)* (v0.4.1) package, and the consensus-based selection retained only genes identified as important by at least six classifiers.

Testing Phase

The testing phase was performed on a merged dataset derived from two independent studies, *PRJNA565216* and *PRJNA985602*, with batch effects accounted for by including the study identifier as a batch variable during preprocessing. This step was conducted using the consensus gene set and comprised PCA with loading analysis, evaluation of AUROC and fold-change with the **pROC***(50)* (v1.18.5) package, and the implementation of a Multi-Layer Perceptron (MLP) classifier. For the MLP analysis, the dataset was divided into a training set containing 70% of the samples and a testing set containing the remaining 30%. The training set was used for hyperparameter tuning of the number of hidden units, number of epochs, and penalty values, employing a 3-fold cross-validation and a grid search of ten parameter combinations with the **tune***(51)* (v1.3.0) package. The optimal configuration was selected based on the highest ROC value. The remaining 30% testing set was then used to evaluate model performance in terms of AUROC, Accuracy, and Brier score, and to compute variable importance scores.

Validation Phase

The validation phase was designed to identify the optimal gene combinations among all possible subsets of consensus genes that maximized class separability in the validation dataset, *PRJNA702434*. To this aim, six unsupervised clustering algorithms were applied, each with the number of clusters set to two. Gaussian Mixture Models (GMM; **mclust***(52)* v6.1.1) and Hierarchical Clustering (HiCl) were performed on normalized gene counts, while K-Means clustering was additionally applied to data projected onto dimensionality reduction embeddings, including PCA, Uniform Manifold Approximation and Projection (UMAP; **umap***(53)* v0.2.10.0), and t-distributed Stochastic Neighbor Embedding (t-SNE; **Rtsne***(54, 55)* v0.17). Clustering performance was evaluated by comparing the resulting partitions with the true class labels in terms of Accuracy, Precision, Recall and F1-score.

Differential Gene Expression Analysis

As a complementary approach, differential gene expression analysis was performed independently on both the training dataset (*PRJNA248469*) and the merged external test dataset (*PRJNA565216* and *PRJNA985602)*. For both datasets, samples with low total read counts (<500,000) were excluded to ensure sufficient coverage, and only genes with 1 CPM in more than 1/10 of the samples were retained to reduce noise from lowly expressed transcripts.

Differential expression analysis was performed using the **DESeq2***(56)* R package (version 1.46.0). **DESeq2** was used to normalize the count data and to compute gene-wise statistics using a negative binomial generalized linear model. Normalized counts were obtained using the variance-stabilizing transformation (VST) for downstream visualization and comparative analysis. For the training dataset, a **DESeq2** object was created using the class variable (CD vs HC) as the design formula. In contrast, the merged test dataset included samples from multiple studies, and the design formula was adjusted to include the study id as batch correction. The resulting differential expression results were filtered for multiple testing using the Benjamini–Hochberg method, and significantly differentially expressed genes were defined based on an adjusted p-value < 0.05.

List of Supplementary Materials:

- Fig. S1 to Fig.S3
- Table S1
- Data File S1 to S2
- Supplementary File 1

References

1. C. McDowell, U. Farooq, M. Haseeb, in *StatPearls*, (StatPearls Publishing, Treasure Island (FL), 2025).

2. S. S. Seyedian, F. Nokhostin, M. D. Malamir, A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. *J Med Life* **12**, 113–122 (2019).

3. S. Nikolaus, S. Schreiber, Diagnostics of inflammatory bowel disease. *Gastroenterology* **133**, 1670–1689 (2007).

4. M. H. D. Nazari, M. Ghorbaninejad, S. Shahrokh, A. Meyfour, Biomarker discovery for non-invasive diagnosis of inflammatory bowel disease using blood transcriptomics. *Front. Immunol.* **16** (2025), doi:10.3389/fimmu.2025.1570374.

5. K. M. Scheurlen, M. A. Parks, A. Macleod, S. Galandiuk, Unmet Challenges in Patients with Crohn's Disease. *Journal of Clinical Medicine* **12**, 5595 (2023).

6. A. H. Syed, H. A. S. Abujabal, S. Ahmad, S. J. Malebary, N. Alromema, Advances in Inflammatory Bowel Disease Diagnostics: Machine Learning and Genomic Profiling Reveal Key Biomarkers for Early Detection. *Diagnostics (Basel)* **14**, 1182 (2024).

7. W. Bao, L. Wang, X. Liu, M. Li, Predicting diagnostic biomarkers associated with immune infiltration in Crohn's disease based on machine learning and bioinformatics. *European Journal of Medical Research* **28**, 255 (2023).

8. K. A. Chen, N. C. Nishiyama, M. M. Kennedy Ng, A. Shumway, C. U. Joisa, M. R. Schaner, G. Lian, C. Beasley, L.-C. Zhu, S. Bantumilli, M. R. Kapadia, S. M. Gomez, T. S. Furey, S. Z. Sheikh, Linking gene expression to clinical outcomes in pediatric Crohn's disease using machine learning. *Sci Rep* **14**, 2667 (2024).

9. S. N. A. Shah, R. Parveen, Differential gene expression analysis and machine learning identified structural, TFs, cytokine and glycoproteins, including SOX2, TOP2A, SPP1, COL1A1, and TIMP1 as potential drivers of lung cancer. *Biomarkers* **30**, 200–215 (2025).

10. M. Abbas, Y. EL-Manzalawy, Machine learning based refined differential gene expression analysis of pediatric sepsis. *BMC Med Genomics* **13**, 122 (2020).

11. H. Ghazal, E.-S. A. El-Absawy, W. Ead, M. E. Hasan, Machine learning-guided differential gene expression analysis identifies a highly-connected seven-gene cluster in triple-negative breast cancer. *Biomedicine (Taipei)* **14**, 15–35.

12. P. Liang, W. Yang, X. Chen, C. Long, L. Zheng, H. Li, Y. Zuo, Machine Learning of Single-Cell Transcriptome Highly Identifies mRNA Signature by Comparing F-Score Selection with DGE Analysis. *Molecular Therapy Nucleic Acids* **20**, 155–163 (2020).

13. KEGG PATHWAY: Inflammatory bowel disease - Reference pathway (available at https://www.kegg.jp/pathway/map05321).

14. Y. Yang, K.-Z. Fu, G. Pan, Role of Oncostatin M in the prognosis of inflammatory bowel disease: A meta-analysis. *World J Gastrointest Surg* **16**, 228–238 (2024).

15. S. Verstockt, B. Verstockt, K. Machiels, M. Vancamelbeke, M. Ferrante, I. Cleynen, G. De Hertogh, S. Vermeire, Oncostatin M Is a Biomarker of Diagnosis, Worse Disease Prognosis, and Therapeutic Nonresponse in Inflammatory Bowel Disease. *Inflamm Bowel Dis* **27**, 1564–1575 (2021).

16. N. R. West, A. N. Hegazy, B. M. J. Owens, S. J. Bullers, B. Linggi, S. Buonocore, M. Coccia, D. Görtz, S. This, K. Stockenhuber, J. Pott, M. Friedrich, G. Ryzhakov, F. Baribaud, C. Brodmerkel, C. Cieluch, N. Rahman, G. Müller-Newen, R. J. Owens, A. A. Kühl, K. J. Maloy, S. E. Plevy, S. Keshav, S. P. L. Travis, F. Powrie, Oncostatin M drives intestinal inflammation

and predicts response to tumor necrosis factor–neutralizing therapy in patients with inflammatory bowel disease. *Nat Med* **23**, 579–589 (2017).

17. S. Weber-Stiehl, J. Taubenheim, L. Järke, C. Röcken, S. Schreiber, K. Aden, C. Kaleta, P. Rosenstiel, F. Sommer, Hexokinase 2 expression in apical enterocytes correlates with inflammation severity in patients with inflammatory bowel disease. *BMC Medicine* **22**, 490 (2024).

18. M. A. Silva, J. Menezes, C. Deslandres, E. G. Seidman, Anti-Inflammatory Role of Interleukin-15 in Crohn's Disease. *Inflamm Bowel Dis* **11**, 219–230 (2005).

19. C. Perrier, I. Arijs, D. Staelens, C. Breynaert, I. Cleynen, K. Covens, M. Ferrante, G. Van Assche, S. Vermeire, G. de Hertogh, F. Schuit, P. Rutgeerts, J. L. Ceuppens, Interleukin-15 receptor α expression in inflammatory bowel disease patients before and after normalization of inflammation with infliximab. *Immunology* **138**, 47–56 (2013).

20. F. Hu, L. Xiong, Z. Li, L. Li, L. Wang, X. Wang, X. Zhou, Y. Zheng, Deciphering the shared mechanisms of Gegen Qinlian Decoction in treating type 2 diabetes and ulcerative colitis via bioinformatics and machine learning. *Front. Med.* **11** (2024), doi:10.3389/fmed.2024.1406149.

21. J. J. Ashton, K. Boukas, J. Davies, I. S. Stafford, A. F. Vallejo, R. Haggarty, T. A. F. Coelho, A. Batra, N. A. Afzal, B. Vadgama, A. P. Williams, R. M. Beattie, M. E. Polak, S. Ennis, Ileal Transcriptomic Analysis in Paediatric Crohn's Disease Reveals IL17- and NOD-signalling Expression Signatures in Treatment-naïve Patients and Identifies Epithelial Cells Driving Differentially Expressed Genes. *J Crohns Colitis* **15**, 774–786 (2021).

22. S. Carnevale, A. Ponzetta, A. Rigatelli, R. Carriero, S. Puccio, D. Supino, G. Grieco, P. Molisso, I. Di Ceglie, F. Scavello, C. Perucchini, F. Pasqualini, C. Recordati, C. Tripodo, B. Belmonte, A. Mariancini, P. Kunderfranco, G. Sciumè, E. Lugli, E. Bonavita, E. Magrini, C. Garlanda, A. Mantovani, S. Jaillon, Neutrophils Mediate Protection Against Colitis and

Carcinogenesis by Controlling Bacterial Invasion and IL22 Production by γδ T Cells. *Cancer Immunol Res* **12**, 413–426 (2024).

23. Y. Haberman, T. L. Tickle, P. J. Dexheimer, M.-O. Kim, D. Tang, R. Karns, R. N. Baldassano, J. D. Noe, J. Rosh, J. Markowitz, M. B. Heyman, A. M. Griffiths, W. V. Crandall, D. R. Mack, S. S. Baker, C. Huttenhower, D. J. Keljo, J. S. Hyams, S. Kugathasan, T. D. Walters, B. Aronow, R. J. Xavier, D. Gevers, L. A. Denson, Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest* **124**, 3617–3633 (2014).

24. N. Loberman-Nachum, K. Sosnovski, A. Di Segni, G. Efroni, T. Braun, M. BenShoshan, L. Anafi, C. Avivi, I. Barshack, D. S. Shouval, L. A. Denson, A. Amir, R. Unger, B. Weiss, Y. Haberman, Defining the Celiac Disease Transcriptome using Clinical Pathology Specimens Reveals Biologic Pathways and Supports Diagnosis. *Sci Rep* **9**, 16163 (2019).

25. A. Mo, C. Krishnakumar, D. Arafat, T. Dhere, H. Iskandar, A. Dodd, J. Prince, S. Kugathasan, G. Gibson, African Ancestry Proportion Influences Ileal Gene Expression in Inflammatory Bowel Disease. *Cell Mol Gastroenterol Hepatol* **10**, 203–205 (2020).

26. A. Garrido-Trigo, A. M. Corraliza, M. Veny, I. Dotti, E. Melón-Ardanaz, A. Rill, H. L. Crowell, Á. Corbí, V. Gudiño, M. Esteller, I. Álvarez-Teubel, D. Aguilar, M. C. Masamunt, E. Killingbeck, Y. Kim, M. Leon, S. Visvanathan, D. Marchese, G. Caratù, A. Martin-Cardona, M. Esteve, I. Ordás, J. Panés, E. Ricart, E. Mereu, H. Heyn, A. Salas, Macrophage and neutrophil heterogeneity at single-cell spatial resolution in human inflammatory bowel disease. *Nat Commun* **14**, 4506 (2023).

27. M. Friedrich, M. Pohin, M. A. Jackson, I. Korsunsky, S. J. Bullers, K. Rue-Albrecht, Z. Christoforidou, D. Sathananthan, T. Thomas, R. Ravindran, R. Tandon, R. S. Peres, H. Sharpe, K. Wei, G. F. M. Watts, E. H. Mann, A. Geremia, M. Attar, Oxford IBD Cohort Investigators, Roche Fibroblast Network Consortium, S. McCuaig, L. Thomas, E. Collantes,

H. H. Uhlig, S. N. Sansom, A. Easton, S. Raychaudhuri, S. P. Travis, F. M. Powrie, IL-1-driven stromal-neutrophil interactions define a subset of patients with inflammatory bowel disease that does not respond to therapies. *Nat Med* **27**, 1970–1981 (2021).

28. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

29. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

30. L. Wang, S. Wang, W. Li, RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).

31. Y. Liao, G. K. Smyth, W. Shi, The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* **47**, e47 (2019).

32. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

33. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).

34. Bioconductor - pvca (available at https://bioconductor.org/packages/release/bioc/html/pvca.html).

35. M. Kuhn, H. Wickham, E. Hvitfeldt, P. Software, PBC [cph, fnd, recipes: Preprocessing and Feature Engineering Steps for Modeling (2025) (available at https://cran.r-project.org/web/packages/recipes/index.html).

36. E. Hvitfeldt, P. Software, PBC, themis: Extra Recipes Steps for Dealing with Unbalanced Data (2025) (available at https://cran.r-project.org/web/packages/themis/index.html).

37. S. Pawley, stevenpawley/colino (2025) (available at https://github.com/stevenpawley/colino).

38. M. B. Kursa, W. R. Rudnicki, Feature Selection with the Boruta Package. *Journal of Statistical Software* **36**, 1–13 (2010).

39. M. N. Wright, A. Ziegler, ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statistical Software* **77**, 1–17 (2017).

40. B. Ripley, W. Venables, nnet: Feed-Forward Neural Networks and Multinomial Log-Linear Models (2025) (available at https://cran.r-project.org/web/packages/nnet/index.html).

41. A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis, kernlab - An S4 Package for Kernel Methods in R. *Journal of Statistical Software* **11**, 1–20 (2004).

42. M. Kuhn, S. Weston, M. Culp, N. Coulter, R. Q. (Author of imported C. code), R. R. (Copyright holder of imported C. code), R. R. P. L. (Copyright holder of imported C. code), C50: C5.0 Decision Trees and Rule-Based Models (2025) (available at https://cran.r-project.org/web/packages/C50/index.html).

43. Y. Shi, G. Ke, D. Soukhavong, J. Lamb, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, N. Titov, Y. Yan, M. Corporation, Dropbox, Inc, A. Ferreira, D. Lemire, V. Zverovich, I. B. M. Corporation, D. Cortes, M. Mayer, lightgbm: Light Gradient Boosting Machine (2025) (available at https://cran.r-project.org/web/packages/lightgbm/index.html).

44. T. Therneau, B. Atkinson, B. R. (producer of the initial R. port, maintainer 1999-2017), rpart: Recursive Partitioning and Regression Trees (2025) (available at https://cran.r-project.org/web/packages/rpart/index.html).

512 45. S. Milborrow, T. Hastie, R. Tibshirani, A. Miller, T. Lumley, earth: Multivariate Adaptive
513 Regression Splines (2024) (available at https://cran.r-
514 project.org/web/packages/earth/index.html).

515 46. T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, H. Cho, K. Chen, R. Mitchell, I. Cano,
516 T. Zhou, M. Li, J. Xie, M. Lin, Y. Geng, Y. Li, J. Yuan, Xgb. contributors (base Xgb.
517 implementation), xgboost: Extreme Gradient Boosting (2025) (available at https://cran.r-
518 project.org/web/packages/xgboost/index.html).

519 47. F. Rohart, B. Gautier, A. Singh, K.-A. L. Cao, mixOmics: An R package for 'omics feature
520 selection and multiple data integration. *PLOS Computational Biology* **13**, e1005752 (2017).

521 48. H. Frick, M. Kuhn, S. Couch, P. Software, PBC [cph, fnd, workflowsets: Create a Collection
522 of "tidymodels" Workflows (2025) (available at https://cran.r-
523 project.org/web/packages/workflowsets/index.html).

524 49. B. Greenwell M., B. Boehmke C., Variable Importance Plots—An Introduction to the vip
525 Package. *The R Journal* **12**, 343 (2020).

526 50. X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, M. Müller, pROC: an
527 open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*
528 **12**, 77 (2011).

529 51. M. Kuhn [aut, cre, P. Software, PBC, tune: Tidy Tuning Tools (2025) (available at
530 https://cran.r-project.org/web/packages/tune/index.html).

531 52. L. Scrucca, C. Fraley, T. B. Murphy, A. E. Raftery, *Model-based clustering, classification,*
532 *and density estimation using mclust in R* (CRC Press, Boca Raton, FL, First edition., 2023).

533 53. T. Konopka, umap: Uniform Manifold Approximation and Projection (2023) (available at
534 https://cran.r-project.org/web/packages/umap/index.html).

54. L. J. P. van der Maaten, G. E. Hinton, Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

55. L. van der Maaten, Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* **15**, 3221–3245 (2014).

56. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).

Acknowledgments: The authors thank Simone Furini (Alma Mater Studiorum – University of Bologna) for fruitful discussions and insightful comments. The authors are also grateful to Mattia Granchi for designing the logo of the TENTACLES package.

