THE UNIVERSITY
of EDINBURGH

Experimental vaccination for onchocerciasis and the identification of early markers of protective immunity

Jessica. A. S. Duprez

Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2017

Declaration

I hereby declare that the work presented in this thesis was composed by myself and that the data collected described herein was primarily carried out by myself, unless stated otherwise in the text, such as data collection carried out by a project student (Marjorie Besançon) and the gene expression data collected by E PIAF members and processed by Fios (Edinburgh, UK). This work has not previously been submitted for any other degree of personal qualification.

Jessica Anais Sybille Duprez

July 2017

Acknowledgements

The research carried out during the last four years would not have been possible without the continued help and support from many people. Firstly, I would like to thank both of my supervisors Prof David Taylor, and Dr Simon Babayan, for their guidance and expertise throughout my PhD that have culminated in this thesis. To Prof David Taylor, thank you for your continuous help, support and motivation and to Dr Simon Babayan and Wei Liu for introducing me to the world of machine learning and guiding me through it. I would also like to thank the Principals career fund (University of Edinburgh) and E PIAF (EU) for providing funding for my PhD.

Having started my PhD in Edinburgh and then moved to Glasgow, there are many people that contributed to my thesis. In Edinburgh: Nick Gray for providing lab support; Alison Fulton for keeping the *L. sigmodontis* life cycle going and for providing my experiments with *L. sigmodontis* parasites (which often coincided with her holidays, for that I apologies); to members of the Maizel lab, Yvonne Harcus and Stephanie Ryan for teaching me new techniques in the lab and for being great support and friends. In Glasgow, I would like to thank Elizabeth Kilbride for her help throughout most of my PhD and to the rest of the technical staff, and to members of Graham Kerr who have welcomed me and helped me throughout my PhD.

A huge thanks Lorna and Laura, who have pretty much been my psychologist and motivational coaches throughout my PhD, and to Maude, Sophie's, Chrissy, Nardus, Julie, Darryl, and many more, not only for their support and solidarity, but also providing many coffee and tea breaks, climbing/yoga/hiking adventures, and many fond memories.

Last but not least, my family and friends out of my PhD life, who have provided more support than I probably deserve. I would like to say a massive thank you / merci / gracias, to my parents, who have believed in me despite not really understating why I did a PhD, and because of them I have made it this far; to my grandparents who have always been there to shoulder me on; to my sister who not only came up and looked after me after breaking my ankle but equally is always there to listen to my nonsense. A massive thank you to everyone that has helped, supported me or provided me with coffee, I can honestly say I couldn't have done it without you all.

Abstract

Onchocerciasis, caused by *Onchocerca volvulus* remains a major public health and socio-economic problem across the tropics, despite years of mass drug administration (MDA) with Ivermectin to reduce disease burden. Through modelling, it has been shown that elimination cannot be achieved with MDA alone and additional tools are needed, such as vaccination, which remains the most cost-effective tool for long-term disease control. The feasibility behind vaccination against *O. volvulus* can be demonstrated in the *Litomosoides sigmodontis* mouse model, which shows that vaccine induced protection can be achieved with immunisation using irradiated L3, the infective stage of *L. sigmodontis* and with microfilariae (Mf), the transmission stage of the parasite. There is further evidence of protective immunity in humans, with individuals living in endemic areas that show no signs of infection despite being exposed to the parasite (endemic normal).

The protective efficacy of promising vaccine candidates were evaluated using an immunisation time course in the *L. sigmodontis* model, using either DNA plasmid or peptide vaccines. In immunisation experiments in *L. sigmodontis,* Mf numbers are used as a measure of protection and marks the end of an immunisation time course. However, when changes in gene expression were measured at the end of an immunisation time course, in attempts to identify gene signatures that could be used as markers of protection (correlates of protection) in the blood, no gene signatures were found to be associated with protection. This suggest that at the end of an immunisation time course, when protection is measured (change in Mf numbers), it is too late in infection to measure changes in immune pathways being triggered.

Changes in gene expression were therefore measured in blood samples collected throughout an immunisation time course in the *L. sigmodontis* model, in order to identify the time point in an immunisation experiment which are the most indicative of protection. Two independent immunisation time courses were used, either using irradiated L3 or Mf as vaccine against *L. sigmodontis*, as these elicit the greatest protection. This generated a large high dimensional dataset, that was too large and complex for a differential fold-change analysis. Therefore, an analysis pipeline was created using machine learning algorithms, to detect changes in gene expression throughout the time courses to detect markers of protection.

The 6 hour time point following immunisation showed the greatest change in gene expression, with the analysis pipeline identifying known pathways associated with vaccine-induced immunity. The pipeline was applied to gene expression data from human samples obtained from individuals living in endemic areas who were either infected with *O. volvulus* or endemic normal (naturally protected), this was to identify pathways associated with protective immunity in humans. When comparing vaccine induced immunity seen in mice and natural protective immunity in humans there was some overlap in pathways being triggered, suggesting that similar pathways are needed for protection and that if a vaccine can trigger the right pathways in mice, it is likely to be effective in humans.

Overall the machine learning analysis of the gene expression data, not only shows that it is feasible to measure change in gene expression in blood during filarial infections, but that during an immunisation time course it is the early time points following immunisation that are the most predictive of vaccine efficacy (protection outcome).

One of the vaccine candidates, cysteine protease inhibitor-2 (CPI), is a known immuno-modulator that inhibits MHC-II antigen presentation on antigen presenting cells such as dendritic cells (DC). This candidate has consistently been shown to induce protection if its immuno-modulatory active site was modified. In *in vitro* studies, it was shown that modification of the active site of CPI rescues antigen presentation in DC. This shows the importance of DC activation before the onset of infection, demonstrating the importance of triggering protective responses early in infection, and provides insight on how one of the vaccine candidates achieves protection.

Lay summary

Onchocerciasis, also known as river blindness is a major public health and socio-economic problem in Africa and in small areas in South America, and despite years of effort from control programmes this debilitating disease is still present. Therefore, an alternative control method is needed, such as a vaccine, if one day this disease it to be eliminated. In this study, potential vaccine candidates were tested in an animal model for onchocerciasis, and at the end of the vaccination time course presence or absence of parasites were measured. Although this measure is a good indicator of protective efficacy, it does not give an indication of what the immunology behind this protection is. Further studies were done to look at indicators in the blood of animals vaccinated with known protective vaccines. Since this produced a large dataset, computational tools were used to identify what the kind of protective immunity was being triggered throughout this time course. Samples taken at time points immediately after the vaccine was given, provided the most information, of the immune pathways being triggered. Overall this thesis has validated certain vaccine candidates against onchocerciasis and shown that markers of protection can be measured in the blood during a vaccination time course. This will be useful in the future to measure the efficacy of vaccine after vaccination instead of having to wait till the end of vaccination time course to know if a vaccine is protective.

Contents

List of Abbreviations

AAM – Alternatively activated macrophages

APC – Antigen presenting cell

BMDC – Bone marrow derived dendritic cells

BP – Biological processes

CPI – Cysteine protease inhibitor

CPIm – Modified cysteine protease inhibitor

DC – Dendritic cells

EN – Endemic normal

ES – Excretory / secretory

GEO – Generalised onchocerciasis

GLM – Generalised linear model

GO – Gene ontology

ICOS – Inducible T-cell costimulator (CD278)

IFN-γ – Interferon gamma

Ig – Immunoglobulin

IL – Interleukin

L3 – Larval stage 3

LF – Lymphatic filariasis

LPS – Lipopolysaccharide

LN – Lymph node

MDA – Mass drug administration

Mf – Microfilariae

MHC – Major histocompatibility complex

MIP1α – Macrophage inflammatory protein 1 alpha

ML – Machine learning

OOB – Out-of-bag error

OVA – Ovalbumin

pEmpty – Empty pcDNA3.1 plasmid

PD-1 – Programmed cell death-1

RF – Random forest

RFE – Recursive feature elimination

ShK – A 6-cysteine domain protein, vaccine candidate

SVM – Support vector machine

TCR – T cell receptor

Teff – Effector T cell

TGF-β – Transforming growth factor beta

Tgh-2 - Transforming growth factor beta homologue in mice

Th – T helper cell

TLR – Toll like receptor

TpD – A chimeric MHC class II peptide containing epitopes form tetanus and diphtheria toxin

TPX – Thioredoxin peroxidase

Treg – Regulatory T cells

Tr1 – Type 1 regulatory T cells

WGCNA – Weighted gene correlation network analysis

WT – Wild type

List of Figures

List of Tables

Chapter 1. General Introduction

1.1 Overview of Onchocerciasis

Onchocerciasis, also known as river blindness, is caused by the filarial nematode *Onchocerca volvulus*, which is transmitted from person to person by a blackfly (*Simulium* spp). Onchocerciasis is endemic in 30 African countries, as well as in Yemen and localised foci in six Latin American countries (Figure 1.1). It is suggested that 95% of *O. volvulus* infected individuals were living in Africa, and with the use of REMO (rapid epidemiological mapping of onchocerciasis) it is estimated that 37 million people are infected, and 90 million are at risk in Africa (APOC, 2013; World Health Organization, 2016b). Onchocerciasis is a debilitating neglected tropical disease, imposing a global health burden of 0.5 million disability-adjusted life-years (DALYs) (Murray *et al.*, 2012). Although onchocerciasis is mostly associated with a chronic infection, with varying degrees of mild pathology, a proportion of individuals do develop severe symptoms including visual impairment, which can lead to irreversible blindness; severe itching with secondary infections if scratched; and hyperactive dermatitis which is termed sowda (blackening of the skin). As a consequence of this debilitating pathology there is stigma and fear of this disease, historically causing people to abandon fertile lands creating an even bigger socio-economic impact (Houweling *et al.*, 2016).

The prevalence of onchocerciasis infection and disease in a community is correlated to the proximity to rivers in which the blackflies breed, with microfilariae (Mf) densities, the transmission stage of *O. volvulus*, increasing with age until around 30

years old, after which infection profiles vary between geographical regions and age (Duerr *et al.*, 2004; Filipe *et al.*, 2005). Endemicity is split depending on prevalence of infection, with areas hyper-endemic for onchocerciasis having more than 60% of the population present with Mf in the skin, with high Mf density levels (<50 Mf/mg skin). A prevalence between 35-65% of onchocerciasis within a population is classified as meso-endemic, and hypo-endemic areas have less than 35% of the population with detectable Mf in the skin.

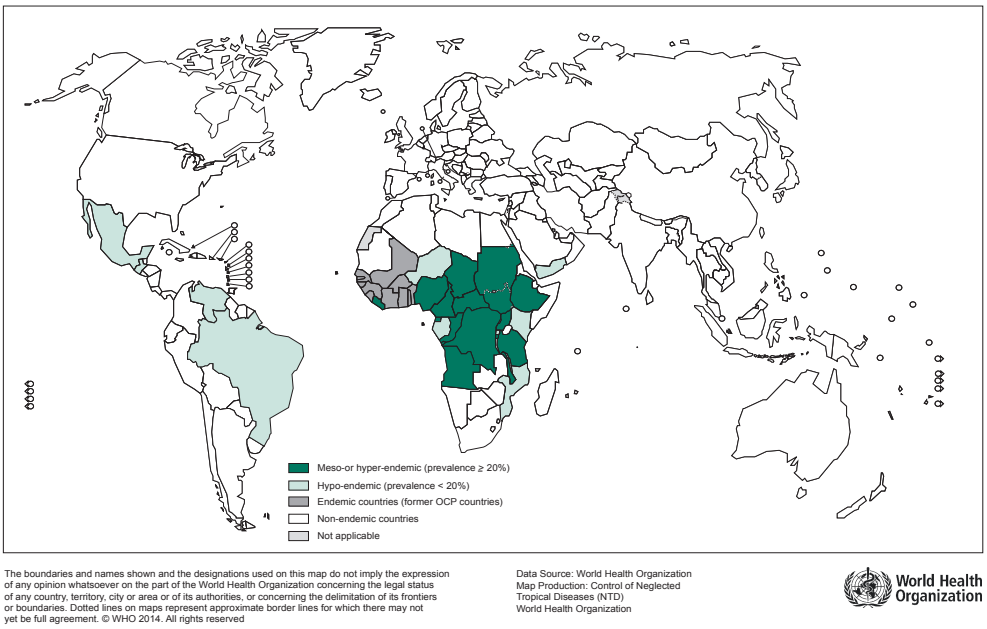

Figure 1.1. Worldwide distribution of onchocerciasis, 2013 (WHO). Map showing different distribution of onchocerciasis around the world, dark green represents meso-or hyper-endemic areas, classified as a prevalence of greater then 20%; light green as hypo-endemic areas, classified as a prevalence less than 20%. Map obtained from World Health Organisation http://www.who.int/onchocerciasis/epidemiology/en/.

1.2 Overview of related human filarial nematodes

Other filarial species that have a human host include *Wuchecheria bancrofti*, *Brugia malayi*, *Brugia timori* which are the causative agents of lymphatic filariasis; *Loa loa* which causes Loiasis; and other less known but well distributed species are *Mansonella streptocerca*, *Mansonella perstans* and *Mansonella ozzardi* (Table 1.1).

Lymphatic filariasis is another cause of chronic morbidity with a negative social impact, and as its name suggests it affects the lymphatics, with severe symptoms such as hydroceles, an accumulation of fluid in the scrotum; and elephantiasis, an obstruction of lymphatic vessels resulting in extreme swelling of skin and tissues, typically in the legs (Otabil, Tenkorang, 2015).

Loiasis is characterised by migration of adult worms across the eye and "Calabar swelling". Calabar swelling are subcutaneous oedemas that are often associated with localised and general itching, frequently found on the limbs, especially on the forearms. These oedemas can cause restricted movement of the nearest joint, and can disappear spontaneously and re-appear at irregular intervals. Compared to onchocerciasis and lymphatic filariasis, loiasis has been much less studied, despite having been known for a long time, most likely due to loiasis being less wide-spread, restricted to equatorial west and central Africa, and severe pathology is rare (Boussinesq, 2013). More recently, it has emerged as a disease of significant public health importance because of its potential negative impact on onchocerciasis control programmes. Individuals with high levels of *L. loa* Mf in their blood (<30,000 Mf/ml), that take the drug Ivermectin, the treatment used for onchocerciasis, have an increased risk of developing severe adverse neurological reactions (Chesnais *et al.*, 2017).

Therefore, in areas with high intensity *L. loa* infections co-endemic for onchocerciasis (Figure 1.2), Ivermectin cannot be used for mass drug administration (MDA), hence hindering control campaigns.

Mansonella species have received even less attention, despite being widespread in many parts of sub-Saharan Africa, notably *M. perstans* which is found in 33 countries within this region, often with a very high prevalence within the population. *Mansonella perstans* is responsible for serous cavity filariasis in humans; infections are often asymptomatic, but when symptoms are seen these include subcutaneous swelling, aches, pains, skin rashes, hormonal disturbances and hypereosinophilia. Despite it being considered as one of the most prevalent parasites in tropical Africa it has received very little attention, mainly because it causes few clinical symptoms, and no severe adverse reactions are detected with drugs used for lymphatic filariasis or onchocerciasis. However, *M. perstans* might be interfering with the host regulatory mechanisms, influencing the outcome of other infections such as malaria, tuberculosis, and HIV, but also potentially influencing vaccine mediated immunity, however this had yet to be proven (Simonsen *et al.*, 2011).

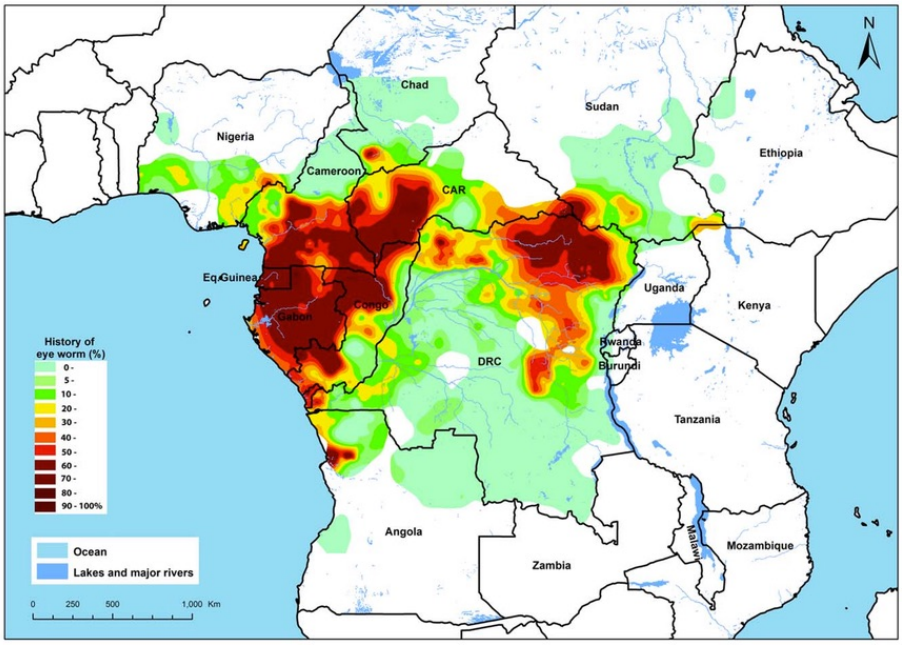

Figure 1.2. Map of estimated prevalence of *Loa loa* in Africa. This map represents an estimate of the distribution of loiasis based on survey data (RAPLOA). Darker areas represent areas of hyper endemicity, demonstrating that there are two zones of high endemicity (dark red). Areas with a prevalence of 40% were classified as high risk areas for Ivermectin treatment. Map taken from Zouré et al. (Zouré *et al.*, 2011)

Table 1.1. Filarial species with human host

Disease & filarial nematode		Burden	Life cycle	Pathology	Treatment / Control	References
Onchocerciasis	*Onchocerca volvulus* Africa, Latin America and Yemen 95% of infected people live in 30 African countries	• Estimated 187 million people at risk in 37 countries. • 37 million people infected. • 4 million suffer from severe itching or dermatitis • 265, 000 are blind	Adult worms reside within subcutaneous nodules, and Mf migrate to the skin and eyes. Vector: Blackfly (*Simulium spp)*	Onchocerciasis can cause severe skin itching (dermatitis) and onchocercomas (subcutaneous nodules), as well as ocular lesions that can progress to blindness. • Severely debilitating pathology	• APOC in Africa • OEPA in the Americas Their main strategy is MDA with Ivermectin, were possible	(World Health Organization, 2016b) (Hotez, Kamath, 2009) (Taylor *et al.*, 2010)
Lymphatic Filariasis	*Wuchereria bancrofti* Throughout the tropics & accounts for 90% of the LF cases in Africa *Brugia malayi* Southeast Asia *Brugia timori* Southeast Indonesia	• 1 billion at risk in 73 countries • 67 million infected • 35 million suffering from severe pathology: • 25 million men with genital disease • 15 million elephantiasis of the leg	Adult worms reside within the lymphatic vessels, and Mf migrate to the blood, in a periodical pattern to match vector feeding habit. Vector: Various mosquito species	Lymphedema: an accumulation of lymphatic fluid generally in limbs. Hydrocele: fluid accumulation in the scrotal sac	Annual MDA with combination of either albendazole and Ivermectin or DEC[1] depending on co-endemicity	(World Health Organization, 2013) (World Health Organization, 2016a) (Hotez, Kamath, 2009) (Taylor *et al.*, 2010)

Loiasis	• *Loa Loa* Present in 10 counties in Africa: restricted to equatorial rain forest of Central and West Africa	• 30 million people at risk • 13 million infected	Adults live in subcutaneous tissues. Mf have diurnal periodicity (in the blood during the day, and non-circulating Mf reside in the lungs. Vector: Deerfly (*Chrysops spp*)	Adult worms pass through the sub-conjunctiva of the eye, migration of adult worm can cause oedema (Calabar swelling) and skin itching.	DEC[1] is given to patients with low Mf loads. In patients with high Mf densities DEC may produce severe adverse effect, therefore albendazole can be given.	(Zouré *et al.*, 2011)
Mansonella species	<u>*Mansonella ozzardi*</u> New world, humid, warm regions from Mexico till north Argentina	• Not known	L3 are transmitted during vector blood meal. Adults appear to live in the serous body cavities, females produce Mf, which migrate to the blood Vector: Biting midges (*Culicoides spp*)	In rare cases can cause symptoms: headaches, fever, pulmonary symptoms, swelling of lymph nodes and liver, and skin itching.	Single dose of Ivermectin, DEC[1] has no effect, doxycycline has no data on efficacy.	(Basano *et al.*, 2014)
	<u>*Mansonella perstans*</u> Humid warm regions in West and Central Africa, South and Central America	• 580 million live in endemic areas* • 114 million possibly infected*		In rare cases from tourist or expatriates in endemic areas, show signs of abdominal pain and some allergic reactions resembling *L. loa* 'Calabar swellings'	If *Wolbachia* is present then treated with doxycycline, if no *Wolbachia* then treated with DEC[1]	(Simonsen *et al.*, 2011)
	<u>*Mansonella streptocerca*</u> Nigeria, Ghana, Cameroon, Congo	• Not known		Mostly asymptomatic, but can cause skin manifestations including itching, papular (nodules) and pigmentation changes	A single round of Ivermectin can suppress Mf loads	(Fischer *et al.*, 1997)

* These are most likely gross estimate since not many epidemiological studies have been carried out (Simonsen *et al.*, 2011), similarly CDC treatment recommendations have very little published data to back them up; [1]DEC, diethycabamazine

1.3 Human filarial nematode life cycles

1.3.1 Onchocerciasis

Onchocerciasis is transmitted by an infected bite of a blackfly vector (*Simulium* spp) (Blacklock, 2016), which can be found around fast-flowing and well-oxygenated streams and rivers. Blackflies transmit the infective third-stage larvae (L3) onto the skin of the human host when taking a blood meal. The L3 then enter the vertebrate host through the bite wound, and once inside the host the L3 reside within the subcutaneous tissue, where they mature into fourth-stage larvae (L4) and then into adults. Adult females reside in subcutaneous nodules called onchocercomas, whereas adult males migrate through the subcutaneous tissues between nodules fertilising different females. Within the nodules male and females mate, and subsequently female worms will release microfilariae (Mf). Mf migrate throughout the body via the subcutaneous tissues, and the lymphatic vessels of connective tissue (Murdoch, Murdoch, 2016), with most the disease symptoms associated with onchocerciasis, induced by the death of Mf passing through the skin or eyes (Taylor *et al.*, 2010). A blackfly will take a blood meal from an infected vertebrate host and ingest Mf from the dermis.

Within the blackflies, the Mf penetrate the midgut and migrate through haemocoel (circulatory system in arthropods) to the thoracic muscles, were they moult first into the first-stage larvae (L1), then the second-stage larvae (L2) and finally into L3. The L3 migrate to the blackflies head and finally emerge in the proboscis, from where they may be transmitted to the vertebrate host during a blood meal (Figure 1.3).

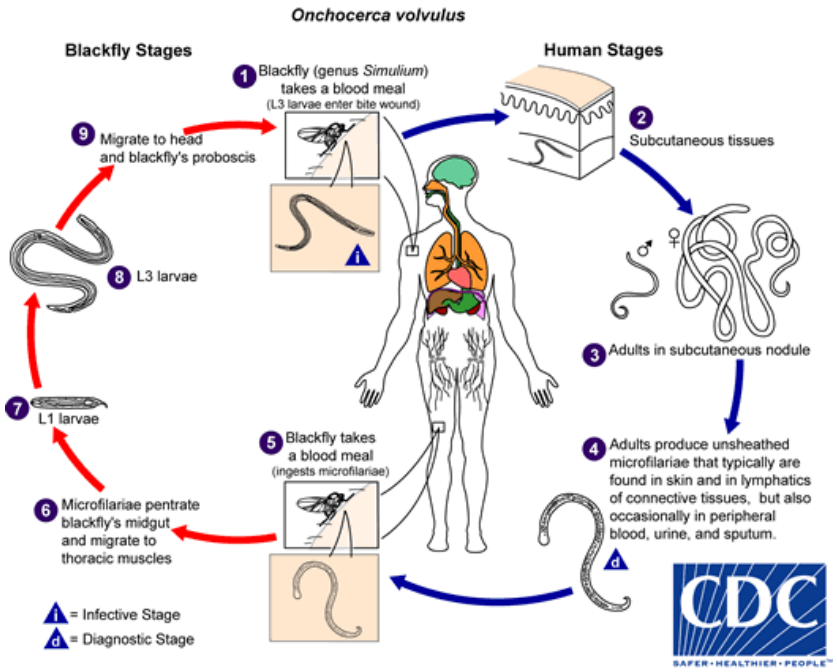

Figure 1.3. ***Onchocerca volvulus*** **life cycle**. 1) During a blood meal, an infected blackly (*Simulium* spp) will introduce the infected third stage larvae (L3) to the skin of the human host, and these will enter the host via the bite wound, as blackflies take a blood meal by cutting into the skin and feeding on the blood pool that forms. 2) Once in the subcutaneous these L3, will molt fourth stage larvae (L4) and then again into adult worms. 3) The adults reside in nodules within the subcutaneous connective tissue, for up to 15 years. 4) In these nodules, the females and males mate, and females produce thousands of microfilariae (Mf), these are the transmission stages of the parasite. Mf can live up to 2 years, and migrate through the skin and lymphatics of connective tissues, and in some occasion travel to the eyes. 5) Blackflies will ingest Mf during a blood meal. 6) Within the blackflies, the Mf migrate to midgut, 7) where they will molt into the first larval stage (L1), following with the second larval stage (L2), 8) and then into L3 infective stage. 9) The L3 migrate to the blackflies mouth parts, 10) and during the subsequent blood meal will infect another human host. Diagram taken from CDC website: https://www.cdc.gov/dpdx/onchocerciasis/index.html

1.3.2 Lymphatic filariasis

Transmission of lymphatic filariasis whether *W. bancrofti, B. malayi* or *B. timori*, begins with the inoculation of infective larvae (L3) into the skin of the human host during a mosquito bite. The L3 larvae enter through the puncture wound and migrate through the lymphatics towards the lymph nodes. The parasites reside within the lymphatics and lymph nodes, where they mature and molt first into the fourth larval stage (L4) and then adult worms. Adult worms mate, and females release live Mf. Mf periodicity in the blood, coincides with the time of feeding of their vector, as a wide range mosquito species act as vectors of lymphatic filariasis (e.g. *Anopheles gambiae*, *Culex quinquefasciatus* and *Aedes polynesiensis*). Once ingested by the respective vector, the Mf mature to form L2 and then L3, ready to infect another host (Figure 1.4).

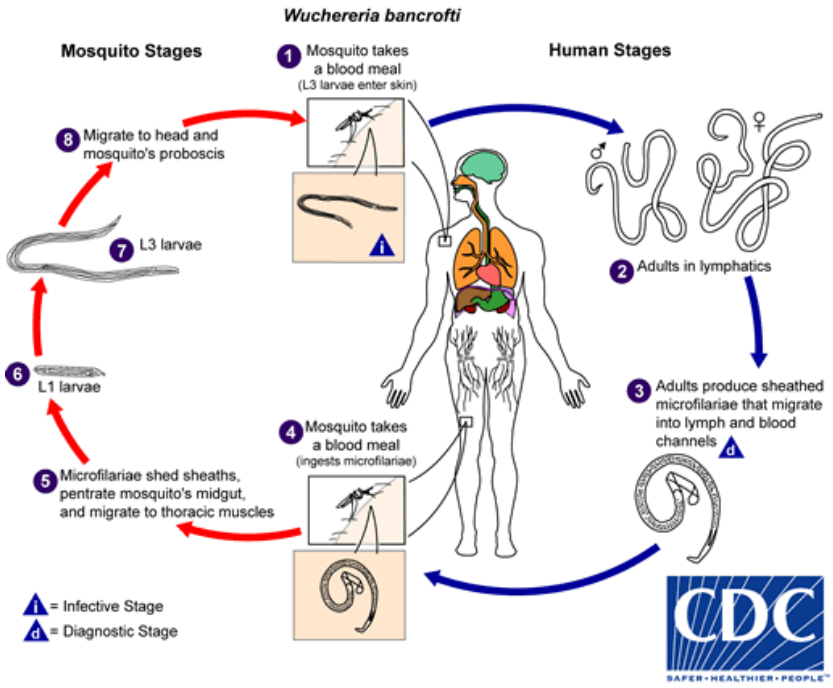

Figure 1.4. *Wuchereria bancrofti* life cycle. *W. bancrofti, B. malayi* and *B. timori* have very similar life cycles. They are all transmitted by mosquitoes, with different mosquito species depending of the geographical distribution. 1) During a blood meal, an infected mosquito transmits the infected third-stages larvae (L3) onto the skin of a human host, these will then penetrate the bite wound. 2) The L3 migrate through the subcutaneous tissue to the lymphatic vessels, were they develop into fourth-stage larvae (L4) and then into adults. 3) Adults mate, and females produce microfilariae. These Mf are sheathed and have a nocturnal periodicity, meaning that Mf migrate between the lymph and blood, and are found in the blood at night. 4) When Mf are in the blood, a mosquito will ingest them after a blood meal. 5) After ingestion, the Mf lose their sheaths and migrate to the thoracic muscles. 6) Mf then develop into first-stage larvae, second-stage larvae, 7) and finally into L3. 8) The L3 migrate to the mosquitoes proboscis, 9) and infect another human host in the subsequent blood meal. Diagram taken from CDC website: https://www.cdc.gov/dpdx/lymphaticfilariasis/index.html.

1.3.3 Loiasis

Loa loa is transmitted by a tabanid fly (*Chrysops* spp). During a blood meal, L3 escape from the proboscis of the fly and are deposited on the surface of the skin. The L3, then enter the skin via the bite wound, and within the subcutaneous tissue, the L3 moult to L4, and then to adults. The adult worms live between the layers of loose connective tissue under the skin, where they mate, and female worms release Mf. The Mf have a diurnal periodicity to coincide with their vectors feeding habit, during the day they are found in the peripheral blood and at night they are found in the lungs. During a subsequent blood meal, the tabanid flies will ingest the Mf. In the fly the Mf will mature into L3, ready to infect a new host following a subsequent blood meal (Figure 1.5) (Boussinesq, 2013).

Although, *O. volvulus* and *L. loa* are transmitted by different vectors, their modes of transmission and life cycles share similarities. In both cases, L3 are deposited onto the skin and migrate into the bite wound, once in the human host, they mature and develop into adults within the subcutaneous tissue. In both nematodes, adults mate, and females produce thousands of Mf. However, *O. volvulus* produce unsheathed Mf that can be found in the skin at any time of the day, as they do not exhibit any form of periodicity, whereas in *L. loa*, the Mf are sheathed and found in the peripheral blood during the day and in the lungs at night.

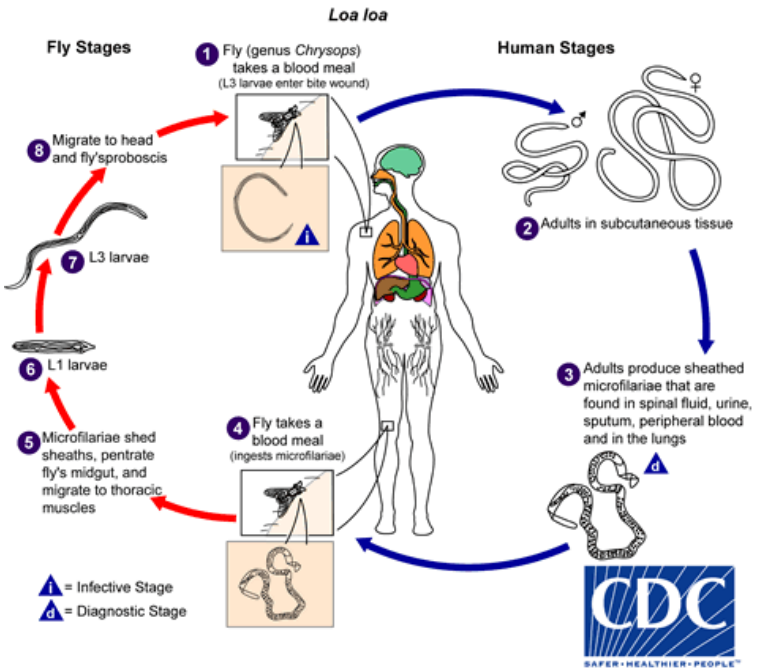

Figure 1.5. *Loa loa* life cycle. 1) *Loa loa* is transmitted by a tabanid fly (*Chrysops spp*), during a blood meal, third-stages larvae (L3) escape the proboscis of the flies and are deposited onto the skin of a human host, these will then penetrate the bite wound. 2) The L3 develop into fourth-stage larvae (L4), by day 9 post-infection and then into adults by day 19 post infection. The adults reside in the subcutaneous tissue for up to 17 years. 3) Adults mate, and females produce microfilariae. These Mf are sheathed and pass into the host lymphatic system before accumulating the lungs. From this reservoir, the Mf migrate to the peripheral blood during the day, to coincide with the tabanid fly feeding habit. In some cases, Mf can be found in urine, saliva and spinal fluid. 4) When Mf are in the blood, the tabanid fly will ingest them. 5) After ingestion, the Mf lose their sheaths and migrate to the thoracic muscles. 6) Mf then develop into first-stage larvae, second-stage larvae, 7) and finally into L3. 8) The L3 migrate to the fly's proboscis, 9) and infect another human host in the subsequent blood meal. Diagram taken from CDC website: https://www.cdc.gov/parasites/loiasis/biology.html.

1.4 *Wolbachia*

Wolbachia is an endosymbiotic bacterium living in many species of insects (52% of terrestrial arthropods) and some nematodes. It is restricted to living within the host cells and in nematodes it is transovarially transmitted (from females to progeny) (Makepeace, Gill, 2016). *Wolbachia* has gained attention in the recent years, because they cause reproductive alterations in insect, which can be manipulated for vector control purposes, and because they are present in nematodes, and can therefore be targeted for treatment (antibiotics). *Wolbachia pipientis* is the only formally recognised species within the genus, however *W. pipientis* has been divided into different strains, termed "supergroups" (Makepeace, Gill, 2016).

Most filarial species that infect humans co-exist with *Wolbachia* within their tissue, such as *B. malayi, B. timori*, *O. volvulus*, *O. ochengi* and *W. bancrofti*, but not *L. loa* (Slatko *et al.*, 2014). In *O. volvulus*, *Wolbachia* is found in the adult worms body wall, oocytes, all embryonic stages and in the Mf (Hoerauf *et al.*, 2000). This bacterium is not only essential for growth and development of the nematodes, but also plays a key role in the fecundity of the adult female worms, and triggers inflammatory responses within the host. Depletion of *Wolbachia* by antibiotics can not only block embryogenesis but also kill adult worms, providing a new route for therapeutic strategies for filariasis control (Slatko *et al.*, 2014).

1.5 Diagnostic tools

For onchocerciasis

Detection of onchocerciasis mostly still relies on identification of Mf in skin snips (small biopsies that are restricted to the upper dermis). Other diagnostic methods include the use of diethylcarbamazine (DEC) patch test, in which itchy patches appear where topical DEC kills Mf, these are however invasive methods with low sensitivity (Eberhard *et al.*, 2017).

An immunodiagnostic test based on the detection of human IgG4 antibodies to the *O. volvulus* specific antigen *Ov*-16 also exist, and is one of the recommended tools for determining whether transmission has been interrupted (WHO guidelines, (WHO, 2016). However, the diagnostic accuracy of *Ov*-16 antibody test has mostly been assessed in populations with high onchocerciasis prevalence, and therefore might not be as sensitive in low-prevalence areas, such as at the end-stage of control programmes or in hypo-endemic areas (Lipner *et al.*, 2006).

More sensitive and less invasive alternative diagnostics are desperately needed, to determine when to stop control interventions, in particular in areas which have become hypo-endemic following MDA (Vlaminck *et al.*, 2015; Golden *et al.*, 2016; Eberhard *et al.*, 2017).