Funding:

This work was supported by:
Department of Medical Biotechnologies of the University of Siena
NextGenerationEU project PNRR MUR Extended Partnership Initiative on Emerging Infectious Diseases project(PNRR PE13 INF_ACT—CUP:B63C22001400007)
Calmette & Yersin PhD Grant from the Pasteur Network (AFC).

Author Contributions:

Conceptualization: GM, MN, HIN
Data curation: AFC, GM
Formal analysis: GM, GDSM, AFC, AF, SL, CS
Methodology: GM, HIN, GDSM, AFC, MN
Project Administration: HIN, DM, AC, FS
Resources: HIN, AC, DM
Software: GM, GDSM, AFC, MN, AF
Supervision: DM, HIN
Validation: GM, GDSM, AFC, MN, SL
Writing – original draft: GM, MN, HIN, GDSM, AFC, AF, SL, CS, AC, FS, DM
All authors read and approved the final manuscript

Competing Interests:

Authors declare they have no Competing Interests.

Data and materials availability:

All datasets analyzed in this study have been deposited on Zenodo and are publicly available at https://zenodo.org/records/17482307. The full code used to perform the analyses is provided in **Supplementary File 1**. The TENTACLES R package, including all core functions and vignettes, is available on GitHub at https://github.com/Giomu/TENTACLES.

Figures

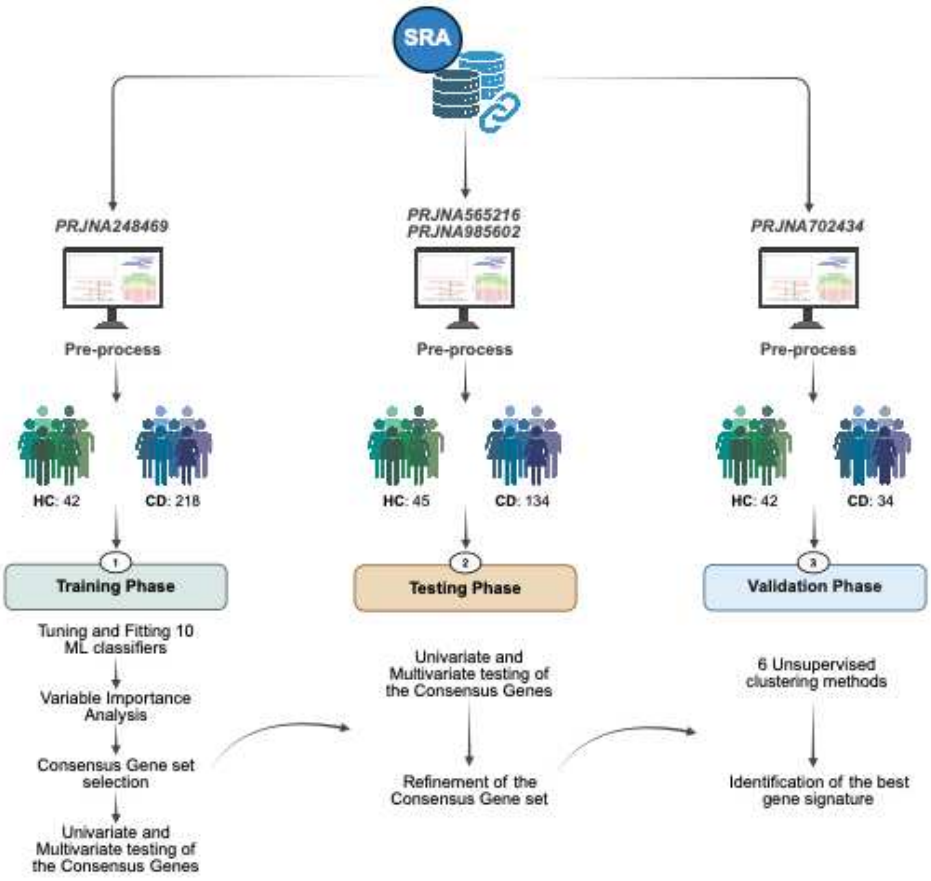


Fig. 1. TENTACLES workflow for gene-signature selection and validation in Crohn's disease. The pipeline has three stages: (1) Training & consensus gene selection on Dataset 1 (*PRJNA248469*; HC = 42, CD = 218) with uniform read-level QC and preprocessing, training of multiple ML classifiers, variable-importance analysis, and derivation of a consensus gene signature; (2) Testing on Dataset 2 derived by merging two independent cohorts —*PRJNA565216* and *PRJNA985602* (total HC = 45, total CD = 134)— for gene-signature evaluation, including univariate and multivariate analyses to refine the signature; (3) External validation on Dataset 3 (*PRJNA702434*; HC = 42, CD = 34) using the same QC/preprocessing pipeline and final performance assessment through different methods of unsupervised clustering. *PRJNA* identifiers denote NCBI SRA BioProjects. HC, healthy controls; CD, Crohn's disease.

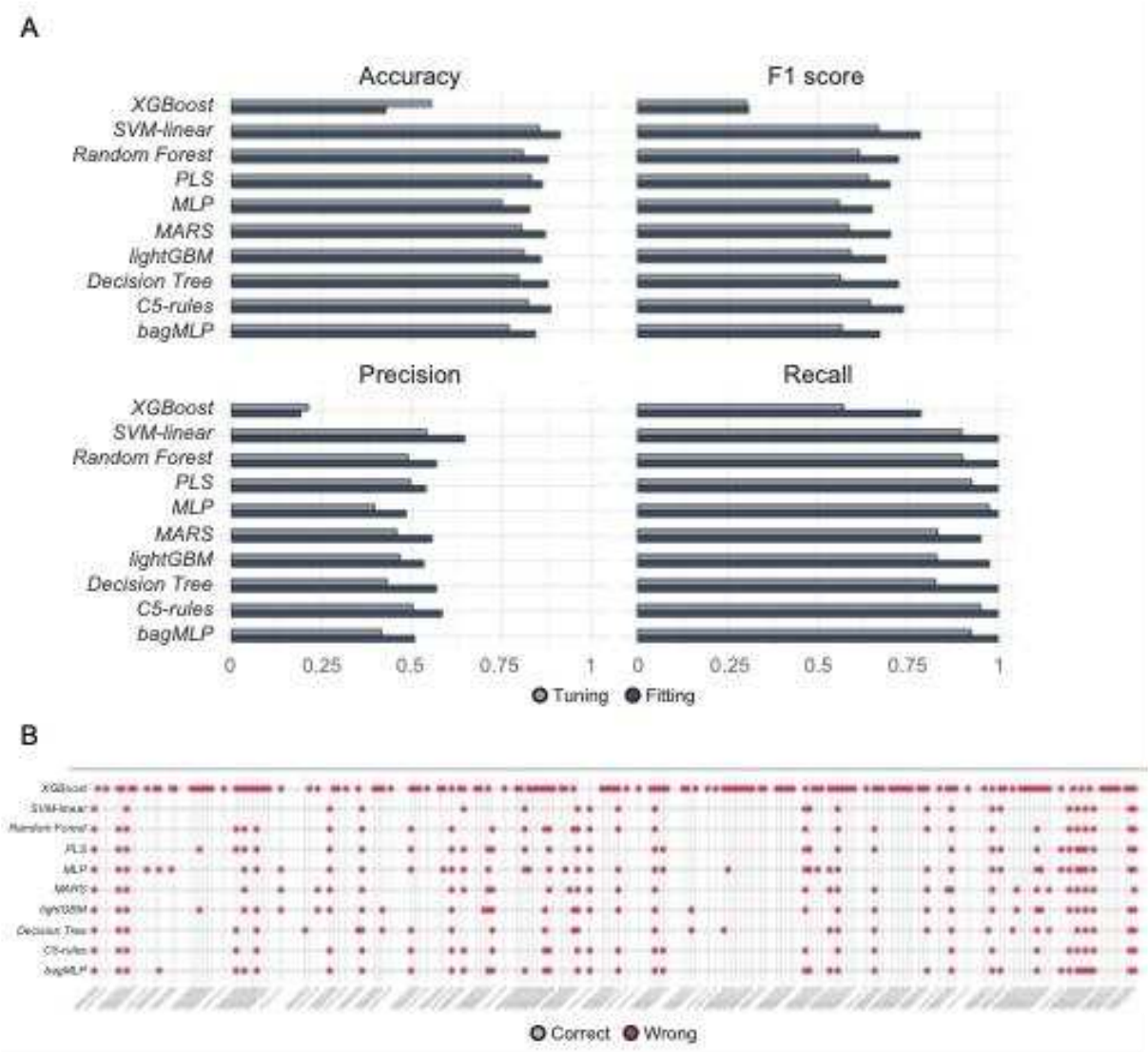


Fig. 2. Performance benchmarking of ten machine learning classifiers on the training dataset. (A) Barplots reporting cross-validated tuning and full-model fitting performance for ten supervised classifiers —Random Forest, Multi-Layer Perceptron (MLP), bagMLP (Bagged-MLP), linear Support Vector Machine (SVM-linear), C5-rules, LightGBM, Decision Tree, Multivariate Adaptive Regression Splines (MARS), XGBoost, and Partial Least Squares (PLS)– evaluated on the training dataset (CD = 218, HC = 42). Metrics include Accuracy, F1-score, Precision, and Recall. All models were trained with optimized hyperparameters and the feature selection strategy specified in Table S1. Light bars ("Tuning") summarize cross-validated performance during hyperparameter search; dark bars ("Fitting") show performance after refitting the selected configuration on the full training set. **(B)** Error sample-level concordance across models. Each column represents a sample and each row a classifier. Red and grey dots indicate misclassified and correctly classified samples, respectively. Most errors originated from the XGBoost model, while the majority of samples were robustly predicted by all other classifiers**.**

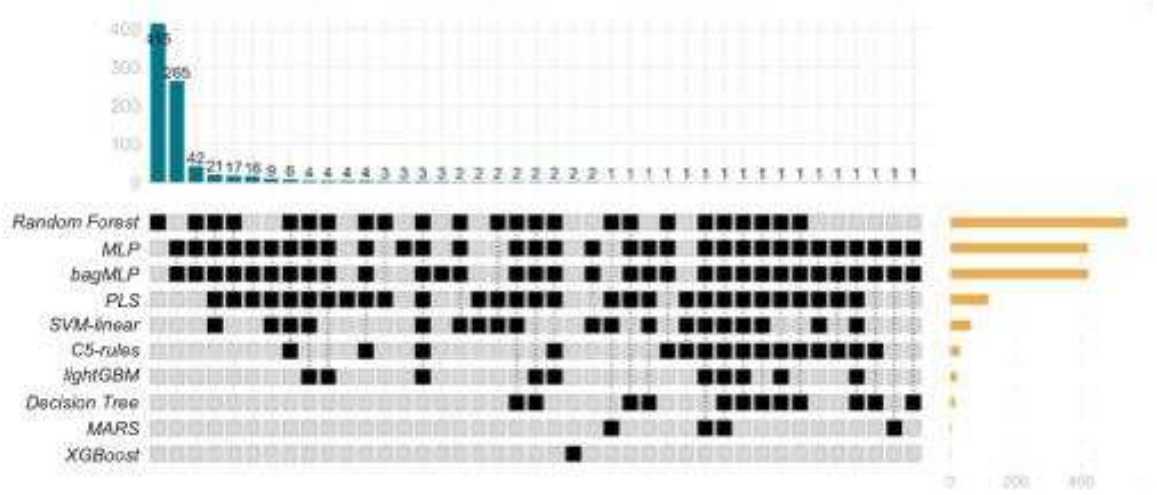


Fig. 3. Consensus-based gene selection UpSet plot of important variables of the ten classifiers. UpSet plot summarizing overlap among genes prioritized by each of the ten classifiers –Random Forest, Multi-Layer Perceptron (MLP), bagMLP (Bagged-MLP), linear Support Vector Machine (SVM-linear), C5-rules, LightGBM, Decision Tree, Multivariate Adaptive Regression Splines (MARS), XGBoost, and Partial Least Squares (PLS)– based on variable-importance analyses; Bars indicate the size of intersecting gene sets; rows correspond to individual models, with black cells denoting model inclusion in the intersection.

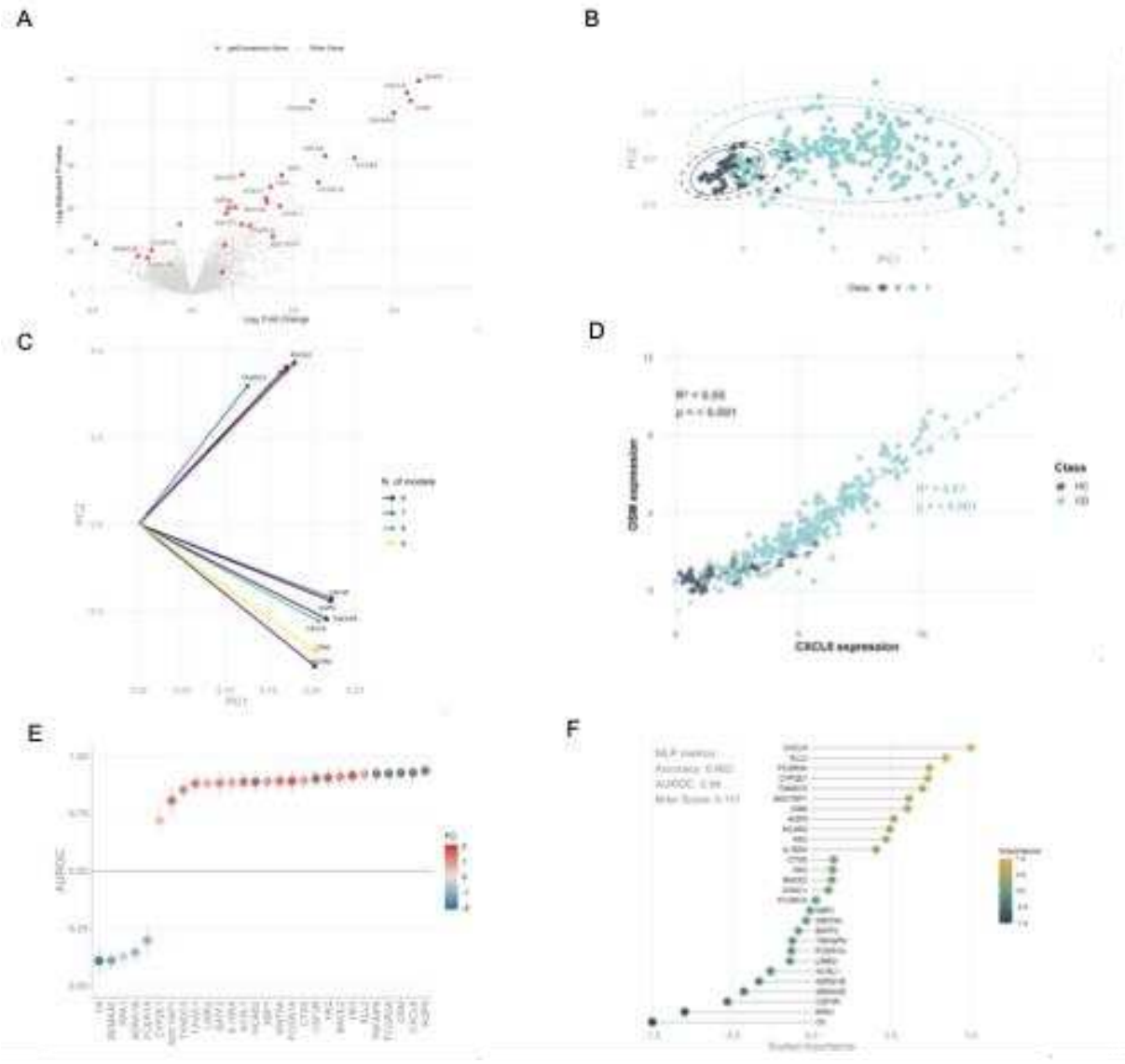


Fig. 4. DEG Analysis and univariate and multivariate analysis of the 28-consensus gene panel on the training dataset. **(A)** Volcano plot displaying differentially expressed genes (DEGs) between CD patients and healthy controls. Red points highlight genes from the consensus 28-gene panel. **(B)** Principal Component Analysis (PCA) of the training cohort using only the 28 consensus genes, showing separation between Healthy and IBD along the first components (points colored by class; axes report variance explained). **(C)** PCA loading plot for the 28 genes; arrow length reflects loading magnitude, and color encodes how many classifiers selected each gene in the training stage (higher values indicate recurrent selection across models). **(D)** Pairwise scatterplot of CXCL8 versus OSM expression (log-normalized counts); points are colored by class. **(E)** Per-gene discriminative performance reported as AUROC, with point color indicating log2 fold change (CD vs Healthy). **(F)** Performance of an MLP trained on the 28-gene panel (Accuracy 0.962, AUROC 0.99, Brier score 0.111) and variable-importance ranking for the same model (higher bars indicate stronger contribution to the classifier's decision).

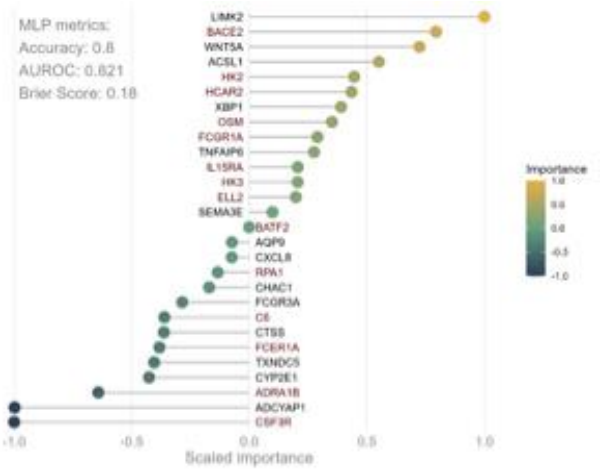


Fig. 5. External testing of the 28-gene consensus panel on an independent dataset. Performance and variable importance of the multilayer perceptron (MLP) classifier trained on the 28 genes and tested on the external dataset. Reported metrics include accuracy, AUROC, and Brier Score. The plot on the right displays the scaled variable importance values for all 28 genes, with color indicating the direction and magnitude of contribution. Genes highlighted in red represent the 14-gene signature showing concordant importance direction across both training and testing phases.

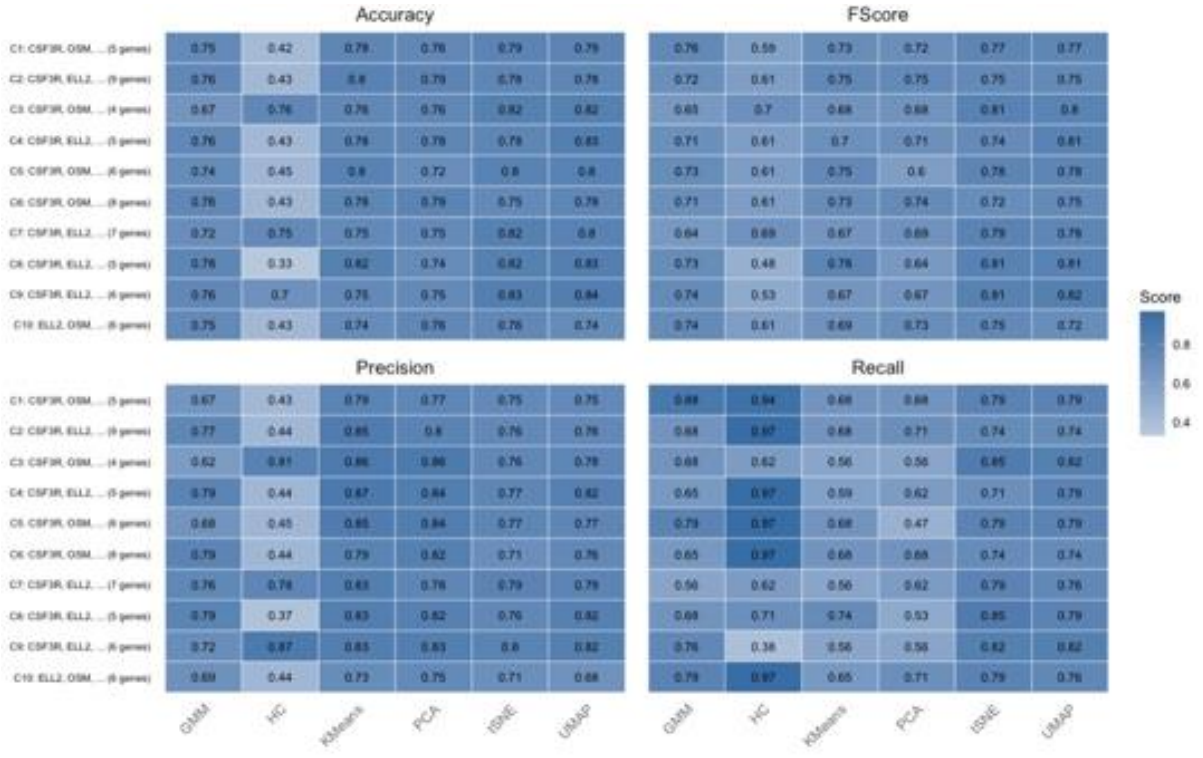


Fig. 6. Performance metrics of the top 10 gene combinations derived from the 14-gene refined signature across six unsupervised clustering methods on an independent validation dataset. Each heatmap panel reports a different evaluation metric: Accuracy, F1-score, Precision, and Recall. Rows correspond to the top 10 gene combinations (C1–C10), selected based on the highest average F1-score across all clustering methods. Columns represent the clustering algorithms applied: Gaussian

Mixture Model (GMM), Hierarchical Clustering (HC), k-means, and k-means on the low-dimensional embeddings from PCA, t-SNE, and UMAP. The color scale reflects the magnitude of each metric, with darker shades indicating better performance. All combinations included between 4 and 9 genes from the 14-gene pool that showed consistent directional importance across training and testing phases.

Tables

Table 1. Datasets used across the model-development workflow. The training cohort (Dataset 1; *PRJNA248469*) was used for model fitting and consensus gene selection. Two independent studies in the testing phase (Dataset 2; *PRJNA565216* and *PRJNA985602*) were used for held-out evaluation, and an external validation cohort (Dataset 3; *PRJNA702434*) was used for final assessment. For each dataset, the table reports the count and proportion of healthy controls (HC) and Crohn's disease (CD) samples.

Stage	Composing Dataset	Study ID	Health Control (% of total)	Crohn's Disease (% of total)
Training phase and Consensus gene selection	*Dataset 1*	*PRJNA248469*	42/260 (16%)	218/260 (84%)
Testing phase	*Dataset 2*	*PRJNA565216*	37/149 (25%)	112/149 (75%)
	Dataset 2	*PRJNA985602*	8/30 (27%)	22/30 (73%)
Validation phase	*Dataset 3*	*PRJNA702434*	42/76 (55%)	34/76 (45%)

Appendix C

rVSVΔG-ZEBOV-GP in-vitro triggers immunomodulatory response in human cells associated with adverse events

rVSVΔG-ZEBOV-GP *in-vitro* triggers immunomodulatory response in human cells associated with adverse events

Paola Martinez-Murillo[1*], Catia Alvarez[2*], Francesco Santoro[3], Maria Novedrati[3], Chiara Sonnati[3], Giorgio Montesi[3], Simone Lucchesi[3], Donata Medaglini[3], Isabella Eckerle[2,4], Claire-Anne Siegrist[1,5]

on behalf of the VEBCON, VSV-EBOVAC and VSV-EBOPLUS Consortia

[1]Center of Vaccinology, Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, Geneva, Switzerland.

[2] Department of Medicine, Faculty of Medicine, University of Geneva, Geneva, Switzerland

[3] Department of Medical Biotechnologies, University of Siena, Siena, Italy.

[4] Geneva Centre for Emerging Viral Diseases, Geneva University Hospitals, Geneva, Switzerland.

[5] University of Geneva Medical School, Geneva, Switzerland.

* These authors contributed equally to this study

Corresponding author: Paola Martinez-Murillo, Center of Vaccinology, Department of Pathology and Immunology, Faculty of Medicine, University of Geneva, 1205 Geneva, Switzerland, paola.martinez@unige.ch

§ Consortia members:
VEBCON Consortium:
Gabon Selidji Todagbe Agnandji, Sanjeev Krishna, Peter G Kremsner, Jessica S Brosnahan (Centre de Recherches Médicales de Lambaréné).
Germany Selidji Todagbe Agnandji, Sanjeev Krishna, Peter G Kremsner, Jessica S Brosnahan (Institut für Tropenmedizin, Universitätsklinikum Tübingen); Marylyn M Addo (University Medical Center Hamburg); Stephan Becker, Verena Krähling (Institute of Virology, Marburg).
UK Sanjeev Krishna (St George's University of London).
Kenya Philip Bejon, Patricia Njuguna (Kenya Medical Research Institute, Kilifi).
S*witzerland* Claire-Anne Siegrist, Angela Huttner (Geneva University Hospitals, Geneva); and Marie-Paule Kieny, Vasee Moorthy, Patricia Fast, Barbara Savarese, Olivier Lapujade (World Health Organization, Geneva).
VSV-EBOVAC Consortium:
Selidji Todagbe Agnandji, Rafi Ahmed, Sravya S Nakka, Floriane Auderset, Philip Bejon, Luisa Borgianni, Jessica Brosnahan, Annalisa Ciabattini, Olivier Engler, Mariëlle C Haks, Ali M Harandi, Donald Gray Heppner, Alice Gerlini, Angela Huttner, Peter Gottfried Kremsner, Donata Medaglini, Thomas Monath, Francis Ndungu, Patricia Njuguna, Tom H M Ottenhoff, David Pejoski, Mark Page, Gianni Pozzi, Francesco Santoro, Claire-Anne Siegrist.
VSV-EBOPLUS Consortium:
Selidji Todagbe Agnandji, Sravya S Nakka, Luisa Borgianni, Annalisa Ciabattini, Sheri Dubey, Michael Eichberg, Olivier Engler, Patrícia Gonzalez-Dias, Peter Gottfried Kremsner, Ali M Harandi, Alice Gerlini, Angela Huttner, Donata Medaglini, Helder Nakaya, Tom HM Ottenhoff, Gianni Pozzi, Sylvia Rothenberger, Francesco Santoro, Eleonora Vianello, Claire-Anne Siegrist.

Abstract (max 250 words)

The recombinant vesicular stomatitis virus–vectored Zaire Ebola virus glycoprotein (rVSVΔG-ZEBOV-GP) vaccine, while effective and well-tolerated, exhibits notable reactogenicity, manifesting in expected adverse events (AEs) such as fever, headache, and pain, along with unexpected AEs like skin lesions, cutaneous vasculitis, and transient arthritis. The presence or absence of AEs following rVSVΔG-ZEBOV-GP vaccination is associated with a specific innate plasma signature. This study aims to elucidate *in-vitro* the tropism of the vaccine for different cell types involved in the unexpected AEs previously reported. Upon *in-vitro* infection with rVSVΔG-ZEBOV-GP, various cell types such as synoviocytes, fibroblast, keratinocytes and endothelial cells (except chondrocytes) demonstrate productive infection, which in dermal fibroblast triggered the release of many innate plasma signature markers, including keratinocytes' proinflammatory and proapoptotic cytokines such as OSM and TRAIL. Infected monocytes from buffy coats, activated by infection, produce most innate plasma signature markers. In co-culture, rVSVΔG-ZEBOV-GP-infected monocytes are the source to synoviocytes infection, resulting in a distinct kinetics delay in innate biomarkers (transcription and secretion) and upregulation of specific genes such as *NEDD8* and *SIGLEC-1* associated with inflammatory arthritis. Altogether, we provide insights into the mechanisms of rVSVΔG-ZEBOV-GP observed reactogenicity by showing tropism of the vaccine for off target cells from AE affected compartments (skin, joints, vessels) that leads to a delayed response when synoviocytes interact with infected monocytes.

Importance (max 150 words)

Our study expanded on previous knowledge about cellular tropism of rVSVΔG-ZEBOV-GP vaccine towards peripheral blood mononuclear cells (PBMCs), and cell lines from compartments affected by the unexpected adverse events (skin, vessels, and joints), and their possible contribution to the innate plasma signature. Using *in-vitro* infection and co-culture techniques, we show that rVSVΔG-ZEBOV-GP infected monocytes can transmit the virus to synoviocytes. Additionally, we demonstrated *in-vitro* how rVSVΔG-ZEBOV-GP affects human monocytes and synoviocytes at both the protein and transcriptomic levels.

Our findings provide insights into the *in-vitro* off-target infection dynamics and innate immune response triggered by rVSVΔG-ZEBOV-GP vaccine. While these results enhance our understanding of rVSV-based vectors, their *in-vivo* relevance requires further study. These findings support the evaluation of off target effects in rVSV-based vaccine candidates, relevant for those currently in development for hemorrhagic fever viruses, including filoviruses (Marburg virus MARV and Sudan ebolavirus SUDV) and Lassa virus (LASV).

Introduction

Ebola virus (EBOV) is the cause of a severe viral haemorrhagic disease of zoonotic origin. Human-to-human transmission of the virus occurs through direct contact with blood, secretions and other bodily fluids of infected individuals [1]. Ebola virus disease (EVD) is endemic to Central and Western Africa [1, 2]. Since its discovery in 1976, EBOV has caused geographically limited outbreaks with several to a few hundred cases; however, two large outbreaks have occurred in West Africa (2014 - 2016) and the Democratic Republic of Congo (2018 - 2020), driving both interest in the disease as well as development of vaccines and therapeutics [1].

The rVSVΔG-ZEBOV-GP vaccine is a replication-competent, live-attenuated, recombinant vesicular stomatitis virus vaccine [3, 4], which has been genetically engineered to express the gene of the glycoprotein from the Zaire ebolavirus, which replaces the gene of the native VSV glycoprotein [5, 6]. Initial ring vaccination efficacy trials indicated the rVSVΔG-ZEBOV-GP vaccine has high efficacy (100% in a controlled environment) [7, 8]. Recent real-world data confirms that a single dose of rVSV-ZEBOV is highly protective against Ebola virus disease 10 days or more after vaccination (84% [95% credible interval 70–92]) and therefore remains an important tool for outbreak response to EBOV [9].

The rVSVΔG-ZEBOV-GP vaccine has been shown to be safe and immunogenic, but reactogenic [5, 7, 10, 11]. Unexpected adverse events (AEs), including transient arthritis, vesicular lesions as well as cutaneous vasculitis occurred during the first two-three weeks after vaccination in a proportion of vaccinees [5, 11]. In the Geneva phase I clinical trial, transient arthritis was observed in 24% of vaccinees with a median of 11 days after vaccination, with joints affected asymmetrically by swelling, tenosynovitis and bursitis [11]. These results were replicated in other studies, in which the onset of arthritis occurred at the same median time and duration but at a lower rate (5%) [12]. In addition, in some of these cases, rVSVΔG-ZEBOV-GP RNA was detected by qRT-PCR in synovial fluid, vesicles and in skin lesions a few weeks after vaccination, suggesting rather a direct effect on these compartments instead of a solely immune-mediated aetiology [5, 11, 12]. rVSVΔG-ZEBOV-GP viremia was common but transient and dose-dependent, and viremia did not correlate with the transient arthritis observed [5]. These findings suggested that in some patients the innate immune response was not able to control virus replication in immune-privileged sites. Reinforcing the role of the innate response, we have recently reported a monocyte innate plasma vaccine signature that correlates with viremia, AEs and haematological changes [13, 14].

To better understand the cell and tissue tropism of this vaccine, it is crucial to evaluate target cells and body compartments affected by expected and unexpected AEs. Given the successful implementation of the rVSVΔG-ZEBOV-GP vaccine, it is expected that other rVSV-derived vaccines for emerging viruses would be

implemented in the future, therefore investigating the pathophysiological mechanism behind these unexpected adverse events is relevant.

In this study, we have investigated the *in-vitro* cellular tropism (off target infection and replication) and the production of innate immune biomarkers after rVSVΔG-ZEBOV-GP infection in a panel of human primary cells from tissues previously associated with unexpected adverse events as well as peripheral blood mononuclear cells (PBMCs). We also set up co-culture assays of monocytes and synoviocytes to assess whether monocytes could be the source of synoviocytes infection and to study *in-vitro* the kinetics of the response at the protein and transcriptomic level.

Material & methods

Healthy PBMCs and CD14+ monocytes isolation

We isolated healthy PBMCs and CD14+ monocytes from buffy coats. PBMCs were isolated by density-gradient centrifugation from ethylenediaminetetraacetic acid (EDTA) buffy coats by Ficoll-Paque PLUS (GE Healthcare), washed extensively in phosphate-buffered saline (PBS) treated with red-blood cell lysis buffer. Cells were counted and frozen in 90% heat-inactivated fetal bovine serum (FBS) and 10% dimethyl sulfoxide (DMSO) (Sigma-Aldrich). CD14+ monocytes were recovered from freshly isolated PBMCs by using EasySep Human Monocyte isolation kit (Stemcell Technologies).

Cell culture of immortalized and primary cells

We used immortalized cell lines and primary cells (Table S1). The immortalized cell lines: Monkey kidney cells (Vero E6) and human synovial sarcoma cells (SW982) were cultured in Dulbecco's modified eagle medium (DMEM) and GlutaMAX (Gibco, Thermo Fisher Scientific) supplemented with 50 mL FBS, 5 mL non-essential amino acids and 5 mL penicillin-streptomycin. The primary cells: human knee articular chondrocytes (NHAC-Kn, Lonza) were cultured in chondrocyte basal medium (Lonza) supplemented with 25 mL FBS, 0.5 mL GA-1000, 2.5 mL bFGF, 1 mL R3-IGF, 0.5 mL transferrin and 1 mL insulin. Human synoviocytes (HS, ScienCell) were cultured in synoviocyte medium (SM, ScienCell) supplemented with 10 mL FBS, 5 mL synoviocyte growth supplement (SGS) and 5 mL penicillin-streptomycin. Human dermal fibroblasts (NHDF-Ad, Lonza) were cultured in fibroblast growth basal medium (FBM, Lonza) supplemented with 10 mL FBS, 0.5 mL rhFGF-B, 0.5 mL GA-1000 and 0.5 mL insulin. Human epidermal keratinocytes (NHEK-Ad, LZ-192627, Lonza) were cultured in keratinocyte cell basal medium (KBM-Gold, Lonza) supplemented with 0.5 mL hydrocortisone, 0.5 mL transferrin, 0.25 mL epinephrine, 0.5 mL GA-1000, 2 mL BPE 0.5 mL hEGF and 0.5 mL insulin. Human dermal microvascular endothelial (HDMEC, ScienCell) were cultured in endothelial cell medium (ECM, ScienCell) supplemented with 25 mL FBS, 5 mL ECGS and 5 mL penicillin-streptomycin. Human dermal lymphatic endothelial (HDLEC, Cell Biologics) were cultured in human endothelial cell medium (ECM, Cell Biologics) supplemented with 25 mL FBS, 0.5 mL VEGF, 0.5 mL heparin, 0.5 mL EGF and FGF, 0.5 mL hydrocortisone, 5 mL L-Glutamine, 5 mL penicillin-streptomycin. Human umbilical vein endothelium (HUVEC, Lonza) were cultured in endothelial cell growth medium (EGM, Lonza) supplemented with 10 mL FBS, 0.5 mL GA-1000, 2 mL hFGF-B, 0.2 mL hydrocortisone, 0.5 mL R3-IGF-1, 0.5 mL VEGF, 0.5 mL asorbic acid, 0.5 mL hEGF and 0.5 mL heparin. Cells were passaged when confluent and grown at 37°C with 5% CO2.

rVSVΔG-ZEBOV-GP vaccine stock

The rVSVΔG-ZEBOV-GP vaccine was used to infect fresh confluent VeroE6 cells for 24h. The next day, the supernatant was collected and concentrated by centrifugation in a Vivaspin columns (Sigma-Aldrich) for 30 min at 3000 x *g*. Aliquots were stored at -80°C and the stock was titrated by plaque assay.