For lymphatic filariasis

For lymphatic filariasis, traditional diagnostic methods involve the examination of blood slides for Mf, collected at night, due to the periodicity of lymphatic filariasis. Compared to onchocerciasis, in the recent years rapid diagnostics have been developed

for lymphatic filariasis, these include a *W. bancrofti* antigen-based test and a *Brugia* spp antibody tests, which have been tremendously useful for national mapping surveys of lymphatic filariasis (Cano *et al.*, 2014).

For loiasis

L. loa is distinguished by the examination of blood, for the presence of Mf. Determining the number of circulating Mf is important for the successful implementations of control programmes for onchocerciasis. MDA programmes have been limited in areas with *O. volvulus* and *L. loa* co-endemicity, because of risk of developing severe adverse events (SEA) following treatment with Ivermectin (used for onchocerciasis control) in individuals harbouring high levels of *L. loa* Mf in the blood (<30, 000 Mf/ml) (Twum-Danso, 2003; Boussinesq *et al.*, 2003; Boussinesq, 2013). A majority of the population infected with *L. loa* have Mf levels below the threshold (Pion *et al.*, 2006). Therefore, a simple and fast diagnostic methods capable of identifying individuals at risk of SEA is needed, so that these individuals with high Mf counts can be excluded from Ivermectin treatment, as this would aid in increasing MDA coverage for onchocerciasis.

Mf burdens in the blood of *L. loa* infected individuals, are detected by counting Mf numbers in blood smears under a microscope, however this is time consuming and not useful in large scale MDA programmes, where fast diagnostic tools are preferred. Alternative diagnostic methods are under development, including:

- The use of loop-mediated isothermal amplification (LAMP), which is a nucleic acid-based method that can detect individuals with high Mf loads, but requires training (Drame *et al.*, 2014);

- A cell counter chamber, which works by allowing blood to pass through but not Mf, making them easy to quantify (Bennuru *et al.*, 2014);

- A cell phone microscope, which uses a software that can determine Mf numbers (D'Ambrosio *et al.*, 2015).

These diagnostic tools under development all provide a rapid method of identifying individuals with high Mf loads, to exclude them from Ivermectin treatment, however these all require blood samples. A less invasive diagnostic tool is under development, with the possibilities to quantifying Mf loads, using an antigen detection assay in urine samples (Drame *et al.*, 2016).

1.6 Treatments available

Current treatments for filarial infections involve the use of diethylcarbamazine (DEC), Ivermectin, albendazole and doxycycline. For lymphatic filariasis, control programmes rely on the use of DEC, albendazole or Ivermectin depending on whether the treatment area is co-endemic for onchocerciasis, whereas onchocerciasis primarily relies of the use of Ivermectin, but doxycycline could be used in areas co-endemic with *L. loa* (Wanji *et al.*, 2009).

DEC is the mainstay drug for the global programme to eliminate lymphatic filariasis (GELF) in areas not co-endemic for onchocerciasis (i.e. all Asian and Pacific regions; and some African and South American countries). A single dose of DEC is effective in reducing acute and chronic microfilaremia, although a single does not clear all Mf nor kill all adults, a regimen for 12 consecutive days will lead to absence of Mf in the blood (Gyapong *et al.*, 2005). Some adverse reaction can occur, due to the rapid killing of adult worms and Mf caused by the release of *Wolbachia*.

Albendazole, is an anthelmintic effective against a broad spectrum of nematodes and flatworms (i.e. Cestodes). The combination of albendazole with DEC or Ivermectin in lymphatic filariasis infected individuals, reduced Mf loads in the peripheries for a longer period, compared to a single treatment, however the combination of Ivermectin and albendazole has no effect in onchocerciasis patients (Awadzi *et al.*, 2013).

DEC was recognised early on to induce severe adverse reaction in individuals infected with *O. volvulus* (Greene *et al.*, 1985; Lariviere *et al.*, 1985), with adverse events correlating with Mf densities in the skin and with the presence of Mf in the eye (Francis

et al., 1985). However, it continued to be the drug of choice for onchocerciasis, because better alternatives are not available.

In the 1980, a single dose of Ivermectin was found to be more effective than DEC at clearing *O. volvulus* Mf, and therefore became the drug of choice for onchocerciasis (Albiez *et al.*, 1988). Ivermectin can be given to both onchocerciasis and lymphatic filariasis infected individuals. Ivermectin is a microfilaricidal (kills Mf), it is effective at killing the Mf stage, but not the adults, although repeated treatment has been suggested to render adult females infertile. However, it cannot be given to children <15kg; to pregnant woman; nor in areas co-endemic with *L. loa* due to sever adverse reactions of Mf dying (Basáñez *et al.*, 2008; Chesnais *et al.*, 2017).

Since lymphatic filariasis and onchocerciasis nematodes carry the *Wolbachia* bacterium, doxycycline and other antibiotics can be used as treatments (Gilbert *et al.*, 2005; Hoerauf, 2008; Wanji *et al.*, 2009). In lymphatic filariasis, patients treated with doxycycline resulted in long-term sterility and eventual death of adult worm, and further decreased pathology. In onchocerciasis, treatment with doxycycline resulted in sterilisation of adult females and an absence of Mf in the skin (Abegunde *et al.*, 2016). Doxycycline is the first drug to show macrofilaricidal (kills adult stages) activity in onchocerciasis, with no severe adverse events, but it is a long treatment (6 week), and cannot be given to children under 9 years old nor pregnant woman. Therefore, doxycycline is a good treatment strategy on an individual basis, but not practical for mass drug administration campaigns (Slatko *et al.*, 2014), unless greater investment is made to train community workers to ensure compliance with treatment regime (Wanji *et al.*, 2009).

1.7 Control programmes for onchocerciasis and lymphatic filariasis

1.7.1 Lymphatic filariasis control

Most control programmes for filarial diseases, rely on mass drug administration (MDA). For lymphatic filariasis, the global programme to eliminate lymphatic filariasis (GPELF) was launched in 2000, with the aim to eliminated lymphatic filariasis by 2020. GPELF strategy is to interrupt transmission of infection using MDA of DEC or Ivermectin combined with albendazole, as well as managing morbidity and preventing disability (Ottesen *et al.*, 2008; Ramaiah, Ottesen, 2014). In 2014, 73 countries were classified as being endemic for lymphatic filariasis, with an estimated 71% MDA coverage by August 2015, and since the creation of GPELF, 18 of these countries have now entered post-MDA surveillance, due to infection having being reduced to below target threshold, following high coverage of five or more rounds of MDA (World Health Organization, 2015). Lymphatic filariasis is arguably easier to control and perhaps eliminate compared to onchocerciasis, principally because 59% of the lymphatic filariasis burden is found in 9 Asian countries, where MDA coverage is high (66%) (SEARO, 2010; World Health Organization, 2015). Moreover, in Asia there is no co-endemicity with onchocerciasis, meaning that MDA can rely on the use of DEC alone or in combination with albendazole, this treatment regime has proven the most effective and so far, has averted 175 million DALY's, and possibly the key factor that makes lymphatic filariasis control more successful compared to onchocerciasis (Turner *et al.*, 2016). Treatment coverage, sustainability, and compliance are key to the success to control programme, however these are

particularly challenging in poor resource settings of sub-Saharan Africa. Mathematical modelling suggest that lymphatic filariasis transmission could be interrupted using MDA alone, but over a longer period of time then the 4-6 years that were initially suggested, due to some lack of compliance in certain areas (Stolk *et al.*, 2015a).

1.7.2 Onchocerciasis control

For onchocerciasis, the story is more complicated, with 95% of the global burden in Africa, where compliance and sustained MDA coverage is notoriously low due to:

- civil strife and conflict;
- insufficient health infrastructure and resources;
- low political commitment and insufficient funds to sustain national programmes for years;
- growing public fear of using Ivermectin due to the risk of developing severe adverse reaction, in *L. loa* infected patients (Gardon *et al.*, 1997; Boussinesq *et al.*, 2003; Wanji *et al.*, 2015b).

These factors have prevented long-term and sustained programmes throughout much of Africa (Cheke, 2017).

The first large scale control of onchocerciasis started in 1974, with the creation of Onchocerciasis Control Programme (OCP) in West Africa, which employed vector control through aerial spraying of organophosphate larvicide to target *Simulium* spp breeding sites. In 1986 Merck & Co donated Ivermectin, a veterinary anthelminthic

effective against *O. volvulus* microfilariae, which greatly contributed to the OCP's activities (Cupp *et al.*, 2011). This initial wave of control proved successful, with tens of millions of cases of onchocerciasis prevented, according to the WHO.

With the end of OCP approaching (2002), a second major initiative emerged the African Programme for Onchocerciasis Control (APOC), which started in 1995 focusing on community-directed mass drug administration (MDA) of Ivermectin in 19 countries throughout Africa. Between 1995 and 2010, APOC prevented more than 8.2 million disability-adjusted life years (DALYs) (Coffeng *et al.*, 2013).

In parallel, the Onchocerciasis Elimination Program for the Americas tackled onchocerciasis with Ivermectin treatment twice a year, with consistently high rates of coverage which has led to the interruption of transmission or elimination in 11 of 13 foci (in Latin America). By 2015, 60% of the population needing Ivermectin received treatment, which was slightly lower than the minimum therapeutic target of 65% (World Health Organization, 2016b).

Although onchocerciasis control has shown much less success in Africa compared to the Americas, due to the differences in size of endemic foci, vector competence, duration and consistency of treatment coverage (Cupp *et al*., 2011; Cheke, 2017). There has been some success in Africa, with interruption of seasonal transmission in three hyper-endemic foci in Mali and Senegal, where only a few infections remained in the human population, after 15 to 17 years of annual or bi-annual Ivermectin treatment (Diawara *et al.*, 2009). In follow up studies (22 months after last treatment), skin snips revealed that if adult worms were present in individuals they were no longer

producing Mf. Furthermore, in two foci in Nigeria, which had meso-endemic pre-control endemicity, elimination was also achieved following by 15-17 years of annual Ivermectin treatment (Tekle *et al.*, 2012). Although this provides evidence of elimination could work in certain foci, it does not imply that elimination is feasible in the other onchocerciasis endemic areas in Africa.

The pre-control endemicity level is an important factor that influences the likelihood of interruption transmission and eliminating of onchocerciasis (Basáñez *et al.*, 2016). Although in Mali and Senegal pre-endemicity was classified as hyper-endemic (22-48 mf/mg skin), there are many foci in Africa with significantly higher endemicity levels. Furthermore, in the foci in Mali and Senegal there was seasonal transmission, which allows for treatment strategies to be optimised, i.e. distributing Ivermectin before the rainy season. In the Nigerian foci, not only was the initial endemicity relatively low but there was also high Ivermectin treatment coverage (<75%), all contributing factors to elimination.

The studies in Mali, Senegal and Nigeria provided some evidence that it could be possible stop transmission and ensure onchocerciasis elimination after 15-17 years of annual treatment, under certain conditions. However, studies in Ghana (Lamberton *et al.*, 2015), Cameroon (Wanji *et al.*, 2015a; Kamga *et al.*, 2016), and north-western Uganda (Katabarwa *et al.*, 2013), show that despite 15 years of MDA with Ivermectin and in some cases vector control, the burden of onchocerciasis might have been reduced but transmission is still ongoing.

Several mathematical models have been created to determine the minimum duration of Ivermectin mass treatment and the number of treatment rounds required to reach a defined threshold of Mf prevalence (<1.4%) below which treatment can be stopped (Coffeng *et al.*, 2014; Turner *et al.*, 2014a; Basáñez *et al.*, 2016). Several simulations have been conducted under a variety of scenarios, such as using different:

- Pre-control endemicity levels – With higher endemicity levels having longer programme durations (Coffeng *et al.*, 2014; Turner *et al.*, 2014a).

- Treatment frequency – Switching to biannual treatment would reduce the duration of the programmes (30-40% reductions) but high compliance would need to be maintained (Coffeng *et al.*, 2014), which is not always feasible twice a year especially in hard to reach communities (Turner *et al.*, 2014b).

- Treatment coverage – Typically 15% of the population are non-eligible for Ivermectin treatment, because of their age (under5 5 years old), weight (<5 Kg), pregnancy or illness (Basáñez *et al.*, 2016). The effect of three coverage levels, 40% (poor), 60% (moderate) and 80% (high), were investigated, and maintaining a high level of coverage is necessary to reach elimination within a reasonable time (15 years).

There are however scenarios were elimination may not be feasible within 15-18 years, even under biannual Ivermectin MDA with high coverage (Coffeng *et al.*, 2014; Stolk *et al.*, 2015a), particularly in areas with intense transmission, most often due to high blackly biting rates.

Furthermore, there is now emergence of possible resistance to Ivermectin in foci in Cameroon (Pion *et al.*, 2013) and Ghana (Osei-Atweneboana *et al.*, 2011), with adult females from areas having received multiple Ivermectin treatments, recovering Mf production earlier, compared to female worms from areas having received only one dose of Ivermectin.

Furthermore, MDA with Ivermectin also cannot be given in areas co-endemic with loiasis, due to the risk of sever adverse reaction associated with *L. loa* death (Chesnais *et al.*, 2017). If Ivermectin treatment were to be implemented in loiasis co-endemic areas, each individual person would need to be tested for the prevalence of *L. loa* infection, as individuals with high Mf burdens would need to be excluded to avoid the risk of adverse reactions. Even by excluding individuals at risk, additional measures would need to be put in place in case individuals with low Mf burdens developed adverse reaction, and this would raise the cost of Ivermectin MDA campaigns (Turner *et al.*, 2015). The alternative treatment doxycycline, cannot be given to children under 9, and is also a long treatment, that needs more trained community workers, also increasing the cost of campaigns. Therefore, this leaves a reservoir of onchocerciasis infections which gives the opportunity for reintroduction of onchocerciasis in neighbouring areas with ongoing MDA treatment or in areas where MDA has been stopped due successful in elimination.

Now that the London Declaration on Neglected Tropical Disease and the WHO have set goals for onchocerciasis elimination by 2020 in selected African countries (World Health Organization, 2012; Uniting to Combat NTDs, 2012), a novel health intervention such as an alternative treatment (macrofilaricide) and/or a vaccine

administrable to children and individuals co-infected with *L. loa*, is needed. Mathematical modelling has suggested that a vaccine complimentary to MDA efforts, would not only reduce onchocerciasis burden in the populations that cannot receive Ivermectin, but also decrease the chance of re-emergence in areas where Ivermectin MDA has been successful and treatment has stopped (Turner *et al.*, 2015).

Table 1.2. Comparison of onchocerciasis and lymphatic filariasis control.

	Onchocerciasis	**Lymphatic filariasis**
Treatments used in control programmes	Annual or biannual treatment with: • Ivermectin (in areas not co-endemic with *L. loa*) (Doxycycline can be used in areas co-endemic with loiasis, but it is a long treatment so not used in control programmes)	Annual MDA: • DEC & albendazole • Ivermectin & albendazole (in areas co-endemic with onchocerciasis) • Albendazole alone (in areas co-endemic with loiasis)
Majority of disease burden	95% in 30 African countries	58% in 9 Asian countries
Success	• 4 countries in Latin America were acknowledged by WHO as achieving onchocerciasis elimination: Colombia (2013), Ecuador (2014), Mexico (2015) and Guatemala (2016) • 10 countries in Latin America have interrupted transmission	• 6 countries were acknowledged by WHO as achieving lymphatic filariasis elimination: Colombia (2013), Ecuador (2014), Mexico (2015) and Guatemala (2016) • 13 countries in Asia, Latin America and Caribbean do not require preventative chemotherapy anymore and are under surveillance to demonstrate elimination has been achieved
Population left untreated in control programmes	• Children under 5 years old • Pregnant women • Individuals with high density *L. loa* infections	• Children under 2 years old • Pregnant women
Recommended period of annual MDA	15-17 years, although due to low coverages more than 30 years would be needed.	In many case's annual MDA for 5-6 years is enough to interrupt transmission.

1.7.3 Using *Onchocerca volvulus* vaccines in *Loa loa* co-endemic regions

Severe side effects following Ivermectin treatment for onchocerciasis in areas co-endemic with loiasis, has side tract the work of APOC. Although other onchocerciasis treatment exist that could be used in this region, they have limitations that make them impractical to use in MDA campaigns. Vaccine against onchocerciasis provides a good alternative or complementary control strategy, and with modelling it was shown that using a vaccine in onchocerciasis and loiasis co-endemic areas, would have a beneficial impact for onchocerciasis control, especially reducing Mf burden in under 20 year olds (Turner *et al.*, 2015).

Although there is a lot of cross-reactivity between *O. volvulus* and *L. loa*, there is 52-72% similarity in amino acid, between the potential vaccine candidates (*Ov*-Ral-2, *Ov*-103) identified for *O. volvulus* in the *L. loa* counterparts, and was suggested that it was unlikely that there would be enough cross-efficacy to cause adverse effects in *L. loa* infected individuals (Turner *et al.*, 2015). However, this issue has yet to be tested in animal models, more so because of the lack of suitable animal to study *L. loa* infections. Through the use to knockout mice it is now possible to look at effects of *L. loa* in murine models (Tendongfor *et al.*, 2012), and therefore measure cross-efficacy of vaccine candidates.

1.8 Animal models used to study immune responses to filarial parasites

A major constraint in vaccine development, is the need of an animal model, not only to understand the immune response to onchocerciasis, but there also needs to be a way to evaluate vaccine candidates. Since it is difficult and unethical to test in humans, research turns to animal models, however a major obstacle is that *O. volvulus* only infects humans and in some cases primates. Experimental infections have been attempted in several animals with only chimpanzees (*Pan troglodytes)* and mangabey monkeys (*Cercocebus atys)* developing a successful infection, both of which are impractical and unethical to screen vaccine candidates (Eberhard *et al.*, 1991; Abraham *et al.*, 2002).

One approach to overcome this problem was the development of the diffusion chamber model, which allows immunity against early larval stages (L3) of *O. volvulus* or *B. malayi* in mice to be investigated (Lange *et al.*, 1993). In this model live L3 are implanted subcutaneously in diffusions chambers, which can be recovered after a period of time, to analyse the parasites survival and microenvironment (cytokines, cells recruited) (Abraham *et al.*, 2004). The advantage of this system is that immune responses to human parasites can be investigated. However, the disadvantages are that only the early stages of infections (larval stages) can be investigated, and the insertion of the chamber (14mm in diameter) itself does cause inflammatory responses, biasing subsequent immune responses.

Alternatively, other filarial species may be used, such as *B. malayi* which is permissive in jirds (a type of gerbil), cats, ferrets, different gerbil species and *Mastomys coucha*

(African mouse). The advantage of these systems is that vaccine efficacy against a human filarial nematode can be tested in a fully permissive host, but immunological readouts are difficult to measure due to the lack of reagents available for the animals (Morris *et al.*, 2013).

A close relative to *O. volvulus* is *Onchocerca ochengi*, which naturally infects cattle. These two *Onchocerca* spp share many similarities, such as transmitted by the same group of insect vector (*Simulium damnosum* complex); they both form collagenous nodules with similar histological structure, although *O. ochengi* forms intradermal nodules, whereas *O. volvulus* are subcutaneous nodules (Makepeace, Tanya, 2016); in both human and in cattle there are individuals that are naturally protected from infection (EN) (Tchakouté *et al.*, 2006). *O. volvulus* and *O. ochengi* are sympatric species (evolved from a single ancestral species, in the same geographic region) (Morales-Hojas *et al.*, 2006), with cross-reactive immunity between cattle and human species (Renz *et al.*, 1994). Therefore, *O. ochengi* model is useful for investigating vaccines efficacy against onchocerciasis under conditions of natural exposure, however cattle are impractical and too expensive for screening large numbers of vaccine candidates.

The most attractive alternative for screening vaccine candidates is the use of *Litomosoides sigmodontis* which is the only filarial species in which the full development cycle can take place in BALB/c mice (Petit *et al.*, 1992), with extensive immunological cross-reactivity with *Onchocerca* spp (Manchang *et al.*, 2014), allowing important immunological as well as parasitological readouts to be measured during any vaccination trial. The murine *L. sigmodontis* model of filarial infection, has

helped elucidate many of the immunological mechanisms that determine susceptibility or resistant to filarial parasites.

Table 1.3. Filarial animal models used to evaluate vaccine candidates

Animal model	Advantages	Disadvantages
***O. volvulus* or *B. malayi* in diffusion chamber in mice** (Used to look at early larval stages (L3/L4))	Effect of vaccine can be determined directly on human parasite, and since chambers are in mice reagents exist to measure immunological readouts.	Not a permissive model, so vaccine effect can only be examined on larval stages of parasite. Further chamber itself may cause inflammatory responses, which may influence the development of specific filarial responses.
***B. malayi* in gerbils** (Permissive model)	Effect of vaccine can be determined directly on human parasite, and on all stages of parasite life cycle	Not many reagents exist for gerbils therefore it is difficult to measure immunological effects of vaccines
***L. sigmodontis* in mice** (Permissive model)	Effect of vaccine can be determined on all stages of parasite life cycle, and since it is in mice reagents exist to measure immunological readouts.	It is not a human filarial nematode
***O. ochengi* in cattle** (Permissive model)	Can measure effect of vaccine efficacy with natural infections, closest relative to human onchocerciasis parasite and allows to study effect of vaccine on all life stages	Can be logistically difficult to trial many vaccine candidates.

1.8.1 *L. sigmodontis* model of infection

In susceptible BALB/c mice, infective third-stage larvae (L3) are transmitted to the mouse during a blood meal by the mite, *Ornitonyssus bacoti*. Once in the subcutaneous tissue, L3 migrate to the lymphatics, reaching the pleural cavity around day 4 post infection. In the thoracic cavity, the L3 moult into fourth-stage larvae (L4) by day 10 post infection, and subsequently develop into adults around day 26-28 post infection.

These adults mate, and females start producing Mf around day 55 post infection, which are detected in the peripheral blood, where they can be taken up by a mite during a blood meal. Within the mite, the Mf moult into L3 completing the life cycle (Figure 1.6). Although BALB/c mice are classified as having susceptible infection, not all mice have detectable Mf found circulating in the blood stream (47% presenting patent microfilaremia) and the adults get cleared around day 90 post infection (Petit *et al.*, 1992).

Whereas, in CBA/Ca and C57BL/6 mice *L. sigmodontis* worm mature, but patency is never reached. In CBA/Ca mice Mf are present in the uteri of female worms and sometimes in high densities but they are not found circulating in the blood (Table 1.4) (Petit *et al.*, 1992). These differences in mouse strain susceptibility to *L. sigmodontis* infections, provides a great model to investigate the immunological and genetic determinants of susceptibility or resistant to infection. Furthermore, Hoffmann et al showed that independent to the mouse strain, survival of *L. sigmodontis* Mf is dependent on the presence of adult female worms (Hoffmann *et al.*, 2001).

The *L. sigmodontis* animal model remains arguable the best experimental system for developing anti-filarial vaccines, as it has similar patterns of infection including larval migration, with some immunological cross-reactivity with human filariasis (*Brugia* spp*, W. bancrofti, L. loa and O. volvulus*) (Allen *et al.*, 2008).

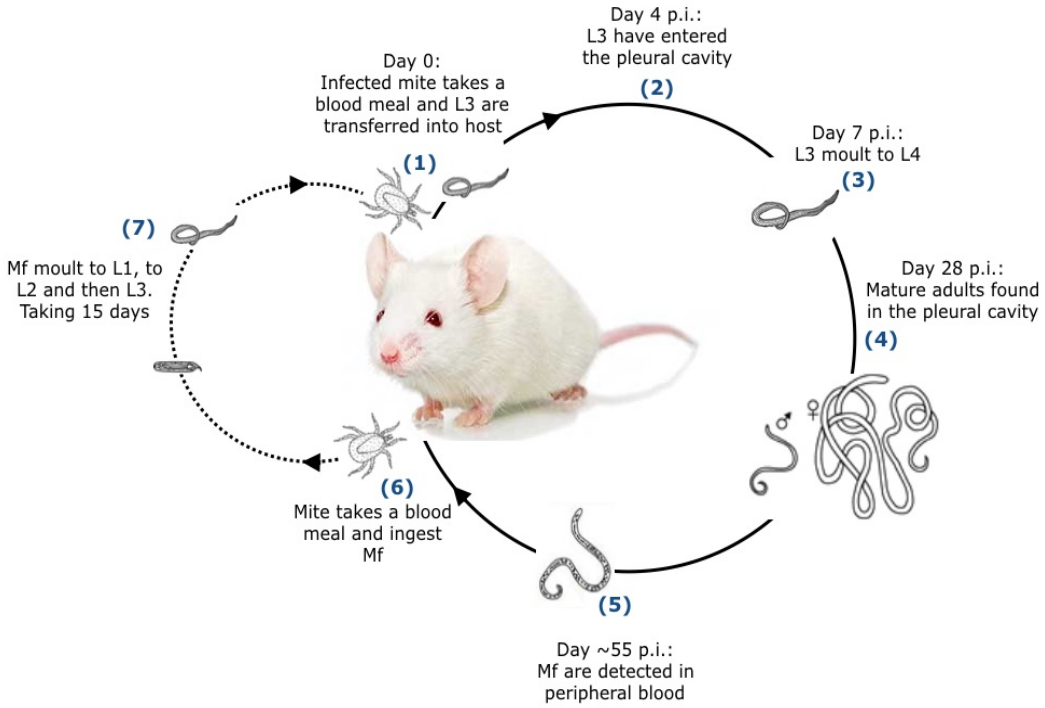

Figure 1.6. *Litomosoides sigmodontis* life cycle. *L. sigmodontis* can complete a full life cycle in either BALB/c mice or gerbil host, using *Ornitonyssus bacoti* (mite) as the vector. (1) An infected mite takes a blood meal transferring third-stage larvae (L3) into the subcutaneous tissue of mice or gerbils (Day 0). (2) The L3 migrate to the lymphatics and arrive to the pleural cavity by day 4 post infection (Day 4 p.i.). (3) By day 7 these have moulted into fourth-stage larvae (L4), (4) and then develop into adults reaching sexual maturity by day 28 p.i. (5) Females release 1000 of microfilariae (Mf), which can be detected in peripheral blood by day 55 p.i., ready to be (6) taken up by the vector during a blood meal. (7) In the vector the Mf develop into L1 and subsequently moult into L3 taking about 15 days. The L3 are then ready to infect a new host through a blood meal completing the life cycle. (Adapted from EPIAF website: http://www.filaria.eu/fil/bio/ls.html)

Table 1.4. Susceptibility to *L. sigmodontis* in different mice strain (Petit *et al.*, 1992)

Mouse Strain	Susceptibility to *L. sigmodontis* infection	
	Presence of Adults	Patent infection
BALB/c	+	+ (47%)
B10.D2	–	–
CBA/J	+	–
C3H/He	+	–
C57BL/6	– *	–
DBA/1	– *	–

*(worms never reach full adult form)

1.9 *O. volvulus* infections in humans

Human studies show that type 2 immune responses are associated with protection, although in a small proportion of individuals it can also lead to severe pathology (Murdoch, Murdoch, 2016). However, filarial nematodes have developed mechanisms to avoid theses immune responses, by regulating the immune system. The exact mechanisms by which filarial nematodes are killed *in vivo* and how these parasites avoid these mechanisms are still being investigated.

1.9.1 Pathology induced by *O. volvulus*

Disease pathogenesis for onchocerciasis is linked to host inflammation invoked by the death of the parasite, especially the Mf stage. In onchocerciasis, much of the pathology is found in the skin due to inflammation induced by Mf dying, resulting in intense itching, starting with acute papular dermatitis (swelling, itching and inflammation), followed by chronic itching (pruritus), more papular dermatitis and scarring, which over time, affected skin may begin to loose elasticity and structure, with signs of

premature ageing, such as lichenification or hanging groin (Taylor *et al.*, 2010). Patchy depigmentation of the legs can lead to a condition known as leopard skin and in extreme cases, there is severe papular dermatitis with hyperpigmentation (darkening of the skin), termed sowda (Taylor *et al.*, 2010). In some cases, *O. volvulus* infections can cause ocular lesion, due to Mf migrating to the posterior and anterior regions of the eye and evoking an in inflammatory responses with their death, this leads to severe visual impairment, with sclerosing keratitis (inflammation of the cornea) and iridocyclitis (inflammation of the iris), and finally blindness (Hise *et al.*, 2003; Brattig, 2004; Taylor *et al.*, 2010). In highly endemic regions, rates of skin disease increase with age until 20 years and then plateau, with the younger individuals exhibiting itching and chronic papular onchodermatitis, whereas the more severe form of depigmentation, visual impairment and blindness is more common in older individuals (<40 years) (Murdoch *et al.*, 2002).

1.9.2 Spectrum of immune responses to *O. volvulus* in humans

The parasites complex life cycle leads to a complicated host immune response, which is thought to explain the spectrum of clinical manifestations of onchocerciasis. *O. volvulus* infected individuals who have not received treatment exhibit a spectrum of disease manifestations (Table 1.5), from generalized onchocerciasis (GEO) a hyporesponsive response to *O. volvulus*, to a hyperreactive response showing severe pathology (Figure 1.8). GEO individuals tend to have palpable nodules (onchocercomas, the subcutaneous nodules containing adult *O. volvulus* worms) under the skin but no strong pathology despite having high Mf density in the skin, whereas

hyperreactive individuals have severe pathology but lower Mf density in the skin (Hoerauf, Brattig, 2002). Around 1-5% percentage of the population are naturally protected, because despite living in endemic areas, they show no signs of infection, neither clinical pathology nor detectable parasites, these individuals are termed endemic normal (EN) (Hoerauf, Brattig, 2002; Brattig, 2004).

1.9.2.1 Hyporesponsive response to *O. volvulus* - GEO

GEO individuals make up most the infected population, they have chronic infection with little pathology and this is thought to be due to *O. volvulus* ability to modulate immune responses, a common strategy of filarial nematodes (Hoerauf *et al.*, 2005). A study in rural Nigeria endemic for onchocerciasis, found that the population with low prevalence of visual impairments was associated with low $CD4^+$ T cell counts (Nmorsi *et al.*, 2007). A reduced antigen-specific T cells responses and T cell proliferation is common trait of onchocerciasis patients with heavy infections, as well as low production of interferon gamma (IFN-γ), interleukin 13 (IL-13) and IL-5, which decreases with increasing Mf densities (Brattig *et al.*, 2002), on the other hand these patients have elevated production IL-10 and TGF-β (Hoerauf *et al.*, 2005; Korten *et al.*, 2010).

This immunosuppression in GEO individuals is associated with regulatory T cells, and several subsets have been identified in onchocerciasis infected patients (Figure 1.7). Analysis of onchocercomas obtained from GEO patients, confirmed the presence of $Foxp3^+$ regulatory T cells (Treg) (Korten *et al.*, 2008) and TGF-β^+ T cells (Th3), (Korten *et al.*, 2009). Furthermore, a third type of regulatory T cells were generated *in*

vitro from onchocercomas of GEO patients, these had typical characteristics of type 1 regulatory T cells (Tr1), such as elevated IL-10 production, variable amounts of IL-5 and IFN-γ, but zero to low production of IL-4 and IL-2 (Doetze *et al.*, 2000; Satoguina *et al.*, 2002). These Tr1 cells have an upregulated expression of CTLA-4 following stimulation, which were able to suppress proliferation of other T-cell clones in co-cultures (Satoguina *et al.*, 2002). CTLA-4 is a known marker of down-regulation and these CTLA-$4^{+}$$CD4^{+}$IL-10 producing Tr1 cells have often been found in higher number in GEO individuals (Steel, Nutman, 2003; Katawa *et al.*, 2015). The Tr1 and Th3 cells have been associated with isotype switching to IgG4 production by B cells, involving IL-10 and TGF-β (Satoguina *et al.*, 2008).

Elevated levels of IgG4 is a hallmark of GEO patients. IgG4 is an antibody subclass that does not fix complement but binds rather weakly to effector cell Fc receptors, and is therefore able to clear antigen without strong stimulation of effector cells. In onchocerciasis, it is a marker of patency as IgG4 levels correlate with peripheral Mf loads (Hoerauf *et al.*, 2005). IgG4 may help Mf survival by binding to the Mf and preventing antibody-dependent cell-mediated cytotoxicity (ADCC).

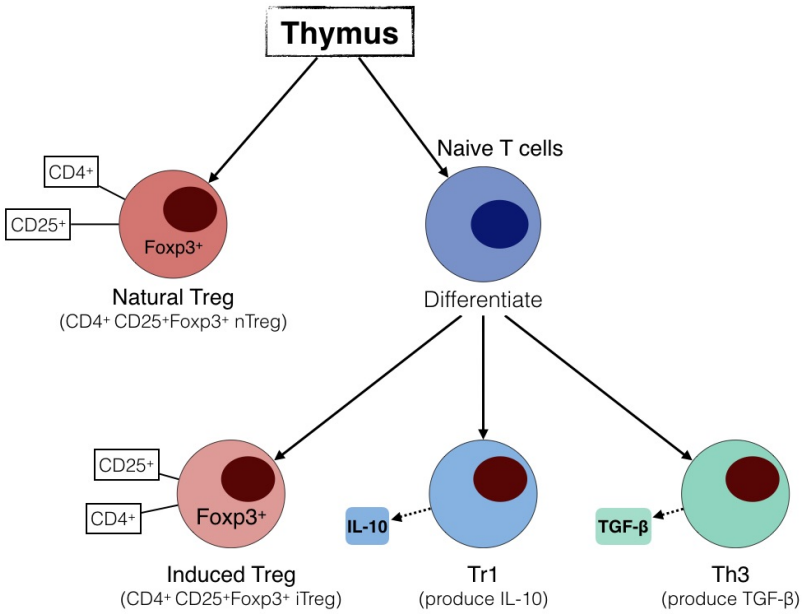

Figure 1.7. Regulatory T cells identified in filarial infections. There are different regulatory T cell subsets found in filarial infections, these were found in both human infections and the *L. sigmodontis* mouse model. Two $CD4^{+}CD25^{+}Foxp3^{+}$ Tregs subsets were identified, the Tregs derived from the thymus and are sometimes termed naturally Tregs and those induced in the periphery from naïve T cells are induced Tregs. In addition to the Foxp3 expressing Tregs, two other subsets of regulatory T exist which do not express Foxp3. These can be classified based on the regulatory cytokines they produce. Type 1 regulatory T cells (Tr1) express mainly IL-10, and Th3 regulatory T cells express TGF-β (Metenou, Nutman, 2013).

1.9.2.2 Hyperreactive responses to *O. volvulus* - Sowda

Hyperreactive form of onchocerciasis is characteristic of low parasite burden, but protective responses tend to lead to severe pathology sometimes referred to as the sowda form of onchocerciasis. These individuals exhibit a dominant Th17/Th2 phenotype with: high levels of IgE; elevated levels of eosinophils and mast cells in nodules; increased activated $CD4^{+}$ T helper; dense infiltrates with T cells and B cells and extensive fibrous nodules termed onchocercomas; but low parasite burden (Korten *et al.*, 1998; Brattig, 2004). Comparing cytokine profiles of activated PBMCs

(peripheral blood mononuclear cell) from infected individuals showed that hyperreactive individuals had elevated levels of IL-13, IL-5 and IFN-γ, but low IL10 and/or TGF-β, and had cutaneous pathology associated with pronounce systemic Th2 type responses to *O. volvulus* (Brattig *et al.*, 2002; Hoerauf *et al.*, 2005; Korten *et al.*, 2010). As well as elevated Th2 responses hyperreactive individuals have a pronounced Th17 phenotype, with greater number of $CD4^{+}$IL-17A secreting T cells compared to GEO individuals but decreased regulatory T cells ($CD4^{+}$ $CD25^{hi}$ $Foxp3^{+}$) (Katawa *et al.*, 2015). Environments which trigger IL-17/IL-17R signalling favour alloreactivity and autoreactive T cells by inhibiting regulatory T cells. This agrees with the reduced local expression of TGF-β, MHC-II and IgG4 detected around pathology and the reduced numbers of regulatory T cells (both $Foxp3^{+}$ T cells and Tr1 cells) in patients exhibiting strong pathology (Korten *et al.*, 2010; Katawa *et al.*, 2015).