Viral titration by plaque assay

The plaque assay was performed in duplicate infecting Vero E6 cells with serial dilutions (from 10^{-1} to 10^{-6}) of the viral stock. Cells were washed twice and covered with a 1:1 mixture of 2xDMEM medium (Gibco) and 2.4% Avicell (Dupont) to prevent the virus infection from spreading indiscriminately. Infected cells were fixed 24 h post-infection with 4% paraformaldehyde (PAF) for 30 min at room temperature (RT), and stained with crystal violet (Sigma-Aldrich) for 30 min at RT. Viral plaques were counted manually. The plaques count together with the dilution factor were used to calculate the number of plaque forming units per sample unit volume (PFU/mL).

rVSVΔG-ZEBOV-GP infection of cell lines and PBMCs

We infected primary human cells and cell lines derived from the respective compartments (Table S1). Primary and immortalized cells were plated in a 24 well-plate at a density of $3x10^5$ cells per well, while thawed PBMCs were plated in round bottom 96-well untreated polystyrene Corning™ plates at a density of 2 million per well. Cell lines were infected next day, while PBMCs were infected after thawing. Infection with rVSVΔG-ZEBOV-GP was done in serum-free medium at different multiplicity of infection (MOI): for primary cells MOI 0.1 and for PBMCs MOI 1, as well as mock control in serum-free medium. Incubated for 1h at 37°C, then extensively washed and fresh growth medium was added, supernatant was collected to quantify viral RNA by RT-qPCR at multiple time points post infection.

RNA extraction and real-time quantitative polymerase chain reaction (RT-qPCR)

RNA was extracted from infected cells using the NucliSens easyMAG (BioMérieux), according to manufacturer's instruction. The viral load was determined from RNA by RT-qPCR using SuperScript™ III Platinum™ One-Step qRT-PCR Kit (Invitrogen) in CFX96 Thermal Cycler (Bio-Rad Laboratories). The RT-qPCR was performed using specific set of primers (Forward, 5'-CGGAGGATTGACGACTAATGC-3', Microsynth / Reverse, 5'-CGAGCCATTCGACCACATC-3', Microsynth) and probe (5'-CGCCACAAGGCAG-3', Microsynth).

Western blot of Ebola receptors expression

We looked at two different receptors: T-cell immunoglobulin and mucin domain 1 (TIM-1) presents on the cell surface serves as a receptor/cofactor for entry of Ebola virus *in-vivo*, enhancing viremia and pathogens [15, 16], and Niemann-Pick C1 (NPC1) is an intracellular lysosomal cholesterol transporter required for Ebola RNA release through endosomal membrane [17-19].

Cells were lysed with lysis buffer (NP-40) containing a protease inhibitor (Roche). Samples were sonicated and centrifuged 10min at 20'000 x *g* at 4°C and incubated 5 min at 95°C. Protein lysates (10 µl for immortalized cells and 20 µl for primary cells) were loaded in a 10% polyacrylamide gel and run at 150 V for 2h. Transfer was done on a PVDF membrane at 150 mA for 1h and blocked with milk 5% in washing buffer for 30min. The membrane was incubated with anti-NPC1 antibody (rabbit, Abcam) or anti-TIM-1 antibody (mouse, MAB1750-100, Bio-Techne AG) overnight at 4°C (dilutions 1:2000, 1µg/mL respectively). After washing the membranes were incubated with the secondary antibodies' goat anti-rabbit IgG HRP-conjugate (Bio-Rad Laboratories) or goat anti-mouse IgG HRP- conjugate (Bio-Rad Laboratories) (dilution 1:5000) for 1h at RT. After washing, the membrane was developed with ECL system (Advansta) and pictures were taken. Membranes were re-incubated with an HRP anti-beta Actin antibody (Abcam) for 1h at RT. After washing, the membrane was developed with ECL system and pictures were taken. The actin expression was used to normalize the protein expression levels using ImageJ program.

rVSVΔG-ZEBOV-GP infection in co-culture

Before the co-culture, human synoviocytes (HS) primary cells were plated in a 24 well-plate at a density of $3x10^5$ cells incubated at for at least 24h at 37°C with 5% CO_2. Freshly isolated, CD14+ purified monocytes from healthy donors were plated in a separate 96-well round bottom plate at a density of $3x10^5$/well, infected with rVSVΔG-ZEBOV-GP at different multiplicity of infection (MOI 0, 1, 5, 10) and incubated for 1h at 37°C with 5% CO_2. For the co-culture, infected monocytes were extensively washed and plated together with the HS cells in fresh growth medium. Supernatant was collected to quantify viral RNA by qRT-PCR at multiple time points post infection.

Luminex assay

Cryopreserved supernatants recovered after rVSVΔG-ZEBOV-GP *in-vitro* infection were analysed by Luminex (Magnetic Luminex assay, R&D Systems) to detect the concentration of 17 markers in culture supernatants. Assays were performed according to the supplier's instructions. Briefly, beads conjugated with the biomarker-specific capture antibodies were incubated at room temperature for 2h with samples, controls or standards. Biotinylated detection antibodies and R-phycoerythrin–conjugated streptavidin were subsequently added. The mean fluorescence intensity of each marker was read on the Bio-Plex 200 array reader (Bio-Rad Laboratories) using the Luminex xMAP Technology (Luminex Corporation). Five-parameter logistic regression curve (Bio-Plex Manager 6.0) was used to calculate sample concentrations. Assessed cytokines and chemokines were IL-1Ra, MCP1, IL-6, IL-10, MIP1b, TNF-alpha, MCP2, MCP3, MCP4, CXCL10, CXCL11, CX3CL1, OSM, MCSF, TRAIL, RANKL and IL15 [14].

Flow cytometry

After *in-vitro* infection with rVSVΔG-ZEBOV-GP of PBMCs, fluorescence-activated cell sorting (FACS) was done to phenotype PBMCs populations that were or not infected and to detect activation markers in gated monocyte populations. In the co-culture, FACS was used to detected infected synoviocytes. To assess the infection, cells were fixed and permeabilized (Thermofisher) and intracellular staining was done with anti-VSV nuclear protein (NP).

After *in-vitro* infection with rVSVΔG-ZEBOV-GP of healthy PBMCs. We did phenotyping of PBMCs populations. Cells were stained in PBS with LIVE/DEAD™ fixable aqua dead cell stain kit (Thermofisher) and FcR binding inhibitor (Miltenyi). Surface staining of the cells with a cocktail of anti-human antibodies allowed the identification of the different populations. Monocytes (CD14+, CD16+/-, HLADR+, CD11C+, CD3-, CD20-, CD56-), T cells (CD3+, CD20-, CD14-, CD56-), NKT cells (CD3+, CD20-, CD14-, CD56+), B cells (CD3-, CD20+, CD14-), NK cells (CD14-, CD3-, CD20-, CD56+), pDCs (CD123+, CD11C-, CD3-, CD20-, CD56-) and mDCs (CD14-, CD16-, HLADR+, CD11C+, CD3-, CD20-, CD56-). In the same stain, the detection of activation markers was done on gated total monocytes populations (infected and non-infected) that were then gated for positive activation markers (CD86, CD169, CD163), gating was defined by using negative controls (non-infected PBMCs) and isotype controls.

For detection of synoviocytes in the Co-culture: Cells were stained in PBS with LIVE/DEAD™ fixable aqua dead cell stain kit (Thermofisher) and FcR binding inhibitor (Miltenyi). Surface staining of the cells with a cocktail of anti-human antibodies allowed the identification of synoviocytes as CD14-, HLA-DR-, CD90+, CD55+.

Transcriptomic analysis of the co-culture

Quantification of gene expression in cell cultures was performed by RNA sequencing. The time point evaluated was 4h after infection. Cells were then harvested by centrifugation at 1500rpm for 5 minutes and resuspended in lysis buffer RLT buffer with β-mercaptoethanol (Qiagen) for RNA extraction. Total RNA was extracted from human monocytes, synoviocytes or co-culture using the Qiagen RNeasy Mini Kit, following the manufacturer's protocol. Cells were lysed, homogenized, and RNA was purified using spin columns. Total RNA was quantified with Qubit (Thermofisher) and about 50ng of total RNA were processed using Illumina Stranded mRNA Prep Ligation kit (Illumina) to prepare cDNA dual-indexed libraries according to manufacturer's recommendations. Dual-indexed libraries were purified using Agencourt AMPure XP Magnetic Beads (Beckman Coulter) and quantified with NEBnext Library Quant Kit for Illumina (New England Biolabs) on a QuantStudio 5 Dx Real-time PCR System (Thermofisher) and their size was estimated on a BioAnalyzer DNA 1000 chip (Agilent

Technologies, Germany). Libraries were diluted to 0.5 nM concentration and pooled in final volume of 100 µL for sequencing on a SP flow cell (Illumina) which was loaded according to manufacturer's protocol. Sequencing reactions were performed for 200 cycles on the Novaseq 6000 platform (Illumina) to obtain 100 bases paired-end reads. Base called data were quality controlled with FASTQC, trimmed using TRIMMOMATIC, and then aligned to the human reference genome (GRCh38.p7) using STAR (v2.4.2a) or to the rVSVΔG-ZEBOV-GP genome using BWA (v0.7.17). HTSEQ-COUNT tool was used to count the number of aligned reads for each gene.

Data analysis

Graphs and statistics were generated using GraphPad Prism version 9.2.0. All data was log10 transformed and then two-way ANOVA analysis was performed, to correct for multiple comparisons we used the Sidak's test implemented in GraphPad Prism. P values ≤ 0.05 are considered significant. We used the following designation of p values in graphs: ns, not significant; **** p≤0.0001; *** p≤ 0.001; ** p≤ 0.01; * p≤ 0.05.

Differential gene expression analysis was performed with the R package DESeq2 [20] (v1.40.2). Since most genes did not show a normal distribution, differences in gene expression were assessed with the Wilcoxon test, corrected for multiple testing with the Benjamini-Hochberg method, on *vst* (variance stabilizing transformation) normalized data from DeSeq2. Enrichment analysis was performed with the clusterProfiler R package with the enrichKEGG and enrichGO functions. The background genes were those mapping to our dataset. Statistical significance was assessed with the hypergeometric test, P-values were adjusted using the Benjamini-Hochberg method to control the false discovery rate, and significance was set at an adjusted p-value <0.01.

Data availability

The RNA-seq data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO) under the accession number GSE296388. The data will be publicly available upon publication. All raw and processed files, including sample metadata and normalized expression matrices, are included in the submission.

Results

In PBMCs rVSVΔG-ZEBOV-GP infects mainly monocytes and induces their activation

In-vitro infection of PBMCs with rVSVΔG-ZEBOV-GP (detected as anti-VSV-NP positive cells by flow cytometry) was used to assess the *in-vitro* tropism of the vaccine to the different immune cell populations present in PBMCs (Figure S1A, B). We tested two different time points for the *in-vitro* infection of PBMCs and found that at 18h there was no increase in percentage of total infected cells compared with 6h [mean 0.69% (95% CI 0.49 to 0.89) and 1.39% (95% CI 0.27 to 2.51), p=0.42] (Figure 1A, left). We also showed that most of the infected cells are monocytes (overall mean 67% at 6h and 75% at 18h; p=0.073) followed by myeloid dendritic cells (mDCs) (overall mean 7% at 6h and 10% at 18h; p>0.999) (Figure 1A left, Figure S1D). Our data show no infection in other PBMC subsets (data not shown), meaning the infection was largely restricted to monocytes and DCs. Additionally, we observed stable infection kinetics in PBMC cultures, where infected cells were detected by 6h, and their levels remained unchanged at 18h and 24h (Figure S1E), suggesting that most susceptible cells were infected early, with no further increase over time. When analysing the total monocytes and DCs populations, we saw that the percentage of infected cells was higher than in total PBMC: infected monocytes were on average 6.7% at 6h and 6.8% at 18h; (p=0.999) and infected DCs were on average 4.6% at 6h and 4.5% at 18h (p=0.999) (Figure 1A right, Figure S1B and S1C). There was no significant difference in the percentage of monocytes and DCs infected cells between 6h and 18h, suggesting that the infection in those populations was well established before our first time point of 6h.

Because monocytes and mDCs are the main subsets infected, we then analysed the expression of activation markers in these subsets 18h after *in-vitro* infection and used the non-infected mock condition to evaluate whether the expression was induced by the infection. All monocytes population (Classical monocytes (CM), Intermediate monocytes (IM) and Non-classical Monocytes (NCM)) showed significant increase on the median fluorescence intensity (MFI) for CD169 after infection compared to non-infection. Only CM and IM showed a significant increase on the MFI for CD86 after infection compared to non-infection, while CD163 MFI was not statistically significant in any monocyte subset (Figure 1B). We also compared infected monocytes to non-infected monocytes within the same culture and found no significant difference in activation marker expression between the monocyte's subset (data not shown).

These results confirmed that, in PBMCs, rVSVΔG-ZEBOV-GP has a tropism for monocytes and mDCs and that activation of the different monocyte populations is not strictly dependent on direct infection but rather a result of exposure to the virus or the infected environment.

CD14+ monocytes produce plasma innate signature markers after in-vitro rVSVΔG-ZEBOV-GP infection

We evaluate whether isolated CD14+ monocytes were producers of innate plasma signature biomarkers (i.e. 17 cytokines and chemokines) reported recently [14]. After rVSVΔG-ZEBOV-GP *in-vitro* infection of CD14+ monocytes all plasma signature markers were detected in culture supernatant as soon as 6h post infection reaching a peak at 24h and are still detectable at 48h. The background level in non-infected CD14+ monocytes was high, and the levels were very heterogeneous, therefore some markers were significantly higher only at 24h compared to the background: MCP-1, MCP-4, MCSF, CXCL10, CXCL11, CX3CL1, IL-10, OSM and TRAIL (Figure 2). This result demonstrated that isolated CD14+ monocytes infected with rVSVΔG-ZEBOV-GP produce most biomarkers previously identified in the innate plasma signature.

Cellular tropism of rVSVΔG-ZEBOV-GP vaccine virus in primary human cells and human cell lines

We next assessed the *in-vitro* cellular tropism of rVSVΔG-ZEBOV-GP in human primary cells from tissues previously associated with unexpected AEs such as synoviocytes, fibroblast, keratinocytes and endothelial cells and chondrocytes.

In joint cells, *in-vitro* viral replication increased rapidly over time, with an increase of almost 4 $\log_{10}$ RNA copies/mL 48h after infection in human synovial sarcoma (cell line) and synoviocytes (primary cells) (Figure 3A, left). In skin tissue primary cells, we observed *in-vitro* viral replication in both human dermal fibroblasts and human epidermal keratinocytes that was higher for fibroblasts (Figure 3B, left). All human endothelial primary cells tested (dermal microvascular, umbilical vein and dermal lymphatic endothelial cells) had an increase of 4 $\log_{10}$ RNA copies/mL after *in-vitro* infection (Figure 3C, left). The plaque assay confirmed that infectious particles were formed after *in-vitro* infection in all these primary cells, except human knee articular chondrocytes, in which no viral replication nor infectious viral particles production was observed (Figure 3A, 3B, 3C middle).

In the supernatant, we observed secretion of innate plasma signature markers after rVSVΔG-ZEBOV-GP *in-vitro* infection that increase over time mainly in lymphatic endothelial cells (MCP3, MCP4, MCP1, IL-6, CXCL10, OSM, TRAIL and TNF-alpha) (Figure 3C right), followed by skin fibroblast (CXCL10, CXCL11, CX3CL1, MCP2, IL10, IL15, TNF-alpha, RANKL and OSM) (Figure 3B right), while synoviocytes only secreted CXCL10 and MCP3 in a lower amount (Figure 3A right and Figure S2). Microvascular and vein endothelial cells produced mainly CXCL10 48h after *in-vitro* infection (Figure S3A). We also found that chondrocytes did not produce these markers over time, all arguing against viral replication in these cells (Figure 3A right and Figure S3A).

Next, we determined whether the permissiveness of these primary cells is linked to the presence of Ebola virus receptors T-cell immunoglobulin and mucin domain 1 (TIM-1) and Niemann-Pick C1 (NPC1). NPC1 protein was present in all primary cell lines tested, while TIM-1 detection was low or absent in some. However, in the non-infectable chondrocytes, both NPC1 and TIM-1 were clearly detected, suggesting that the presence of these receptors alone may not be sufficient to confer susceptibility to infection (Figure S3B).

We showed that several cell types of the tissue previously associated with AEs such as synoviocytes, fibroblast, keratinocytes and endothelial cells were productively infected *in-vitro* by rVSVΔG-ZEBOV-GP, and the *in-vitro* infection induced the secretion of several innate plasma signature biomarkers especially in dermal fibroblast and dermal lymphatic endothelial cells.

Infected monocytes transmit infection into synoviocytes and their interaction delays innate response

Then, we evaluated whether infected monocytes (the main blood population permissive to rVSVΔG-ZEBOV-GP infection) could infect synoviocytes (Figure S4A). In the co-culture, the kinetics of viral replication differed from that observed in monocytes and synoviocytes. Specifically, replication was slower compared to synoviocyte infection alone, but faster compared to monocyte infection alone. Viral RNA was first detected 6 hours post-infection and peaked at 48 hours (Mean $3.56x10^6$ copies/mL of supernatant). This replication pattern was not influenced by the viral dose, as similar results were observed at MOI 1 (Figure 4A), MOI 5, and MOI 10 (Figure S4B). To identify the percentage of infected synoviocytes in the co-culture, synoviocytes were defined by FACS at the morphological gate as live cells, CD14-, HLA-DR-, CD90+, CD55+ (Figure S4C) and we defined infected synoviocytes as the gated synoviocytes that were VSV-NP+ (Figure S4D). In the co-culture, we saw that synoviocytes were infected, although with a slower kinetics compared to synoviocytes-alone culture at 24h (mean 2.9%, SD 2.2)(Figure 4B), but then infected synoviocytes in the co-culture reached similar levels as synoviocytes-alone culture at 48h (mean 16.5%, SD 3.5), which were maintained at 72h (Figure 4B) (mean 11.61%, SD 3.1), similar results were seen when we increased the viral inoculum (MOI 5 and 10) (Figure S4D).

We also evaluated whether the interaction of the monocytes and the synoviocytes in the co-culture had an impact on the secretion of the innate signature markers. In the co-culture, we observed that the secretion kinetics of the innate signature markers was intermediate, it was slower compared to monocytes alone and faster compare to the synoviocytes alone at both 6h and 24h post infection, and it was reestablished to similar levels as to the single culture cells at 48h for most of the markers except MCP1 (Figure 4B).

These results indicate that *in-vitro* rVSVΔG-ZEBOV-GP infected monocytes produced infecting particles, which served as the source for synoviocytes infection in co-culture, and this also delays the secretion of the innate response signature biomarkers.

Transcriptomic analysis indicates an inflammatory response to rVSVΔG-ZEBOV-GP infection

We then analysed the transcriptomic profile of monocytes, synoviocytes and their co-cultures in response to 4h rVSVΔG-ZEBOV-GP *in-vitro* infection. Viral gene expression was about 100-fold and 2-fold higher in the synoviocytes and monocytes, respectively, compared to the co-culture (Figure S5). Differential gene expression analysis identified relatively few significant (FDR<0.05) differentially expressed (DE) genes in infected monocytes and co-cultures compared to the uninfected cells, while more than 10% of the expressed genes were DE in infected synoviocytes (Table S2). Gene enrichment analysis showed that there was a conserved response to infection involving modules related to innate immunity activation, including RIG-I signalling, activation of dendritic cells, interferon response and chemokines (Figure S6). Infected synoviocytes were enriched in modules regarding regulation of cell proliferation and death.

A comparison of DE genes in the different cell lines identified 12 genes that were unique for the co-culture (Figure S7). Among these genes it is worth to mention *NEDD8*, coding for a ubiquitin-like protein possibly involved in the inflammatory arthritis pathogenesis [21], and *SIGLEC-1*, coding for the CD169 macrophage surface marker (Figure 5A-B). Both genes are only upregulated in the infected co-culture, indicating that the interaction of infected monocytes with synoviocytes triggered their activation. Coherently with the results of the Luminex assay, the expression levels of cytokine/chemokine genes were generally lower in the co-culture compared to monocytes (Figure S8), except for *TNFSF11*, coding for RANKL, *CX3CL1*, and *IL6* which were more abundant in the synoviocytes and in the co-culture, in both infected and uninfected conditions (Figure 5C-D). Interestingly, the genes coding for EBOV receptors, *NPC-1* and *TIM-1* were more expressed in the co-culture compared to both monocytes and synoviocytes, and their expression did not increase upon infection with rVSVΔG-ZEBOV-GP (data not shown).

Discussion

In this study, we investigate the complex events that unfold after vaccination by studying *in-vitro* the cellular tropism of the rVSVΔG-ZEBOV-GP vaccine *in-vitro*. Our findings revealed that in cell culture, endothelial cells, fibroblasts, keratinocytes, synoviocytes, and peripheral monocytes, supported the infection and production of infectious viral particles. Among these permissive cells, peripheral monocytes were found to secrete most innate signature biomarkers, indicating their significant role in the early immune response to the vaccine viral infection. Upon rVSVΔG-ZEBOV-GP infection, monocytes got activated and when co-cultured with synoviocytes were able to transmit the infection to synoviocytes *in-vitro*. Furthermore, the interaction between monocytes and synoviocytes led to a kinetics delay in the secretion of innate response signature biomarkers and upregulation of 12 genes, including *NEDD8* and *SIGLEC-1*, which are associated with inflammatory arthritis.

We evaluated *in-vitro* the tropism of rVSVΔG-ZEBOV-GP in PBMCs and showed that monocytes and dendritic cells were the only permissive cells for rVSVΔG-ZEBOV-GP infection in PBMCs. These results confirm that the vaccine's tropism in PBMCs is limited to these two populations and as expected, it mirrors the tropism of the Ebola virus glycoprotein (GP). Even if NK cells are modulated after rVSVΔG-ZEBOV-GP vaccination [22], they were not permissive for the infection. This is consistent with previous *in-vitro* studies, which have shown that the Ebola virus GP is a ligand for TLR-4 and induces activation of uninfected monocytic cell lines, monocyte-derived DCs and macrophages to produce cytokines [23, 24].

We also showed that infection with rVSVΔG-ZEBOV-GP induced activation of both classical and intermediate monocytes, but neither of myeloid dendritic cells nor non-classical monocytes, similar to what has been observed after vaccination in humans, where CD86 was significantly increased already at one day after vaccination in monocytes [25]. CD169 was upregulated in all three monocytes subsets upon infection implying a possible differentiation of monocytes into macrophages, consistent with the transcriptional upregulation [26, 27] and the increased surface expression [5] of CD169 seen after rVSVΔG-ZEBOV-GP vaccination in humans, which is also consistent with the critical role of CD169+ macrophages to induce protective innate and adaptive response after rVSVΔG-ZEBOV-GP vaccination in mice [28]. In contrast to what has been shown for SARS-CoV-2 infection, where monocytes aborted the replication and infected particles are not detected [29], here we showed that monocytes allowed the replication of the rVSVΔG-ZEBOV-GP and also produced infecting viral particles, reinforcing their role in the vaccine innate response [13, 14].

We saw production of many innate signature biomarkers after rVSVΔG-ZEBOV-GP *in-vitro* infection of dermal lymphatic endothelial cells and dermal fibroblast, mainly monocyte chemotactic proteins such as MCP1, MCP3, MCP4 as well as CXCL10, TRAIL and OSM. Interestingly, TRAIL has been shown to promote apoptosis of

keratinocytes [30] and OSM is a potent inductor of skin inflammation [31]. Given that unexpected AEs such as dermatitis and cutaneous vasculitis reported after rVSVΔG-ZEBOV-GP vaccination were not always linked with the presence of the virus vaccine in the skin biopsy [5], we suggested that dermatitis and cutaneous vasculitis might be the result of indirect effects of these secreted cytokines as has been shown in EBOV disease disruption of the vascular endothelium [32, 33].

We showed that synoviocytes *in-vitro* infection with rVSVΔG-ZEBOV-GP allowed replication and competent virus production, this is in line with the capacity of EBOV to infect monkey's synoviocytes *in-vivo* and *in-vitro* [34]. We also showed that the synoviocytes secretion of innate signature markers was dependent on the viral inoculum. Although arthritis was reported in patients who survived EBOV disease [35], the transient arthritis reported after rVSVΔG-ZEBOV-GP vaccination was an unexpected AE [5, 12]. Previously, we have shown that transient arthritis after rVSVΔG-ZEBOV-GP vaccination develops in individuals with a lower level of innate inflammatory plasma signature response after rVSVΔG-ZEBOV-GP vaccination, and we believe those individuals have a less effective early control of viral dissemination [14], which may in turn lead to viral presence in privileged sites such as joints. In the co-culture, our data suggest that rVSVΔG-ZEBOV-GP infected monocytes may be a source of infected particles that infected synoviocytes, and we also showed a delay in the secretion of the innate signature markers. Altogether, suggests that the infected monocytes can act as a trojan horse to allow the virus to reach immune-privileged sites, such as the joints, which together with the delayed innate response, could propitiate an environment for viral replication and transient viral arthritis. We also saw in the co-culture, the reduction of bone resorption markers such as RANKL (secretion level) as well as molecules that promote osteoclast differentiation such as TRAIL (at both transcriptional and secretion levels) [36], which is in line with the absence of bone resorption lesions in the transient arthritis after rVSVΔG-ZEBOV-GP vaccination [5], in contrast to chikungunya arthritis [37]. However, the observed upregulation of *NEDD8* in the co-culture and infected synoviocytes, points towards a transcriptional program that has been involved in post-transcriptional modifications in the inflammatory processes of arthritis [21].

Our study had some limitations. The number of samples was low, ranging from 2 to 5, mainly due to the complexity of the experiments, which involved measuring different time points, different viral inoculum and different readouts. Monocytes are well known to be easily activated *in-vitro* upon contact with plastic materials. To minimize this activation, we used round-bottom well plates made of low-adherent treated plastic. Although, we observed a clear increase in monocytes activation markers after infection through FACS, the uninfected condition showed a high basal background, which was also present in the supernatant quantification of innate markers. We also used different MOIs for the infection of PBMCs/monocytes and primary cells because our

readout for monocyte infection was through FACS, and a MOI of less than 1 required longer incubation times, increasing the risk of nonspecific activation. It is important to note that our study was conducted in an artificial *in-vitro* system using an elevated infectious dose (MOI 1 for PBMCs and MOI 0.1 for primary cells) to maximize infection efficiency and assess off-target cell susceptibility. While this approach allowed us to characterize *in-vitro* cellular tropism and innate immune responses, further studies are needed to determine whether these findings translate to *in-vivo* conditions.

In conclusion, we identified *in-vitro* off-target cellular tropism of rVSVΔG-ZEBOV-GP, infecting several cell types from tissues previously associated with unexpected adverse events such as synoviocytes, fibroblast, keratinocytes and endothelial cells. These cells were productively infected by rVSVΔG-ZEBOV-GP and secrete specific innate pro-inflammatory biomarkers. Monocytes were the main population supporting rVSVΔG-ZEBOV-GP *in-vitro* infection in PBMCs, became activated, and secreted most innate plasma signature markers upon *in-vitro* infection, and potentially may serve as a source of synoviocytes infection *in-vitro*. Additionally, co-culture revealed that interactions between infected monocytes and uninfected synoviocytes modulates the innate immune response delaying the secretion of biomarkers.

Altogether, our study provides an example of how *in-vitro* models can be used to study off-target vaccine virus tropism and the interplay between immune cells and primary cells. However, further studies are needed to determine *in-vivo* relevance of these off-target infections, their role in early and sustained immune responses to the vaccine, and strategies to reduce them.

Acknowledgements

The authors thank Chantal Tougne for her technical support, Benjamin Meyer for the viral discussions, and Lucas, Alex and Mariela for their inspiration. This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 116068 (VSV-EBOPLUS project) and No 115842 (VSV-EBOVAC project). This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation program and EFPIA.

Figure legends

Figure 1. PBMCs infection and activation markers after rVSVΔG-ZEBOV-GP *in-vitro* inoculation. (A) Percentage of infected cells after rVSVΔG-ZEBOV-GP *in-vitro* inoculation (left), total infected cells (grey circle), monocytes in total infected cells (white square) and dendritic cells (DCs) in total infected cells (white triangle). Percentage of monocytes and DCs infected after rVSVΔG-ZEBOV-GP *in-vitro* inoculation (right), gated total infected monocytes (grey square), gated total infected DCs (grey triangle). **(B)** Geometric mean (GM) media fluorescence intensity (MFI) value for different activation markers expressed on total (infected and uninfected) monocyte populations (intermediate, classical and non-classical) after rVSVΔG-ZEBOV-GP *in-vitro* infection of PBMCs. Each symbol represents a different donor (n=5). Mock (white diamond). MOI=1 for all conditions. Two-way ANOVA analysis was performed, to correct for multiple comparisons we used the Sidak's test implemented in GraphPad Prism. Asterisk represents p-value: less than 0.05 (*), less than 0.01 (**), less than 0.001 (***).

Figure 2. Pure monocytes cytokine production after rVSVΔG-ZEBOV-GP *in-vitro* inoculation. Luminex concentration in supernatants (pg/ml) for each marker was plotted at each time point (h) after rVSVΔG-ZEBOV-GP *in-vitro* inoculation in the different groups: mock (white diamond) and MOI 1 (pink circle). Bars represent the median concentrations with the CI (n=3: 6h and 48h; n=5: 24h). Red dotted lines indicate the limit of detection for each marker. Samples below the limit of detection were assigned a value corresponding to 50% of the last standard dilution value. Two-way ANOVA analysis was performed, to correct for multiple comparisons we used the Sidak's test implemented in GraphPad Prism. Asterisk represents p-value: less than 0.05 (*), less than 0.01 (**), less than 0.001 (***).

Figure 3. Viral replication, plaque assay and cytokine response in joint (A), skin (B) and vessels (C) primary cells. Viral loads were measured in supernatant by RT-qPCR at 0h, 6h, 24h and 48h after *in-vitro* rVSVΔG-ZEBOV-GP infection MOI 0.1 and are expressed as mean ± SEM (n=4). Infectious titers were calculated by performing a plaque assay in Vero cells at 24h after *in-vitro* infection and the cytokine response was analysed

at 6h, 24h and 48h by Luminex assay by doing a ratio between infection at MOI 0.1 versus Mock and are expressed as mean ± SEM of two biological replicates.

Figure 4. Co-culture viral replication and cytokine response. (A). Viral loads (left) were measured in supernatant by RT-qPCR at 0h, 6h, 24h, 48h and 72h after *in-vitro* rVSVΔG-ZEBOV-GP infection (MOIs 0 and 1) of synoviocytes (black), monocytes (blue) or co-culture (red) and are expressed as mean ± SEM (n=2). **(B)** Percentage of infected synoviocytes after *in-vitro* rVSVΔG-ZEBOV-GP infection (MOIs 0 and 1) of synoviocytes alone (black) or synoviocytes in co-culture synoviocytes (red). MOI 0 (circle), MOI 1 (square) **(C)** Heat map shows the mean (n=3) of the ratio (MOI 1 over Mock control) for each marker except RANKL (n=1), at different time points (6h, 24h and 48h) in different populations after *in-vitro* rVSVΔG-ZEBOV-GP infection (MOIs 1). Markers with values higher at mock are in blue, markers with similar value between MOI and mock and in light grey and markers with values higher at MOI than in mock are shown from (low) yellow to red (high).

Figure 5. Normalized expression (log_2expression) levels of selected genes in different cell cultures. (A) NEDD8, **(B)** SIGLEC1, **(C)** CX3CL1, and **(D)** TNFSFS11 expression levels in infected and uninfected cell cultures. Data are reported as box and whiskers plot, where the marked line inside the box represented the median value, the box the interquartile range (IQR), and whiskers the minimum and maximum values in the range ±1.5*IQR. Individual values are reported as black dots. Differences in gene expression between infected and uninfected samples (A and B) or among different cell culture types (C and D) were assessed with the Wilcoxon test (** $p<0.01$, *** $p<0.001$).