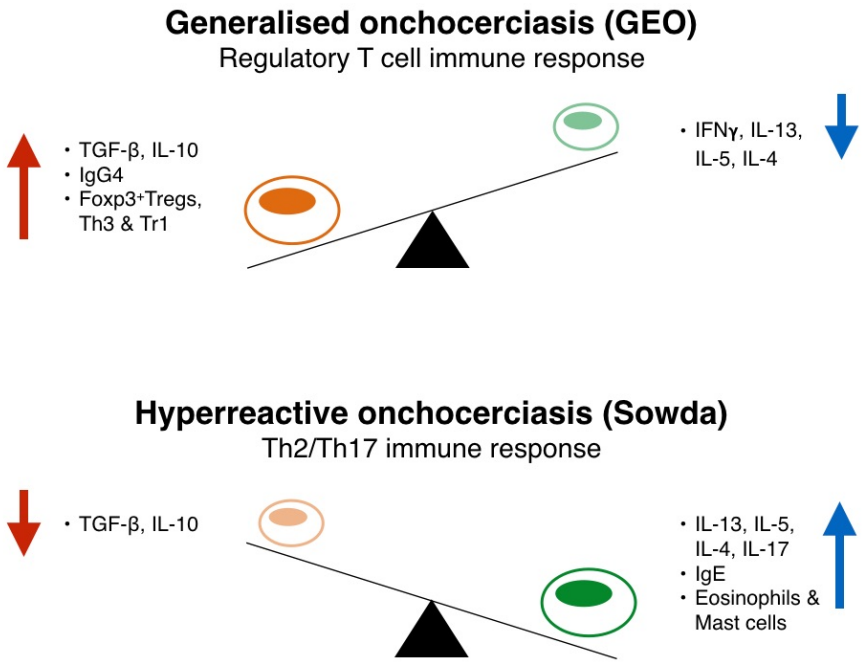

Figure 1.8. Model of balance between effector and suppressor mechanism, seen in GEO and hyperreactive individuals. Generalised onchocerciasis (GEO), has a dominant regulatory T cells immune responses, with increased numbers of regulatory T cells ($Foxp3^{+}$, Th3 and Tr1 cells); increased levels of IL-10 and TGF-β cytokines; and higher levels of IgG4. Counterbalancing this, there is lower levels of IL-13, IL-5 and IL4. Whereas in hyperreactive there are strong Th2/Th17 responses, with increased eosinophil and mast cell numbers; increased IL-13, IL-5, IL-4 and IL-17 cytokines levels; and elevated levels of IgE, but with lower suppression mechanism seen with lower levels of IL-10 and TGF-β. (Diagram adapted from (Hoerauf, Brattig, 2002) and (Katawa *et al.*, 2015))

Dermatitis and ocular keratitis immunopathology seen in onchocerciasis patients is associated with dying Mf in skin and eyes, respectively. Neutrophils are the major components of the early inflammatory infiltrate around damaged Mf in the cornea and skin, their recruitment and activation is dependent on the release of *Wolbachia* derived antigen from Mf present at those sites (Turner *et al.*, 2009). A lipoprotein on the surface of *Wolbachia* has been identified as being the main trigger of neutrophil inflammatory responses via the activation of Toll-like receptor (TLR) -2/6, during ocular keratitis (Turner *et al.*, 2009; Tamarozzi *et al.*, 2014). Neutrophils are also present in nodules

surrounding adult worms, these are found attached to worms containing *Wolbachia*. Nodules derived from doxycycline treated patients where *Wolbachia* would have been depleted, neutrophils can still be found but in lower numbers (Tamarozzi *et al.*, 2016). Neutrophils in nodules do not appear to be detrimental to living adults, and recently the formation of neutrophil extracellular traps (NETs) have been identified surrounding the worms. NETs exact role is still undetermined, it has been hypothesised to be the host protective mechanisms to any *Wolbachia* release from the worm's uteri or in excretory / secretory products; from *Wolbachia* released from Mf; or as method of trapping Mf and consequently limiting the Mf being released each day.

1.9.2.3 Protective immunity to *O. volvulus* – Endemic normal

Evidence of protective immunity in onchocerciasis is derived principally from the existence of non-infected individuals living in endemic areas, known as endemic normal. When comparing GEO patients to EN individuals, the latter display stronger antigen-specific proliferation and a mixed Th1/Th2 response. Immune protected individuals living in Ecuador presented with increased IFN-γ to *O. volvulus* antigen compared to Mf^+ individuals, but lower levels of IL-5 and IL-10 (Elson *et al.*, 2008). Similar results were found in individuals living in onchocerciasis hyperendemic areas in Cameroon. When comparing specific anti-larval and anti-adult cytokine levels, endemic normal had higher levels of GM-CSF (granulocyte-macrophage colony-stimulating factor) and IL-5 compared to infected individuals, and high levels of IFN-γ which were not detected in infected individuals (Turaga *et al.*, 2000). Overall EN individuals had stronger Th1 phenotype, with IFN-γ producing T cells expressing T-bet, with some IL-4 secreting $CD4^+$ T cells but in lower frequencies compared to

infected individuals (Katawa *et al.*, 2015). Suggesting that naturally resistant individuals are protected from *O. volvulus* by their ability to produce a mixed Th1/Th2 type response, although little is known about how an individual is predisposed to be protected.

1.9.2.4 Genetic determinants of susceptibility in humans

Onchocerca volvulus like many nematode parasites, are capable of influencing the host immune response, through secretion of potent immunomodulators (proteins). The extent of any immunomodulation largely depends on the intensity of infection, but the host genetics may also play a role. Genetic studies have linked HLA-DQ expression with the level of immune response (cutaneous reactions) the host mounts to parasite antigens (Murdoch *et al.*, 1997). Further, the promoter haplotype of IL-10 was found to influence *in vitro* PBMC proliferative response to *O. volvulus* antigen (Timmann *et al.*, 2004). Hyperreactive patients were association with a polymorphism in the IL-13 gene, which could lead to enhanced induction of the Th2 pathway. Studies, showed that the Th2 dominant sowda form of onchocerciasis was associated with the same mutation in the IL-13 gene (Arg110Gln) that is linked to allergic hyper-sensitivity (atopy, asthma). The presence of Arg110Gln variant of IL-13 is a significant risk factor for the development of sowda, although these individuals had higher IgE levels, these were found to be two independent factors. This IL-13 variant is thought lead to higher signalling via the IL-4Ra receptor (receptor for both IL-4 and IL-13) (Hoerauf *et al.*, 2002).

1.9.2.5 Immune responses in Ivermectin treated individuals

Since the introduction of MDA with Ivermectin, another group of patience can be categorized, these are patients that have been treated with Ivermectin, and thus have adult worms and nodules present but no detectable dermal Mf (Mf -ve). These individuals have been referred to as amicrofilaremic or as having an occult infection (Lechner *et al.*, 2012; Arndts *et al.*, 2014). Ivermectin treatment triggers an immediate immune response, with elevated chemokine (eotaxin and MCP-4) profiles responsible for promoting the migration of effector cells such as eosinophils and macrophages into patient's skin, prompting a purported cellular mediated Mf death. During the year following treatment there is a gradual rise in chemokines associated in neutrophil recruitment, suggesting that these are recruited to the adult worms, possibly due to the release of *Wolbachia* derived molecules, resulting in a reduction in numbers of Mf released into the dermal tissue (Lechner *et al.*, 2012). Following repeated Ivermectin treatment, there is a decrease in circulating and tissue eosinophils linked to a decrease in eotaxin; with a decrease in Th2 promoting cytokines (IL-5), and regulatory makers such as IL-10, since there is an expiring *O. volvulus* infection (Lechner *et al.*, 2012; Arndts *et al.*, 2014).

Table 1.5. Summary of spectrum of disease manifestation in individuals living in onchocerciasis endemic areas.

Group	Adult worms	Mf densities	Pathology	Immune response
Hyporesponsive: Generalised onchocerciasis	Present	High	Low / no pathology	Modulated immune response. High levels: IL-10, TGF-β, Tregs, AAMs and IgG4 Low levels: IgE, IFN-γ
Hyperreactive: Sowda	Present	Low	Severe pathology	Strong Th2 and Th17 response High levels: IgE and eosinophilia
Naturally protected: Endemic normal	None detected	None detected	No pathology	Stronger Ag-specific response, and a mixed Th1/Th2 response. Balanced: IFN-γ /IL-4 Low levels: IL-10 & IgE,
Ivermectin treated: Amicrofilaridermic	Present	None	No pathology	High levels: neutrophils Low levels: eosinophilia, low IL-10 and IL-5.

1.10 Immune responses to filarial infections in animal models

Animal models such as *L. sigmodontis* infection in mice are used to elucidate some of the complex interactions between the filarial nematode and its host.

1.10.1 Protective immune responses to filarial infection in animal models

It is accepted that the host immune response to filarial parasites in both animal models and humans is of the T helper 2 (Th2) type, which promotes parasite killing of the different life stages (Allen *et al.*, 2008).

Innate immune responses to L3

Innate responses to infection have been associated with the rapid destruction of L3 larvae within two days' post-infection, with neutrophils recruited to the invading L3 in the skin (Pionnier *et al.*, 2016). Neutrophil recruitment following a primary infection was shown to be in response to *Wolbachia,* an endosymbiotic bacterium living within *L. sigmodontis* worms. Mice deficient for NOD2, an intracellular pattern recognition receptor (PRR) to gram-negative and -positive bacteria, had increased worm burden at the early stage of infection caused by impaired neutrophil recruitment (Ajendra *et al.*, 2016). Furthermore, using CXCR4 gain in function mutant, on C57BL/6 mice background who are naturally more resistant to *L. sigmodontis* infections, it was demonstrated that resistance to infection is associated with an elevated number of dermal neutrophils, and where L3 were able to promote an oxidative bursts response and there was the release of neutrophil extracellular traps (NET) (Pionnier *et al.*, 2016).

This NET formation was also found in human filarial infection with *O. volvulus* around nodules containing adults worms (Tamarozzi *et al.*, 2016).

Although, in *in vitro* studies, eosinophils were found to adhere to the L3 and Mf of *O. volvulus* (Brattig *et al.*, 1991), in *in vivo* studies eosinophils were not present in innate immune responses to invading L3. As mice lacking eosinophils due to a deficiency in IL-5 had no change in worm burdens compared to wild types (Volkmann *et al.*, 2003).

The chemokine, CCL17, has been shown to be involved in early immune responses, limiting parasite invasion in the host. Deficiencies in CCL17, induced a recruitment and degranulation of mast cells, which led to increased vascular permeability, facilitating *L. sigmodontis* larval migration through the lymphatics to the pleural cavity. This mechanism was dependent on the presence of *Wolbachia* in L3 and promoted by TLR2 signalling (Specht *et al.*, 2011a).

Immune responses in the pleural cavity (to L3, L4 and adults)

To evade the innate immune responses, L3 migrate to lymphatics within hours of infection, through the heart and lung before reaching the pleural cavity (Babayan *et al.*, 2003; Karadjian *et al.*, 2017). In the pleural cavity of BALB/c mice, L3 can be detected as early as 2 hours post-infection, with 20% of the L3 reaching the pleural cavity between day 4-8 post-infection (Karadjian *et al.*, 2017). Following arrival of L3 within the pleural cavity, there is a recruitment of neutrophils to the pleural cavity via the CXCL1 chemokine, as well as eosinophils and macrophages (Karadjian *et al.*, 2017).

The L3 will moult into L4 around day 9-10 post-infection. Then depending on the mouse genetic backgrounds, L4 will develop into adults and then mature or will stay as immature adults. The differences in susceptibility to *L. sigmodontis* infections, can be detected by day 30 post infection between susceptible BALB/c mice and resistant C57BL/6 mice (Table 1.6), with less that 15% of the parasites at the L4 stage in BALB/c, compared 30% of the parasites in C56BL/6 (Petit *et al.*, 1992; Babayan *et al.*, 2003). This difference leads to BALB/c mice producing patent infections around day 55-60, while in C57BL/6 mice, worms are progressively encysted in granulomas and destroyed from day 40 post-infection. The lack in parasite maturation seen in C57BL/6 mice has been associated with a mixed Th1/Th2 immune response early on infection (day 10) compared to BALB/c mice who already have predominant Th2 response. The exact mechanisms dictating C56BL/6 resistant are not completely understood, but increased cellular recruitment and elevated CXCL12 concentration in the pleural cavity have been implicated. Blockage of the CXCL12/CXCR4 axis caused a decrease in cellular recruitment to the pleural cavity and an increase in worm burden, suggesting that indeed CXCL12 is important in resistance (Bouchery *et al.*, 2012a)

Granuloma formation within the pleural cavity

Different cell types are recruited to the pleural cavity in *L. sigmodontis* infection, and although the exact trigger or sequence of events that causes granuloma formation around the worms that leads to them eventually being cleared, are not fully known. What is known is that granuloma formation is an important part of the immune response to adult worms.

In *L. sigmodontis* time course studies, in both resistant and susceptible mice showed that granulomas were not present around larval stages, but were formed around the shed cuticle from these larval stages and consist mostly of eosinophils (Attout *et al.*, 2008).

In BALB/c mice, granulomas are found around ageing adult worms consisted mostly of neutrophils, as neutralizing IL-5 or G-CSF (chemokine for neutrophils) led to failure of neutrophil recruitment to the pleural cavity and hence the survival of adult worms for a longer period of time (Al-Qaoud *et al.*, 2000; Volkmann *et al.*, 2003). Similar results can be seen in IFNγ deficient mice, which showed that neutrophils were essential for granuloma formation (Saeftel *et al.*, 2001). Corroborating what is seen in *O. volvulus* infections where neutrophils form the inner layer around live adult worms (Tamarozzi *et al.*, 2016), and high levels of IL-5 are found in hyperreactive form of onchocerciasis (Brattig *et al.*, 2002; Korten *et al.*, 2010).

In resistant CBA/Ca and C57BL/6 mice, granulomas formed around the young adults and consisted mostly of eosinophils, that would degranulate in the presence of the host antibodies (Martin *et al.*, 2000b; Attout *et al.*, 2008).

Granulomas formed in different mouse strains contain different proportion of neutrophils, the percentage of neutrophils was greater in granulomas around worms in susceptible BALB/c mice compared to resistant C57BL/6, 47% and 17.8% respectively. The difference in percentage of neutrophils could be due to Mf release, as adult worms in BALB/c mice will release Mf which is another source *Wolbachia* and therefore contributing to extended recruitment of neutrophils. Since adult worms

in C57BL/6 do not reach patency there is no release of Mf, and therefore no recruitment of neutrophils. The link between *Wolbachia* and neutrophils can be further demonstrated using the *O. ochengi* cattle models, in which nodules formed around adult worms have an inner layer of neutrophils, and with the depletion of *Wolbachia* by antibiotic treatment, there is switch from neutrophils to degranulating eosinophils, found adjacent to the worms surface (Nfon *et al.*, 2006; Hansen *et al.*, 2011). Eosinophils made up 70% of the cell population in the granulomas surrounding young adults in C57BL/6. The origin of the different granulomas and how cells attach to the worms is not well known, so far it is hypothesised that neutrophils are recruited by a LPS-like dependent mechanism, whereas for eosinophils a glycan dependent process could be implicated, but it has been suggested that for cellular recruitment and granulomas to formation alteration/damage of the filarial worm is needed. (Attout *et al.*, 2008).

Altogether there seems to be a balance between eosinophils and neutrophils, as they are both associated with filarial infections, with eosinophils involved in parasite killing whereas neutrophils appearing to be involved in the nodule formation around the adults, attracted to the *Wolbachia* derived molecules secreted by the worms. Following antibiotic chemotherapy, neutrophils are replaced with eosinophils that degranulate on the worm cuticle (Hansen *et al.*, 2011).

Immune responses to microfilariae

Around day 55 post infection with *L. sigmodontis* in BALB/c mice, adult females will start releasing Mf, which can be detected in peripheral blood. Protective immunity to Mf, has been linked to Th1 and Th2 responses.

Infections in IFN-γ deficient mice caused an increase in Mf in the peripheral blood compared to wild type controls (Saeftel *et al.*, 2001), and in BALB/c mice an increase in IFN-γ RNA levels in splenocytes can be seen following onset of patency (Taubert, Zahner, 2001). Furthermore, following Mf immunisation, IFN-γ was found associated with vaccine mediated protection (Ziewer *et al.*, 2012), and in the related filarial species *B. malayi,* Mf injections in BALB/c mice induced IFN-γ production as well IgG2a production which was not seen when adult worms were implanted (Lawrence *et al.*, 2000). Th2 cytokines such as IL-4 and IL-5 have also been associated with control of Mf, however the action of both of these cytokines are more associated with a decreased in adult female fertility or adult worm containment than Mf killing (Volkmann *et al.*, 2001).

Organs associated with clearance of Mf in *L. sigmodontis* infections are the lungs, spleen and liver (Pfaff *et al.*, 2000), with more resistant mice strains being able to clear Mf faster than in BALB/c mice due to a rapid accumulation of Mf in spleen, liver and lungs (Bouchery *et al.*, 2012b). Although the exact mechanism of Mf clearance is relatively unknown, the receptor for IL-33 (ST2) was found to be important in splenic clearance of Mf (Ajendra *et al*., 2016). In the non-permissive *B. malayi* mouse model, where Mf were either inoculated in the blood stream or female adults producing Mf were implanted into the pleural cavity, clearance of Mf was associated with antibodies (Gray, Lawrence, 2002) and eosinophils (Simons *et al.*, 2005). However, *L. sigmodontis* infection in the ST2 deficient mice which saw a impaired splenic clearance, levels of IgM, IgG1, IgG2a and IgG2b as well as Th2 cytokine levels, did not differ between ST2 KO and wild types. Suggesting that increased Mf in ST2-KO

mice is not antibody mediated nor did it correlate with impaired Th2 cytokine responses, but was due to impaired splenic clearance of Mf (Ajendra *et al*., 2016).

Antibody responses in *L. sigmodontis* infections

B-cells have an important role in *L. sigmodontis* immunity, with mice lacking of B1 cells showing increased susceptibility to filarial infections seen as a higher Mf and adult worm burdens (Carter *et al.*, 2007). B1 cells are implicated in resistance in both *L. sigmodontis* (Al-Qaoud *et al.*, 1998) and human filariasis (*B. malayi*) (Mishra *et al.*, 2014). Although IgE is an important antibody in human helminth infections, and has been linked to parasite death and pathology, mice eosinophils however do not express the surface receptor that bind IgE, and further overexpression of IgE in *L. sigmodontis* infections had no effect on *L. sigmodontis* infections (Martin *et al.*, 2000b). When measuring levels of IgE, IgG1 (Th2 antibody) and IgG2 (Th1 antibody) in plasma of *L. sigmodontis* infected BALB/c mice, total IgE levels were increased compared to non-infected mice as early as day 5 and levels increased throughout the course of infection. Levels IgG1 (Th2 marker in mice) increased around day 14 of infection and remained elevated up to patency, whereas IgG2 (Th1 marker in mice) were low and no difference were seen between non-infected mice (Boyd *et al.*, 2015). In the same infection time course, levels of IL-5 were detected in plasma at day 42 and remained high at day 60. Therefore, the high levels of IL-5 and IgG1 detected in plasma illustrate that *L. sigmodontis* infection induce a systemic Th2 responses which peaks prior to the onset of patency.

Role of Th1 / Th2 in *L. sigmodontis* infections

Th2 responses are absolutely necessary for resistance to infection, as full parasite development and patency was achieved in IL-4 deficient C57BL/6 mice, which would otherwise be resistant (Le Goff *et al.*, 2002). Furthermore, in susceptible BALB/c mice, deficiencies in IL-4, IL-5 or IL-4Rα (receptor for IL-4 or IL-13) led to an increase in Mf numbers following *L. sigmodontis* infection compared to wild type controls (Volkmann *et al.*, 2003), and upon administration of anti-CD4 antibodies in BABL/c there was increase in worm burden and circulating Mf associated with reduced Th2 responses (Al-Qaoud *et al.*, 1997). This fits with human onchocerciasis, where Th2 induction in sowda patients leads to string reduction of both Mf and adult worm nodules (Hoerauf, Brattig, 2002).

This is not to say that Th2 responses are the only effector mechanism against filarial parasites. Pro-inflammatory Th1 responses have been associated with adult worm death and Mf clearance (Babu *et al.*, 2000). BALB/c mice deficient in IL-4Rα, had accelerated death of the adult stages compared to wild type controls, due to a change to a Th1 phenotype at the site infection (pleural cavity) (Volkmann *et al.*, 2001) and IFNγ was found to be essential for encapsulation of adult worms (Saeftel *et al.*, 2001). Furthermore, IFNγ and IL-5 appear to act synergistically to destroy the adult parasites and this effect is mediated by neutrophils (Saeftel *et al.*, 2003). Therefore, both Th1 and Th2 responses can act synergistically to control parasite loads.

Initiation of Th2 responses

The exact sequence of events that initiate a Th2 responses in filarial infections is still largely unknown, it has been hypothesized that dendritic cells, basophils and type 2 innate lymphoid cells (ILC2) have a role in inducing Th2 immunity. Basophils in *L. sigmodontis* infections seem to be more involved in the amplification of type 2 responses such as increased eosinophil levels and IgE production (Torrero *et al.*, 2010; 2013).

Recently, the role of ILC2 in initiating Th2 responses has been suggested, with evidence of local ILC2 expansions in the pleural cavity of infected mice from as early as day 5 post infection, with increasing numbers during the course of infection peaking in the pre-patent stage (day 36 and 44 post-infection) (Boyd *et al.*, 2015). ILC2 act primarily by initiating and maintaining Th2 responses at site of infection (pleural cavity) in *L. sigmodontis* infections, with the majority of ILC2 producing IL-5, possibly driving eosinophil recruitment to the pleural cavity seen in *L. sigmodontis* infections (Boyd *et al.*, 2015). IL-33 is known to promote Th2 responses, and has been suggested to do this through the activation ILC2. However, mice lacking the IL-33 receptor do not have an impaired Th2 response following a *L. sigmodontis* infection, suggesting that IL-33 is not responsible for driving the localized Th2 response seen in *L. sigmodontis* infections (Ajendra *et al*., 2016).

The absence of IL-33 signalling had no effect on adult worm burden, but led to a higher Mf burden. This increase in circulating Mf did not correlate with any change in Th2 response, but was shown to be due to impaired splenic clearance of Mf. When macrophages in pleural cavity of IL-33R deficient were analysed following a *L.*

sigmodontis infection, there was no change in macrophage proliferation, but alternatively activated macrophages failed to be induced (Jackson-Jones *et al.*, 2016). Both IL-4Ra and IL-33 are needed to alternative activation of macrophages. Therefore, IL-33 does have a role in parasite clearance but not in Th2 induction and further studies are need to determine the mechanisms responsible for the induction of ILC2s in *L. sigmodontis* infections and subsequently Th2 responses.

1.10.2 Regulatory immune responses induced by filarial infections

Helminth infections including *Onchocerca spp*, are master regulators of the host immune response, this allows them to maintain chronic infection within its host, and in the case of *O. volvulus* this can be up to 15 years. Immunosuppressive responses induced by filarial parasite are considered responsible for these chronic infections, and the lack of pathology seen in some individuals. This immune suppression has been demonstrated in both mice and cattle models of infection (Maizels *et al.*, 2001b; Hoerauf *et al.*, 2005).

Litomosoides sigmodontis infections in mice are known to induce inflammatory responses, with protection mediated by the Th2 arm of immunity. *Litomosoides sigmodontis* also induce regulatory immune responses which allows them to establish patent infections in mice, these immune regulatory responses happen early on infection. While invading L3 induce innate immune responses, they also cause a rapid recruitment and increased proliferation of a natural population of $CD4^+$ $CD25^+Foxp3^+$ regulatory T cells (Tregs), and within 7 days of infection these can be found in the

pleural cavity where the larval stages migrate to (Taylor *et al.*, 2009). These $Foxp3^+$ Treg cells continue to be present in the pleural cavities during the adult stage of *L. sigmodontis* infections in BALB/c mice.

Alongside the $Foxp3^+$ Treg responses in the pleural cavity there is also loss of antigen-responsiveness by $CD4^+CD25^-$ effector T (Teff) cells, these hypo-responsive Teff cells exhibit increased expression of the co-inhibitory receptor CTLA-4 and GITR (Taylor *et al.*, 2005). Both the $Foxp3^+$ Tregs and hypo-responsive Teff cells contribute to parasites survival, as depletion/neutralisation of these cells types alone in the pleural cavity of infected mice had little effect on worm survival, but the combined depletion of $Foxp3^+$ Tregs and neutralization of CTLA-4 on the hypo-responsive Teff cells enhanced parasite killing (Taylor *et al.*, 2007).

Although the depletion of $Foxp3^+$ Treg in the pleural cavity during infection had no effect on parasites survival, their depletion prior to infection caused a reduction in adult parasite burdens, as well as causing an anti-fecundity effects on the surviving female adults, but no effect on larval stages. This increase in protection was associated with an early increase in expression of GITR on $CD4^+$ Teff cells, followed with an increase in *L. sigmodontis* specific Th2 responses seen around day 60 post infection (Taylor *et al.*, 2009). Suggesting that early priming of T cells is important in determining the outcome of infection.

The difference in $CD4^+$ Teff cell priming can also explain why difference in protection is seen between different mice strain (Table 1.6). Although both the resistant (C57BL/6) and susceptible (BALB/c) mice have similar initial $Foxp3^+$ Treg responses,

$CD4^+CD25^-$ Teff cells from the resistant mice have greater expression of GITR within the first 12 days of infection which was associated with greater *in vivo* proliferation of $CD4^+$ T cells, compared to susceptible mice. Implying that C57BL/6 mice are either better at priming $CD4^+$ T cells or better at overcoming the initial suppressive effect of the filarial L3 parasites (Taylor *et al.*, 2009).

In susceptible mice strain hypo-responsive Teff cells loss of function was mediated by programmed cell death protein-1 (PD-1, a cell surface receptor, which suppresses T cells) co-inhibition by PD-L2, and induced a progressive loss of IL-4, IL-5 and IL-2 cytokine production, which could be recovered *in vivo* by blocking of PD-1/PD-L2 pathway (van der Werf *et al.*, 2013). The mechanisms that induce these intrinsic changes in Th2 effector cells are thought to be mediated by immune cells, it has been suggested that B cells might be inducing these changes as opposed to alternatively activated macrophages (AAM), however this still has not been proven (Taylor *et al.*, 2006). Another possibility is through dendritic cells (DC), as demonstrated during *Schistosoma japonicum*, which induces TLR-2 signalling in DC that leads to PD-L2 expression and through PD-1/PD-L2 interaction inhibits T cell response to *S. japonicum* (Gao *et al.*, 2013).

When a single immature female *L. sigmodontis* worm was implanted in susceptible BALB/c mice it promotes the survival of co-injected Mf, however in IL-10 deficient mice Mf survival was drastically reduced, suggesting that adult females aid survival of Mf in a IL-10 dependent manner (Hoffmann *et al.*, 2001). The role of IL-10 can be further demonstrated in resistant C57BL/6 mice, which are rendered susceptible by knocking out IL-4, but an addition knockout of IL-10 reverts them to being resistant,

however the addition of IFNγ in these double IL-4/-10 deficient mice had no effect and these mice remained resistant (Specht *et al.*, 2004). The cytokines, TGF-β and IL-10 play an important role in the suppression, as in *L. sigmodontis* infections the absence of functional TGF-β and IL-10 receptor signalling rescued T cell proliferation (Hartmann *et al.*, 2015).

In *L. sigmodontis* infections, T cell derived IL-10 is particularly associated with suppression of $CD4^+$ T cell proliferation, whereas B cell derived IL-10 is not (Haben *et al.*, 2013). These IL-10 producing T cells, are termed T regulatory type 1 (Tr1) cells, and although they are regulatory T cells they do not express Foxp3 or CD25. Tr1 cells are derived from naïve T cells in the periphery following antigen challenge, one of their characteristics is that they produce high levels of IL-10, compared to naturally occurring $CD4^+CD25^+Foxp3^+$ Tregs that emerge from the thymus (Roncarolo *et al.*, 2006). These Tr1 can be found in onchocercomas of *O. volvulus* infected humans, and *ex vivo* studies show that they are able to induce B cells to secrete the "regulatory" isotype IgG4 (Satoguina *et al.*, 2005), in a GITR/GITR-L dependent mechanism with both IL-10 and TGF-β required (Satoguina *et al.*, 2008).

T cell are not the only source of IL-10, transgenic overexpression of IL-10 by macrophages in resistant *L. sigmodontis* infected mice leads to an increase susceptibility with higher number of adult worms and converted resistant FVB mice towards a patent phenotype, supporting the suppressive role of IL-10 (Specht *et al.*, 2011b). These macrophages overexpressing IL-10 in the pleural cavity of infected mice had characteristics of alternatively activation. Alternatively activated macrophages (AAM) are gaining attention in their role in Th2 responses, and are

defined as a macrophage population that rely on IL-4 and IL-4Rα signalling to initiate proliferation and express arginase-1, Ym1 and Fizz1(Rückerl, Allen, 2014). In *L. sigmodontis* infection proliferation of AAM is restricted to the sites of parasites migration and the pleural cavity before the onset of patency, however following release of Mf, AAM are also found in the draining lymph node (Taylor *et al.*, 2006). The specific contribution of AAM in killing and expulsion of helminths is still unknown. It has been suggested that *L. sigmodontis* infections induce suppressive AAM, that can block T cell proliferation (Taylor *et al.*, 2006), however these are not the cause of the PD-1 associated Teff cell hypo-responsiveness in the pleural cavity (van der Werf *et al.*, 2013). Therefore, AAM are not thought to be the drivers of T cell hypo-responsiveness. An alternative theory is that B cells or dendritic cells are conditioning Th2 T cells to a hypo-responsiveness phenotype.

Altogether immunosuppression during *L. sigmodontis* infection consists of several independent overlapping mechanisms, from the different $CD4^+$ T cell regulation ($Foxp3^+$ Tregs and Tr1 Tregs), the intrinsically hyporesponsive effector T (Teff) cells, and alternatively activated macrophages.

Table 1.6. Immune responses differences between susceptible, resistant and vaccinated mice.

Mouse phenotype	Innate response	D10-12 response (pleural cavity)	Response to adults	Response to Mf	Granulomas
BALB/c (patent infection)	Neutrophils	Th2 response	IL-4 and IL-5	IL-4, IL-5, IFN-γ, IgG2a	Mostly neutrophils, surrounding senescing adults. (< day 70 p.i.)
C57BL/6 (non-patent infection)	Neutrophils	Mixed Th1/Th2, CXCL12 and increased cellular recruitment, strong $CD4^+$ T effector cell priming	Only young adults are found	N/A	Mostly eosinophils surrounding young adults. (~ day 40 p.i.)
Irradiated L3 induced protection	Eosinophils	The L3 that have escaped immune responses (~10%), will have the same immune responses as non-immunised mice.			

1.10.3 Role of dendritic cells in filarial infections

Dendritic cells (DC) are professional antigen-presenting cells which are often the first cells to encounter foreign antigens, and therefore play a crucial role in presenting antigen to T cells to initiate immune responses. Their exact role in initiating Th2 protective responses or regulatory responses have been relatively overlooked in *L. sigmodontis* infection.

In helminth infections, depletion of DC severely impaired Th2 immunity to *Heligmosomoides polygyrus* a gastrointestinal nematode and *Schistosoma mansoni* a trematode (Smith *et al.*, 2011; Méndez-Samperio, 2016), suggesting that DC have a potential role in induction of Th2 immunity. Parallel to this, helminths and their excretory and secretory (E/S) products have been linked to functional impairment of DC, which has been suggested as a potential mechanism for nematode induced suppression. In non-helminth models, DC were able to expand $Foxp3^+$ Tregs (Yamazaki *et al.*, 2006; Na *et al.*, 2016), and certain parasite antigen were able to modulate DC function (Silva *et al.*, 2006; Segura *et al.*, 2007; Sun *et al.*, 2013). In human, *B. malayi* L3 were able to induce cell death in human dendritic cells, inhibit their ability to produce IL-10 and IL-12, as well as inhibiting their ability to activate $CD4^+$ T cells (Semnani *et al.*, 2003). Similar results were seen in live infections in mice, where *B. malayi* L3 induced different patterns of maturation and activation in DC subsets, correlating with impaired antigen uptake and presentation, and to some degree the attenuation of T cell proliferation (Sharma *et al.*, 2016). Despite the small number of studies investigating the role of DC in filarial infections, it is being suggested that filarial parasites are able modulate the early immune responses to

infection, by impairing DC antigen presentation and therefore rendering DC ineffective in initiating strong adaptive immune response (Segura *et al.*, 2007; Sun *et al.*, 2013; Sharma *et al.*, 2016).

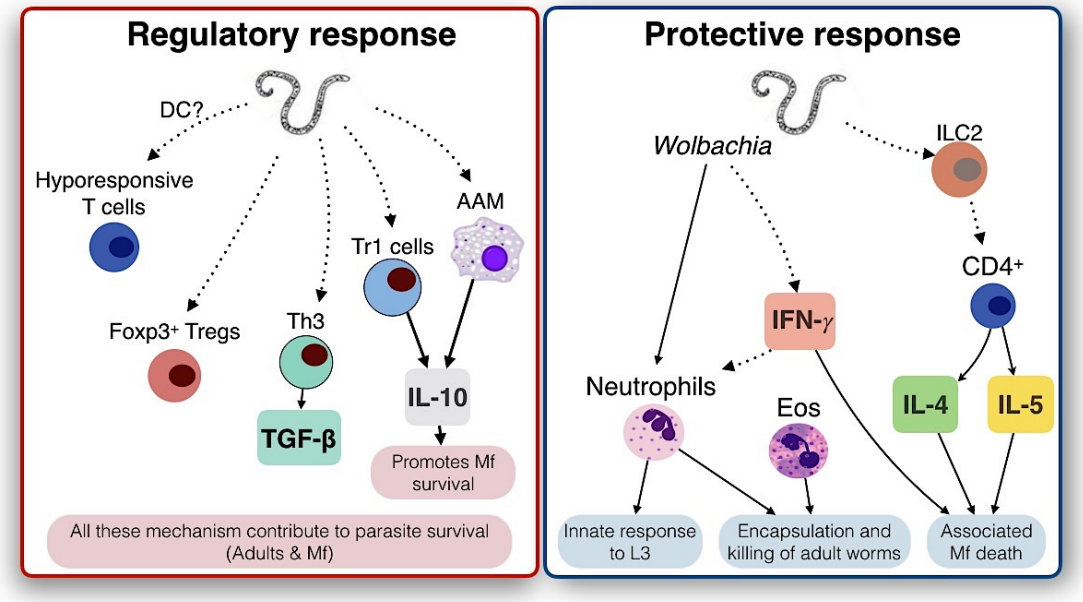

Figure 1.9. Summary of the main immune responses triggered during *L. sigmodontis* infections in BALB/c mice.