References

1. WHO, *Ebola situation reports: archive.* 2016.
2. Na, W., et al., *Ebola outbreak in Western Africa 2014: what is going on with Ebola virus?* Clinical and experimental vaccine research, 2015. **4**(1): p. 17.
3. Garbutt, M., et al., *Properties of replication-competent vesicular stomatitis virus vectors expressing glycoproteins of filoviruses and arenaviruses.* Journal of virology, 2004. **78**(10): p. 5458-5465.
4. Jones, S.M., et al., *Live attenuated recombinant vaccine protects nonhuman primates against Ebola and Marburg viruses.* Nature medicine, 2005. **11**(7): p. 786-790.
5. Huttner, A., et al., *The effect of dose on the safety and immunogenicity of the VSV Ebola candidate vaccine: a randomised double-blind, placebo-controlled phase 1/2 trial.* Lancet Infect Dis, 2015. **15**(10): p. 1156-1166.
6. Marzi, A., et al., *Vesicular stomatitis virus–based Ebola vaccines with improved cross-protective efficacy.* The Journal of infectious diseases, 2011. **204**(suppl_3): p. S1066-S1074.
7. Henao-Restrepo, A.M., et al., *Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!).* Lancet, 2017. **389**(10068): p. 505-518.
8. WHO. *Preliminary results on the efficacy of rVSV-ZEBOV-GP Ebola vaccine using the ring vaccination strategy in the control of an Ebola outbreak in the Democratic Republic of the Congo: an example of integration of research into epidemic response.* 26.09.2019]; Available from: https://www.who.int/csr/resources/publications/ebola/vaccines/en/.
9. Meakin, S., et al., *Effectiveness of rVSV-ZEBOV vaccination during the 2018-20 Ebola virus disease epidemic in the Democratic Republic of the Congo: a retrospective test-negative study.* Lancet Infect Dis, 2024.
10. Heppner, D.G., Jr., et al., *Safety and immunogenicity of the rVSVG-ZEBOV-GP Ebola virus vaccine candidate in healthy adults: a phase 1b randomised, multicentre, double-blind, placebo-controlled, dose-response study.* Lancet Infect Dis, 2017. **17**(8): p. 854-866.
11. Agnandji, S.T., et al., *Phase 1 trials of rVSV Ebola vaccine in Africa and Europe.* New England Journal of Medicine, 2016. **374**(17): p. 1647-1660.
12. Halperin, S.A., et al., *Six-month safety data of recombinant vesicular stomatitis virus–Zaire Ebola virus envelope glycoprotein vaccine in a phase 3 double-blind, placebo-controlled randomized study in healthy adults.* The Journal of infectious diseases, 2017. **215**(12): p. 1789-1798.
13. Huttner, A., et al., *A dose-dependent plasma signature of the safety and immunogenicity of the rVSV-Ebola vaccine in Europe and Africa.* Sci Transl Med, 2017. **9**(385).
14. Martinez-Murillo, P.A., et al., *Refined innate plasma signature after rVSVΔG-ZEBOV-GP immunization is shared among adult cohorts in Europe and North America.* Front Immunol, 2023. **14**: p. 1279003.
15. Kondratowicz, A.S., et al., *T-cell immunoglobulin and mucin domain 1 (TIM-1) is a receptor for Zaire Ebolavirus and Lake Victoria Marburgvirus.* Proceedings of the National Academy of Sciences, 2011. **108**(20): p. 8426-8431.

16. Brunton, B., et al., *TIM-1 serves as a receptor for Ebola virus in vivo, enhancing viremia and pathogenesis.* PLoS neglected tropical diseases, 2019. **13**(6): p. e0006983.
17. Kuroda, M., et al., *Interaction between TIM-1 and NPC1 is important for cellular entry of Ebola virus.* Journal of virology, 2015. **89**(12): p. 6481-6493.
18. Miller, E.H., et al., *Ebola virus entry requires the host - programmed recognition of an intracellular receptor.* The EMBO journal, 2012. **31**(8): p. 1947-1960.
19. Wang, H., et al., *Ebola viral glycoprotein bound to its endosomal receptor Niemann-Pick C1.* Cell, 2016. **164**(1): p. 258-268.
20. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 1-21.
21. Liu, K., et al., *TRAF6 neddylation drives inflammatory arthritis by increasing NF-κB activation.* Laboratory Investigation, 2019. **99**(4): p. 528-538.
22. Pejoski, D., et al., *Rapid dose-dependent Natural Killer (NK) cell modulation and cytokine responses following human rVSV-ZEBOV Ebolavirus vaccination.* npj Vaccines, 2020. **5**(1): p. 32.
23. Okumura, A., et al., *Interaction between Ebola virus glycoprotein and host toll-like receptor 4 leads to induction of proinflammatory cytokines and SOCS1.* J Virol, 2010. **84**(1): p. 27-33.
24. Escudero-Pérez, B., et al., *Shed GP of Ebola virus triggers immune activation and increased vascular permeability.* PLoS Pathog, 2014. **10**(11): p. e1004509.
25. Rechtien, A., et al., *Systems Vaccinology Identifies an Early Innate Immune Signature as a Correlate of Antibody Responses to the Ebola Vaccine rVSV-ZEBOV.* Cell Rep, 2017. **20**(9): p. 2251-2261.
26. Vianello, E., et al., *Transcriptomic signatures induced by the Ebola virus vaccine rVSVΔG-ZEBOV-GP in adult cohorts in Europe, Africa, and North America: a molecular biomarker study.* The Lancet Microbe, 2022. **3**(2): p. e113-e123.
27. Santoro, F., et al., *Human Transcriptomic Response to the VSV-Vectored Ebola Vaccine.* Vaccines (Basel), 2021. **9**(2).
28. Friedrich, S.K., et al., *Usp18 Expression in CD169(+) Macrophages is Important for Strong Immune Response after Vaccination with VSV-EBOV.* Vaccines (Basel), 2020. **8**(1).
29. Junqueira, C., et al., *FcγR-mediated SARS-CoV-2 infection of monocytes activates inflammation.* Nature, 2022. **606**(7914): p. 576-584.
30. Rethi, B. and L. Eidsmo, *FasL and TRAIL signaling in the skin during cutaneous leishmaniasis - implications for tissue immunopathology and infectious control.* Front Immunol, 2012. **3**: p. 163.
31. Pohin, M., et al., *Oncostatin M overexpression induces skin inflammation but is not required in the mouse model of imiquimod-induced psoriasis-like inflammation.* Eur J Immunol, 2016. **46**(7): p. 1737-51.
32. Geisbert, T.W. and H. Feldmann, *Recombinant vesicular stomatitis virus-based vaccines against Ebola and Marburg virus infections.* J Infect Dis, 2011. **204 Suppl 3**(Suppl 3): p. S1075-81.
33. Wahl-Jensen, V.M., et al., *Effects of Ebola virus glycoproteins on endothelial cell activation and barrier function.* J Virol, 2005. **79**(16): p. 10442-50.
34. Cooper, T.K., et al., *Filoviruses Infect Rhesus Macaque Synoviocytes in Vivo and Primary Human Synoviocytes in Vitro.* Am J Pathol, 2020. **190**(9): p. 1867-1880.
35. Howlett, P., et al., *Ebola Virus Disease Complicated by Late-Onset Encephalitis and Polyarthritis, Sierra Leone.* Emerg Infect Dis, 2016. **22**(1): p. 150-2.

610 36. Liao, H.-J., et al., *TRAIL inhibits RANK signaling and suppresses osteoclast*
611 *activation via inhibiting lipid raft assembly and TRAF6 recruitment.* Cell Death &
612 Disease, 2019. **10**(2): p. 77.
613 37. Chen, W., et al., *Arthritogenic alphaviruses: new insights into arthritis and bone*
614 *pathology.* Trends Microbiol, 2015. **23**(1): p. 35-43.
615

Figure 1

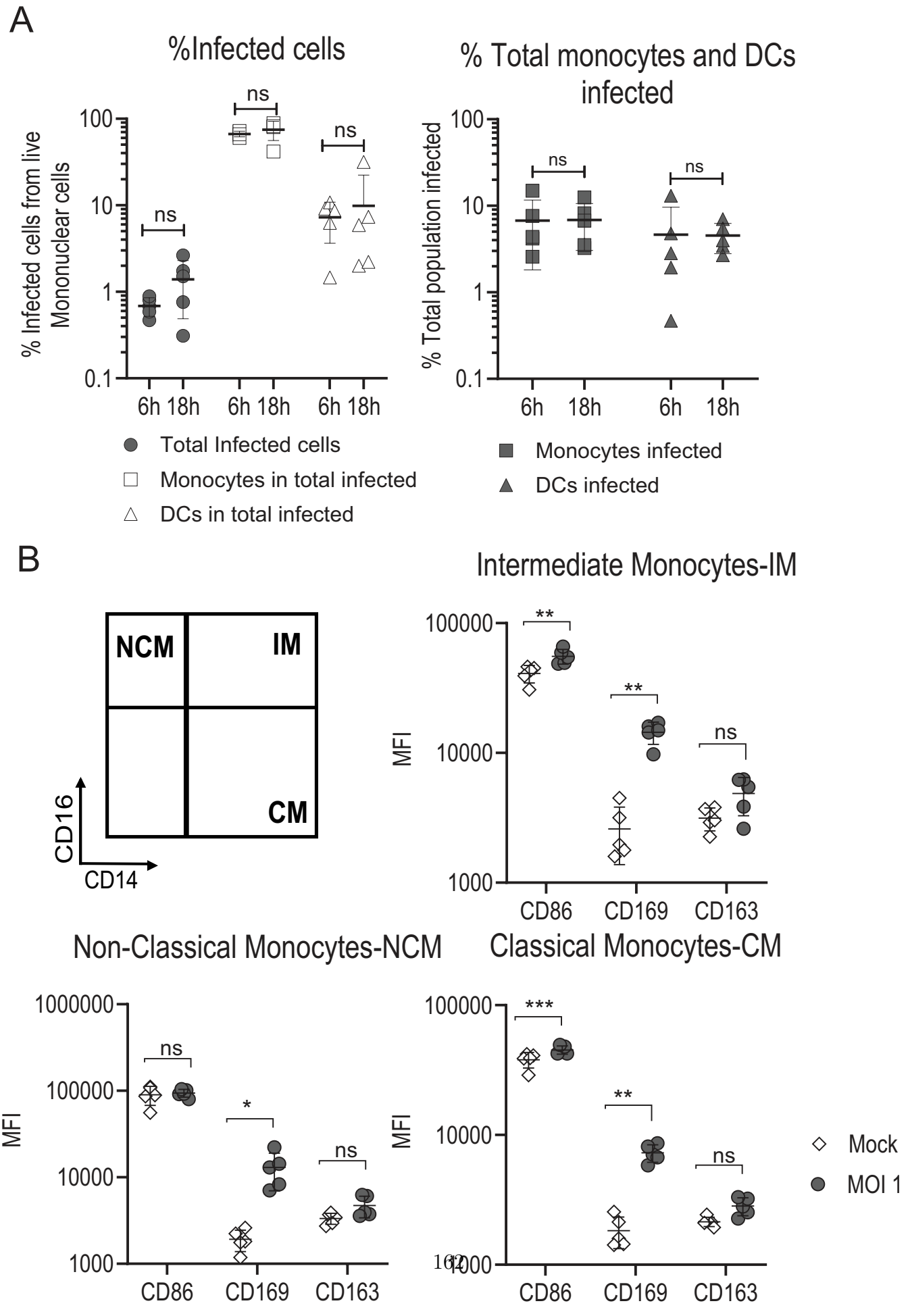
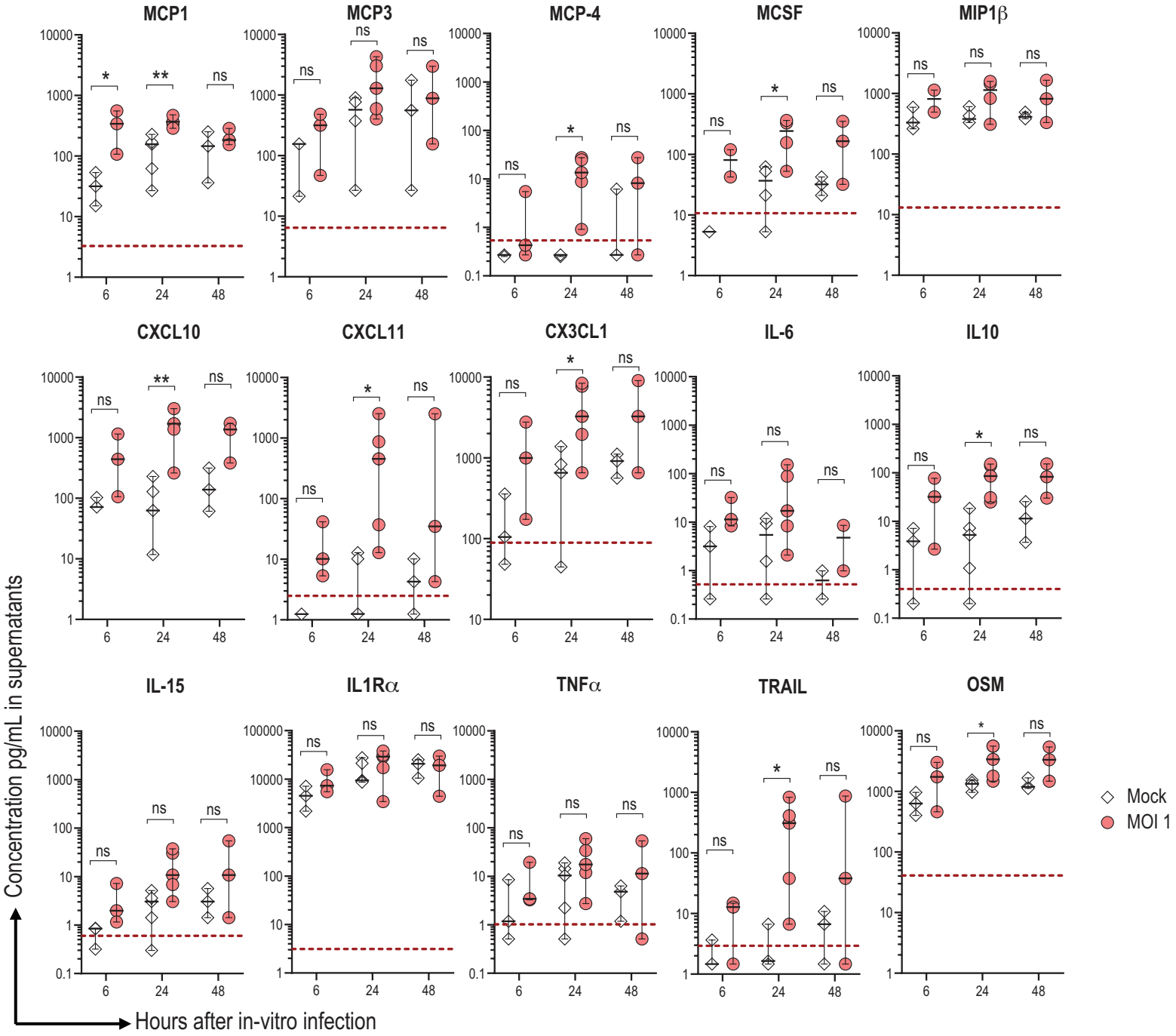

A
%Infected cells
% Infected cells from live Mononuclear cells
100
10
1
0.1
ns
ns
ns
6h 18h 6h 18h 6h 18h
Total Infected cells
Monocytes in total infected
DCs in total infected
% Total monocytes and DCs infected
% Total population infected
100
10
1
0.1
ns
ns
6h 18h 6h 18h
Monocytes infected
DCs infected
B
NCM
IM
CM
CD16
CD14
Intermediate Monocytes-IM
MFI
100000
10000
1000
**
**
ns
CD86 CD169 CD163
Non-Classical Monocytes-NCM
MFI
1000000
100000
10000
1000
ns
*
ns
CD86 CD169 CD163
Classical Monocytes-CM
MFI
100000
10000
1000
***
**
ns
CD86 CD169 CD163
Mock
MOI 1

Figure 2


MCP1
MCP3
MCP-4
MCSF
MIP1β
CXCL10
CXCL11
CX3CL1
IL-6
IL10
IL-15
IL1Rα
TNFα
TRAIL
OSM
Concentration pg/mL in supernatants
Hours after in-vitro infection
Mock
MOI 1