1.11 Vaccine induced immunity

Vaccine induced protection in naïve animals can be induced using immunisation with irradiated L3, as demonstrated in several animal models, including *L. sigmodontis* in mice, *O. ochengi* in cattle and *O. volvulus* in the mice diffusion model (Table 1.7). With irradiated *L. sigmodontis* L3 immunisation, protection is associated with a rapid reduction of invading L3 within the first 2 days, leading to a 70% reduction in worm burden (Le Goff *et al.*, 2000). This rapid L3 death has been linked to eosinophils and antibody mediated degranulation of eosinophils. Following immunisation with irradiated *L. sigmodontis* L3, there is an increase in IL-5 levels and subcutaneous eosinophils compared to non-immunised controls (Martin *et al.*, 2000a). However, in mice with impaired B-cell maturation and antibody production (μMT strain) there is a lack of protection, associated with impaired eosinophils degranulation although these are recruited to the site of infection (Martin *et al.*, 2001). This also helps explain the difference between primary infection and immunisation with respect to eosinophils, as during the primary infection there is a delay in inducing IL-5 dependent mechanisms and production of specific antibodies, and by the time eosinophils arrive at the site of infection and have the ability to degranulate, the L3 have already migrated to the lymphatic vessels to escape the inflammatory responses (Marechal *et al.*, 1996). Furthermore, basophils have been linked to the establishment of the vaccine protective immunity, because when basophils were depleted prior to immunisation this diminished protective efficacy, suggesting early amplification of Th2 before immunisation is also necessary for protection (Torrero *et al.*, 2013).

Repeated vaccination with irradiated L3 followed by repeated exposure to challenge L3 infections was carried out, to investigated if repeated exposure would induce immunological tolerance and reduce Th2 vaccine induced immune responses. However, IgE levels; antigen driven basophils release of IL-4; and Th2 skewing of the cellular immune responses remained the same throughout the repeated exposure to the parasite and protective efficacy of the irradiated vaccine was maintained. Suggesting that vaccines which induce a strong Th2 immune response could maintain efficacy through repeated parasite exposure (Hübner *et al.*, 2010).

Subcutaneous immunisation of BALB/c mice with *L. sigmodontis* Mf, inhibits embryogenesis in females worms causing a reduction in Mf found in the peripheral blood, without any effect on adult worm burden (Ziewer *et al.*, 2012). Protection induced Mf immunisation was associated with a shift towards a Th1 environment, mediated through IFN-γ promoted IgG2a. The role of IFN-γ has been linked to immune responses to Mf several times (Lawrence *et al.*, 2000; Saeftel *et al.*, 2001; Taubert, Zahner, 2001).

O. volvulus is not permissive to mice, therefore implantation of *O. volvulus* larval stages into chambers within mice, are used to measure immunological changes linked to parasite death or survival. By immunising mice with irradiated *O. volvulus* L3, protective immunity was developed to subsequent L3 implanted in subcutaneous diffusion chambers. The initial observations found large numbers of eosinophils present in the chambers, and that immunity was dependent on IL-5 and IL-4, thus a Th2 responses (Lange *et al.*, 1994; Johnson *et al.*, 1998). When similar studies were repeated using mice knockouts, mice deficient in eosinophils, mature B cells or IgE

had impaired protective immunity, whereas no difference was seen in protective immunity when eosinophil peroxidase or B1 cells were absent. Therefore adaptive protective immunity to larval *O. volvulus* in diffusion chambers was due to eosinophil recruitment and IgE, corroborating *L. sigmodontis* irradiated L3 mouse studies (Abraham *et al.*, 2004).

The protective efficacy of irradiated L3 immunisation, was successfully translated into field trials using the *O. ochengi* cattle model (Tchakouté *et al.*, 2006). Further cross-protection between *O. volvulus* and *O. ochengi* filarial species can be demonstrated in cattle using immunisation with *O. volvulus* L3, which can induce partial protection of cattle to challenge infections with *O. ochengi* L3 (Achukwi *et al.*, 2007), suggesting that there is antigenic homology between the two species.

Table 1.7. Immunisation with irradiated L3 in different filarial models

Parasite	Model	Protection following L3 irradiated vaccination
B. malayi	Jirds (permissive host)	56-91% reduction in worm burden following challenge infection. (Yates, Higashi, 1985)
	BALB/c mice (non-permissive host) – Diffusion chamber	95-100% reduction in L3 challenge survival. (Hayashi *et al.*, 1984; Abraham *et al.*, 1989)
L. sigmodontis	BALB/c mice (permissive host)	70% reduction in worm burden following challenge infection. (Le Goff *et al.*, 1997; 2000)
O. ochengi	Cattle (permissive host)	84% protection against natural infections compared to non-vaccinated controls. (Tchakouté *et al.*, 2006)
O. volvulus	BALB/c mice (non-permissive host) – Diffusion chamber	64% reduction in L3 challenge survival. (Lange *et al.*, 1993)

1.12 Vaccine development

Vaccines have undeniably played an important role in improving public health by reducing morbidity and mortality to infectious disease. However, most of the commercially available vaccines for humans are against single cell organisms such as bacteria (Tuberculosis, BCG vaccine) and virus (measles, mumps and rubella, MMR vaccine), with more often or not the use live attenuated or inactivated (killed) vaccines (Table 1.8). For filarial disease, protective immunity can also be induced using live attenuated vaccines in animal models, using irradiated L3 (Table 1.7). Although there is an irradiated L3 vaccine commercially available against the nematode *Dictyocaulus viviparus* (lungworm) in cattle (Bain, 1999), the use on an irradiated L3 vaccine would not be feasible for the use against onchocerciasis in humans. This is because there would be technical and production constraints, ethical consideration when administering to humans, and further protection associated with irradiated L3 vaccine is limited to the infective larval stage (Babayan *et al.*, 2006), which does not stop adults producing Mf and allowing transmission to continue.

Table 1.8. Examples of commercially available vaccines

Vaccine (Pathogen)	Type	Protective immunity / Correlates of protection
Vaccine commercially available for humans		
BCG – *Mycobacterium tuberculosis* (TB)	Live attenuated vaccine	Specific IFN-γ production by $CD4^+$ T cell necessary for protection but not a good correlate of protection. (Mittrücker *et al.*, 2007)
DTaP – Diphtheria, Tetanus and Pertussis	Inactivated toxin (toxoid) vaccine	Possible antibody mediated, but not good correlate of protection
Hepatitis A	Inactivated virus	Antibody levels are used as correlates of protection (Plotkin, 2010)
Hepatitis B	Recombinant protein of hepatitis B surface Ag	Antibody levels are used as correlates of protection (Plotkin, 2010)
HPV –Human papillomavirus	Recombinant protein antigen vaccine	Neutralising antibodies prevent mucosal and skin invasion, but not good correlate of protection (Romanowski, 2014)
IPV –Polio	Live inactivated vaccine	Antibody mediated preventing viremia, correlates of protection not known
MMR – Measles, mumps and rubella	Live attenuated vaccine	cellular immunity preventing viremia, correlates of protection not known
RV –Rotavirus	Live attenuated vaccine	Prevents viral replication, correlates of protection not known
Ty21a - Typhoid	Live attenuated vaccine	Baseline *Salmonella typhi* specific CD8+ responses associated with protection and delayed disease onset, however upregulation regulatory T cells have also been linked, and the correlates of protection are still not clear (Fresnay *et al.*, 2016).
Vi capsular polysaccharide - Typhoid	Inactivated subunit vaccine	Vi-specific IgG antibody levels are used for the assessment of protection, but no correlates of protection have been identified for the Vi-based vaccine (Ochiai *et al.*, 2014).
Yellow fever	Live attenuated vaccine	Type I interferon pathways during the first two weeks post vaccination, as well complement pathways, inflammasomes and some regulatory genes (Gaucher *et al.*, 2008)

Vaccine commercially available for veterinary use		
Dictyocaulus viviparus - Cattle	Irradiated L3 larvae	
Taenia ovis – Sheep *Taenia saginata* – Cattle *Taenia solium* - Pig	Recombinant antigen	Lysis of early developmental stages by antibody and complement (Lightowlers *et al.*, 2016)
Boophilus microplus – against tick for Cattle	Recombinant tick gut antigen (Bm86)	Antibody mediated binding and lysis of intestinal cells interfering with blood feeding activity (Dalton, Mulcahy, 2001)

These live/attenuated vaccines can be referred to as first generation vaccines, although they have been tremendously useful for the control of disease such as smallpox and polio, they cannot be used for onchocerciasis. Therefore, second generation vaccines such as subunit vaccines, referring to recombinant proteins and peptides could be a possible alternative strategy. In immunisation experiment against filarial nematode, the use of recombinant proteins have shown some promise (Table 1.10) (Hewitson, Maizels, 2014). However, they can be difficult to produce and can include unnecessary epitopes, not all epitopes found in a protein may contribute to protective responses, as protective immunity is usually dependent on a few antigenic epitopes. If anything, these extra antigenic epitopes complicated the situation by inducing increased immunogenicity to antigens that could cause hypersensitivity and adverse reactions (Linhart *et al.*, 2014; Skwarczynski, Toth, 2016).

Alternatively, peptides based on protective antigen epitope of a vaccine candidates could be used. These short peptide fragments can be used to induce highly targeted

immune responses, incidentally are also cheaper and easier to produce (Li *et al.*, 2014). In a peptide vaccine, amino acid sequences are synthesised to form an immunogenic peptide molecule representing the specific epitope of an antigen. Since these are the antigenic determinants within larger proteins, they should be considered sufficient for activation of the appropriate cellular and humoral responses. In reality, because they contain one antigenic epitope, immunisation with a single peptide may not be capable of producing an appropriate response against the filarial parasites, that have complex life cycles and are masters at immunomodulation, therefore the use of multivalent vaccines are necessary to achieve protection.

Epitope-based vaccines containing well-characterised immunogenic regions has shown success against human papilloma virus (HPV), with two multivalent peptide based vaccines commercially available against HPV (Romanowski, 2014). Although no peptide vaccines are commercially available for filarial parasites, some are being investigated as potential vaccines candidates (Madhumathi *et al.*, 2010).

Third generation vaccines, such as DNA plasmids could potentially be another approach, they have certain advantages over live attenuated or recombinant proteins. DNA vaccines allow protein expression in mammalian cells after introduction of plasmid (Shedlock, Weiner, 2000), which are then able generate cellular and humoral responses (Li *et al.*, 2004). These are simple to produce and purify in large quantities, with low production cost, and easily transported since they do not require a cold chain, these features make DNA vaccines desirable for large scale use in areas endemic to onchocerciasis. Although no human vaccine is commercially available, two DNA vaccines have been licensed for use in animals, one against the fish rhabdoviruses

Infectious Haematopoietic Necrosis Virus (IHNV) in Atlantic salmon in Canada, and the other against West Nile virus in horses, in the United States (Kurath, 2008). In onchocerciasis research, several immunisation experiments in mice with DNA vaccines, have showed potential with signs of protective immunity (Harrison *et al.*, 2000; Babayan *et al.*, 2012; Joseph *et al.*, 2012; Steisslinger *et al.*, 2015).

1.12.1 Strategies for vaccine candidate discovery

Mathematical modelling has shown that a vaccine would greatly complement ongoing efforts to control and eliminate onchocerciasis (Turner *et al.*, 2015), and would benefit in different control settings:

- A prophylactic vaccine, that targets the incoming infective L3, would prevent the establishment of parasites. Preferably a vaccine targeted to children under the age of 5, as to not only protect a vulnerable set of the population but also eliminate a possible reservoir source.

- Immunoprophylaxis vaccine, similar to a prophylactic vaccine, but given to communities who have successfully achieved elimination and therefore protecting them from re-infection (from neighbouring infected communities).

- Therapeutic vaccine targeting the Mf stage, given to infected individuals to prevent severe pathology as well as reducing transmission, and could be an important tool in areas of co-endemicity with loiasis.

Ideally a vaccine would target different life stages of the parasite such as the infective L3 stage and Mf the transmission stage.

Historically vaccine antigens would have been identified by immunoscreening cDNA libraries using serum of naturally protected individuals. For onchocerciasis, 26 recombinant antigens were identified by immunoscreening, including *Ov*-CPI-2, *Ov*-Ral-2 and *Ov*-103, these were then validated with immunisation experiment in several animal models (Table 1.10) (Lustigman *et al.*, 2002; Manchang *et al.*, 2014).

Due to whole genome sequencing, stage-specific transcriptional profiling, and proteomic analysis (Table 1.9), there is now an abundance of proteomic and genomic data on filarial nematodes. This not only provides valuable biological insight into these parasite, but also allows for the identification of potential vaccine candidates (Seib *et al.*, 2012). Knowing that helminth produce excretory-secretory product (ES) that are critical for parasite establishment within the host, as these have been shown to modulate the immune system (Harnett, 2014), in particular adult female ES (Hoffmann *et al.*, 2001), and therefore these can be investigated as potential vaccine targets.

Targeting immunomodulatory molecules as vaccine candidates has shown some success against *Teladorsagia circumcinta* in sheep and *L. sigmodontis* in mice, and by removing the parasite modulatory function it allows for enhanced immune response to the parasite (Babayan *et al.*, 2012; Nisbet *et al.*, 2013; Arumugam *et al.*, 2014b). Some immunomodulatory targets have already been extensively investigated such as a cysteine protease inhibitor (CPI). Whereas other products have been hypothesised as being important due to being highly abundant in female ES, such as *Ls*-ShK, or being homologues of human immunomodulatory molecules, such as the TGF-β homologue, TGH-2 (Table 1.10) (Gomez-Escobar *et al.*, 1998; McSorley *et al.*, 2009; Armstrong *et al.*, 2014).

Targeting the nematodes immunomodulation is one way of removing its defensive mechanism against the host, but helminths do have other defence mechanism, such as antioxidants and detoxification enzymes (Maizels *et al.*, 2001a). These enzymes were initially studied for their use as chemotherapeutics, however their immunogenicity in animal models suggested they might be potential vaccine candidates, such as thioredoxin peroxidase (TPX) which showed some promise in immunisation experiments (Anand *et al.*, 2008; 2012).

Table 1.9. Genomic, transcriptomic and proteomic data available of different filarial species (Adapted from (Grote *et al.*, 2017))

Species	Genome Sequencing	Transcriptomic analysis of:	Whole body proteomic analysis of:	Proteomic analysis of ES products of:
B. malayi	✓		L3, L4, AF, AM, Immature Mf *Wolbachia* (in all stages) (Bennuru *et al.*, 2011)	L3, L3 to L4 moulting, AF, AM, Mf (Hewitson *et al.*, 2008; Moreno, Geary, 2008; Bennuru *et al.*, 2009)
L. sigmodontis	✓	L3 (Allen *et al.*, 2000)	L3, AM, pre-gravid AF, gravid AF, immature Mf *Wolbachia* (in all stages) (Armstrong *et al.*, 2014)	L3, AM, pre-gravid AF, gravid AF, immature Mf *Wolbachia* (in all stages) (Armstrong *et al.*, 2014)
O. volvulus	✓	L3, L4, AF, AM, Mf, vector derived stages *Wolbachia* (in all stages) (Bennuru *et al.*, 2016)	L3, L4, AF, AM, Embryonic stages, Mf *Wolbachia* (in all stages) (Bennuru *et al.*, 2016)	
O. ochengi	✓		L3, AF, AM, immature Mf *Wolbachia* (in all stages) (Armstrong *et al*., 2016)	L3, AF, AM, immature Mf, intradermal nodules *Wolbachia* (in all stages) (Armstrong *et al*., 2016)

L3, third-stage larvae; AF, Adult Female; AM, Adult Male; ES, Excretory-secretory

Table 1.10. Potential vaccine candidates for onchocerciasis.

Vaccine candidate	Function / potential role	Localisation in *L. sigmodontis* and related species*	Immunogenicity & Vaccination experiments
***Ls*–CPI-2**	Cysteine protease inhibitor, an immunomodulator which impairs function of antigen presenting cells.	*Bm*, *Ov*, *Oo*, *Ls*-CPI-2 secreted by adult female worm (ES) and present in all life stages of the parasite.	*Ov*-CPI-2 – Protective antibody responses increases with age (Cho-Ngwa *et al.*, 2010) *Ls*-CPI-2 (mutated) – DNA vaccine, reduced Mf numbers (Babayan *et al.*, 2012) *Bm*-CPI-2 (mutated) – Recombinant protein vaccine, reduced worm burden (~48%) and worm fecundity. (Arumugam *et al.*, 2014b) *Ov*-CPI-2 – Recombinant protein vaccine, reduced L3 survival in diffusion chamber (~49% reduction) (Hess *et al.*, 2014)
***Ls*–TPX-2**	Thioredoxin peroxidase is an antioxidant and detoxification enzymes, which protects the nematodes from host immune effector mechanisms	*Bm*, *Ov* and *Ls*-TPX-2 Expressed by L3 stages (Allen *et al.*, 2000) *Bm*-TPX-2 also found expressed by Mf and in ES products (Anand *et al.*, 2012)	*Bm*-TPX-2 – Protection against larval stages in *B. malayi* chamber model using DNA vaccine, and enhanced efficacy in a multivalent vaccines (Anand *et al.*, 2008) Vaccination with recombinant *Bm*-TPX-2 induces a Th-2 biased response, with IgG3 and IgG1 being elevated in vaccinated mice (Anand *et al.*, 2012) EN individuals carry IgG1, IgG2 and IgG3 *Bm*-TPX-2 antibodies, with higher levels compared to infected individuals (Anand *et al.*, 2012)
***Ls*–Ral-2**	Unknown function	*Bm*-Ral-2 and *Ov*-Ral-2 localised in hypodermis and cuticle of adult female worms, and the surface of L3 and Mf (Lustigman *et al.*, 1992b; 1992a) *Ls*-RAL-2 and *Oo*-RAL-2 found in ES of adults. (Armstrong *et al.*, 2014)	*Ov*-Ral-2 – Recombinant protein vaccine, reduced L3 survival in diffusion chamber (Hess *et al.*, 2014) *Bm*-Ral-2 – Recombinant protein vaccine in *B. malayi* gerbil showed a reduction in adult worm and Mf numbers seen as a reduction in embryonic development stages in female worms (Arumugam *et al.*, 2016)

Vaccine candidate	Function / potential role	Localisation in *L. sigmodontis* and related species*	Immunogenicity & Vaccination experiments
***Ls*–103**	Microfilariae surface associated protein but function is unknown.	*Bm*-103 and *Ov*-103 localised in hypodermis and cuticle of adult female worms, L3 and Mf (Lustigman *et al.*, 1992b; 1992a)	*Ov*-103 – Recombinant protein vaccine, reduced L3 survival in diffusion chamber (Hess *et al.*, 2014) *Bm*-103 – Recombinant protein vaccine in *B. malayi* gerbil showed a reduction in adult worm but no change in Mf numbers (Arumugam *et al.*, 2016) *Ls*-103 – DNA vaccination in *L. sigmodontis* showed no change in worm numbers but reduction in number of Mf circulating in the blood of mice (Unplublished, J Peace MRes Thesis, 2012)
***Ls*–Tgh-2**	Transforming growth protein 2-like protein, potentially an immunomodulatory by binding to TGF-β receptors	*Oo*-Tgh-2 found secreted by adult worms (found in nodule fluid) (Armstrong *et al*., 2016) *Bm*-Tgh-2 found expressed throughout the *B. malayi* life cycle, and also found secreted by adult worms (Gomez-Escobar *et al.*, 2000)	
***Ls*–ShK**	Binds to Kv1.3 gate channels on memory T cells	*Oo*-ShK (found in nodule fluid) (Armstrong *et a*l., 2016) *Ls*-ShK found in ES of gAF (Armstrong *et al.*, 2014)	*Ls*-ShK immunisation in *L. sigmodontis* model showed a reduction in Mf numbers found in blood after vaccination experiment using DNA plasmid vaccines (Duprez. J, MRes Thesis, 2013, University of Edinburgh).

*Localisation of vaccine candidates and its orthologues.

Endemic normals (EN), *B. malayi* (*Bm*), *O. ochengi* (Oo), *O. volvulus* (*Ov*) and *L. sigmodontis* (*Ls*)

1.13 Systems biology and its role in vaccine development

The goal of vaccination is to confer long-term protection in a population at risk of infection and disease, despite their tremendous success, most commercially available vaccines were designed without knowing the mechanisms by which they mediate protection (Table 1.8), and are only now being investigated. Understanding the immunological mechanism following vaccination, could be used to predict vaccine efficacy, or used to improve the immunogenicity of vaccine, allowing for a more rational vaccine development.

Recent studies, have used a systems biology or sometimes referred to as systems vaccinology approach to decipher the immune responses to vaccination in humans (Pulendran *et al.*, 2010). The advantage of using a system biology approach is that it combines "omics" technology such as transcriptomic, proteomics, metabolomics and genomics, with advanced computation tools, to investigate the complex interactions between all parts of a biological systems. Whereas traditional molecular biology techniques only investigate parts of the systems (a gene, a protein, or a cell type), missing out on interactions seen within a system.

A systems approach has the ability to identify early correlates or biological markers (biomarkers) of protection predictive of vaccine responses (Hagan *et al.*, 2015). Since the effectiveness of a vaccine can only be determined once individuals have been infected, being able identifying biomarkers of protection following vaccination, could be used to predict, optimise and evaluate the immunogenicity of vaccines before

infection or before the end of a vaccine trial (Li *et al.*, 2013a; Marino *et al.*, 2016). Using high-throughput techniques, such as microarrays, it is now possible to measure the expression of tens of thousands of genes simultaneously, allowing for a deeper understanding of the molecular signatures induced by infection or vaccination. Microarrays produce high-dimensional data (tens of thousands of genes but usually few samples), requiring more advanced computational methods, such as machine learning (Nakaya *et al.*, 2011a).

Machine learning is subfield of artificial intelligence, and is concerned with developing computer algorithms (models) that learn from the data and can make predictions on new data (Libbrecht, Noble, 2015). It has been used to decipher the complex interaction within the immune system and predict which responses are associated with protection. For example, when machine learning methods were used analyse high dimensional gene expression data following immunisation with yellow fever vaccine (YF-17D), a deeper understanding on how the YF-17D vaccine induced protection was gained, and was able to predict high and low responders within the population studied (Gaucher *et al.*, 2008; Querec *et al.*, 2008). Similar machine learning methods were used to predict systemic adverse events in smallpox vaccine (Reif *et al.*, 2008), and immunogenicity in seasonal influenza vaccine (Nakaya *et al.*, 2011b)

A further use of gene expression data and machine learning tools, is to predict how translatable a vaccine developed in mice model are to humans. Although mouse models have been tremendously useful in deciphering the cellular and molecular

immune responses to vaccines, these are not always predictive of human vaccine responses (Gerdts *et al.*, 2007).

Chapter 2. Validation of vaccine candidates using either DNA or peptide vaccines in *L. sigmodontis* model

2.1 Background

Mathematical modelling of the impact of vaccination against *Onchocerca volvulus* suggested that a vaccine in addition to current controls effort would lead to reduced disease burden and aid in the elimination of onchocerciasis (Turner *et al.*, 2015). In particular, a vaccine would protect vulnerable populations such as children under five and pregnant mothers who are omitted from mass drug administration (MDA) campaigns. Moreover, a vaccine would provide an answer to control of onchocerciasis in populations where Ivermectin cannot be used because of the risk of adverse reactions. By vaccinating individuals currently not receiving treatment but living in endemic areas, would decrease the chance of re-emergence of the parasite to neighbouring areas where MDA has been stopped (Turner *et al.*, 2015).

Since human filarial parasites cannot undergo a full life cycle in mice, several animal models have been used to test the efficacy of vaccine candidates, with *Litomosoides sigmodontis* being the most attractive model, as it is the only filarial species in which the full development cycle can take place in BALB/c mice (Petit *et al.*, 1992), with high immunological cross-reactivity with *Onchocerca* spp (Manchang *et al.*, 2014), allowing immunological and parasitological readouts to be measured during a vaccination trial. The infective larval stages are inoculated subcutaneously, and migrate to the pleural cavity (4-6 days), where they mature into adults and produce

microfilariae (Mf), which enter the bloodstream around day 60 post infection (Petit *et al.*, 1992). Mf and adult worm burdens are used as a measure of protection.

Vaccine mediated protection was generated with live attenuated vaccines using *Brugia malayi* in gerbils, *L. sigmodontis* in BALB/c mice and *Onchocerca ochengi* in cattle models. The irradiated third-larval (L3) vaccine, mediates protection against the incoming infective L3 and subsequently reduces the adult worm burden, although it may reduce the prevalence of disease if not 100% efficacious some of the parasites may develop to maturity and produce Mf, which may not prevent pathology associated with the microfilariae (Mf) stage nor stop transmission. However, a vaccine that targets the Mf stage, while not preventing infection nor pathology may be more successful in reducing morbidity and transmission. Therefore, an ideal vaccine would be one that targets different life stages of the parasite, protecting individuals not only against infection but also pathology caused by the adult stages and Mf of the parasites and hence stop transmission. Despite these live attenuated vaccines inducing 70-91% reduction in adult worm burden, and providing proof-of-principle that a vaccine could induce protection, they cannot be used beyond laboratory setting due to the logistical difficulties of their production, attenuation, packaging, delivery and ethical considerations. Therefore, new strategies such as recombinant proteins, peptides or DNA vaccines are required.

Recombinant proteins are an alternative to live attenuated vaccines, however these tend to present challenges in terms of safety and mass production (Li *et al.*, 2014). Peptides based on the protective antigen epitope of vaccine candidates are an attractive alternative, as a single protein can contain several antigenic epitopes. Peptide vaccines

not only induce a highly specific immune responses, but are also cheaper and easier to produce (Li *et al.*, 2014). A peptide consists of amino acid sequences representing a specific epitope of an antigen, since these represent the antigenic epitope within larger proteins, it should be sufficient for activation of the appropriate cellular and humoral responses. However, in reality a single antigenic epitope (peptide) vaccination may not be strong enough, as filarial parasites have a complex life cycles with the ability to modulate the immune system. Therefore, the use of multivalent vaccines may be necessary to achieve protection. Two commercially available peptide vaccines are currently licensed against human papilloma virus the causal agent of cervical cancer, these are multivalent vaccine, a bivalent HPV-16/18 and the other a quadrivalent HPV-6/11/16/18 (Romanowski, 2014).

DNA vaccines allow protein expression in mammalian cells after introduction of a plasmid, and subsequent induction of the immune system (Shedlock, Weiner, 2000). Advantages of DNA vaccines are that they are relatively simple and inexpensive to produce, although no DNA vaccines are currently licensed for human use, DNA vaccines have been used in vaccination trials for filariasis in animal models (Joseph *et al.*, 2012; Babayan *et al.*, 2012; Steisslinger *et al.*, 2015).

Different approaches have been used to discover vaccine targets, historically these would have been identified by immunoscreening cDNA libraries using serum of naturally protected individuals (Lustigman *et al.*, 2002; Manchang *et al.*, 2014). Using this approach two promising vaccine antigens *Ls*-103 and *Ls*-Ral-2, were discovered. These two candidates have homologues characterised in *O. volvulus* and *B. malayi*, and have been recently evaluated as vaccine candidates either on their own or in

combination (Hess *et al.*, 2014; Arumugam *et al.*, 2016). Both these proteins can be found on the surface and oesophagus of L3 of *O. volvulus* and *B. malayi*, as well as the hypodermis and cuticle of adult worms and the surface of Mf, although their functional properties are currently unknown. In *O. volvulus* using the mouse diffusion chamber model, immunisation with either recombinant proteins *Ov*-Ral-2 and *Ov*-103 in combination with alum as an adjuvant induced protection, seen as a 39% and 30% reduction in worm survival respectively, and 21% reduction was seen when proteins were administered as a fusion protein (Hess *et al.*, 2014). However, in the absence of alum, protection was absent (Hess *et al.*, 2016). Similar results were obtained using *B. malayi* recombinant proteins *Bm*-Ral-2 and *Bm*-103 in gerbils, with a 22-46% reduction in worm burden when proteins were administered alone, and 49-51% reduction when proteins were administered concurrently and 56-61% reductions when administered as a fusion protein (Arumugam *et al.*, 2016). Although both *Bm*-Ral-2 and *Bm*-103 affect adult worm survival, only *Bm*-Ral-2 had an effect on worm fecundity (Arumugam *et al.*, 2016). Using the *O. ochengi* cattle model, recombinant *Oo*-Ral-2 when in combination with 8 other antigens (including *Oo*-CPI) induced protection, seen as a 42% decrease in dermal Mf following natural infections compared to non-vaccinated controls (Makepeace *et al.*, 2009). In endemic areas for onchocerciasis 77% of the population had *Ov*-Ral-2 antibodies, and individuals with these antibodies were significantly less likely to develop ocular pathology (Gallin *et al.*, 1989). In a separate study, higher quantities of antibodies raised against *Ov*-103 were found in endemic normal (naturally protected individuals) compared to infected individuals (Johnson *et al.*, 1995). So far, the evidence shows that both Ral-2 and 103 could be potential vaccine candidates for both onchocerciasis and lymphatic filariasis.

Another strategy to vaccine candidate discovery, is to use our understanding of filarial parasites, to have a targeted approach in the search of vaccine targets. In filarial infections, type 2 immune responses are necessary for controlling infection within the host, however filarial parasites are able to modulate the host immune system, which allows them to survive and evade the host protective immune responses. Therefore, a strategy that targets immunomodulators, would offer an attractive approach. Identification of potential immunomodulators is achievable through extrapolation of known regulators in other organisms (i.e. related species) or because it is known that a potential source of immunomodulators are the excretory-secretory proteins (E/S) of female worms, these can be investigated. Adult females have been shown to contribute to the parasite's ability to maintain chronic infections, and their ES has been hypothesised to contain immunomodulators (Hoffmann *et al.*, 2001).

One of the most promising vaccine candidates is cysteine protease inhibitor-2 (CPI-2), CPI-2 has been characterised in *O. volvulus* (Lustigman *et al.*, 1992b), *B. malayi* (Manoury *et al.*, 2001), *L. sigmodontis* (Allen *et al.*, 2000). CPI-2 belongs to the cystatin superfamily, which have been described extensively in parasitic nematodes (Vray *et al.*, 2002; Gregory, Maizels, 2008), and functions as an immunomodulator (Pfaff *et al.*, 2002) by blocking mammalian protease activity in antigen processing cells (Manoury *et al.*, 2001). In *B. malayi,* vaccination with recombinant *Bm*-CPI-2 showed no protection in gerbils following subsequent L3 infection, however it did induce a significant antibody response and altered worm distribution, resulting in a decrease of adult worms in lymphatic tissue (31% of the worms) and increased worm migration to the heart and lungs (69% of the worms) (Arumugam *et al.*, 2014a). A

mutation in CPI at amino acid position 66 (Asn66 to Lys66), in two independent vaccination experiments using either recombinant protein in the *B. malayi* gerbil model (Arumugam *et al.*, 2014b) or using plasmid DNA in *L. sigmodontis* BALB/c model (Babayan *et al.*, 2012), affected female worm fertility and resulted in reduced Mf numbers circulating in the blood, proving that targeting immunomodulators and allowing the host to mount protection immune response is a feasible strategy for vaccination.

Thioredoxin peroxidase (TPX-2) is a thiol-specific antioxidant detoxification enzyme, a critical component in the parasite's defence against injury caused by oxygen radicals and characterised in *B. malayi* and *O. volvulus* (Chandrashekar *et al.*, 1998). TPX-2 shows low protection on its own, however it does induce a strong Th2 immune response in mice (Anand *et al.*, 2008); furthermore naturally protected individuals in *B. malayi* endemic areas show strong antibody responses against *Bm*-TPX-2 (Anand *et al.*, 2012). When used in combination with other antigens in multivalent DNA vaccines, TPX-2 does induce protection against *B. malayi* L3, in a chamber model (Anand *et al.*, 2008) and a reduction in circulating Mf in *L. sigmodontis* model (Honglin, 2011), demonstrating that TPX-2 is strongly immunogenic and is a potential filarial vaccine but needs to be used in combination with other antigens to confer protection.

A novel antigen *Ls*-ShK (nLs_04059) was identified in the E/S of *L. sigmodontis* gravid adult female (Armstrong *et al.*, 2014), containing six metridin-like ShK toxin domains. These ShK domains have a wide phylogenetic distribution, but the pattern found in *L. sigmodontis* is conserved in the filarial nematodes *B. malayi, D. immitis,*

O. ochengi, O. volvulus, Acanthocheilonema vitae, Wuchereria bancrofti and Loa loa (and *Ascaris suum* an ascaridid nematode) (Armstrong *et al.*, 2014). Although the exact function of *Ls*-ShK is currently unknown it has been hypothesised to be an immunomodulator, as ShK domains can be found in type 1 toxins which are known to block voltage-gated potassium channels (Kv1.3 channels). Since ShK proteins inhibit Kv1.3 channels, a ShK domain peptide from *B. malayi* is currently under development as therapy for autoimmune diseases, specifically targeting Kv1.3 channels expressed by effector memory T-lymphocytes (Chhabra *et al.*, 2014), suggesting that *Ls*-ShK could be modulating acquired immunity by inhibiting memory T cells. In the *L. sigmodontis* model, DNA immunisation with *Ls*-ShK showed initial promise by inducing protection, with no Mf found circulating in the blood compared to controls (Duprez. J, MRes Thesis, 2013, University of Edinburgh).

In human and *L. sigmodontis* filarial infections, TGF-β has been associated with the parasite's ability to modulate the immune system. Members of the transforming growth factor β (TGF-β) family were identified in *B. malayi*, of particular interest was Tgh-2 which shows close similarity to human TGF-β. Tgh-2 has been postulated to be a immunomodulator, whereby Tgh-2 can ligate to host TGF-β receptors and therefore reduce immune responses (Gomez-Escobar *et al.*, 2000). Homologues of Tgh-2 in *B. malayi*, *L. sigmodontis* and *O. volvulus* all contain the same conserved C-terminal domain, and 9 cysteine residues. Although no vaccination experiments with Tgh-2 have been done so far, in chronic *O. volvulus* infections where immune down-regulation plays a key role in survival of parasites, production of TGF-β family of

cytokines have a direct role in immunosuppression (Korten *et al.*, 2010), therefore Tgh-2 shows potential as a vaccine target against immunomodulatory protein.

Whether it is as recombinant protein, peptide or DNA vaccines, a problem with many vaccine candidates is that they have low immunogenicity, therefore to overcome this, adjuvants can be used to enhance their response. DNA vaccine induced immunity can be enhanced by the co-delivery of plasmids encoding cytokines, chemokines or any co-stimulatory molecule by increasing the magnitude or type of immune responses necessary for protection. Co-delivery of interleukin-4 (IL-4) and the macrophage inflammatory protein 1 alpha (MIP1α) also known as the chemokine CCL3, were demonstrated to enhance DNA vaccine-induced immunity (Honglin, 2011; Babayan *et al.*, 2012). Alum is commonly used as an adjuvant in human vaccines, usually in combination with peptides or recombinant proteins, and elicits a strong Th2 and humoral immune response, primarily mediated by IgG1 in mice (Rubin *et al.*, 1986; Beck, Spiegelberg, 1989).

Another strategy to increase protection and overcome the poor immunogenicity of individual peptides is to increase their membrane trafficking, by attaching to the peptide a palmitoyl group (derived from pamitic acid) to enhance their hydrophobicity (Beekman *et al.*, 1997). However, even with these modifications and as multivalent vaccine, there might not be strong enough immune stimulation to trigger an innate immune response, which leads to a strong adaptive immune response. Therefore, peptide-based vaccines also require adjuvants, such as a peptide containing chimeric MHC class II peptide (TpD) epitopes from tetanus and diphtheria toxoid; this peptide

is able to bind to broad range of MHC II alleles, and has been shown to increase CD4 memory T cell recall responses and robust antibody production (Fraser *et al.*, 2014).

Aims of chapter

- Does *Ls*-ShK induce protection following immunisation?
- Can peptides derived from the promising vaccine candidates *Ls*-102, *Ls*-Ral-2, *Ls*-CPIm, *Ls*-ShK and *Ls*-Tgh-2, induce protection?

The protective efficacy of vaccine candidates were evaluated in independent immunisation time courses, using the permissive *L. sigmodontis* BALB/c model. In these immunisation experiments mice received a challenge infection following immunisation. The first experiment involved validating ShK as a vaccine antigen using DNA vaccines. Since peptides are an alternative strategy, relying on highly targeted immune responses with the potential to avoid allergic responses, peptides derived from ShK and the other vaccine candidates were trialled and their protective efficacy evaluated. By targeting non-immunomodulatory regions of the antigen or with a mutated form of an immunomodulatory peptide, it is hoped that the immunomodulation by *L. sigmodontis* can be overcome, allowing a protective immune response to be triggered and protection to be achieved.

2.2 Methods

2.2.1 Ethics statement

All procedures involving animals were approved by the University of Edinburgh and the University of Glasgow ethical review committees, and performed under license from the UK Home Office in accordance with the Animals (Scientific Procedures) Act 1986. Animal experiments were conducted following the ARRIVE (Animal Research: Reporting of *In Vivo* Experiments) guidelines developed by NC3Rs (National Centre for the Replacement Refinement & Reduction of Animals in Research) (https://www.nc3rs.org.uk/arrive-guidelines), which are intended to maximise information published while minimising unnecessary studies (Kilkenny *et al*).

2.2.2 Mice and parasites

All mice used in the vaccination experiment were female BALB/c mice that were obtained from either the Anne Walker Animal unit of the University of Edinburgh or purchased from Charles River (UK). Mice were housed in individually ventilated cages (IVC) at either the University of Edinburgh or the University of Glasgow and in each vaccination experiment the treatment groups were randomly allocated to avoid any cage effect. All mice were between 6-8 weeks of age before the start of any procedure.

The *L. sigmodontis* life cycle was maintained in gerbils using the mite vector *Ornithonyssus bacoti* by Alison Fulton at the University of Edinburgh. To obtain infective third stage larvae (L3) for challenge infections, mites were infected by allowing them to blood feed on *L. sigmodontis* infected gerbils carrying Mf in their

bloodstream. The infected mites were picked from the gerbils and left to incubate for 12 days at 37°C and 70% humidity, allowing Mf to mature to L3. Infective L3 were extracted from those mites (see challenge 2.2.12).

2.2.3 Preparation of *L. sigmodontis* cDNA

Adult female *L. sigmodontis* worms were harvested from the pleural cavity of infected gerbils and washed in sterile phosphate buffered saline (PBS: 137mM NaCl, 2.7nM KCL, 10mM Na_2HPO_4, 1.8mM KH_2PO_4, pH 7.4).

RNA was extracted using the RNeasy kit (QIAGEN), by first homogenising 30mg of adult worms, in 600µl of RTL buffer, for 30s at 30Hz (using Tissue Lyser II and sterile steel balls, QIAGEN). The homogenised worm suspension was centrifuged for 3mins at 300 x g at room temperature, then up to 700µl of the supernatant fluid was transferred to an RNase free 1.5ml tube (Eppendorf) and 1 volume of 70% ethanol was added to it and mixed using a pipette.

The sample (including the precipitate) was transferred to a RNeasy spin column placed in a 2ml collection tube, centrifuged for 15s at ≥8,000 x g and the flow through discarded. The column membrane was washed 3 times, first with 700µl Buffer RW1 (and centrifuged for 15s, at 8,000 x g and flow though discarded), then twice with 500µl Buffer RPE (centrifuged for 15s the 1st time and 2mins the 2nd time, at 8,000 x g and the flow though discarded each time), the 3rd wash step is to ensure no ethanol is carried over during RNA elution. To elute the RNA from the column, 30-50µl RNase-free water was added directly to the spin column membrane and centrifuged

for 1 min at ≥8,000 x g to elute the RNA into a clean collection tube (store RNA at -80°C).

Genomic DNA contamination was removed from the RNA sample using Ambion DNA-free DNA removal kit (Invitrogen), by adding 0.1 volume of 10x DNase I buffer and 1.5µl rDNaseI was added to the RNA and mixed gently, before incubating for 30 mins at 37 °C. Then 0.1 volume of DNase Inactivation Reagent (making sure it has been re-suspended) was added, mixed and incubated for 2mins at room temperature, occasionally mixing, before centrifuging at 8,000 x g for 1.5mins. The supernatant fluid was transferred to a clean RNase free 1.5ml tube, and the concentration of RNA determined using a Nanodrop 2000 spectrophotometer (Thermo Scientific).

To convert RNA into complementary DNA (cDNA), reverse transcriptase polymerase chain reaction (RT-PCR) was used. To do this 1µg of RNA was diluted into of RNase free water to get a final concentration of 74.1ng/µl, to which 1µl of oligo(dT) primer (Bioscript, Bioline) was added to prime the reverse transcriptase reaction, the mix was incubated at 70°C for 5mins. Then 6.5µl of RT-PCR master mix containing: 4µl of 5x Reaction Buffer (Bioscript, Bioline); 2µl of 1µM dNTPs (Promega); 0.5µl of Ribosafe RNase Inhibitor (Bioscript, Bioline); and 0.25µl of reverse transcriptase enzyme (Bioscript, Bioline), was added to the RNA. The RNA mix was placed in a PCR machine and incubated at 37°C for 1hr and then 70°C for 10 minutes, and cDNA is produced (stored at -20°C).

2.2.4 Preparation of whole *L. sigmodontis* antigen

Whole *L. sigmodontis* antigen was made by homogenising adult worms in 1xPBS using a ground-glass homogeniser. The homogenised *L. sigmodontis* was left on ice for 1hr and then centrifuged at 6,000 x g for 10mins. The supernatant fluid was collected and filtered using a 0.22µm syringe filter (Millipore). The protein concentration was determined using a Bradford assay and stored at -80ºC.

2.2.5 Amplifying ShK from *L. sigmodontis* cDNA

Polymerase chain reaction (PCR) was used to amplify ShK from *L. sigmodontis* cDNA, for either cloning into pcDNA3.1 (Invitrogen) or pET29c (Novagen) plasmid, using specific primers (Table 2.1). Each PCR reaction contained: 1x *Pfu* buffer (from 10x *Pfu* buffer: 200mM Tris-HCl pH8.8 at 25°C, 100mM KCl, 100mM NH_4SO_4, 20nM MgSO4, 1mg/ml nuclease free BSA and 1% Triton X-100); 2mM of dNTP mix (Promega, containing 10mM of dATP, dCTP, dGTP and dTTP in water); 1µM of forward primer; 1µM of reverse primer; 1.25u/50µl of *Pfu* polymerase; 0.1µg of cDNA; and nuclease free water to make up the volume needed.

The PCR programme consisted of an initial denaturation step at 95°C for 2 minutes; followed by 35 cycles of: denaturation step at 95°C for 1min, annealing step at a temperature according to the Tm of the primers for 30s (Table 2.1), and an extension step 72°C for 2mins; with a final extension at 72°C for 5 minutes; if left on hold at 4°C.

2.2.5.1 Agarose Gel separation and gel extraction

The PCR products were separated by gel electrophoresis at 100V on a 2% agarose gel (Fisher scientific) in 1xTBE buffer (Thermo Fisher scientific) containing 0.5x SYBR safe DNA gel stain (Invitrogen), by loading the PCR products with 1x Blue Juice Gel Loading Buffer and run alongside a 100bp ladder (Invitrogen).

The QIAquick gel extraction kit (QIAGEN) was used to extract DNA bands of interest (*Ls*-ShK 808bp). The section of the gel that contained the correct sized band was cut out, and placed in 3 volumes of buffer QC for 1 volume of gel, then 1 volume of isopropanol was added and mixed. The dissolved gel is samples was then added to a QIAquick spin column and centrifuged for 1min at 17,900 x g to bind the DNA to the membrane within the column. The flow through was discarded and the DNA washed by a 5mins incubation with 750µl of Buffer PE, followed by centrifuging 1min at 17,900 x g. The flow through was then discarded before re-centrifuging for another minute to remove any residual wash buffer. The columns were then placed in clean microfuge tubes and DNA eluted by adding 50µl buffer EB to the centre of the membrane in each column and left to stand for 1 min before centrifuging for 1min at 17,900 x g. A spectrophotometer (NanoDrop, Thermo Scientific) was used to quantify the concentration of the DNA, at an optical density of 260 nm.

2.2.5.2 Sanger sequencing

To verify that the amplified PCR product had the correct sequence, Sanger sequencing was used, by following the BigDye Terminator v3.1 Cycle sequencing reaction kit (Invitrogen). Each sample was sequenced with their respective forward primer and in

another reaction their reverse primer (Table 2.1), each sequencing reaction consisted of 2µl of 5x sequencing buffer, 0.32µl of 10µM primer, 3.68µl of nuclease free water, 2µl of BigDye and 2µl of the sample (200-500ng of DNA). The sequencing samples then underwent a PCR program of 25 cycles of 95ºC for 30s, 50ºC for 20s and 60ºC for 5mins 15s, when the cycles were over the samples were given to Genepool (Edinburgh Genomics) for sequencing.

Table 2.1. ShK primers for cloning in either pcDNA3.1 or pET29c

	Primers for ShK to clone in:	Annealing Temp
pcDNA3.1	Forward: 5'-CACCATGTCACCGAGTGTTGAGATTGG-3' Reverse: 5'-TTAACAATAATTACAAGTTTTTTCACAATATTTTTGGG-3'	47 ºC
pET29c	Forward: 5'-TATATGGATCCTAGCCGCCACCATGTCACCGAGTGTTGAGATT-3' Reverse: 5'-ATTGGCGCGCGGTCCGCGTGCGCAAGCTTTTAACAATAATTACAAGT-3'	55ºC

*CACC = sequence needed for cloning into pcDNA3.1; GGATCC = BamHI restriction site; AAGCTT = HindIII restriction site

2.2.6 Cloning of ShK pcDNA3.1 plasmid for DNA vaccine

Ls-ShK was cloned into a pcDNA3.1 plasmid to be used as DNA vaccines, in the vaccination time course experiements. The purified *Ls*-ShK PCR product was cloned into the pcDNA3.1 plasmid using the pcDNA3.1 Directional TOPO Expression Kit (Invitrogen). The kit recommended blunt PCR products, hence the use of *Pfu* polymerase, and a 5'CACC sequence for which the primers were designed accordingly (Table 2.1). The PCR product were added at 0.52ng/µl to 2.5-3.3ng/µl of Topo vector (different ratios of PCR product to vector were tried till an optimal ratio was found), in a 1/6 dilution of salt solution (1.2M NaCl, 0.06M $MgCl_2$) provided, the remaining volume was made up to 50µl with nuclease free water. The mixture was incubated at room temperature for 5mins, and immediately put on ice.

The plasmids containing the inserts were then transformed into competent *Escherichia coli* One Shot TOP10 chemically competent *E. coli* cells (Invitrogen), by adding 2µl

of plasmid to one vial of competent cells, and heat-shocked for 30 seconds at 42ºC. To grow up the cells, 250µl of S.O.C media (provided by Invitrogen) was added to the cells and left to grow in a shaking incubator for an hour at 37ºC and 2.5 x g. After the hour, the cells were plated on Luria Broth (LB) (Fisher Scientific UK Ltd) agar plates (LB with 1.5% agar) containing 100µg/ml of ampicillin (Sigma Aldrich), these were left overnight in a non-shaking incubator at 37ºC.

Colonies were picked and grown overnight in 5ml of LB containing ampicillin (100µg/ml) at 37ºC in a shaking incubator at 2.5 x g. The next day the cells were harvested by spinning 2ml of the overnight cultire at 8,000 x g for 3mins and the supernatant fluids removed. Only cells containing the plasmid will have grown. The plasmids were extracted and purified from the cells using the QIAprep spin miniprep kit (QIAGEN).

2.2.6.1 TAQ PCR for plasmid validation

The presence and orientation of the insert in the purified plasmids were checked using TAQ PCR (Bioline), using either T7 primer as the forward primers and the inserts reverse primer (Table 2.1 & Table 2.2); or the forward primer of the insert (Table 2.1) and BGH primer as the reverse primer (Table 2.2). For each 20µl TAQ PCR reaction the following master mix was used: 2µl of 10x NH4 Buffer, 0.6µl of 50mM MgCl2, 0.2µl of 100mM dNTPs, 0.4µl of the 10µM forward primer, 0.4µl of the 10µM reverse primer, 0.2µl of BIOTAQ DNA polymerase, 15.2µl of Nuclease-free water (to make the volume up to 19µl) and 1µl of template DNA(0.1µg).

The PCR program used was: 95°C for 1min as the initial denaturation step; then 35 cycles of 94°C for 45 seconds for denaturation, 30s at 47°C for annealing, and 72°C for 1min for extension, and after the cycles a final extension of 72°C for 10 minutes, and stored at 4°C. The PCR products were then run on a 2% agarose gel (Thermo Fisher scientific) in 1xTBE buffer (Thermo Fisher scientific) for 45 minutes at 100v. The positive colonies that had the insert in the correct direction, were then sequenced using sanger sequencing (section 2.2.5.2). If the sequence was correct, the cells that contained the correct plasmid were stored in glycerol at -80°C.

Table 2.2. Primers used to extract and verify insert in pcDNA3.1 plasmids

Primer	Sequence
T7	5′-TAATACGACTCACTATAGGG-3′
BGH	5′-TAGAAGGCACAGTCGAGG-3′

2.2.6.2 Plasmid amplification and purification for vaccination experiments

Every DNA vaccination required 80µg of plasmid DNA, therefore plasmid quantities were amplified by growing the cells containing the plasmid of interest overnight. A volume of 10µl of the cells (stored at -80°C) were added to 500ml of LB containing ampicillin (100µg/ml) and cultured for 12-16hrs in a shaking incubator at 35°C and 2.5 x g. The bacterial cells where then harvested by centrifuging the culture media for 20mins at 9,300 x g and 4°C, the supernatant fluid was discarded so that only a bacterial pellet was left.

The HiSpeed Plasmid Maxi Kit (QIAGEN) was used to purify the plasmids from the cells. To use the smallest possible volume of DNA vaccine during immunisation, the plasmid concentration must be high, therefore in the final step of the kit 500µl of elution buffer was used instead of the recommended 1ml. To check the concentration of the plasmid DNA samples a NanoDrop was used.

2.2.7 Cloning and expression of recombinant *Ls*-ShK for ELISA

The pET system (Novagen) was used for the cloning and expression of recombinant *Ls*-ShK in *E. coli*. *Ls*-ShK was cloned into pET29c plasmids, using BamHI and HindIII restriction sites. *Ls*-ShK insert was amplified from adult *L. sigmodontis* with primers spanning the full CDS minus the stop codon and with flanking restriction sites, BamHI at the 5' end and HindIII at the 3'end, following the *pfu* PCR protocol in section 2.2.5). The *Ls*-ShK insert and the pET29c vector were digested separately in a 20µl reaction (all from Promega, Uk): 2µl 10x restriction enzyme buffer E; 0.2µl BSA(10µg/µl); 0.5µl BamHI (10u/µl); 0.5µl HindIII (10u/µl); 1µg of *Ls*-ShK insert or pET2c vector; and made up to a final volume of 20µl using nuclease free water. The reaction was incubated for 1hr at 37°C, followed by 15mins at 65°C to inactivate the enzymes. Digested inserts and plasmids were separated on agarose gel and extracted using the same method as section 2.2.5.1. The digested insert and plasmid were ligased together using T4 DNA ligase (Promega), using a ratio 1:3 of vector to insert, using the following equation:

$$ng\ of\ insert = \frac{ng\ of\ vector\ x\ Kb\ size\ of\ insert}{Kb\ of\ vector}\ x\ molar\ ratio\ of\ \frac{insert}{vector}$$

The T4 DNA ligase reaction (Promega): 100ng of digested vector; 50ng of digested insert, 1µl of 10X Ligase Buffer; 0.1-1U T4 DNA ligase; and made up to a final volume of 10µl using nuclease free water. The reaction was incubated for 3hrs at room temperature (~22°C), followed by 10mins at 70°C to heat inactivate the T4 DNA ligase.

The newly ligated plasmids were transformed into JM109 competent cells (Promega), by adding 2µl of ligated plasmids to 30µl of thawed JM109 cells, and incubated for 20mins on ice. The cells were then heat shocked for 30s at 42°C and transferred on ice for 2mins. The transformed cells were then cultured in 100µl of S.O.C media (Invitrogen) in a shaking incubator for 1hr at 37°C and 2.5 x g. 30µl of the cultures were plated onto a Kanamycin (30µg/ml) agar plates and left to incubate overnight at 37°C in a non-shaking incubator.

The following day colonies were picked from the plates, these cells will contain plasmid as cells with no plasmid will not have grown. These colonies were further grown overnight in 5ml of LB with Kanamycin (30µg/ml), in a shaking incubator at 2.5 x g at 37ºC. The plasmids were extracted and purified from the cells using the QIAprep spin miniprep kit (QIAGEN). The presence and orientation of the insert in the purified plasmids were checked using TAQ PCR (Bioline), using either T7 primer

as the forward primers and the inserts reverse primer (Table 2.2) following the same protocol as section 2.2.6.1.

Plasmids which contained the plasmid with the correct insert, where then transformed into BL21 DE3 competent *E. coli* (Novagen) for recombinant protein expression. These were grown overnight at 37°C on Kanamycin (30µg/ml) agar plates. The following day colonies were picked from the plates, further grown overnight in 10ml of LB with Kanamycin (30µg/ml), in a shaking incubator at 2.5 x g at 37ºC.

5ml of the overnight culture was added to 500ml of LB with Kanamycin (30µg/ml) and cultured at 37°C in a shaking incubator (2.5 x g), when the cell culture reached an OD 600 of 0.6, protein expression was induced by adding 1mM of IPTG (Sigma Aldrich). 3hrs following induction, cells from the culture were harvested by centrifuging the cultures for 20mins at 9,3000 x g, and cell pellets frozen at -20°C (this is to facilitate cell shearing).

First the soluble protein was released from the cells, by thawing the cell pellet and re-suspending in 30ml of 1x Binding buffer (5mM Imidazole; 0.5M NaCl, 20mM Tris, pH7.9), the re-suspended cells were frozen and thawed a second time and sonicated to release the protein with 3 times: 30s On, followed by 30s Off. The cells were centrifuged for 20mins at 9,300 x g, supernatant fluids containing the soluble protein were stored at 4°C.

To collect the insoluble protein, the cells were the re-suspended in 30ml 1x Binding buffer containing 6M of Urea. Similar as for the soluble proteins, the re-suspended cells were frozen and thawed and sonicated to release the protein with 3 times: 30s On,

followed by 30s Off. The cells were then centrifuged for 20mins at 9,300 x g, and supernatant fluids stored as insoluble protein stored at 4°C.

The expression of proteins, was verified using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) in samples of the un-induced cell culture; 1hr, 2hr and 3hr post-induction cell culture; soluble protein; and insoluble protein (Figure 2.1). The samples were prepared by adding 75µl of the samples to 25 µl of 4x loading dye (200mM Tris-Cl pH6.8, 400mM β-mercaptoethanol, 8% sodium dodecyl sulphate, 0.4% bromophenol blue, and 40% glycerol), these were boiled at 100°C for 5minutes, and loaded on the gel (NuPAGE, Novex 4-12% Bis tris gel (NP0335BOX)) with a protein ladder (Biolabs, P77115), for 35mins at 200v. To develop the gel, add the gel to Coomassie blue (5g Coomassie brilliant blue R-250 (Sigma B-0630) in 2.5L of 25% methanol and 7.5% acetic acid) for 1hr to stain, and then 1min in de-stain buffer (25% methanol and 7.5% acetic acid).

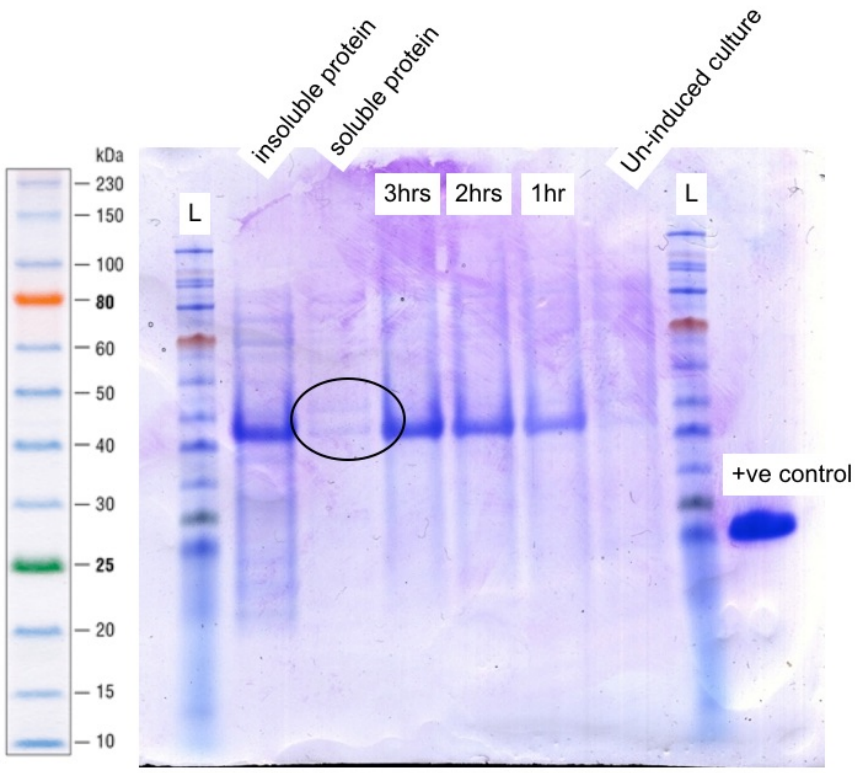

Figure 2.1.SDS-PAGE of *Ls*-ShK at the different stages of expression. 75µl of the samples were added to 25µl of 4x loading dye, and loaded into each well. Samples were from insoluble proteins released from cells; soluble protein released from cells; sample from cell culture 3hrs after induction, 2hrs after induction, and 1hr after induction; and from the un-induced cell culture. Most of the *Ls*-ShK protein was insoluble, with faint bands of soluble protein, that was purified and used to measure *Ls*-ShK specific IgG1 using indirect ELISA.

Unfortunately, most of the protein was insoluble, therefore the insoluble and soluble proteins were isolated separately using His-binding resin columns 1ml Nickel column HiTrap chelating (17-0408-01, GE healthcare) using automated AKTAprime (Amersham Pharmacia) and purified by dialysis using 3500 MeCo membrane (Spectra/Por) in 5L of 1xPBS for soluble proteins or 5L of 1xPBS with 6mM of urea for insoluble proteins. Protein was quantified using a nanodrop, to find protein concentration (mg/ml):

$$A_{280\ (nanodrop)} \times \frac{Molecular\ weight}{Extinction\ coefficient}$$

The soluble protein was used for indirect ELISA to measure *Ls*-ShK specific antibodies.

2.2.8 Vaccination timeline

Two 102-day long vaccination time courses were used to investigate vaccine efficacy. For DNA vaccination experiments, two immunisations were given two weeks apart, followed by a challenge infection four weeks after the second immunisation (Figure 2.2).

For the peptide vaccination experiment, mice received three immunisations, a week apart from each other, and a challenge infection four weeks following the 3rd immunisation (Figure 2.2).

The challenge infection for both experiments consisted of a subcutaneous injection with 23gauge (G) needle of 40 infective L3. For both vaccination experiments the end point was 60-days post challenge infection (D60 p.i) as this coincides with the onset of the patent phase (Mf present in the bloodstream).

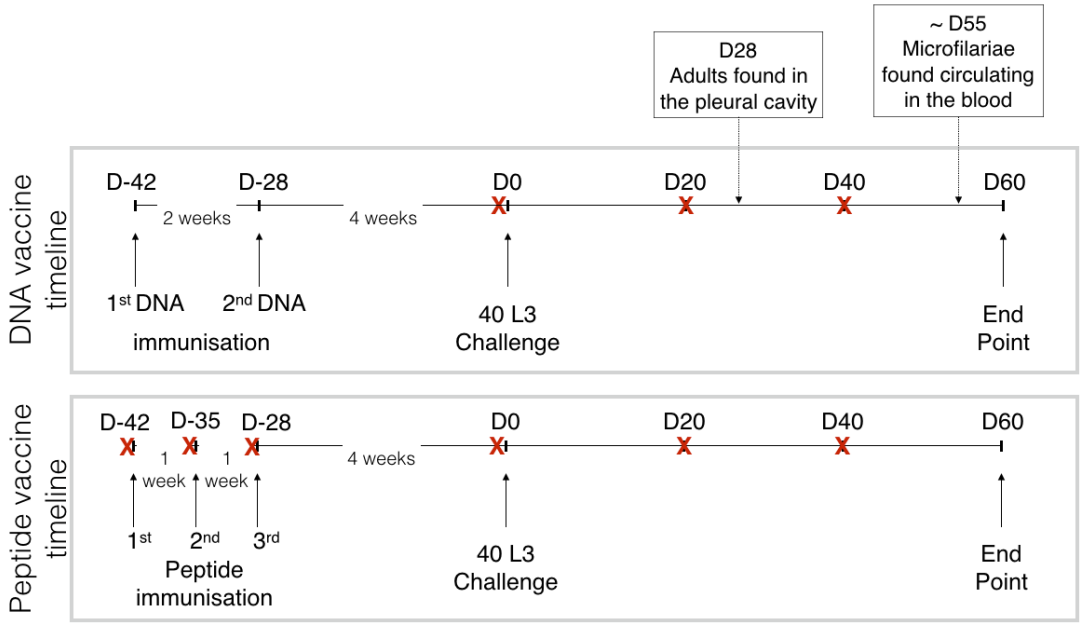

Figure 2.2. Vaccine experiment timeline for DNA and Peptide vaccine candidates. All mice used were female BALB/c, between 6-8 weeks old at the start of the experiment, mice were weighed during the time course before each procedure. For the DNA vaccination experiments immunised mice received two DNA immunisation two weeks apart from each other, and for the peptide vaccination experiment immunised mice received three peptide immunisations a week apart from each other. This was followed by a challenge infection of 40 infective L3 four weeks after last immunisations, so either after 2nd DNA or 3rd peptide immunisation. The experiment was stopped 60-days post challenge infection. During the time course, tail bleeds were performed on all mice (for subsequent antibody analysis) the day before the challenge (D -1), approximately 20 days (D20) and 40 days (D40) post challenge infection. The peptide experiment had extra tail bleeds the day before the peptide immunisation were given (D-43, D-34 and D-29).

2.2.9 DNA vaccine preparation and administration

DNA vaccines used were a cocktail of pcDNA3.1 plasmid with either candidate antigen insert, chemokine or cytokine mouse inserts as adjuvants or empty plasmids as controls (Table 2.3). The pCPIm, pTPX, pIL-4 and pMIP1α DNA vaccines were previously created by members of the Babayan group (Babayan *et al.*, 2012).

Table 2.3. DNA plasmids used in vaccination experiments

Plasmid Name	Plasmid vector backbone	Insert	Role
pCPIm	pcDNA3.1	Mutated form of *L. sigmodontis* CPI-2. (Accession; AF229173.1) Point mutation of asparagine at position 66 to lysine	Vaccine antigen
pTPX	pcDNA3.1	*L. sigmodontis* TPX (WormBase: nLs.2.1.2.g10049)	Vaccine antigen
pShK	pcDNA3.1	*L. sigmodontis* secreted ShK-domain protein (WormBase: nLs.2.1.2.g04059)	Vaccine antigen
pIL-4	pcDNA3.1	*Mus musculus* IL-4 (NCBI Gene ID: 16189)	Adjuvant
pMIP1α	pcDNA3.1	*Mus musculus* MIP1α (Accession: M73061.1])	Adjuvant
pEmpty	pcDNA3.1	-	Non-coding plasmid control

In these experiments 40 BALB/c mice were split into four groups (n=10). The sample size of the treatment groups was calculated based on the vaccine effect on Mf counts of the most promising vaccine candidate *Ls*-CPI (Duprez. J, MRes Thesis, 2013, University of Edinburgh). Using the Biomath website (http://www.biomath.info) the sample size was calculated using a two-sample t-test, with the Mf means from the

pEmpty control group as 55.5 Mf and *Ls*-CPIm vaccination treatment group as 12.8 Mf; an average standard deviation of 73.6, a power 0.8 and a significant level (alpha) of 0.05. This effect size revealed that a sample size of 26 mice per experimental group was needed. However, 26 mice per experiment group would have been too large to process, therefore this was decreased to 10 mice per experimental group, and the immunisation experiment was repeated twice.

For the immunisation experiments, three groups received DNA vaccines or pEmpty control, followed by a challenge infection, whereas the fourth group received neither DNA vaccine nor challenge infection (Table 2.3). Since DNA vaccines (purified plasmids) were stored at -20°C, the day before vaccination the plasmids were placed in a 4°C fridge to allow them to thaw out slowly to avoid any precipitation of DNA. On the day of immunisation, the concentration of the plasmids were quantified using a NanoDrop, and vaccine cocktails made up to total of 80µg of DNA, with equal quantities of each plasmid, this meant that the quantity of each individual plasmid decreased as the number of different plasmid were added to a vaccine cocktail (Table 2.4). The plasmids were delivered in 50µl doses and elution buffer (QIAGEN: 10mM Tris-Cl, pH 8.5) was used to make up the volume.

DNA vaccine cocktails were administered to the tibialis anterior muscle of the left leg with a 27G needle, immediately followed by electroporation with an ECM 830 generator and Tweezertrodes (BTX Harvard Apparatus) using a setting of 8 pulses, 200 V/cm, 40ms duration with 460ms intervals. During the vaccination process mice were under anaesthesia using gas inhalation of isoflurane and placed on heat pads. To wake mice, the gas inhalation was removed and mice recovered naturally on heat pads.

The treatments were randomised per cage so that each cage had at least one of each of the treatment groups.

Table 2.4. Amount of plasmid added to the different vaccine cocktails in a DNA vaccine experiment.

Plasmids	Vaccine A (CPIm+TPX+Adj)	Vaccine B (ShK+Adj)	Vaccine Control (pEmpty+Adj)	Naïve Control (No vaccine)
pCPIm	20μg	-	-	-
pTPX	20μg	-	-	-
pShK	-	40μg	-	-
pIL-4	20μg	20μg	20μg	-
pMIP1α	20μg	20μg	20μg	-
pEmpty	-	-	40μg	-
Total DNA	80μg	80μg	80μg	-

2.2.10 Peptide vaccine preparation and administration

All peptides in the peptide vaccination experiment were designed by Dr. Ben Makepeace (University of Liverpool) (Table 2.5) and produced JPT Peptide Technologies GmbH (Berlin, Germany). These peptides were predicted as being immunogenic using a publicly available website (http://imed.med.ucm.es/Tools/antigenic.pl), using the Kolaskar and Tongaonkar immunogenic scale. This prediction is based on a semi-empirical method, using physicochemical properties of amino acid residues and their frequency of occurrence in experimentally known segmental epitopes, and can efficiently predict immunogenic peptides with about 75% accuracy (Kolaskar, Tongaonkar, 1990). To improve their immunogenicity a palmitoyl group (derived from palmitic acid, a fatty acid) is attached and aids their membrane trafficking.

Table 2.5. Peptides used in vaccination experiment and their amino acid sequence.

Vaccine candidate	Immunogenic peptides	Peptide sequence used in vaccination
Ls-Ral-2	RAL-2_1.1	Palmitoyl-PPFLVGAPPRVVGEFQQLLS-NH2
	RAL-2_1.2	Palmitoyl-QQQQQQQQQQQQQAIPPFLV-NH2
	RAL-2_1.3	Palmitoyl-FVIASLLISCAIAQQQQQQQ-NH2
	RAL-2_2	Palmitoyl-IQRSFQQFKQQAISALQQ-NH2
	RAL-2_3	Palmitoyl-AEHQVAVAKLSP-NH2
	RAL-2_4	Palmitoyl-MQLSAIAKSATLTPVQKQ-NH2
Ls-103	103_1	Palmitoyl-RVMDLLTSIQDKLEPLK-NH2
	103_2	Palmitoyl-KIGSIIS-NH2
	103_3	Palmitoyl-WESLVKKIFVGEGLNAVIPLLKM-NH2
	103_4	Palmitoyl-NGAPAIPITYLLTCVLPLLT-NH2
Ls-ShKm	ShK_1	Palmitoyl-TDANQLCEKADCYAAPNFSQKYCEK-NH2
Ls-CPIm	CPIm_1	Palmitoyl-AGMKYKMEIQVARSD-NH2
	CPIm_2	Palmitoyl-KMEIQVARSDCKKSSC-NH2
	CPIm_3	Palmitoyl-VARSDCKKSSNEKID-NH2
	CPIm_4	Palmitoyl-CKKSSNEKIDLKTC-NH2
	CPIm_5	Palmitoyl-TLEVWEKAWEDFLQV-NH2
	CPIm_6	Palmitoyl-PDQIITLEVWEKAWE-NH2
Ls-Tgh2	Tgh2_1	Palmitoyl-SYQEVCT-NH2
	Tgh2_2	Palmitoyl-KIREVPG-NH2
Adjuvant	TpD	H-ILMQYIKANSKFIGIPMGLPQSIALSSLMVAQ-OH

The sample size for the peptide vaccination was calculated based on the vaccine effect of the same antigens but from DNA vaccines, as this was a pilot study for using peptide vaccines in a *L. sigmodontis* model. Therefore, the same power calculation was used as for DNA vaccine experiments. Using the Biomath website (http://www.biomath.info) the sample size was calculated using a two-sample t-test, with the Mf means from the pEmpty control group as 55.5 Mf and *Ls*-CPIm vaccination treatment group as 12.8 Mf (Duprez. J, MRes Thesis, 2013, University of Edinburgh); an average standard deviation of 73.6, a power 0.8 and a significant level (alpha) of 0.05. This effect size revealed that a sample size of 26 mice per experimental group was needed. However, 26 mice per experiment group would have been too large to process, since the maximum number of mice that can be processed per experiment is around 40 mice, the number of mice was decreased to 7 mice per experimental group.

Therefore, for the peptides vaccination experiment 42 mice were used, which were split into six groups (n = 7). Four groups of mice received peptide vaccines of either: the immunogenic peptides (Vaccine A, Table 2.6); immunomodulatory peptides (Vaccine B, Table 2.6); all peptides (Vaccine C, Table 2.6); or just the TpD adjuvant as control, which is chimera of universal epitopes from tetanus and diphtheria toxoids and is used as an adjuvant (Table 2.6). Two other groups were also present in the experiment and consisted of a challenge infection control (Primary infected) and a naïve control (no vaccine or challenge infection).

Peptide cocktails, including the adjuvant control had a total of 100µg of peptide per dose, with equimolar quantities of each peptide per cocktail, similar to the DNA vaccines the more peptides incorporated into a cocktail, the smaller the quantity of

each peptide. Each peptide cocktail was prepared on the day of vaccination, by mixing with 9% aluminium potassium sulfate dodecahydrate (alum) (Sigma-Aldrich), at a 1:1 ratio, since the final volume per dose per mice was 50µl and there were 7 mice per group, therefore 200µl of peptides were mixed with 200µl of alum. Sterile and filtered PBS was used to make the peptide cocktails to a final volume of 200µl and the 9% alum solution was made in sterile water. Once the peptides and alum solution were added together, 35µl of phenol red (Sigma Aldrich) was added to the mix, this was used as pH indicator, the solution was initially yellow and sodium hydroxide (1M) was added dropwise to the mix (vortexing after each drop) until the solution turned pink. The final pink solutions were then turned on the rotator (Stuart SB2) for 30mins at room temperature before being centrifuged for 10mins at 6000 x g. The supernatant fluid was then discarded and the pellet resuspended in 15ml of sterile PBS. The centrifugation step and resuspension step were repeated 3 times, except that on the 3rd time the pellets of peptides and alum were resuspended to the final vaccine volume with sterile PBS (i.e. 400µl for each peptide cocktail).