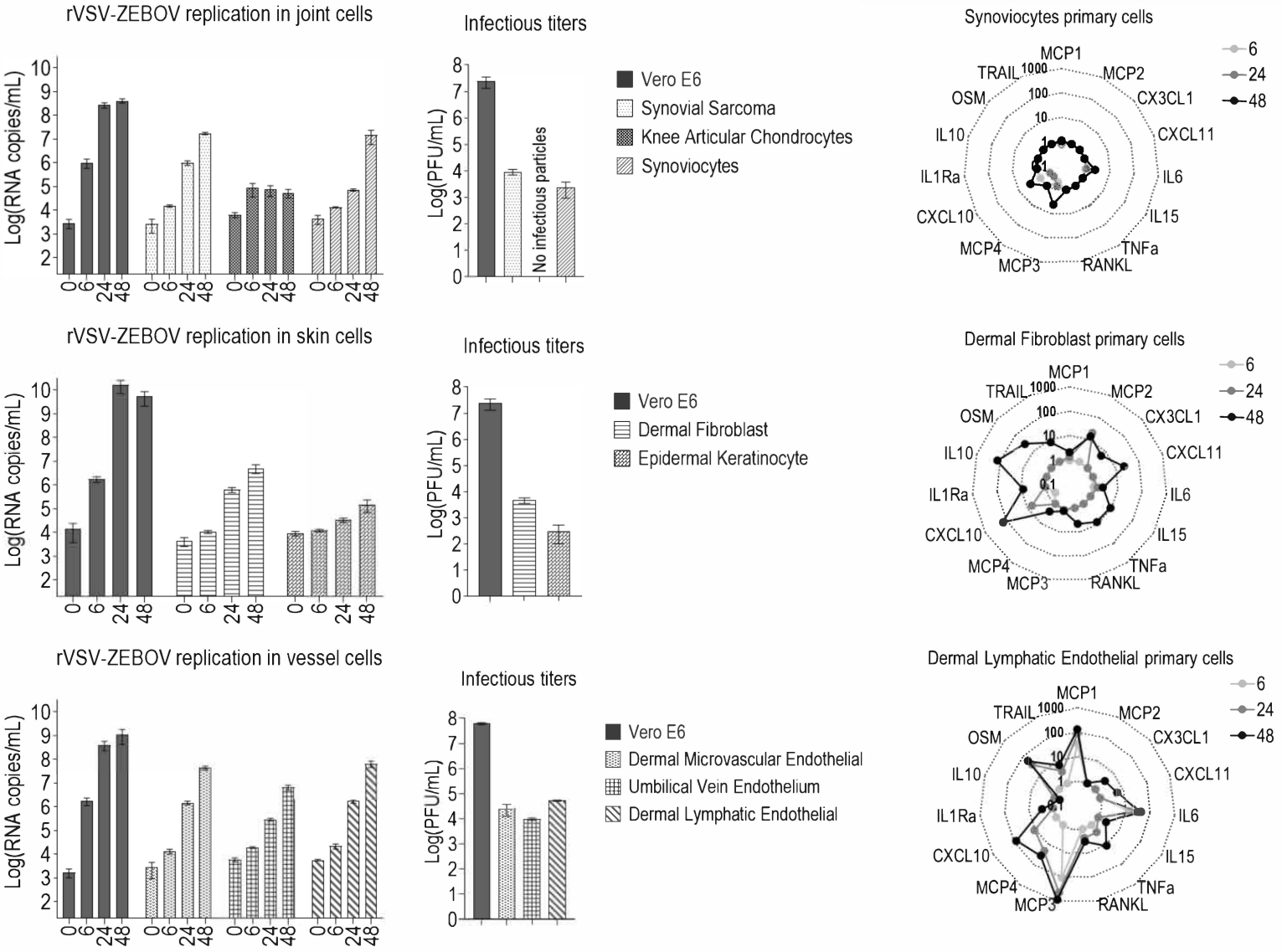
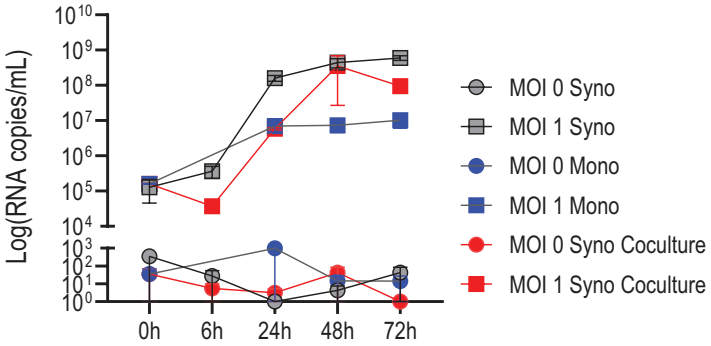
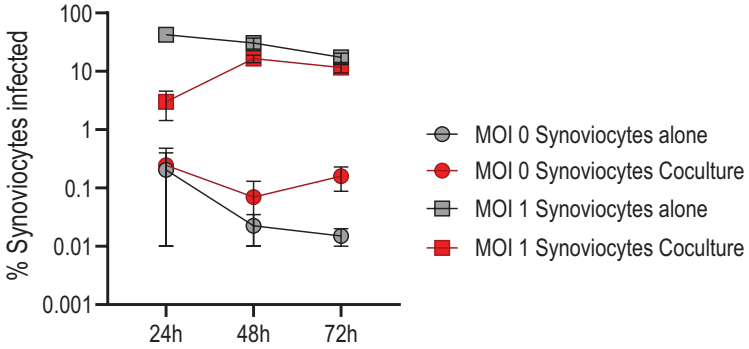

rVSV-ZEBOV replication in joint cells
Log(RNA copies/mL)
0 6 24 48
Infectious titers
Log(PFU/mL)
No infectious particles
Vero E6
Synovial Sarcoma
Knee Articular Chondrocytes
Synoviocytes
Synoviocytes primary cells
MCP1
MCP2
CX3CL1
CXCL11
IL6
IL15
TNFa
RANKL
MCP3
MCP4
CXCL10
IL1Ra
IL10
OSM
TRAIL
6
24
48
rVSV-ZEBOV replication in skin cells
Dermal Fibroblast
Epidermal Keratinocyte
Dermal Fibroblast primary cells
rVSV-ZEBOV replication in vessel cells
Dermal Microvascular Endothelial
Umbilical Vein Endothelium
Dermal Lymphatic Endothelial
Dermal Lymphatic Endothelial primary cells

Figure 4

A


Log(RNA copies/mL)
10^{10}
10^{9}
10^{8}
10^{7}
10^{6}
10^{5}
10^{4}
10^{3}
10^{2}
10^{1}
10^{0}
0h
6h
24h
48h
72h
MOI 0 Syno
MOI 1 Syno
MOI 0 Mono
MOI 1 Mono
MOI 0 Syno Coculture
MOI 1 Syno Coculture

B


% Synoviocytes infected
100
10
1
0.1
0.01
0.001
24h
48h
72h
MOI 0 Synoviocytes alone
MOI 0 Synoviocytes Coculture
MOI 1 Synoviocytes alone
MOI 1 Synoviocytes Coculture

C

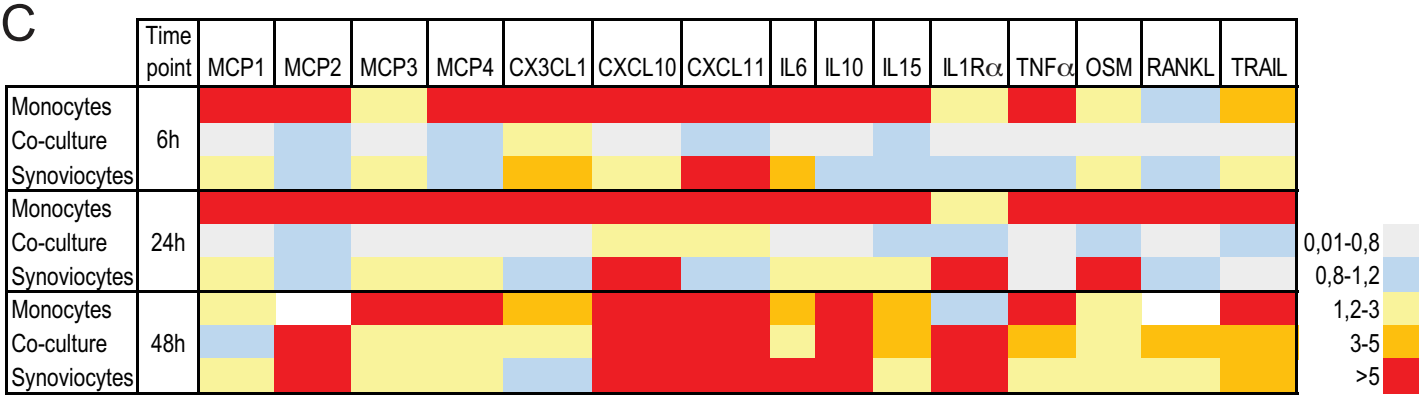
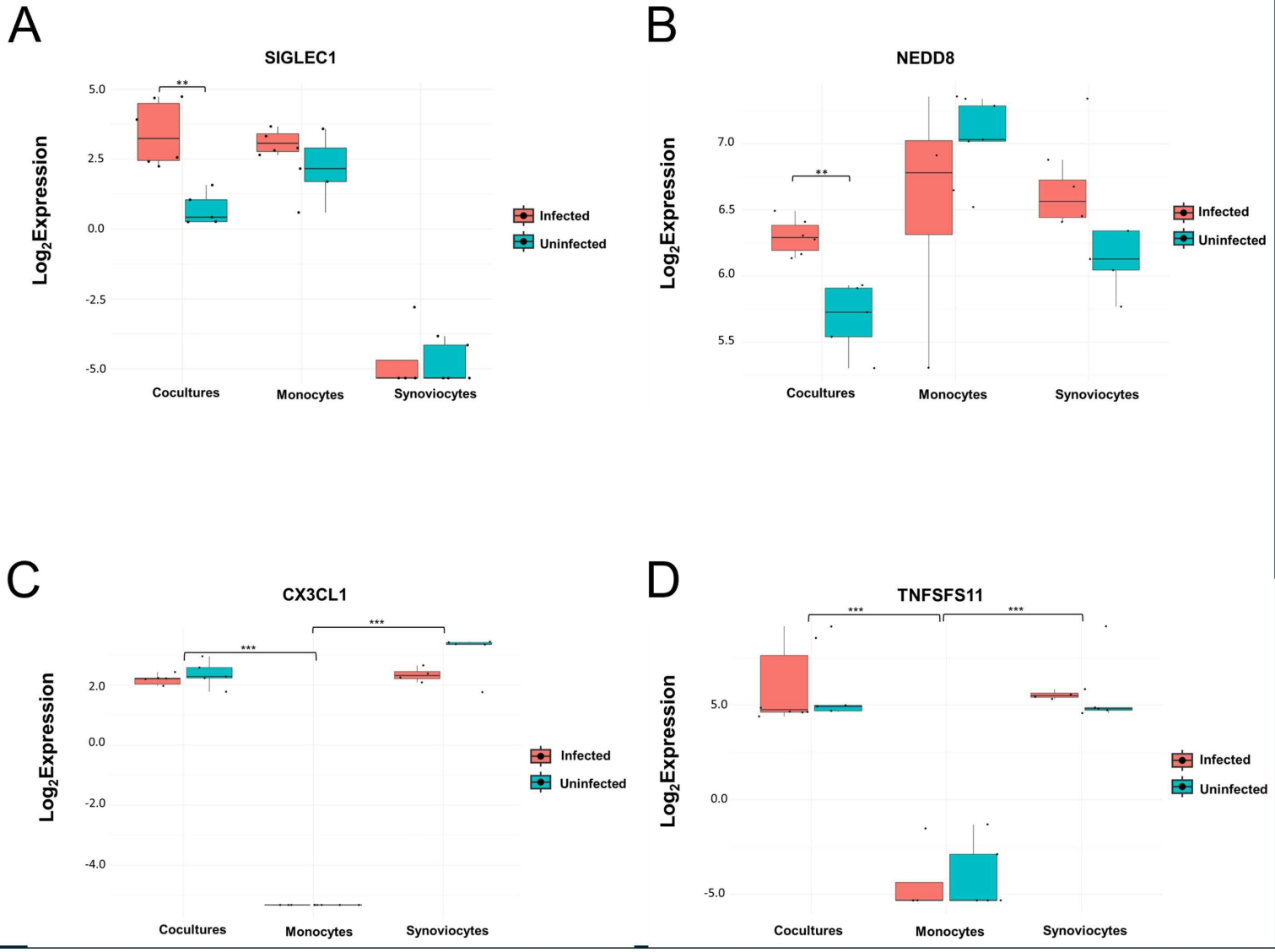

Time point
MCP1
MCP2
MCP3
MCP4
CX3CL1
CXCL10
CXCL11
IL6
IL10
IL15
IL1Rα
TNFα
OSM
RANKL
TRAIL
Monocytes
Co-culture
Synoviocytes
6h
24h
48h
0,01-0,8
0,8-1,2
1,2-3
3-5
>5

Figure 5


A
SIGLEC1
Log2Expression
5.0
2.5
0.0
-2.5
-5.0
**
Cocultures
Monocytes
Synoviocytes
Infected
Uninfected
B
NEDD8
Log2Expression
7.0
6.5
6.0
5.5
**
Cocultures
Monocytes
Synoviocytes
Infected
Uninfected
C
CX3CL1
Log2Expression
2.0
0.0
-2.0
-4.0
***
***
Cocultures
Monocytes
Synoviocytes
Infected
Uninfected
D
TNFSFS11
Log2Expression
5.0
0.0
-5.0
***
***
Cocultures
Monocytes
Synoviocytes
Infected
Uninfected

Appendix D

Acknowledgments

I would like to express my sincere gratitude to all those who have supported and guided me throughout the course of my doctoral studies. This work would not have been possible without their contribution.

First and foremost, I am deeply grateful to Prof.Francesco Santoro and Prof.Francesco Ianelli for their continuous guidance, constructive feedback, and constant encouragement. Their expertise and dedication have been invaluable in shaping both this thesis and my development as a researcher. I would also like to thank the members of the laboratory and research team for their collaboration, technical support, and for creating a stimulating and supportive working environment. Their willingness to share knowledge and discuss ideas has greatly enriched this work.

Finally, I would like to express my deepest gratitude to my family for their unconditional support, patience, and encouragement. Their belief in me has been the foundation that sustained me throughout this journey.

Bibliography

[1] Nicholas Israel Nii-Trebi. "Emerging and Neglected Infectious Diseases: Insights, Advances, and Challenges". In: *BioMed Research International* (2017).

[2] Stephenson Niamh Jackson Yves. "Neglected tropical disease and emerging infectious disease: An analysis of the history, promise and constraints of two worldviews". In: *Global Public Health* (2015).

[3] Marzi A. Furuyama W. "Ebola Virus: Pathogenesis and Countermeasure Development". In: *The Annual review of Virology* (2019).

[4] Alvisi G Salata C. Calistri A. "Ebola Virus Entry: From Molecular Characterization to Drug Discovery". In: *Vol. Viruses.* (2019).

[5] Muhlberger E. "Filuvirus replication and transcription". In: *Future Virology* (2007).

[6] Chertow D. Baseler L. "The Pathogenesis of Ebola Virus Disease". In: *The Annual Review of Pathology Mechanism of Disease* (2017).

[7] Jacob S.T. Crozier I. et al. "Ebola virus disease". In: *Nature Review, Disease Primers* (2020).

[8] Glynn. "Age-specific incidence of Ebola virus disease". In: *The Lancet* (2015).

[9] WHO Ebola Response Team. "Ebola Virus Disease among Children in West Africa". In: *The New England Journal of Medicine* (2015).

[10] Dobbs K. et al. "Ebola virus disease in children: epidemiology, pathogenesis,management, and prevention". In: *Springer Nature* (2023).

[11] Marzi A. et al. "Current Ebola Vaccine Progress". In: *BioDrugs* (2019).

[12] Tomori O. et al. "Ebola Virus Disease: current vaccine solutions". In: *Current Opinion in Immunology.* (2021).

[13] Agnandji S. et al. "Phase 1 Trials of rVSV Ebola Vaccine in Africa and Europe-preliminary report". In: *The New England Journal of Medicine.* (2015).

[14] Huttner A. et al. “The effect of dose on the safety and immunogenicity of VSV Ebola condidate vaccine: a randomised double-blind, placebo-controlled phase1/2 trial”. In: *Lancet Infect Disease.* (2019).

[15] Henao-Restrepo. et al. “Efficacy and effectiveness of an rVSV-vectored vaccine in preventing Ebola virus disease: final results from the Guinea ring vaccination, open-label, cluster-randomised trial (Ebola Ça Suffit!)” In: *Lancet* (2017).

[16] Agnandji S. et al. “Safety and immonogenicity of rVSV-ΔG-ZEBOV-GP Ebola vaccine in adults and children in Lambaréné, Gabon: A phase 1 randomised trial”. In: *PLOS Medicine* (2017).

[17] Scarpini A. et al. “Visceral Leishmaniasis: Epidemiology, Diagnosis, and Treatment Regimens in Different Geographical Areas with a Focus on Pediatrics”. In: *Microorganisms* (2022).

[18] MacLean L.M. et al. “Trafficking and release of Leishmania metacyclic HASPB on macrophage invasion”. In: *Cellular microbiology* (2012).

[19] Kumari D. et al. “Virulence factors of Leishmania parasite: Their paramount importance in unraveling novel vaccine candidates and therapeutic targets”. In: *Life Sciences* (2022).

[20] Zijlstra E. “The immunology of post-kala-azar dermal leishmaniasis (PKDL)”. In: *Parasite and Vectors* (2016).

[21] Zijlstra E. et al. “Post-kala-azar dermal leishmaniasis”. In: *The Lancet infectious Diseases* (2003).

[22] Zijlstra E. “Biomarkers in Post-kala-azar Dermal Leishmaniasis”. In: *Frontiers in Cellular and Infection Microbiology* (2019).

[23] Srivastava S. et al. “Possibilities and challenges for developing a successful vaccine for leishmaniasis”. In: *PLOS Neglected Tropical Diseases* (2016).

[24] Osman M. et al. “A third generation vaccine for human visceral leishmaniasis and post kala azar dermal leishmaniasis: First-in-human trial of ChAd63- KH”. In: *Parasite and Vectors* (2016).

[25] Lacey C. et al. “LEISH2b - A phase 2b study to assess the safety, efficacy, and immunogenicity of the Leishmania vaccine ChAd63-KH in post-kala azar dermal leishmaniasis”. In: *Wellcome Open Research* (2022).

[26] Younis B. et al. “A randomized, double-blind phase 2b trial to evaluate efficacy of ChAd63-KH for treatment of post kala-azar dermal leishmaniasis”. In: *Molecular Therapy Methods & Clinical Development* (2024).

[27] Joshua Fierer. "Invasive Non-typhoidal Salmonella (iNTS) Infections". In: *Clinical Infectious Diseases* (2022).

[28] Mac Lennan C. et al. "Interleukin (IL)-12 and IL-23 are key cytokines for immunity against Salmonella in humans." In: *Journal of infected disease* (2004).

[29] Piccini G. Montomoli E. "Pathogenic signature of invasive non-typhoidal Salmonella in Africa: implications for vaccine development". In: *Human vaccines & immunotherapeutics* (2020).

[30] Gilchrist J. et al. "Invasive Nontyphoidal Salmonella Disease in Africa". In: *EcoSalPlus* (2019).