Each vaccinated mouse received a 50µl dose of peptide cocktail, the treatments were randomised per cage so that each cage had at least one of each of the treatment groups. The doses were injected intramuscularly into the tibialis anterior muscle of the left leg using a 27G needle, since no electroporation was needed for this procedure mice were not anaesthetised.

Table 2.6. Table Peptide vaccine cocktails.

Vaccine A (Immunogenic peptides)	Vaccine B (Immunomodulatory peptides)	Vaccine C (All peptides)	Adjuvant control
Ls-Ral-2 peptides		*Ls*-Ral-2 peptides	
Ls-103 peptides		*Ls*-103 peptides	
	Ls-CPIm peptides	*Ls*-CPIm peptides	
	Ls-ShK peptides	*Ls*-ShK peptides	
	Ls-Tgh peptides	*Ls*-Tgh peptides	
TpD	TpD	TpD	TpD

2.2.11 Tail bleeds

Tail bleeds were carried out by first placing mice in a heat box at 37°C for 20mins, followed by a vein puncture of the tail vein using 27G needle and the blood was collected in a BD Microtainer SSTTM Tube containing a gel matrix. These were then centrifuged at 6,000 x g for 5mins and stored at -20°C.

2.2.12 Challenge

In each vaccination experiment (DNA or peptide), all mice except those in the naïve group received a challenge with 40 infective L3, four weeks after the last vaccination. On the day of challenge, infected mites were crushed using tweezers in RPMI 1640 media (Gibco, UK), under a dissection microscope. A glass pipette (with a heat elongated tip) was used to collect 40 motile L3, which were then transferred to a glass well, where a 1ml syringe with a 23G needle was used to suck up each dose of L3 and some extra RPMI 1640 media to make a total dose of 0.2ml. To prevent the L3 gathering in the needle, some air was sucked up to the first graduation (0.0 ml). These

infective L3 doses were then given as a subcutaneous challenge infection over the shoulders into the loose skin on the back of their necks.

2.2.13 Samples collected at the end of experiment

At the end point of the experiment (D60), mice were sacrificed by exsanguination under overdose by receiving 20µg of medetomidine hydrochloride (Domitor, Pfizer) and 4mg of ketamine (Vetalar, Boehringer Ingelheim), followed by CO_2. Blood was collected from the subclavial artery under the axila of the mice for different purposes: 30µl of it was added to 270µl of FAC's lysis buffer (BD bioscience) diluted in 1 in 10 in distilled water and later used for microfilariae count; 300µl of blood was added to 1.2ml of RNA later (Ambion) for future RNA extractions; and the rest of the blood was collected in SSTTM blood tubes (BD) and processed in the same manner as for the tail bleeds.

Before mice could be dissected they were placed in a CO_2 chamber to make sure they were dead. Firstly, the pleural cavity washes were performed by initially washing the pleural cavities of the mice with 2ml of 1x PBS, and then with 8ml of 1xPBS. To collect any worms if any present for worm counts and cells and cytokines for future analysis. The pleural cavities were checked for any worms that might be left. Then the mediastinal and parathymic lymph nodes were harvested (lymph nodes draining the pleural cavity) and kept in 5ml RPMI 1640 media (containing HEPES) and supplemented with 100 U/ml penicillin-streptomycin (Gibco), 2mM L-glutamine (Gibco), 10% Foetal Bovine Serum (FBS) (Invitrogen), from now on referred to as complete RPMI (cRPMI).

2.2.14 Immunological read-outs

2.2.14.1 Processing of lymph nodes

The lymph nodes harvested at the end of the vaccination experiment were processed on the same day. The lymph nodes were dissociated to obtain a single cell suspension by grinding them through a 70µm nylon mesh (Fisher Scientific) using forceps in a petri-dish. The single cell suspension was centrifuged at 400 x g for 5mins at 4ºC, the supernatant fluid poured off and the remaining cells were re-suspended in 2ml of cRPMI. The cell concentration of each sample was determined using a haemocytometer, and the cells were made up to a concentration of 10^6 cells/ml using cRPMI, which were then used in re-stimulation assays.

For the proliferation assay, 100µl of the cell suspension were added in triplicate for each treatment to a sterile 96 well round bottom plate. In *vitro* re-stimulation was carried out by adding 100µl of either anti-CD3, *L. sigmodontis* whole antigen or cRPMI as a control to the cell suspensions.

Anti-CD3 monoclonal antibody (BioLegend) was used to stimulate T lymphocytes, it was diluted to 2µg/ml in cRPMI, so that when 100µl was added to the cells the final concentration was 1µg/ml. The anti-CD3 cross-links with T-Cell Receptor present on all T cells and therefore induces all types of T cell proliferation, which is an indication of a healthy cell population. *Litomosoides sigmodontis* whole antigen was made from homogenized worms and made to a concentration of 20µg/ml in cRPMI so that when 100µl was added to the cells their final concentration was 10µg/ml in the wells.

The cells were incubated at 37ºC, 5% CO_2 for 24 hours. To measure proliferation, 20µl (10% of volume in wells) of alamar blue (AbD Serotec) was added 24hours after plating the cells out and then placed back in the incubator. Alamar blue is an indicator dye, that quantitatively measures the proliferation of cells, it is an oxidation-reduction (REDOX) indicator. Approximately 16hrs after adding alamar blue, proliferation was measured by measuring absorbance at 540nm, this was repeated every 2-3 hours until saturation (no change in colour). 72hrs after adding the treatment to the cells, the plates were centrifuged at 400 x g for 5mins and 150µl of the supernatant fluids were removed and stored at 20ºC for cytokine analysis.

2.2.14.2 Processing of pleural cavity lavages

From the pleural lavages, 1ml of the initial 2ml lavage was added to clean Falcon tubes making sure no worms were transferred over, these were centrifuged at 400 x g for 5mins and the supernatant fluid aspirated and stored at -80ºC for future cytokine analysis. The remaining cells and worms were separated, by aspirating the worms from the pleural lavages into a clean falcon tubes and fixing them in 70% ethanol, to be used for adult worm counts.

The remaining cells from pleural lavages were pooled together per sample, and given a red blood cells lysis treatment by pelleting the cells at 400 x g for 5mins at 4°C, treating them with 2ml of RBC lysis buffer (Sigma Aldrich) for 4mins and stopping the treatment with 10ml of cRPMI. The treated cells were centrifuged at 400 x g, 4°C for 5 mins, and the supernatant fluids discarded. The concentration of each sample was

then determined using haemocytometer and then made up to 10^7 cells/ml using cRPMI ready for cell analysis using flow cytometry.

2.2.14.3 Flow cytometry on pleural lavage cells

Flow cytometry was used to detect cell populations using cell surface markers. Pleural lavage cells were stained with a cocktail of markers for T cells, B cells, eosinophils, dendritic cells and macrophages using antibodies in Table 2.7.

Table 2.7. Panel of flow cytometry antibodies used in FACS analysis of pleural lavage cells.

Staining antibodies against:	Fluorochrome conjugated to antibodies	Company
CD11b	Peridinin chlorophyll II protein complex (PerCP)	BioLegend
CD11c	allophycocyanin (APC)	BD Pharmingen
CD19	phycoerythrin (PE)	BD Pharmingen
CD3	PE/Cy7	BD Pharmingen
CD4	fluorescein isothiocyanate (FITC)	BD Pharmingen
F4/80	Alexa Fluor 700 (AF700)	BioLegend
MHC II (I-A/I-E)	Violet 500 (V500)	BD Horizon
SiglecF	Brilliant Violet 421	BD Horizon

To 10^6 cells in FACS tubes, non-specific binding was blocked using rat anti-mouse CD16/32 (BD Pharmingen) at 5µg/ml in 1/20 mouse serum for 30mins in the dark at 4°C. These cells were then washed in FACS buffer (PBS with 2% FBS), centrifuged at 400 x g for 5mins at 4°C, supernatant fluid discarded and re-suspended before adding the staining antibodies for the surface markers, all diluted 1/400. These were left for 30mins in the dark at 4°C cells, this was followed by a wash in FACS buffer, centrifuged at 400 x g for 5mins at 4°C, the supernatant fluid discarded and the

remaining cells re-suspended in 170µl FACS buffer. Cells were analysed using an LSR II (BD Biosciences), running FACSDiva software (BD Biosciences). Some cells were left unstained as negative controls and compensation beads (Affymetrix, eBioscience) were stained with each individual staining antibody in a 1/400 dilution as the compensation control. Analysis of the flow cytometry was performed using Flowjo (Tree star).

2.2.14.4 IgG1 and IgG2a ELISA (Indirect ELISA)

To measure parasite specific Ig levels in sera collected from the tail bleeds or the D60 bleeds, indirect enzyme linked immunosorbent assay (ELISA) were used. ELISA flat bottom NUNC plates (Thermo Scientific) were coated with 5µg/ml *L. sigmodontis* antigen in bicarbonate buffer (0.45M $NaHCO_3$, 0.18M Na_2CO_3 (Sigma-Aldrich), pH9.6) and left overnight at 4°C.

Plates were blocked with 200µl of 4% bovine serum albumin (BSA) in PBS and incubated for 1hr at 37°C in the dark. Plates were washed 5 times in TBS (50mM Tris-Cl, 150mM NaCl, pH7.5) containing 0.05% Tween-20 (Sigma-Aldrich). The mouse serum was serially diluted (1 in 2 dilution) in PBS with 1% BSA. For detection of IgG1, the samples were initially diluted at 1/400 and for IgG2a at 1/200. The serially diluted samples were added to the washed ELISA plates (50µl/well), except the last column of the plate to which PBS with 1% BSA was added (Blank control). Plates were incubated overnight at 4°C in the dark.

The next day the plates were washed 5 times in TBS and 0.5% Tween-20 and incubated with detection antibody isotypes. 50µl per well of HRP-conjugated goat anti-mouse

IgG1 (1/400 dilution) or IgG2a (1/200 dilution) (Southern Biotechnology Associates) in PBS with 1% BSA was added to the washed plates and incubated at 37°C in the dark for an hour. Plates were again washed 5 times with TBS and 0.05% Tween-20, before adding 50µl/well of TMB-H_2O_2 (Merk Chemicals) and left to develop in the dark at room temperature for 5-10mins until a gradient of blue is seen, the reaction is stopped by adding 100µl of H_2SO_4 (1M). Absorbance was measured at 405nm and antibody endpoint titres were determined at the highest O.D. values which exceeded three standard deviations above either blank wells or wells belonging to naïve samples from the same plate.

2.2.14.5 Cytokine ELISA (capture ELISA)

Levels of interleukin 4 (IL-4), IL-5, IL-6, IL-10, IL-12p40, IFNɣ and IL-13 were measured by capture ELISA. Flat bottom NUNC plates (Thermo Scientific) were coated with 50µl/well of coating antibody, each cytokine had a specific concentration and buffer (Table 2.8), and left to incubate overnight in the dark at 4°C.

The next day plates were washed 4 times with PBS and 0.05% Tween-20 buffer, and then blocked with 200µl of 4% BSA in PBS by incubating for 1hr at 37°C in the dark. These were then washed four times with PBS and 0.05% Tween-20, and then 50µl/well of samples were added to the plate in triplicate and the standards for each cytokines. The standards were serially diluted (1 in 2) in their appropriate buffer, starting at the appropriate concentration (Table 2.8) and added to the plate (50µl/well), these will be used to determine the concentration of the samples. The plates were then incubated overnight at 4°C in the dark.

The next day the plates were washed four times in PBS and 0.05% Tween-20, and biotinylated detection antibodies were added to each well (50µl) at the recommended concentration and buffer (Table 2.8). These were incubated at 37°C for 1hr in the dark, and then washed four times in PBS and 0.05% Tween-20. Then AMDEX streptavidin-peroxidase (Sigma) was added to each well (50µl) in a 1/6000 dilution made in same buffer as for the detection antibodies, the plates were incubated in the dark for 30 minutes at 37ºC. Once the final incubation was done the plates were washed five times in PBS with 0.05% Tween-20. To develop the plates 50µl/well of TMB-H_2O_2 (Merk Chemicals) was added and left to develop in the dark at room temperature for 5-10mins until a gradient of blue is seen, the reaction was then stopped by adding 100µl of H_2SO_4 (1M). Absorbance was measured at 405nm and the concentration of cytokine levels in samples were determined using the standard curve made using the O.D. values of the standards against their known concentrations (Figure 2.3).

Table 2.8. Concentrations (Conc) and buffers used for the antibodies in the capture ELISA.

Cytokine	Coating Antibody		Standards		Detection Antibody	
	Conc	Buffer	Initial Conc	Buffer	Conc	Buffer
IL-4	1.2µg/ml [1]	1xPBS	8ng/ml [4]	10% NCS in PBS	0.25µg/ml [3]	10% NCS in PBS
IL-5	1.6µg/ml [1]	1xPBS	10ng/ml [4]		0.17µg/ml [3]	
IL-6	1.0µg/ml [1]	0.1M Na_2HPO_4 pH12	25ng/ml [4]		0.5µg/ml [1]	
IL-10	4.0µg/ml [2]	0.2M $NaHPO_4$ pH6.5	20ng/ml [4]		0.25µg/ml [2]	
IL-12p40	2µg/ml [1]	0.2M Na_2HPO_4 pH6.5	50ng/ml [4]		0.5 µg/ml	
IFN-ɣ	1.6µg/ml [3]	0.1M $NaHCO_3$ pH9.6	50ng/ml [4]		1.0 µg/ml [3]	
IL-13	4.0µg/ml [2]	1xPBS	20ng/ml [4]	1% BSA in PBS	1.0 µg/ml	1% BSA in PBS

For the detection of IL-4, IL-5, IL-6, IL-10, IL-12p40, IFN-ɣ and IL-13 cytokines ([1]BD Pharmingen, [2]eBioscience, [3]BioLegend, [4]Peprotech).

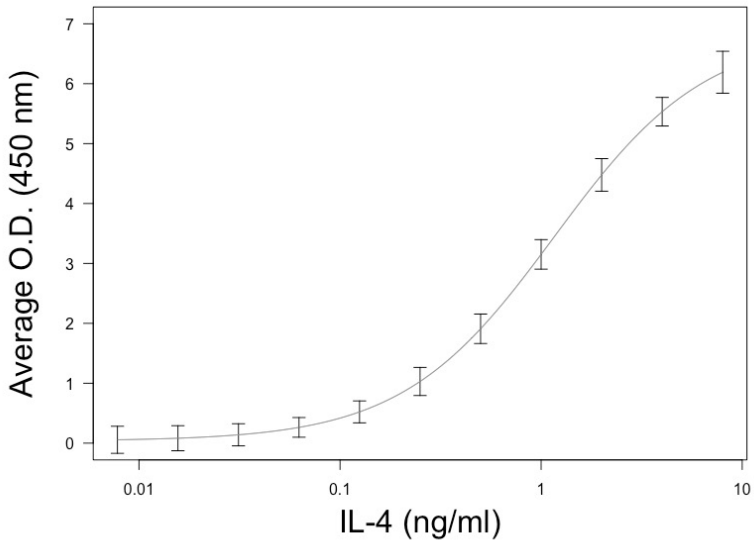

Figure 2.3. Cytokine ELISA Standard curve. An example of a standard curve used to calculate cytokine concentration in samples, this was the standard curve for IL-4, and was used to determine the IL-4 levels in the wells.

2.2.15 Parasitological read-outs

Parasite survival was measured by counting female and male worms that were removed from the pleural cavity during the pleural lavages. The fixed worms were poured into a petri-dish, counted and sexed under a dissecting microscope. Worm survival was calculated as the percentage worms grown to adult stage from the 40 infective L3 challenge, and protection was calculated where possible using the equation below:

$$Protection = \frac{Mean\ burden\ in\ primary\ infected\ animals - Mean\ burden\ in\ vaccinated\ animals}{Mean\ burden\ in\ primary\ infected\ animals} \times 100$$

For microfilariae counts, the 30µl (in RBC lysis buffer) blood samples collected at D60, were centrifuged at 400 x g for 5mins and 150µl of the supernatant fluid discarded. The pellet was re-suspended in the remaining 150µl of supernatant fluid left and then spread on a microscope slide. To count microfilariae an optical microscope using phase contrast and x40 magnification was used.

Adult female fertility was assessed by mounting females onto microscope slides with 40% glycerol and analysed under bright field microscopy. To measure fertility, one would start at vulva and follow the uteri back identifying the different embryonic stages (Stages 1-3) (Figure 2.4). If within the female uteri there were the elongated Mf present then that female was classified as having stage 3 Mf (Figure 2.4). If no elongated Mf were seen then pretzel shaped Mf were looked for (Figure 2.4), if present then that female was classified as having stage 2 Mf. If neither the elongated nor pretzel

shape Mf were present, then the presence of fertilized or non-fertilized eggs were looked for, as it was hard to distinguish between the two, they were both classified as stage 1 Mf (eggs).

When the elongated Mf (stage 3) were present, then the density at which they were found was recorded, density scores ranged from 1<4. A density of 4 was given if elongated Mf were tightly packed and lined up next to each other, as in Figure 2.4, whereas a density of 1 would be given if only a few elongated Mf were found, with lots of pretzel shaped Mf or eggs present with them at the end of the uteri.

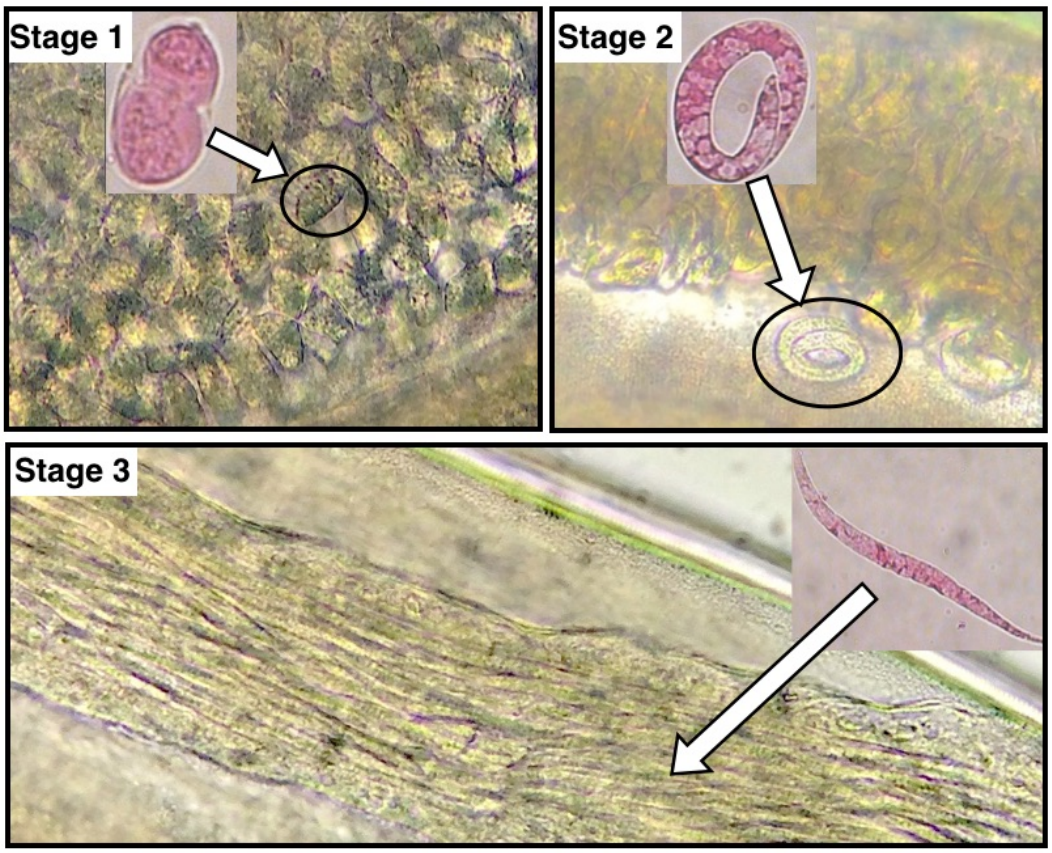

Figure 2.4. Representation of the different embryonic stages found in the female uteri under light microscope at x40 magnification. Within the female embryonic stages were enumerated. In the pictures above the big pictures represents what can be seen under the microscope in the female uteri, when many Mf are close together, whereas the pink pictures were taken from Ziewer et al (2012) and represent the different life stages on their own, once removed from the females. Stage 1 represents fertilised eggs, although sometime hard to distinguish between fertilized and non-fertilised so both were classified at stage 1. Stage 2 represented the developing Mf, sometimes referred to as a pretzel shaped Mf. Stage 3 represented the developed Mf (/elongated Mf), these are sometimes hard to identify as the line up next to each other in the uteri as seen in the picture, the pink picture represents on its own. (Ziewer *et al.*, 2012).

2.2.16 RNA isolation and RT2 profiler PCR array

Blood collected at ~Day 60 post vaccination was stored in RNA later. Total RNA was extracted from blood using the RiboPure^{Tm} Blood Kit (Ambion) following their

recommendations. Briefly cells were collected by centrifuging the blood in RNA later for 3mins at 8000 x g and the supernatant fluid discarded. The pelleted cells were re-suspended in 2ml of lysis buffer (provided by Ambion, with 1% β mercaptoethanol), once re-suspended 200µl of sodium acetate (3M, Ambion) was added and mixed, this was then made up to 3.8ml by adding more lysis solution. Once cells are lysed and all clots dissolved, RNA was extracted by adding 1.5mL of acid-phenol:chloroform (Sigma) and shaking vigorously for 30s. The mix was left for 5 mins and centrifuged for 10mins at 16,000 x g and the aqueous phase was recovered. To recover total RNA from the aqueous phase, 0.5 volume of 100% ethanol (Sigma) was added and vortexed for 10s, which can then be purified using Ambion filter cartridges. To elute the purified RNA from the cartridges, 50µl nuclease free water was used.

RNA quality and quantity was determined using a Tape Station (Agilent Technologies) and a NanoDrop, samples with a RIN values between 7-10 and a A260/280 ratio between 1.8-2.1 were determined as having passed quality control. Therefore 0.5-1µg of QC passed RNA samples were send to Tepnel Pharma Services (Manchester, UK), for processing and quantitative RT-PCR. Briefly Tepnel reverse-transcribed each sample using the RT^2 First Strand kit (Qiagen) mixed with RT^2 qPCR Master Mix containing SYBR Green (Qiagen), to then be added to a customised Mouse Innate & Adaptive Immune Responses RT² Profiler™ PCR Array CAPM13455 (QIAGEN). Each array measured the expression of 84 genes (Appendix A: Table S1); five housekeeping genes (β-actin (Actb); β-2-microglobulin (B2m); glyceraldehyde-3-phosphate dehydrogenase (Gapdh); glucuronidase beta (Gusb); and heat shock protein 90 alpha class B member 1 (Hsp90ab1)); and controls to monitor DNA contamination,

first strand synthesis (RTC) and real-time PCR efficiency (PPC). Quantitative RT-PCR was performed using the ABI 7900HT Fast System (Life Technologies), the data was obtained as threshold cycles (Ct), these values denote the cycle number at which the increasing fluorescence signal of target DNA crosses the threshold set in the logarithmic phase of amplification, therefore lower Ct values indicate greater concentration of target DNA. Genes with low expression (C_t <35) were removed and relative gene expression to Actb were calculated to use to compare changes in expression between vaccinated treatment groups. Actb was used, as this was the only housekeeping gene not to have a statistically significant relationship between treatment and expression (Ct values).

$$\text{Relative gene}_i \text{ expression} = \text{Actb Ct} - \text{gene}_i \text{ Ct}$$

2.2.17 Statistics

Generalised linear models were used to assess the differences between the vaccination groups and the parasitological and immunological readouts. Worm counts were modelled with a Poisson distribution, and Mf counts were log transformed and then modelled with a negative binomial distribution. Differences in gene expression between vaccination groups measured by qPCR arrays was assessed using GLM. Since 84 genes were measured, multiple testing was accounted for using Bonferroni correction. Residuals were tested for normality using Shapiro test and results with a *P*-value lower than 0.05 were considered statistically significant. All statistical analysis and graphs were done on Rstudio v0.97.318 statistical software, and graphs were drawn using ggplot2.

2.3 Results

2.3.1 Validating *Ls*-ShK antigen in DNA plasmid vaccination experiments

Initial vaccination with *Ls*-ShK p.c.DNA3.1 plasmid showed protection, with a decrease in worm burdens in the pleural cavity and numbers of microfilariae found circulating in the blood (Figure 2.5 ShK Vacc Exp 1, from Duprez. J, MRes Thesis, 2013, University of Edinburgh). This initial vaccination experiment showed that *Ls*-ShK was significantly protective, by decreasing Mf numbers when used on its own, but when combined to CPIm which is a promising vaccine candidate, the protective effect was abolished. Therefore, vaccination experiments using ShK were repeated and compared to different vaccine combinations. In these experiments 40 BALB/c mice were split into four groups (n=10), with three groups receiving two intra-muscular injections with plasmid DNA containing the antigen inserts of interest two weeks apart from each other, this was then followed by a challenge infection of 40 infective L3 four weeks after the last immunisation. The three immunised groups were:

- CPIm_TPX (Adj) – These mice received a cocktail of pcDNA3.1 plasmids containing the CPIm and TPX antigens, with IL-4 and MIP1α as adjuvant plasmids. This was used as a control for a protective vaccination, as it has been previously shown to be a promising vaccination cocktail (Honglin, 2011).

- ShK (Adj) – These mice received a cocktail of ShK pcDNA3.1 plasmid, with IL-4 and MIP1α as adjuvant plasmids.

- pEmpty (Adj) – These mice receive pcDNA3.1 plasmid with no antigen insert, as well as the IL-4 and MIP1α as adjuvant plasmids. This was to control for the effect of the plasmid backbone and adjuvants.

The fourth group were mice that received no vaccination nor challenge infection and termed naïves.

To assess the protective effect and immunogenicity of the ShK plasmid vaccine, blood was collected at day 60 post challenge for microfilariae counts and antibody levels in the serum; and the pleural cavity of mice were washed to collect adult worms to determine worm survival. However, the repeated ShK vaccination experiments (ShK Vacc Exp 2 and ShK Vacc Exp 3), showed no sign of protection with similar percentages of worm survival between the ShK and the pEmpty control (Figure 2.5. A), similarly no decrease in Mf numbers was seen in vaccinated groups in the repeat ShK experiments (Figure 2.5. B), contrary to what seen in the first ShK experiment.

Antibodies raised to *L. sigmodontis* antigen were found in challenged mice, with significantly increased levels of IgG1 (Figure 2.6. A) in infected mice compared to naïve controls, although no differences were seen between the immunised groups. IgG2a which is normally found in low levels in *L. sigmodontis* infections, were raised in infected groups in the third ShK experiment compared to Naïve controls, suggesting a Th1 response was raised (Figure 2.6. B, ShK Vacc Exp 3).

To determine if the ShK plasmid used in the repeat DNA vaccinations experiments still had the correct antigen insert, the plasmids were amplified using PCR, and the products showed that indeed an insert of the correct size (808bp) was present,

suggesting that the change in protection is not due to the absence of the vaccine antigen in the immunisation (Figure 2.7).

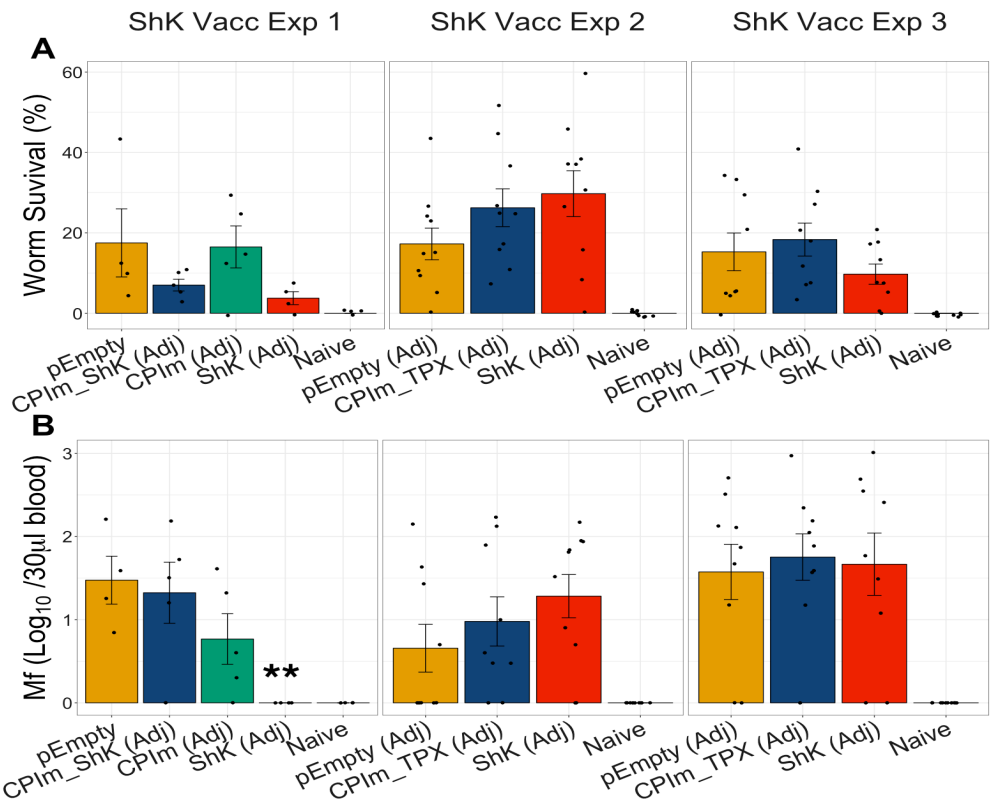

Figure 2.5 Percentage worm survival and microfilariae counts from three independent ShK vaccination experiments. Mice were vaccinated using different combinations of DNA plasmids, using pEmpty as the vaccination control. 4 weeks after the last vaccination mice were challenged using 40 L3, and then at day 60 post-challenge infection (p.i.) mice were sacrificed and 30µl of blood was collected for B) Mf counts, these were counted under a phase contrast microscope. Pleural lavages were performed to collect the worms from the pleural cavity, these were counted under a dissection microscope, and A) worm survival was calculated as the percentage of worms present at day 60 p.i. from the 40 L3 used in challenge infection. ShK Vacc Exp 1 data was collected in a previous study (Duprez. J, MRes Thesis, 2013, University of Edinburgh) and differs in terms of vaccination groups compared to the ShK Vacc Exp 2 and ShK Vacc Exp 3. After the ShK Vacc Exp 1 vaccination experiment, the CPIm_ShK with adjuvant and CPIm with adjuvant groups were replaced with CPIm_TPX with adjuvant, and was intended to be used as a positive control for vaccination. Dots represent individual mice (N=5 per treatment group for 1st experiment, N=9-10 for the 2nd and 3rd experiment), and bars represent the mean per treatment group with error bars as the standard error of the mean. GLM's were used to analyse the statistical difference between the pEmpty group and the vaccinated groups, since the Naïve group was not infected it omitted from analysis (** P-value <0.01).

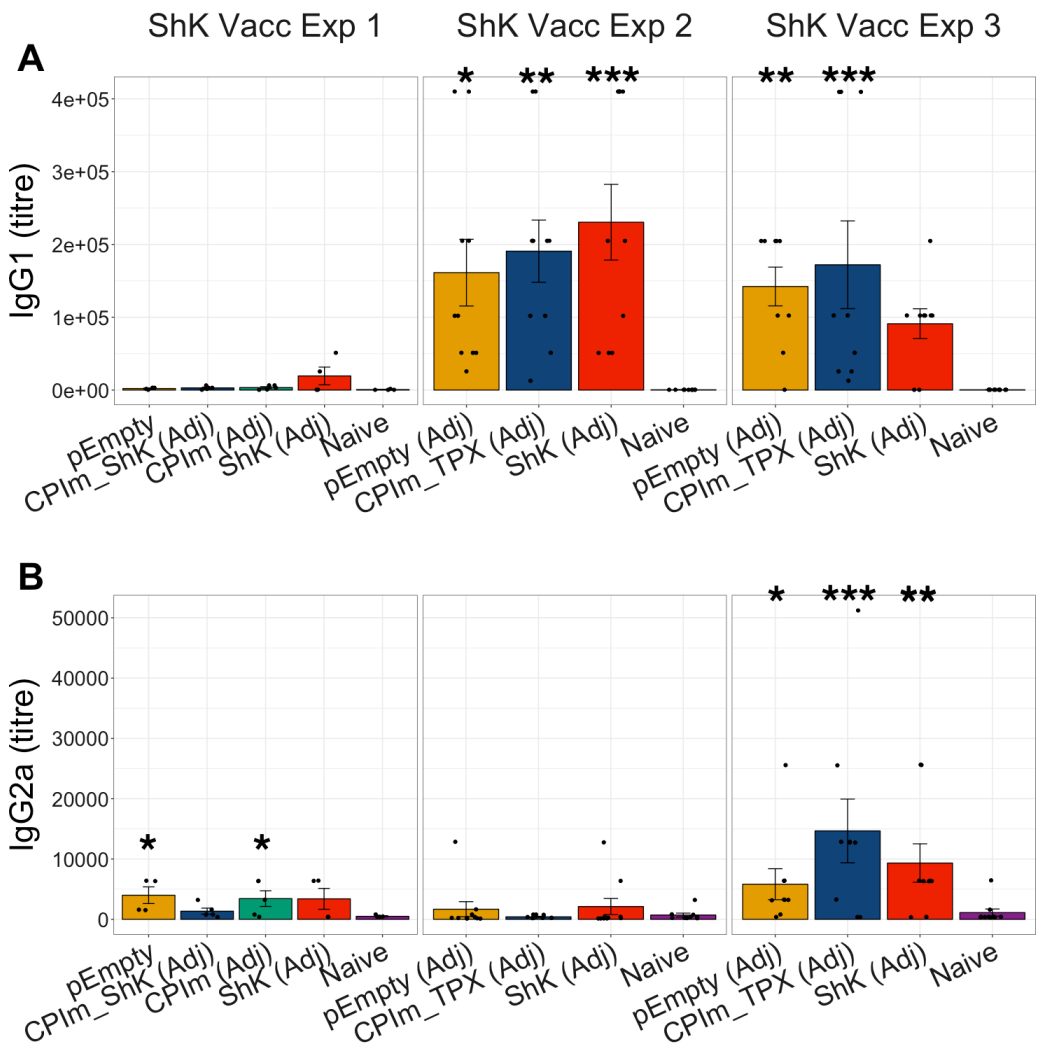

Figure 2.6. Levels of IgG1 and IgG2a specific to *L. sigmodontis* whole antigen, represented as titres in blood serum at day 60 p.i. in three independent ShK vaccination experiments. Serum from blood collected at day 60 p.i. were analysed for *L. sigmodontis* antigen (*Ls*-Ag) specific to A) IgG1 and B) IgG2a in three independent ShK vaccination experiments. Antibody levels were measured using an indirect ELISA and are represented as titres. Each dot represents individual mice, split into their vaccination groups (N=3-5 per group for ShK Vacc Exp 1, N=9-10 for the ShK Vacc Exp 2 and ShK Vacc Exp 3), and bars represent the mean per treatment group with error bars as the standard error of the mean. GLM's were used to analyse the statistical difference between the näive group and the other vaccinated groups, as no difference was seen between infected groups (* P-value <0.05, ** P-value <0.01, *** P-value <0.005).

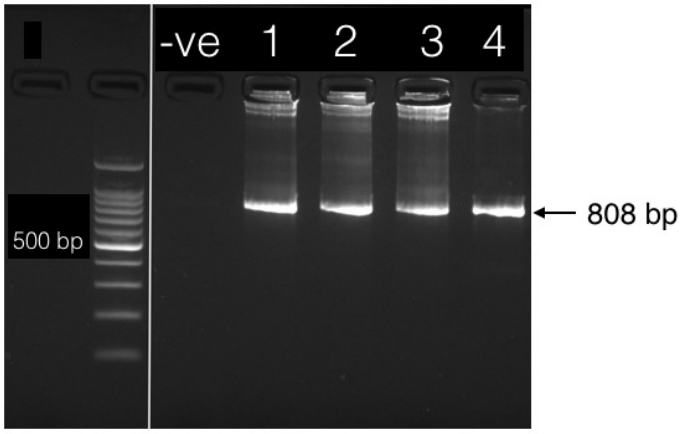

Figure 2.7. Gel of PCR products of ShK plasmid. To determine if the ShK pcDNA3.1 plasmid used in the vaccination experiments still contained the antigen insert, a PCR using a ShK primer and a T7/BGH primer (Table 2.2) was done on four different ShK plasmids batch that were used to make up the vaccine cocktails. The PCR product show a band of the correct size representing the ShK antigen (808bp), each band on the gel marker corresponds to 100bp.

2.3.1.1 Measuring changes in gene expression at day 60 post infection

Blood samples were collected for RNA extraction at the end of the DNA vaccination experiments (day 60 post challenge infection), to identify changes in gene expression that might correlate with protection, with hope to use these as markers of protection. RNA was extracted from the blood and the quality of the samples were analysed before being send to Tepnel Pharma Services (Manchester, UK) for qPCR array analysis, to measure the expression of 84 genes involved in immune pathways. Not all samples passed the quality control and of those that did, only 6 mice were classified as protected showing no signs of Mf in the blood following immunisation. Of the 6 protected mice two mice were from antigen immunised groups (ShK or CPIm_TPX immunisations) and four from the control immunised group (Figure 2.8). Thus, it was impossible to

investigate changes in gene expression associated with protection induced by different vaccine candidates.

Changes in gene expression were therefore compared to Mf numbers in the blood (regardless of immunisation), and genes involved in interferon (IFN) signalling (Irf7, Stat1, Jak2, Tyk2) or downstream effects of IFN (Ccl5) were found to be significantly positively correlated with Mf numbers (Figure 2.9), as well as H2-T23 and Myd88 (Figure 2.10). H2-T23 is a MHC class I presentation gene and Myd88 is an adaptor protein involved in most toll like receptor signalling (except TLR3).

It was not possible to determine whether changes in gene expression was associated with vaccine-induced protection (only 2 samples from immunised mice had 0 Mf). However, Ccr4 and Cd80 were found to have higher expression in vaccinated mice compared to primary infections, but there was no correlation with Mf or worm numbers (Figure 2.11). Ccr4 (C-C chemokine receptor type 4) encodes a protein which is a receptor for various CC chemokines known to regulate cell trafficking of various types of leukocytes and aid the interaction of antigen primed T cells and DC (Wu *et al.*, 2001). Furthermore, Cd80 encodes a protein for a co-stimulatory molecule found on dendritic cells and activated B cells, and is necessary for T cell activation, this possibly indicates that in vaccinated mice there was an increase in DC and T cell interaction compared to primary infected mice.

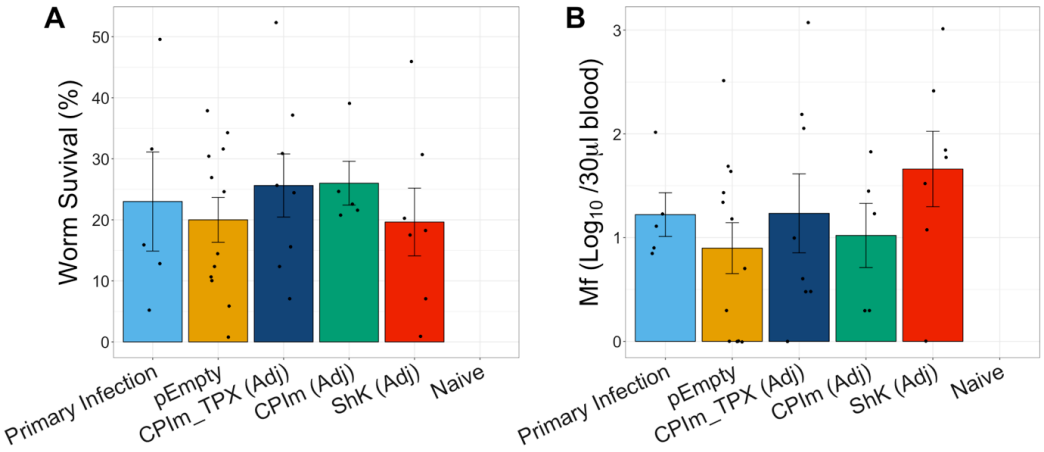

Figure 2.8. Percentage worm survival and microfilariae counts from three independent DNA vaccination experiments, used in the qPCR array. Worm survival and log transformed Mf numbers found circulating in the blood at day 60, in three independent DNA vaccination experiments from which RNA was extracted from blood at day 60 for qPCR analysis. The vaccination groups were: *Ls*-CPIm and *Ls*-TPX with adjuvants (n=8); *Ls*-CPIm with adjuvants (n=5); *Ls*-Shk with adjuvants (n=7); empty pcDNA3.1 control (n=12), these all received a challenge infection. The primary infected group received no vaccine and were only challenged infected (n=5), and the naïve group received neither vaccine nor challenge infection (n=11). Many RNA samples, failed quality control, hence the unbalanced groups.

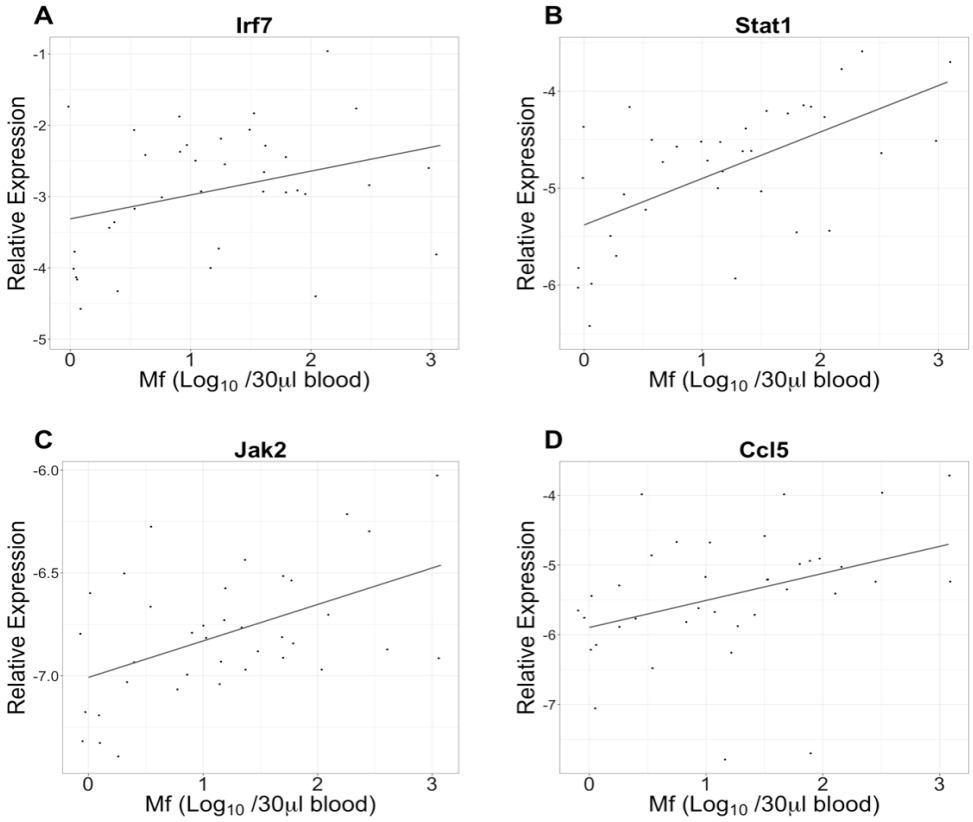

Figure 2.9. IFN signalling genes were found to have a significant association with microfilariae numbers in the blood. RNA was extracted from blood at day 60 post challenge infection. and qPCR arrays were used to measure changes gene expression between vaccination groups. Relative expression of genes to Actb were calculated and expression of genes involved in IFN signalling (Irf7, Stat1, Jak2 and Ccl5) were found to have a positive correlation with Mf numbers in the blood (represented as log transformed Mf data). These associations all had P-values <0.05 (GLM).

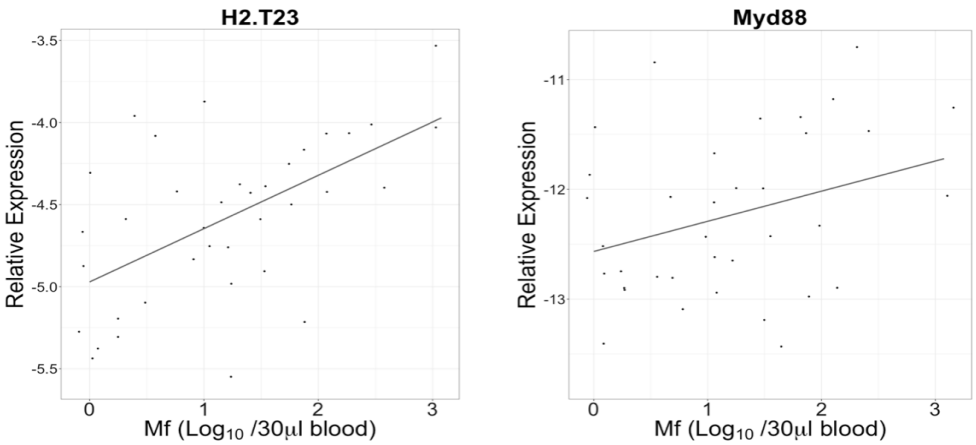

Figure 2.10. H2-T23 and Myd88 gene expressions were significantly associated with microfilariae numbers in the blood. RNA was extracted from blood at day 60 p.i. and qPCR arrays were used to measure changes gene expression between vaccination groups. Relative expressions of genes to Actb were calculated and expression of H2-T23 and Myd88 were found to have a positive correlation with Mf numbers in the blood (represented as log transformed Mf data). H2.T23 is a component of MHC-class II molecules and Myd88 is an adaptor connecting proteins that receive signals from outside the cell to the proteins that relay signals inside the cell. These associations all had P-values <0.05 (GLM).

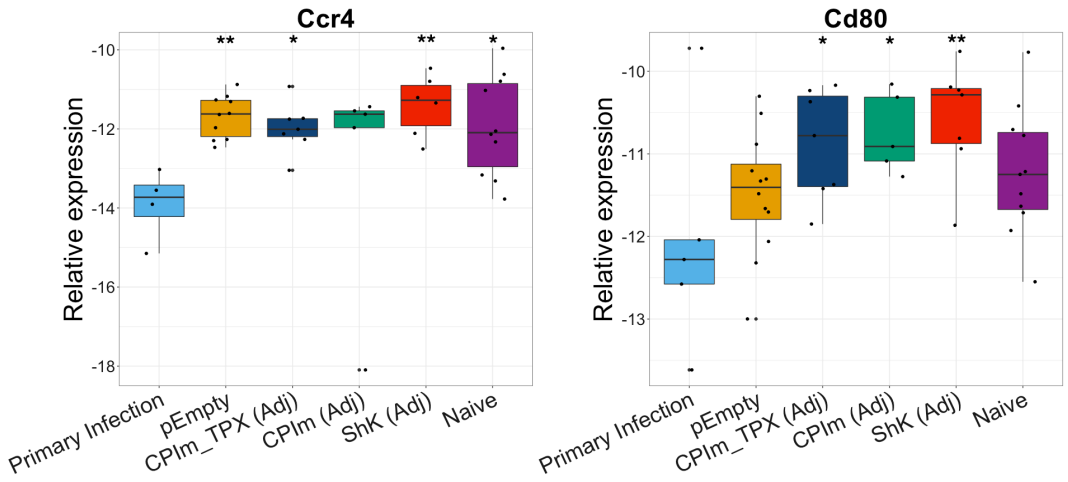

Figure 2.11. Relative expression of Ccr4 and Cd80 across treatment groups. RNA was extracted from blood at day 60 p.i. and qPCR arrays were used to measure changes gene expression between vaccination groups. Relative expression genes to Actb were calculated, Ccr4 and Cd80 were the only genes in which there was a difference in expression between vaccinated groups and primary infections. Ccr4 is a receptor for the CC chemokines MIP-1, RANTES, TARC and MCP-1. Cd80 is a receptor found on dendritic cells, activated B cells and monocytes, and induces T cell proliferation and cytokine production. Each dot represents individual mice split into their vaccination groups (n=5-12 per treatment) and error bars represent standard error of the mean. GLM's were used to assess the statistical difference between the vaccinated groups and the primary infection. *** P-value <0.001, ** P-value <0.01 and * P-value <0.05

2.3.2 Vaccination with peptide antigens

Peptides derived from the vaccine candidates *Ls*-Ral2, *Ls*-103, *Ls*-CPIm, *Ls*-ShK and *Ls*-Tgh2, were predicted to be immunogenic using the Kolaskar and Tongaonkar antigenicity scale (Kolaskar, Tongaonkar, 1990). This prediction algorithm makes use of physicochemical properties of amino acid residues and their frequencies of occurrence in experimentally known epitopes, to predict potential antigenic epitopes in protein sequences. Application of this method to a large number of proteins was shown to predict antigenic determinants with 75% accuracy which is higher than most known methods (Kolaskar, Tongaonkar, 1990).

The vaccine candidates can be split into two groups based on their known function (or hypothesised function), peptides derived from proteins such as CPI, ShK and Tgh2 which are known to have immunomodulatory properties were grouped together, whereas Ral2 and 103 functions remain unknown but show to be immunogenic and induce protection in various vaccination experiments formed the second group (Table 2.5).

For the immunomodulatory proteins, six peptides were derived from *Ls*-CPI, whereas *Ls*-ShK (nLs_04059) had 19 immunogenic peptides predicted, however only the peptide that corresponded to the putative active C-terminal domain was used, and this amino acid sequence was mutated to remove the active site, by replacing both Lys-528 and Tyr-529 with Ala. The rationale behind the mutation is similar to the mutation of CPI by Babayan et al (2012), in which mutating the active site of the molecules would neutralise the immunoregulatory function molecule. This mutation would not only

avoid immunising with a potential functional molecule, but also the antibodies raised against the molecule following immunisation would have the ability to neutralise the native molecule in the following infection and hence prevent potential future immune regulation, allowing the immune response to mount a protective response.

As for *Ls*-Tgh2 (nLs.2.1.2.g04798), which is highly conserved with mammalian TGF-β especially in the region of the active domain in the C-terminal portion, two peptides were predicted within that region and synthesised.

For the second group of peptides, *Ls*-RAL2 (nLs_01747) had four peptides predicted, however one of them was too long to synthesise (50 amino acid long) and was therefore broken down into three, whereas *Ls*-103 (nLs_03356) had five antigenic peptides predicted and were all synthesised.

All peptides had a palmitoyl group added to the N-terminal of the peptide, to increase their immunogenicity by increasing their membrane trafficking. To increase the immune responses to the peptides, a chimera of universal epitopes from tetanus and diphtheria toxoids (TpD) was co-administered.

Since there were 20 peptides predicted from the antigens of interest, assessing the immunogenicity and protection of individual peptide would not only be pointless as one peptide on its own is not likely to be immunogenic enough, but also logistically it would require too many mice. Therefore, peptides were grouped by their functional role and vaccination experiments immunising BALB/c mice against groups of peptides was carried out. Forty-two mice were split into six groups (n=7), four of these groups

were given three intra muscular injections a week apart from each other with either a cocktail of:

- Ral2_103_TpD: A combination of highly immunogenic *Ls*-Ral-2 and *Ls*-103 peptides with the TpD adjuvant in alum (total of 11 peptides).
- CPI_ShK_Tgh2_TpD: A combination of immunomodulatory *Ls*-CPI, *Ls*-ShK and *Ls*-Tgh2 peptides with the TpD adjuvant in alum (total of 10 peptides).
- All_peptides: A combination of all the 19 peptides including TpD adjuvant in alum (a total of 20 peptides).
- TpD: The chimera of universal epitopes from tetanus and diphtheria toxoids (TpD) was give on its own in alum, to measure its effect as a control.

Four weeks after the last immunisation, the immunised groups and a fifth group which had not received an immunisation (Primary infection), received a challenge infection with 40 infective L3. The sixth remaining group did not receive any immunisation nor a challenge infection, and was termed naïve control.

2.3.2.1 No change in adult worm survival but a trend towards decreased Mf in peptide immunised mice

To assess protection efficacy of the peptide vaccines, worm survival was calculated as the percentage of adult worm's present in the pleural cavity out of the 40 infective L3 used in the challenge. Overall around 20% of the worms survived and matured into adults, for all the challenged (/infected) groups showing no difference between the different immunisations groups (Figure 2.12. A).

Numbers of microfilariae were counted from 30µl of blood collected at the end of the vaccination experiment. Groups vaccinated with Ral2_103_TpD and CPI_ShK_Tgh2_TpD showed a slight decrease in Mf numbers compared to primary infection (Figure 2.12. B), but this small effect is abolished when all peptides are combined.

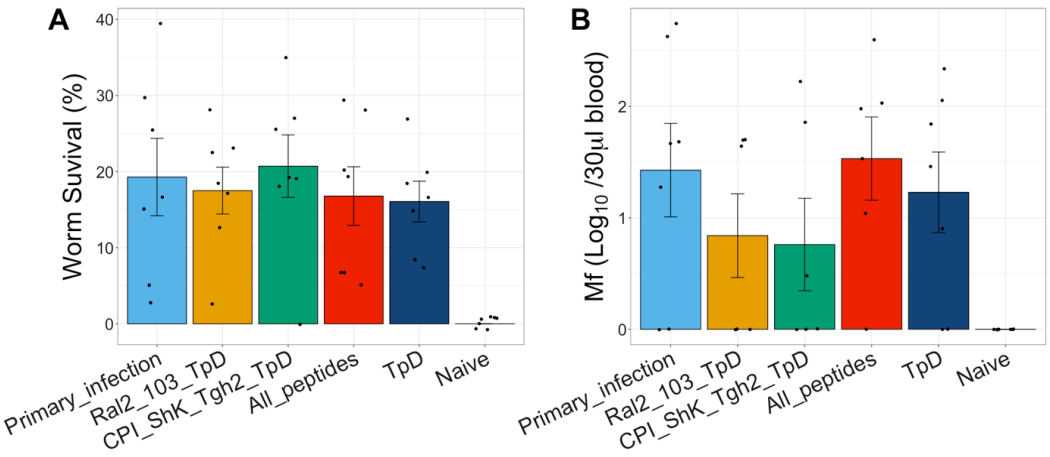

Figure 2.12. Worm survival and microfilariae count at day 60 post-challenge infection. At day 60 post-challenge infection (p.i.) mice were sacrificed and 30µl of blood was collected for B) Mf counts, these were counted under a phase contrast microscope. Pleural lavages were performed to collect the worms from the pleural cavity, these were counted under a dissection microscope, and A) worm survival was calculated as the percentage of worm's present at day 60 post infection from the 40 L3 used in challenge infection. Dots represent individual mice (N=7 per treatment group), and bars represent the mean per treatment group with error bars as the standard error of the mean. GLM's were used to analyse the statistical difference between the challenged groups (Naïve group was omitted from the analysis since it was not challenged), but no significance was found.

To further investigate the impact of the peptide vaccinations on the fertility of female worms, the female uteri of each worms were looked at under a microscope and a score was given depending on the Mf development stage present in its uteri: oocyte or divided egg (stage 1), pretzel Mf (stage 2) and elongated/mature Mf (stage 3) (Figure

2.4). Empty females or females that were too degraded to measure were excluded from the analysis, leaving 76% of females to be analysed.

When comparing the effect of the different immunisations, on the development of Mf found within worms, mice immunised with CPIm_ShK_TpD had a greater percentage of female worms with only eggs (stage 1) present within their uteri (i.e. missing the fully developed Mf). All other groups had a greater percentage of female worms with mature Mf.

When mature Mf were present in the uteri of female worms, the density at which they were found was also recorded, as in some worms where mature Mf were present, these were found in low numbers with lots of dead eggs or dead "pretzel" Mf stages. The density of mature Mf was given as a score (1-4), where 1 is given when few mature Mf are present with lots of dead eggs or pretzel shaped Mf, and 4 when the majority of the Mf present were mature Mf.

Immunisation with Ral2_103_TpD had no effect on the development of Mf (Figure 2.13A), however this peptide vaccine did reduce the density at which the Mf were produced (P-value<0.05), with more dead eggs and debris found in these female worms, compared to primary infections (Figure 2.13.B).

In conclusion, CPIm_ShK_TpD immunisation is affecting the development of Mf within the uteri. Ral2_103_TpD immunisation does not impair the ability of females worms to produce mature Mf, but the density at which they are produced is reduced, compared to those found in the worms recovered from mice that had not been vaccinated (primary infection). When no Mf were detected in the blood of mice,

analysis of adult female uteri revealed that only stages 1-2 were present (no fully mature Mf). Therefore, immunisation with either immunogenic or immunomodulatory peptides is affecting the fecundity of female worms, hence why there was a slight decrease in circulating Mf in those vaccinated mice.

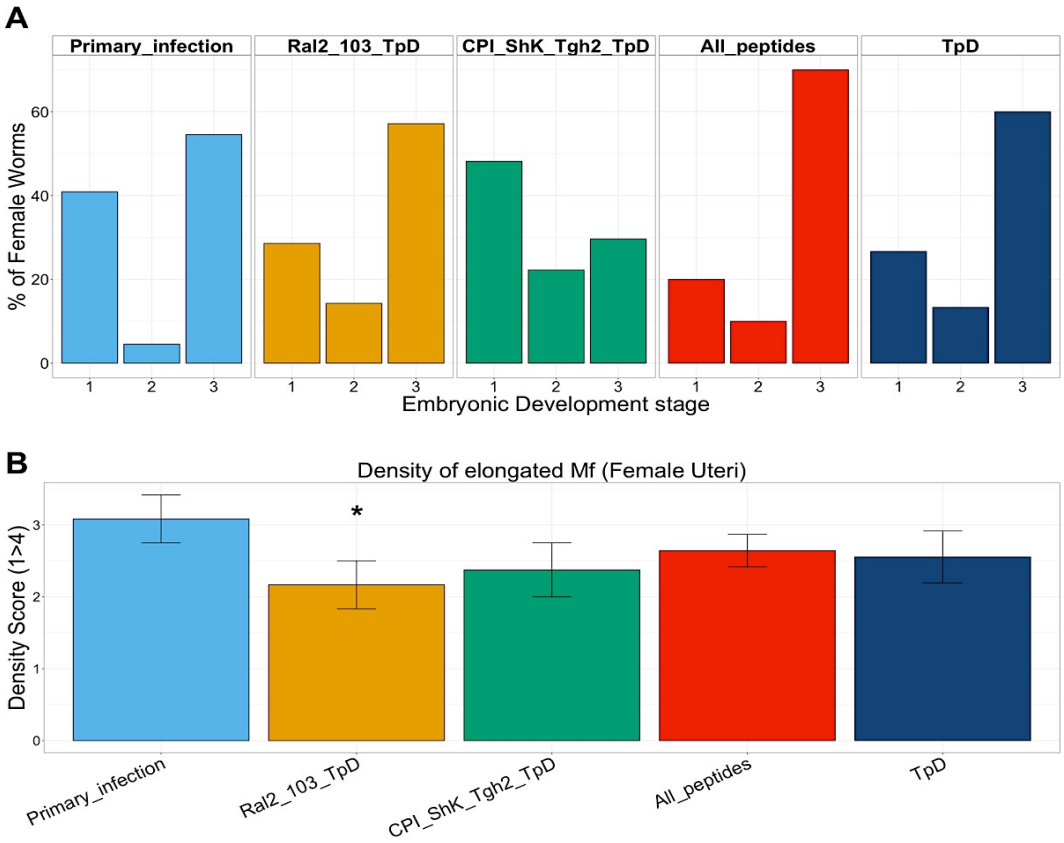

Figure 2.13. Embryonic stages in female uteri between treatment groups. The uteri of adult females from the vaccination experiments were analysed under an inverted microscope and the worms were scored from 1 to 3 depending on the developmental stage of Mf present. Stage 1, Fertilised egg; Stage 2, Pretzel shaped Mf; and Stage 3, Elongated Mf (mature Mf). A) represents the percentage of females harbouring the different life stages, between the different treatment groups. B) When the fully mature Mf were found in the female uteri, the density (1<4) at which they were found was also recorded, the low density of 1 is when lots of dead fertilised eggs or pretzel shaped Mf were present with the elongated Mf, and the higher density (4) is when Mf were tightly packed together. Therefore, the graph above represent the average density per immunisation group. Density of elongated Mf were compared between the vaccinated groups and primary infection using GLM (* P-values <0.05).

2.3.2.2 Ral2 and 103 peptide combination raises a Th2-type response

To assess the immunogenicity of the peptides, blood was collected from mice at days 14, 42 and 60 post challenge infection to measure antibody levels. The kinetics of IgG1 was assessed by measuring IgG1 levels throughout the infection, showing that IgG1 levels specific to *L. sigmodontis* antigen (*Ls*-Ag) significantly increased over time, with a 5-fold increase in antibody titre at each time point in infected mice (Figure 2.14A). By day 60 post infection levels of IgG1 titers were higher in immunised groups compared to primary infection with only Ral2_103_TpD vaccination being significantly increased (P-value < 0.01, Figure 2.14B), whereas IgG2a responses were low, with only a few mice in the CPI_Shk_Tgh2_TpD and TpD groups having increased levels of IgG2a (Figure 2.14C).

IgG1 antibodies specific to CPI were only slightly raised in mice vaccinated with the CPIm, suggesting some cross-reactivity with other peptides (Figure 2.15A). ShK specific IgG1 was significantly increased in the CPI_Shk_Tgh2_TpD vaccinated (Figure 2.15B), however since ShK specific antibodies were only measured in the primary infection, CPI_Shk_Tgh2_TpD and naïve group it is impossible to tell if there is cross-reactivity with the other peptides used.

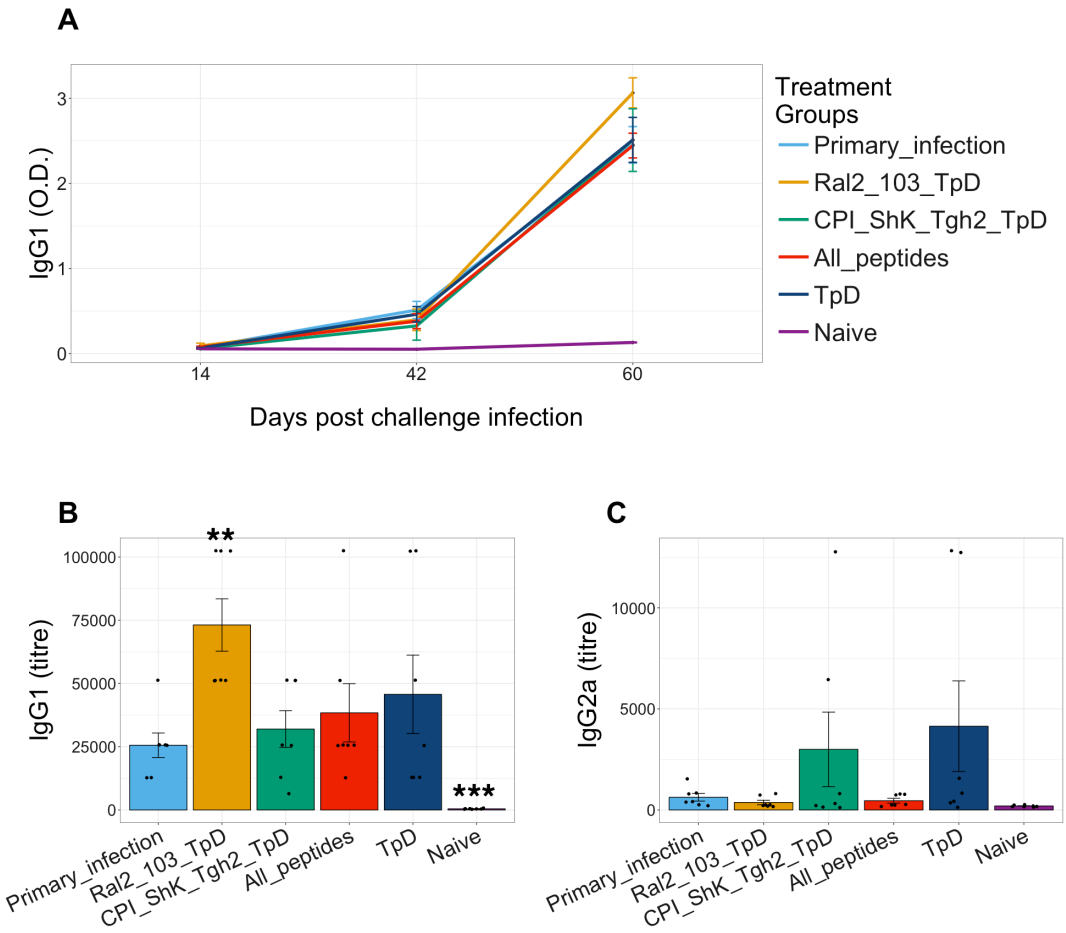

Figure 2.14. IgG1 specific to *L. sigmodontis* whole antigen in blood serum at day 14 and day 42 post infection. A) Kinetics of IgG1 specific to *L. sigmodontis* antigen levels in blood serum at days 14, 42 and 60 post challenge infection, these responses were measured using indirect ELISA and are shown as the O.D. values (450nm) for the 1/800 dilution. B) *Ls*-Ag specific IgG1 titers at day 60 p.i. for which all challenged groups had significantly higher IgG1 compared to naïve controls (P-value <0.05), and vaccination with RAL-2_103_TpD peptides induced significantly IgG1 compared to the non-vaccinated primary infection control (** P-value <0.01, *** P-value <0.005). C) *Ls*-Ag specific to IgG2a titres, only the CPI_ShK_Tgh2_TpD and TpD control vaccination had increased IgG2a compare to naïve controls (P-value <0.05), but none of the vaccinated groups had any statistical difference with primary infection. Each dot represents individual mice, split into their vaccination groups (N=3-7 mice per group, some mice did not have enough blood for the ELISA), and error bars represent standard error of the mean. GLM's were used to assess the statistical difference between the vaccinated groups and the primary infection.

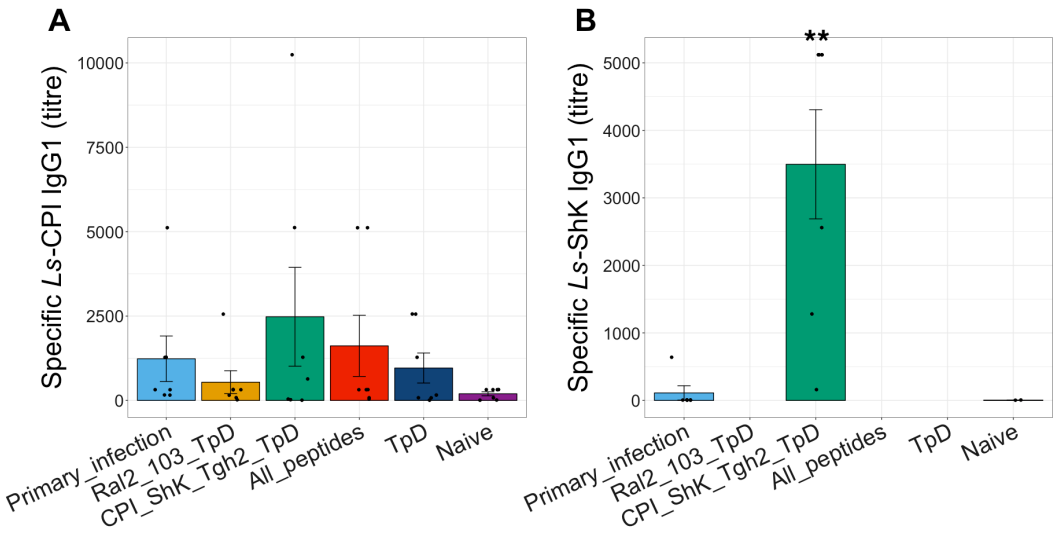

Figure 2.15. IgG1 titres specific to *Ls*-CPI and *Ls*-ShK in blood serum at day 60 p.i. Serum from blood collected at day 60 p.i. were analysed for IgG1 antibodies specific A) *Ls*-CPI and B) *Ls*-ShK by indirect ELISA and are represented as titres, however only the CPI_ShK_Tgh2_TpD, primary infection and naïve control groups were measured for *Ls*-ShK specific IgG1, and showed that the CPI_ShK_Tgh2_TpD mice had increased IgG1 compared to the primary infection (** P-value <0.005) and naïve control (P-value <0.05). Each dot represents individual mice, split into their vaccination groups (N= 7 mice per group), and error bars represent standard error of the mean. GLM's were used to assess the statistical difference between the treatment groups.

2.3.2.3 Lower numbers of macrophages, eosinophils and DC recruited to pleural cavity after peptide vaccination compared to infected controls

To determine what cells were being recruited to the pleural cavity. Cells were collected at the end of the vaccination experiment (day 60 p.i.) by pleural lavages with PBS. Overall infected mice had greater number of cells present in the pleural cavity compared to naïve controls (Figure 2.16A), analysis of the cell population with flow cytometry showed that eosinophils are the predominant cell population found in infected mice although the numbers of macrophages, eosinophils and activated DC

($CD11c^+$ MHC^+ cells) were found in lower numbers in peptide vaccinated mice compared to primary infected, especially when vaccinated with all peptides and the TpD control (Figure 2.16. B, C and D). For the number of lymphocytes, B cells showed little difference between the groups even compared to non-infected naïve mice, suggesting that B cells have little role in infection in the pleural cavity (Figure 2.16E), whereas $CD4^+$ T cells numbers were increased in infected mice with higher numbers in mice vaccinated with Ral2_103_TpD (Figure 2.16F).