[31] Marchello. et al. "Incidence of non-typhoidal Salmonella invasive disease: A systematic review and meta-analysis". In: *J Infect* (2021).

[32] Balasubramanian R. et al. "The global burden and epidemiology of invasive non-typhoidal Salmonella infections". In: *Human vaccines & immunotherapeutics* (2019).

[33] ramasamy M. et al. "Salmonella Vaccine Study in Oxford (SALVO) trial: protocol for an observer-participant blind randomised placebo-controlled trial of the iNTS-GMMA vaccine within a European cohort". In: *BMJ Open* (2023).

[34] Scorza B. et al. "High Yield Production Process for Shigella Outer Membrane Particles". In: *PloS ONE* (2012).

[35] Rossi O. et al. "Toll-Like Receptor Activation by Generalized Modules for Membrane Antigens from Lipid A Mutants of Salmonella enterica Serovars Typhimurium and Enteritidis". In: *Clinical and Vaccine Immunology* (2016).

[36] Fiorino F. et al. "Long-Term Anti-Bacterial Immunity against Systemic Infection by Salmonella enterica Serovar Typhimurium Elicited by a GMMA-Based Vaccine". In: *Vaccines* (2021).

[37] Luciani F. et al. "Next generation deep sequencing and vaccine design: today and tomorrow". In: *Trends in Biotechnology* (2012).

[38] Lee A. et al. "Immunogenicity and Vaccine Shedding After 1 or 2 Doses of rVSVΔG-ZEBOV-GP Ebola Vaccine (ERVEBO®): Results From a Phase 2, Randomized, Placebo-controlled Trial in Children and Adults". In: *Clinical Infectious Disease* (2023).

[39] Ayodele A. et al. "Replication, safety and immunogenicity of the vectored Ebola vaccine rVSV-ΔG-ZEBOV-GP in a sub-Saharan African paediatric population: A randomised controlled, open-label trial in children aged 1-12 years living in Lambaréné, Gabon". In: *Journal of Infection* (2024).

[40] Simon Andrews. *Babraham Bioinformatics*. URL: `https://www.bioinformatics.babraham.ac.uk/projects/fastqc/`.

[41] Anthony M Bolger. et al. "Trimmomatic: a flexible trimmer for Illumina sequence data". In: *Bioinformatics* (2014).

[42] Simon Anders. et al. "HTSeq—a Python framework to work with high-throughput sequencing data". In: *Bioinformatics* (2014).

[43] Andres S. et al. "Count-based differential expression analysis of RNA sequencing data using R and Bioconductor". In: *Nature protocols* (2013).

[44] Love M. et al. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2". In: *Genome Biology* (2014).

[45] Li S. et al. "Molecular signatures of antibody responses derived from a systems biological study of 5 human vaccines". In: *Nature Immunology* (2014).

[46] Weiner J. et al. "tmod: an R package for general and multivariate enrichment analysis". In: *PeerJPreprints* (2016).

[47] Guangchuang Yu. et al. "clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters". In: *OMICS A Journal of Integrative Biology* (2012).

[48] Kanehisa M.A. et al. "KEGG biological systems database as model of the real world". In: *Nucleic Acids Research* (2025).

[49] Tomczak A. et al. "Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations". In: *Nature scientific reports* (2018).

[50] Russo P. et al. "CEMiTool: a Bioconductor package for performing comprehensive modular coexpression analyses". In: *BMC Bioinformatics* (2018).

[51] Korotkevich G. et al. "Fast gene set enrichment analysis". In: *bioRxiv* (2021).

[52] Scrucca L. et al. "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models". In: *RJ* (2016).

[53] Nakaya HI. et al. "Systems biology of immunity to MF59-adjuvanted versus non-adjuvanted trivalent seasonal influenza vaccines in early childhood". In: *Proc Natl Acad Sci* (2016).

[54] Nakaya HI. et al. "Systems Scale Interactive Exploration Reveals Quantitative and Qualitative Differences in Response to Influenza and Pneumococcal Vaccines". In: *Immunity* (2013).

[55] Lei Ye. et al. "Profiling of Early Immune Responses to Vaccination Using THP-1-Derived Dendritic Cells". In: *Int J Mol Sci* (2024).

[56] Macdonald A. Richards K. H. "Putting the brakes on the anti-viral response: negative regulators of type I interferon (IFN) production". In: *Microbes Infect* (2011).

[57] Elena Winheim. et al. "Interferon-induced activation of dendritic cells and monocytes by yellow fever vaccination correlates with early antibody responses". In: *Proc Natl Acad Sci U S A* (2025).

[58] Douglas R Mathern . Peter S Heeger. "Molecules Great and Small: The Complement System". In: *Clinical Journal of the American Society of Nephrology* (201).

[59] Yuqing Zhang. et al. "ComBat-seq: batch effect adjustment for RNA-seq count data". In: *NAR Genomics and Bioinformatics* (2020).

[60] Alexander Dobin. et al. "STAR: ultrafast universal RNA-seq aligner". In: *Bioinformatics* (2012).

[61] McInnes L. et al. "McInnes L, Healy J, Saul N, Großberger L." In: *J Open Source Software* (2018).

[62] Li S. et al. "Molecular signatures of antibody responses derived from a systems biology study of five human vaccines". In: *Nat Immunol* (2014).

[63] Zyla J. et al. "Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms". In: *Bioinformatics* (2019).

[64] Russo P. "CEMiTool: a Bioconductor package for performing comprehensive modular coexpression analyses". In: *Bioinformatics* (2018).

[65] Anders S. Huber W. "Differential expression analysis for sequence count data". In: *Genome Biology* (2010).

[66] Bolotin DA. et al. "Antigen receptor repertoire profiling from RNA-seq data". In: *Nat Biotechnol* (2017).

[67] Bolotin A. et al. "MiXCR: software for comprehensive adaptive immunity profiling". In: *Nat Methods* (2015).

[68] Zhu H. et al. "Molecular correlates of vaccine-induced protection against typhoid fever". In: *The Journal of Clinical Investigation* (2023).

[69] Wickham H. "ggplot2: Elegant Graphics for Data Analysis". In: *Springer-Verlag* (2016).

[70] Wong WK. et al. "Comparative Analysis of the CDR Loops of Antigen Receptors". In: *Front Immunol* (2019).

[71] Tsuchiya Y. et al. "The diversity of H3 loops determines the antigen-binding tendencies of antibody CDR loops". In: *Protein Sci* (2016).

[72] Mancini F. et al. “The Known and The Unknown”. In: *Front Immunol* (2021).

[73] Mancini F. et al. “OMV Vaccines and the Role of TLR Agonists in Immune Response.” In: *Int J Mol Sci.* (2021).

[74] Rossi O. et al. “Modulation of Endotoxicity of Shigella Generalized Modules for Membrane Antigens (GMMA) by Genetic Lipid A Modifications”. In: *J Biol Chem* (2014).

[75] O’Connor D. et al. “Gene expression profiling reveals insights into infant immunological and febrile responses to group B meningococcal vaccine”. In: *Mol Syst Biol* (2020).

[76] Calabro S. et al. “Vaccine adjuvants alum and MF59 induce rapid recruitment of neutrophils and monocytes that participate in antigen transport to draining lymph nodes”. In: *Vaccine* (2011).

[77] Galson JD. et al. “BCR repertoire sequencing: different patterns of B-cell activation after two Meningococcal vaccines”. In: *Immunol Cell Biol* (2015).

[78] Mitchell R. et al. “Polysaccharide-specific B cell responses to vaccination in humans”. In: *Hum Vaccines Immunother* (2014).

[79] Zhu H. et al. “Molecular correlates of vaccine-induced protection against typhoid fever”. In: *Hum Vaccines Immunother* (2023).

[80] Forgacs D. et al. “Convergent antibody evolution and clonotype expansion following influenza virus vaccination”. In: *Plos One* (2021).

[81] Liu KJ. et al. “The Longitudinal Analysis of Convergent Antibody VDJ Regions in SARS-CoV-2-Positive Patients Using RNA-Seq”. In: *Plos One* (2021).

[82] Rao VN. et al. “Public antibodies: convergent signatures in human humoral immunity against pathogens”. In: *Yount J, editor. mBio* (2025).

[83] Dahora LC. et al. “Salmonella Typhi Vi capsule prime-boost vaccination induces convergent and functional antibody responses”. In: *Sci Immunol* (2021).

[84] Youis B. et al. “Safety and immunogenicity of ChAd63-KH vaccine in post-kala-azar dermal leishmaniasis patients in Sudan”. In: *Molecular Therapy* (2021).

[85] Zijlstra E.E El-Hassan A.M. “Post kala-azar dermal leishmaniasis”. In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* (2001).

[86] Zijlstra E.E et al. “Post-kala-azar dermal leishmaniasis”. In: *The Lancet Infectious Diseases* (2003).

[87] Ganguly S. et al. “Post-kala-azar dermal leishmaniasis–an overview”. In: *International Journal of Dermatology* (2010).

[88] Saurabh S. et al. “Changing clinico-epidemiology of post-kala-azar dermal leishmaniasis (PKDL) in India”. In: *Journal of Vector Borne Diseases* (2020).

[89] Zijlstra E.E et al. “Post-kala-azar dermal leishmaniasis in the Indian subcontinent: A threat to the South-East Asia Region Kala-azar Elimination Programme.” In: *Plos neglected tropical diseases* (2017).

[90] Le Rutte E A. et al. “The potential impact of human visceral leishmaniasis vaccines on population incidence”. In: *Plos neglected tropical diseases* (2020).

[91] Jorge A. et al. “Towards the elimination of visceral leishmaniasis as a public health problem in east Africa: reflections on an enhanced control strategy and a call for action”. In: *Lancet Global Health* (2020).

[92] Musa A M. et al. “Treatment-Based Strategy for the Management of Post-Kala-Azar Dermal Leishmaniasis Patients in the Sudan”. In: *Journal of Tropical Medicine* (2013).

[93] Datta A. et al. “Therapeutic Modalities in Post Kala-azar Dermal Leishmaniasis: A Systematic Review of the Effectiveness and Safety of the Treatment Options.” In: *Indian Journal of Dermatology* (2021).

[94] Zijlstra E.E et al. “PKDL and other dermal lesions in HIV co-infected patients with Leishmaniasis: review of clinical presentation in relation to immune responses.” In: *Plos neglected tropical diseases* (2014).

[95] Musa A M. et al. “Paromomycin and Miltefosine Combination as an Alternative to Treat Patients With Visceral Leishmaniasis in Eastern Africa: A Randomized, Controlled, Multicountry Trial.” In: *Clinical infectious Diseases* (2023).

[96] Younis B M. et al. “Safety and efficacy of paromomycin/miltefosine/liposomal amphotericin B combinations for the treatment of post-kala-azar dermal leishmaniasis in Sudan: A phase II, open label, randomized, parallel arm study”. In: *Plos neglected tropical diseases* (2023).

[97] Desjeux P. et al. “Report of the Post Kala-azar Dermal Leishmaniasis (PKDL) Consortium Meeting, New Delhi, India, 27-29 June 2012”. In: *Parasite Vectors* (2013).

[98] Ghalib H. et al. “Consultation meeting on the development of therapeutic vaccines for post kala azar dermal leishmaniasis”. In: *Kinetoplastid Biology and Disease* (2007).

[99] Zijlstra E.E et al. “Report of the Fifth Post-Kala-Azar Dermal Leishmaniasis Consortium Meeting,Colombo, Sri Lanka, 14-16 May 2018”. In: *Parasite Vectors* (2020).

[100] Das VNR. et al. "Development of post-kala-azar dermal leishmaniasis in AmBisome treated visceral leishmaniasis: a possible challenge to elimination program in India". In: *Journal of Postgraduate Medicine* (2013).

[101] Mukhopadhyay D. et al. "Post kala-azar dermal leishmaniasis: an unresolved mystery". In: *Trends in Parasitlogy* (2014).

[102] Ismail A. et al. "The pathogenesis of post kala-azar dermal leishmaniasis from the field to the molecule: does ultraviolet light (UVB) radiation play a role?" In: *Medical Hypotheses* (2006).

[103] Musa A. et al. "The natural history of Sudanese post kala-azar dermal leishmaniasis: clinical, immunological and prognostic features". In: *Ann. Trop. Med. Parasitol* (2002).

[104] Mukherjee S. et al. "Impaired activation of lesional CD8(+) T-cells is associated with enhanced expression of Programmed Death-1 in Indian Post Kala-azar Dermal Leishmaniasis". In: *Scientific reports* (2019).

[105] Mukhopadhyay D. et al. "Impaired activation of lesional CD8(+) T-cells is associated with enhanced expression of Programmed Death-1 in Indian Post Kala-azar Dermal Leishmaniasis". In: *Plos neglected tropical disease* (2015).

[106] Zijlstra E E. et al. "The immunology of post-kala-azar dermal leishmaniasis (PKDL)". In: *Parasites Vectors* (2016).

[107] Ramesh V. et al. "Histopathology of Post Kala-azar Dermal Leishmaniasis". In: *Indian Journal of Dermatology* (2020).

[108] Khalil EAG. et al. "Safety and immunogenicity of an autoclaved Leishmania major vaccine". In: *Indian Journal of Dermatology* (2000).

[109] Khalil EAG. et al. "Autoclaved Leishmania major vaccine for prevention of visceral leishmaniasis: a randomised, double-blind, BCG-controlled trial in Sudan". In: *Lancet* (2000).

[110] Satti I N. et al. "Immunogenicity and safety of autoclaved Leishmania major plus BCG vaccine in healthy Sudanese volunteers". In: *Vaccine* (2001).

[111] Kamil A A. et al. "Alum-precipitated autoclaved Leishmania major plus bacille Calmette-Guerrin, a candidate vaccine for visceral leishmaniasis: safety, skin-delayed type hypersensitivity response and dose finding in healthy volunteers". In: *Transaction of the Royal Society of Tropical Medicine and Hygiene* (2003).

[112] Khalil EAG. et al. "Safety and immunogenicity of a candidate vaccine for visceral leishmaniasis (Alum-precipitated autoclaved Leishmania major + BCG) in children: an extended phase II study". In: *Ann.Trop. Paediatr* (2006).

[113] Musa AM. et al. "Leishmaniasis Research Group/Sudan (2008). Immunochemotherapy of persistent post-kala-azar dermal leishmaniasis: a novel approach to treatment". In: *Transaction of the Royal Society of Tropical Medicine and Hygiene* (2008).

[114] Osman M. et al. "A third generation vaccine for human visceral leishmaniasis and post kala azar dermal leishmaniasis: First-in-human trial of ChAd63-KH." In: *PLoS Neglected Trop. Dis* (2017).

[115] Youni B. et al. "A third generation vaccine for human visceral leishmaniasis and post kala azar dermal leishmaniasis: First-in-human trial of ChAd63-KH." In: *Molecular Therapy* (2021).

[116] Zijlstra E E. et al. "Endemic kala-azar in eastern Sudan: post-kala-azar dermal leishmaniasis." In: *The american Journal of Tropical Medicine and Hygiene* (1995).

[117] Hagan T. et al. "Transcriptional atlas of the human immune response to 13 vaccines reveals a common predictor of vaccine-induced antibody responses". In: *Nature Immunology* (2022).

[118] Gasim S. et al. "High levels of plasma IL-10 and expression of IL-10 by keratinocytes during visceral leishmaniasis predict subsequent development of post-kala-azar dermal leishmaniasis". In: *Clinical and experimental Immunology* (1998).

[119] Volpedo G. et al. "The History of Live Attenuated Centrin Gene-Deleted Leishmania Vaccine Candidates." In: *Pathogens* (2022).

[120] Maini M K. et al. "Tissue T cells in prophylactic and therapeutic vaccination responses". In: *Seminars in Arthritis and Rheumatism* (2023).

[121] Dey N S. et al. "arly reduction in PD-L1 expression predicts faster treatment response in human cutaneous leishmaniasis". In: *J.Clin.Invest* (2021).

[122] Dey N S. et al. "IL-32 producing CD8+ memory T cells and Tregs define the IDO1/PD-L1 niche in human cutaneous leishmaniasis skin lesions." In: *medRxiv* (2024).

[123] Hill A V S. et al. "Prime-boost vectored malaria vaccines: progress and prospects". In: *Human Vaccine* (2010).

[124] Ramasamy M N. et al. "Safety and immunogenicity of ChAdOx1 nCoV-19 vaccine administered in a prime-boost regimen in young and old adults (COV002): a single-blind, randomised, controlled, phase 2/3 trial". In: *The Lancet* (2021).

[125] Rotrosen E. et al. "Assessing the generation of tissue resident memory T cells by vaccines." In: *Nature reviews immunology* (2023).

[126] Wise J. et al. “Covid-19: Two rare vaccine side effects detected in large global study”. In: *BMJ* (2024).

[127] Lacey j et al. “LEISH2b - A phase 2b study to assess the safety, efficacy, and immunogenicity of the Leishmania vaccine ChAd63-KH in post-kala azar dermal leishmaniasis”. In: *Welcome Open research* (2022).

[128] Wolf-Henning Boehncke. Michael P Schön. “Psoriasis”. In: *Lancet* (2015).

[129] Huang R. et al. “An Exploration of the Role of MicroRNAs in Psoriasis”. In: *Medicine* (2015).

[130] Guzmán-Martín CA. et al. “miR-16-5p, miR-21-5p, and miR- 155-5p in circulating vesicles as psoriasis biomarkers”. In: *Scientific Reports* (2025).