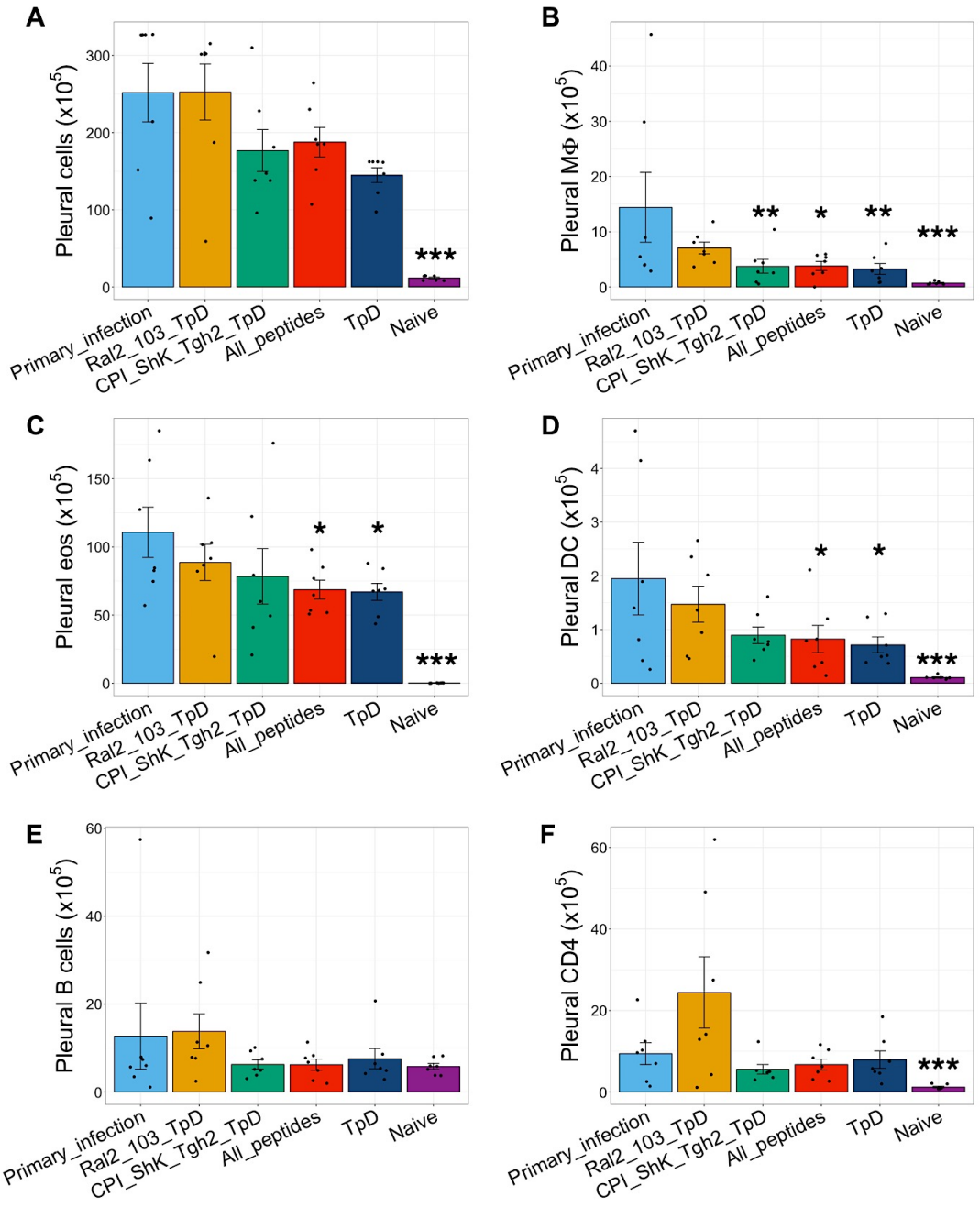

Figure 2.16. Cell recruitment to pleural cavity at day 60 p.i. At day 60 p.i. mice were sacrificed and their pleural cavity washed to recover cells, A) represent the total number of pleural cells harvested from the pleural cavity, showing that challenged mice irrespective of vaccination or not had a significant increase in cells present in the pleural cavity compared to naïve controls (*** P-value <0.001 compared to primary infection, but all challenged groups have significantly higher number of cells compared to naïve, P-value <0.05). Flow cytometry was used to differentiate between cell populations found in the pleural lavages, B) macrophages (Mϕ) classified as $CD11b^{+}F4/80^{+}$ ($SiglecF^{-}CD19^{-}$) cells, C) eosinophils as

SiglecF^{+}CD11c^{-}, D) activated dendritic cells (DC) as CD11c^{+}MHCII^{+}, E) B cells as CD3^{-} CD19^{+} and F) CD4^{+} T as CD3^{+}CD4^{+} cells. Asterisks denotes significant difference between primary infection and the other groups (*** P-value <0.001, ** P-value <0.01, * P-value <0.05). Each dot represents individual mice, split into their vaccination groups (N= 7 mice per group), and error bars represent standard error of the mean. GLM's were used to assess the statistical difference between the treatment groups.

2.3.2.4 No change in cytokine responses in the pleural cavity with peptide vaccinations

No change in cytokine levels were measured in the pleural cavity of mice at day 60 between the different peptide vaccination groups and the primary infected. Using capture ELISA, Th2 cytokines IL-4, IL-5, IL10 (Figure 2.17. A, B and C) and Th1 cytokines IFNγ (Figure 2.17.D) were measured. Only changes in IL-4 and IFNγ levels were detected compared to naïve controls, but overall low levels of cytokines were detected in the pleural cavity.

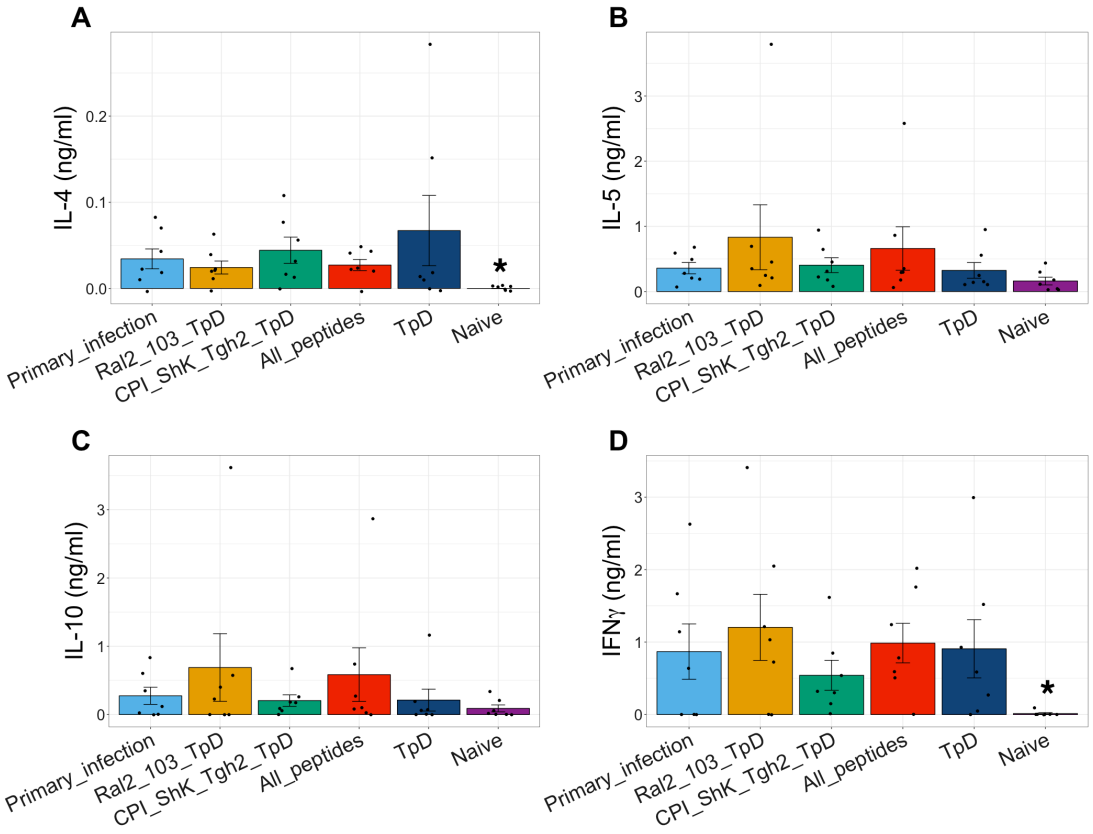

Figure 2.17. Cytokine concentrations found in the pleural cavity at day 60 p.i. At day 60 p.i. mice were sacrificed and their pleural cavity washed to analyse cytokine concentrations of A) IL-4, B) IL-5, C) IL-10 and D) IFNγ concentrations in the pleural cavity. These were measured using capture ELISA, and each dot represents the mean of triplicated samples per individual mice, and each treatment group (N=7) is represented as the mean with error bars representing the standard error of the mean. Asterisks denotes significant difference between primary infection and the other treatment groups (* P-value <0.05). GLM's were used to assess the statistical difference between the treatment groups.

2.3.2.5 No difference in lymph node cell proliferation between vaccinated mice

To investigate the immune responses within the draining lymph nodes of the pleural cavity, the mediastinal and parathymic lymph nodes were harvested at day 60 post infection and used in re-stimulations assays. Lymph node cells from infected mice had increased proliferation compared to the naïve controls in response to re-stimulation

with either *Ls*-Antigen, anti-CD3 (to stimulate T cells) or left unstimulated with media control (Figure 2.18. A, B and C). There was no striking difference in proliferation between the immunised groups and with the primary infection control.

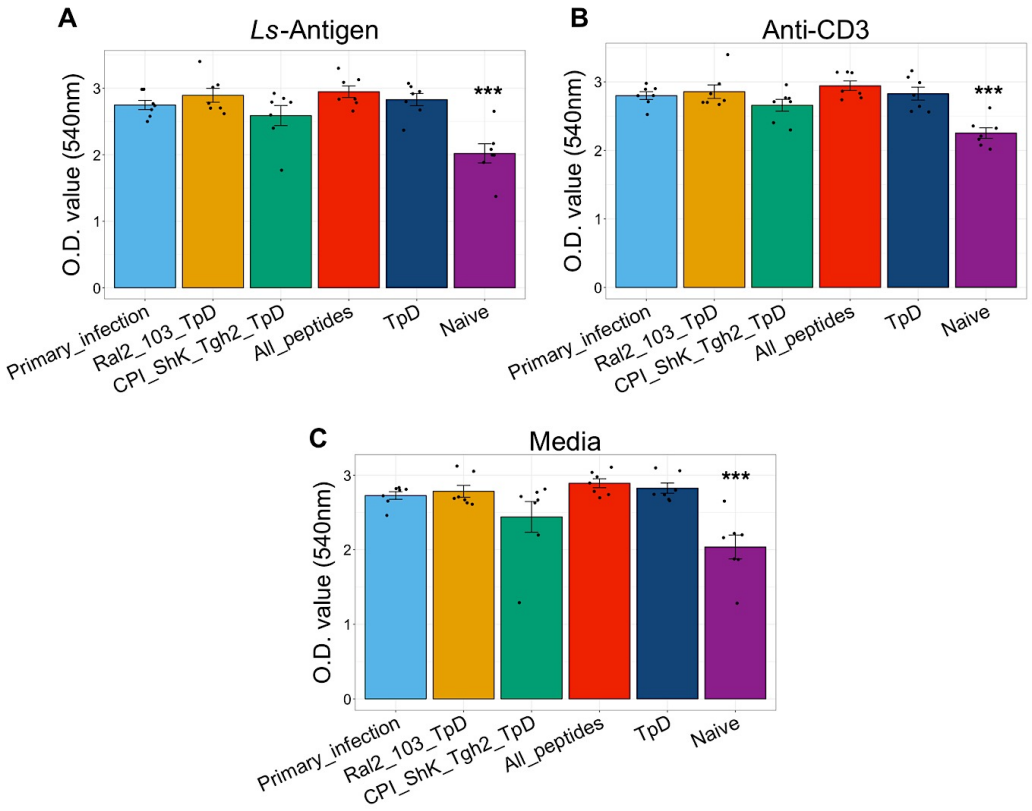

Figure 2.18. Proliferation of re-stimulated lymph node cells. The mediastinal and parathymic lymph nodes were harvested at day 60 post challenge infection, and their cells were re-stimulated with either A) *L. sigmodontis* whole antigen to stimulate proliferation of Ls-Ag specific cells, with B) anti-CD3 to stimulate T cell proliferation and with C) RPMI 1640 as the media control. Proliferation is represented as O.D. values as 540nm, the higher the value the more cells have proliferated. Each dot represents individual mice, split into their vaccination groups (N= 7 mice per group), and error bars represent standard error of the mean. GLM's were used to assess the statistical difference between the treatment groups, where asterisks denote significant difference between primary infection and the other treatment groups (*** P-value <0.001).

2.3.2.6 Decreased IL-4 production by re-stimulated lymph nodes after vaccination with CPIm, ShK, Tgh2 peptides

Levels of cytokines produced by re-stimulated lymph nodes were assessed using a capture ELISA. Overall, lymph nodes re-stimulated with either *Ls*-Ag or anti-CD3 lymph nodes had an increased production of IL-4 (Figure 2.19. A and C) and IL-5 (Figure 2.19. B and D) in infected mice compared to naïve controls. However, induction of IL-4 production was significantly lower in mice vaccinated with the CPI_ShK_Tgh2 peptides, which are known to be immunomodulatory. No change in IL-10 nor IFNγ production was detected between the different treatment groups (Figure 2.20).

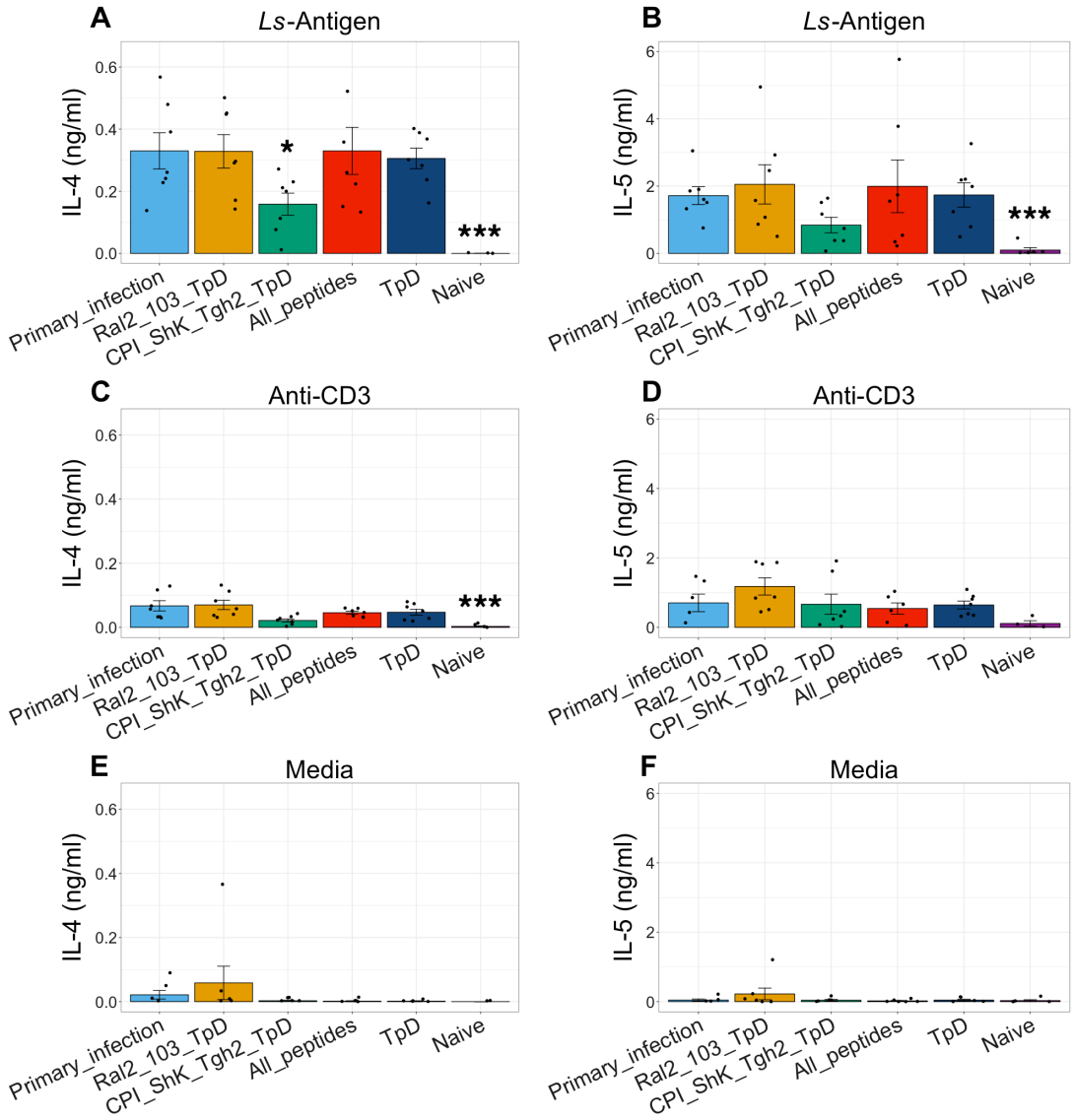

Figure 2.19. IL-4 and IL-5 production by re-stimulated lymph node cells. Cells from lymph nodes harvested at day 60 p.i. were stimulated with either *L. sigmodontis* whole antigen, anti-CD3 or RPMI 1640 as a media control for 72 hours and supernatant fluids were collected to measure cytokine productions by re-stimulated cells using capture ELISA. The graphs show concentrations of IL-4 (ng/ml) after re-stimulation with either A) Ls-Antigen, C) anti-CD3 or E) media; and show concentrations of IL-5 (ng/ml) after re-stimulation with either B) Ls-Antigen, D) anti-CD3 or F) media. Asterisks denote significant difference between primary infection and the other treatment groups (*** P-value <0.001, ** P-value <0.01, * P-value <0.05). Each dot represents individual mice, split into their vaccination groups (N= 7 mice per group), and error bars represent standard error of the mean. GLM's were used to assess the statistical difference between the treatment groups.

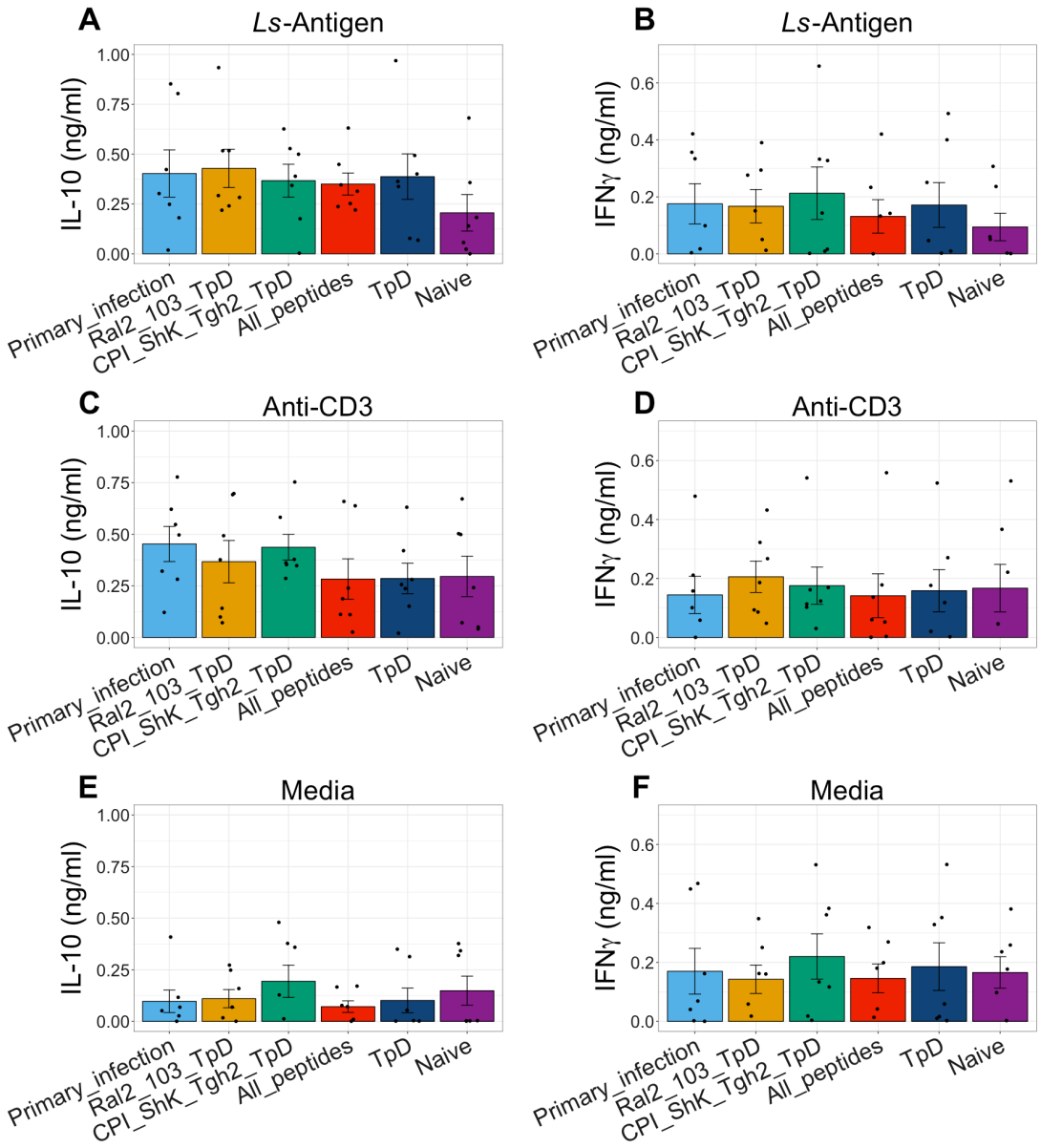

Figure 2.20. IL-10 and IFNγ production by re-stimulated lymph node cells. Cells from lymph nodes harvested at day 60 p.i. were stimulated with either *L. sigmodontis* whole antigen, anti-CD3 or RPMI 1640 as a media control for 72 hours and supernatant fluids were collected to measure cytokine productions by re-stimulated cells using capture ELISA. The graphs show concentrations of IL-10 (ng/ml) after re-stimulation with either A) Ls-Antigen, C) anti-CD3 or E) media; and show concentrations of IFNγ (ng/ml) after re-stimulation with either B) Ls-Antigen, D) anti-CD3 or F) media. Each dot represents individual mice, split into their vaccination groups (N= 7 mice per group), and error bars represent standard error of the mean. GLM's were used to assess the statistical difference between the treatment groups, but no difference between the treatment groups were detected.

2.4 Discussion

Vaccine induced immunity against filarial parasites was shown to be possible using live attenuated vaccines, such as immunisation with irradiated L3 or Mf (Le Goff *et al.*, 2000; Ziewer *et al.*, 2012), these induce antigen-specific responses that protect the host against subsequent infection. Although they provide proof that vaccination is possible for filarial diseases, due to constraints with production and ethical issues, it is unfeasible to use them in humans, therefore the use of DNA plasmids or peptides of filarial antigens as vaccines is proposed.

Several potential vaccine candidates have already been identified and can demonstrate protection and/or immunogenicity in various filarial animal models, and novel ones are continuously being identified. A relatively novel vaccine candidate in filarial research is *Ls*-ShK. It was identified in the E/S (secretome) of *L. sigmodontis* gravid females (Armstrong *et al.*, 2014), and although it was moderately abundant in the female E/S, it was present in the E/S of all mammalian derived stages. Homologues to *Ls*-ShK, such as ones found in *B. malayi* have been hypothesised to be immunomodulators, by modulating memory T cells (Chhabra *et al.*, 2014), therefore *Ls*-ShK stood out as being a potential immunomodulator vaccine candidate.

In initial DNA vaccination experiment, *Ls*-ShK showed promise (Duprez. J, MRes Thesis, 2013, University of Edinburgh) inducing protection seen as reduction in Mf numbers, however in repeat experiments this failed to show any protective response. Since *Ls*-ShK is hypothesised to be an immunomodulator, the lack of protective immunity could be similar to what was seen with the vaccine candidate CPI-2, where

its immunomodulatory properties reduced its efficacy as vaccine candidate (Arumugam *et al.*, 2014a). With CPI-2, when modified to remove its immunomodulatory sequence protective immunity was achieved (Babayan *et al.*, 2012; Arumugam *et al.*, 2014b), therefore in a subsequent vaccination experiments ShK was modified to removed its putative immunomodulatory side, in hopes that this would allow the vaccinated mice immune system to mount a protective response following challenge infection. Overall DNA vaccines (*Ls*-CPIm_TPX and *Ls*-ShK) showed low immunogenicity and protection even with the control immunisation (*Ls*-CPIm_TPX), therefore another approach to vaccination was used.

Peptide vaccines have several advantages over live attenuated, DNA or recombinant protein vaccines. Peptides are relatively inexpensive, due to the ease of production and simple composition, these can also be safer by avoiding the inclusion of unnecessary components possessing high reactogenicity to the host. Therefore, the efficacy of peptides derived from *Ls*-ShK, *Ls*-CPI, *Ls*-103, *Ls*-Ral2 and *Ls*-Tgh2, were investigated using the *L. sigmodontis* model. These peptides were split into two groups; peptides derived from *Ls*-Ral2 and *Ls*-103 were termed highly immunogenic peptides, as these antigens in recombinant form have induced protection against *B. malayi* in gerbils (Arumugam *et al.*, 2016), *O. volvulus* using the chamber model (Hess *et al.*, 2014) and *O. ochengi* in cattle model with natural infections (Makepeace *et al.*, 2009). The second group of peptides were derived from known immunomodulatory proteins: *Ls*-CPI, which has been shown to induce protection, especially when modified (Babayan *et al.*, 2012; Arumugam *et al.*, 2014b; Hess *et al.*, 2014); and two novel candidates *Ls*-ShK and a TGF-β homologue *Ls*-Tgh-2.

Immunisation with different combinations of peptides induced a reduction in Mf number but did not inhibit L3 development, as worm burdens did not differ. This is not so surprising, as immunisation with the antigens CPIm, Ral2, 103 have previously been shown to affect Mf numbers but not worm burdens (Babayan *et al.*, 2012; Hess *et al.*, 2014; Arumugam *et al.*, 2016). Although not significant, there was a reduction in Mf observed in the blood in mice that were immunised with either a combination of *Ls*-Ral3 and *Ls*-103 peptides (highly immunogenic); or *Ls*-CPI, *Ls*-ShK and *Ls*-Tgh2 peptides (immunomodulatory), but not when all peptides were used together. This decrease in Mf was linked to an intrauterine inhibition of embryogenesis, as female worms derived from the mice vaccinated with either the immunogenic or immunomodulatory peptides, either had no viable Mf found in the uteri of the female worms (these correlated with no Mf found in blood) or had low viable Mf density with lots of fertilised eggs or pretzels shaped Mf found in the uteri alongside with the elongated Mf. This vaccine induced reduction in fecundity has been demonstrated several times in the *L. sigmodontis* models (Babayan *et al.*, 2012; Ziewer *et al.*, 2012), *B. malayi* in gerbils (Arumugam *et al.*, 2014b) and with an ES vaccine in *Brugia pahangi* gerbil model (Zipperer *et al.*, 2013).

Immunisations with immunogenic or immunomodulatory groups of peptides showed some signs of protection, however when all peptides were combined, all protective effect was abolished. There are two possible reasons why immunisations with a combination of all peptides may not to have worked; either the combinations of peptides interacted with each other, or because the quantity to peptide used in each

immunisation dose was controlled for, the group with all peptides had less of each individual peptide and not enough to induce an effective response.

Filarial parasites are complex infections triggering different cellular and humoral responses, as well as inducing a suppressive mechanism, with different life stages varying in both antigenic make-up and tissue location, therefore pinpointing effector mechanisms of protection has been difficult and the exact mechanisms remain unknown. Early research using passive transfer experiments of serum from *B. pahangi* infected cats, suggested that serum antibodies could mediate Mf killing (Medeiros *et al.*, 1996). The role of antibodies in *in vivo* killing of Mf, was later confirmed using mice deficient in B cells, where Mf-specific antibodies were indeed necessary to clear *B. malayi* Mf in mice (Gray, Lawrence, 2002). However, in the permissive *L. sigmodontis* model of infection, there are contrasting results depending on the B cell deficiency. Mice with a μMT mutation, which lack mature B cells were amicrofilariamic (Mf-ve) (Martin *et al.*, 2001), whereas mice lacking B1 cells in BALB/c Xid mice had a higher microfilaremia compared to wild types (Al-Qaoud *et al.*, 1998), but great care needs to be taken in interpretation of knockout mice as often mutations can induce important modification. Antibodies have been implicated in antibody-dependent de-granulation of eosinophils in irradiated L3 immunisation in the *L. sigmodontis* model, responsible for the death of incoming L3 following challenge infection (Martin *et al.*, 2001), but since murine eosinophils do not express receptors that bind IgE, the antibody responsible is most likely IgG as murine eosinophils do express IgG receptor and this is sufficient to activate degranulation (de Andres *et al.*, 1997).

In a study by Ziewer et al, where mice were immunised with *L. sigmodontis* Mf, Mf-specific IgG1 and IgG2a antibodies were present throughout the infection (Ziewer *et al*., 2012), in this study it was suggested that protection was mediated by Mf-specific IgG entering the female worm uterus or interfering with developing stages of Mf, hence why adult female where intact despite low Mf numbers found circulating in the blood (Ziewer *et al.*, 2012). Although it is impossible to say that this is what is happening with the peptides immunisation, since no significant protection was achieved with any of the vaccines combinations, there are some similarities.

Using flow cytometry the number of B cells were found to be similar between the immunised groups and the primary infection. Further, when measuring adult *L. sigmodontis* specific antibodies, IgG1 was found elevated in the mice vaccinated with the *Ls*-Ral2 and *Ls*-103 peptides (these mice had low densities of elongated Mf). This difference in IgG1 was only detected at day 60 (patency) and no differences between the infected mice whether immunised or not were seen at day 14 and day 42 post-challenge suggesting a cumulative increase. However, in the future it would be necessary to measure Mf-specific antibodies, as well as using immunohistochemistry, to establish if antibodies are binding to the embryonic stages.

Cytokine profiles during infections can give an indication of the type of immune responses being triggered. Th-2 cytokines, IL-4 and IL-5 in *L. sigmodontis* infections are necessary for parasite containment (Al-Qaoud *et al.*, 2000; Volkmann *et al.*, 2001; Le Goff *et al.*, 2002; Volkmann *et al.*, 2003), as well as the Th-1 cytokine IFN-γ (Lawrence *et al.*, 2000; Saeftel *et al.*, 2003). Whereas IL-10, which was initially described as a Th-2 associated cytokine, was found to be produced by regulatory T

cells and is now mostly associated with immunomodulation and parasite survival (Specht *et al.*, 2004; Haben *et al.*, 2013; Hartmann *et al.*, 2015). Therefore, levels of IL-4, IL-5, IFN-γ and IL-10 were measured in the lavages of the pleural cavity, but no changes were detected between the different immunisation groups.

Cytokines produced by lymph node cells re-stimulated with either *L. sigmodontis* antigen, anti-CD3 to stimulate T cells or culture media were also measured, with only cells from mice immunised with immunomodulatory peptides showing a change in IL-4 production, producing lower levels of IL-4 compared to other immunised groups and the primary infection group. Although cytokines and cells present in the pleural cavity and lymphoid system were collected at the time of necropsy, earlier time points would be more informative, during the migration of larval parasites which are the likely major targets of the protective immune response and that by day 60 immune responses have returned to baseline levels.

The evidence so far shows that the peptide vaccines are targeting female fertility and not the circulating Mf, and in onchocerciasis, dermal and ocular pathology is associated with innate immune responses to *Wolbachia* released from dying Mf in the skin and eyes (Hise *et al.*, 2003; Tamarozzi *et al.*, 2011). Therefore, a vaccine against the Mf stage would be beneficial to the affected population, not only to decrease transmission but also to ameliorate pathology. However, because immunity to filarial parasites is mostly associated with Th2-type responses, there may be a risk that a vaccine might induce strong eosinophil and IgE mediated responses which could lead to induction of pathology (Babu, Nutman, 2012). Therefore, vaccine candidates tested in mice would need to be carefully monitored during and future clinical trials (Diemert

et al., 2012; Hotez *et al.*, 2013), especially since in the *L. sigmodontis* model mice do not exhibit pathology.

Hypersensitivity (allergic reactions) following vaccination are a cause for concern. Phase I trials of the hookworm (*Necator americanus*) vaccine, against the ancylostoma-secreted protein 2 (*Na*-ASP-2), resulted in generalised urticarial (allergic) reactions in volunteers and the suspension of the trial. This reaction was associated with pre-existing *Na*-ASP-2 specific IgE, most likely from previous hookworm infections, as negligible levels of *Na*-ASP-2 IgE were found in hookworm naïve-adults living in non-endemic areas (Diemert *et al.*, 2012). Surveys in hookworm endemic foci in Brazil showed that a significant percentage of the population had increased levels of IgE to *Na*-ASP-2. This has caused serious concern and has implications for the development of vaccines against helminths (Diemert *et al.*, 2012).Therefore, one way to prevent this from happening in onchocerciasis vaccine trials, is firstly to test the potential vaccine antigens against serum from individuals living in endemic foci, to make sure the antigens are not recognised by IgE antibodies. Since one of the strategies to identify vaccine candidate has been to target immunomodulators, this decreases the risk of having IgE specific antibodies prior to vaccination.

The use of a peptide vaccine is an attractive vaccine strategy against filarial disease, firstly because of the ease of production and secondly because specific antigen epitopes can be chosen and easily modified, so that they do not elicit to IgE. However, because they usually only contain one antigenic epitope, these would need to be combined and different combinations would need to be investigated, such as

combining an immunogenic peptide with an immunomodulatory peptide (Babayan *et al.*, 2012). The duration of protection elicited by any vaccine would also need to investigated, as this was modelled to have a long-term impact in vaccination campaigns (Turner *et al.*, 2015).

Protection in vaccination experiments are measured around day 60 post challenge infection, as this is the point at which adult worms are known to be present and Mf are being produced. Although immunological readouts can be informative of what type of immune responses are being triggered, at day 60 immunological readouts may not be always indicative of protection. Quantitative RT-PCR arrays were used to measure changes in gene expression in vaccinated mice, to determine if gene signatures at the end of vaccination experiments could be more informative than antibody, cytokine and cellular measures. However, most gene signatures identified were associated with a response to parasites burden and not predictive of protection.

Review of the data obtained and published results of other studies suggest that investigative associations at day 60 may be too late to measure changes associated with vaccination, and events at the onset of infection may determine infection outcome (Taylor *et al.*, 2009). Identifying gene signatures or biomarkers predictive of vaccine efficacy or immunogenicity would greatly aid vaccine trials, by reducing costs if a vaccine shows signs of not being protective early on (Mastelic *et al.*, 2013; Hagan *et al.*, 2015). Therefore, one would need to look at changes in gene signature throughout a vaccination time course to pinpoint which time point is the most predictive of its efficacy.

Summary

- Immunisation with Ls-ShK using DNA vaccines in the *L. sigmodontis* model does not confer protection.

- Immunisation with peptides derived from immunogenic antigens (*Ls*-Ral2 and *Ls*-103) and from immunomodulatory antigens (*Ls*-CPI, *Ls*-ShK and *Ls*-Tgh2), affects adult female fecundity.

- Immunisations with a combination of peptides derived from all vaccine antigens shows no protection.

- Peptides have potential as vaccines but correct formulation needs to be determined.

- Measuring gene signatures potentially associated with protection at the end of a vaccination time course is too late in the infection time course.

Chapter 3. Using machine learning techniques to dissect gene expression patterns in filarial infection from whole blood data

3.1 Background

The possibility of vaccination against filarial infections such as onchocerciasis is based on the evidence that protective immunity is observed both in humans and animals naturally exposed to filarial infections, and the ability to induce partial protection in multiple laboratory models of filarial infection. In humans, protective immunity is seen in a small proportion of the population living in hyper-endemic areas which show no clinical or parasitological signs of infection even though they have had a life-long exposure (Hoerauf, Brattig, 2002).

In the *Litomosoides sigmodontis* mouse model, such apparent natural immunity/resistance is host strain-dependent, e.g. BALB/c mice are susceptible to the full development of adults and production of blood-circulating microfilariae (Mf), while C75BL/6 mice, termed resistant, eliminate the parasite within 40 days of infection, i.e. before the onset of patency (Petit *et al.*, 1992; Marechal *et al.*, 1996; Babayan *et al.*, 2003). Furthermore, in BALB/c mice, protective immunity can be induced by vaccination with the infectious stage (L3) larvae or the transmissible filarial offspring (microfilariae, Mf) (Le Goff *et al.*, 2000; Ziewer *et al.*, 2012). Protection induced by vaccination in the *L. sigmodontis* models is determined as a decrease in parasite burden, either in adult worm number seen in the pleural cavity, or in the number of circulating microfilariae (Mf), the transmission stage of the parasite, in the blood.

Identifying markers of early immune responses to immunisation that are predictive of vaccine efficacy, would provide a surrogate endpoint to vaccine trials, but could also ultimately help predict the efficacy of a vaccine in humans. Because mice differ from humans, with respect to the development and activation of innate and adaptive immune response (Mestas, Hughes, 2004; Seok *et al.*, 2013), successful translation of any vaccine from murine models to clinical application is not always straightforward, and vaccines that showed promise in murine models, have had a lack of efficacy in humans (Gray *et al.*, 2011; Tameris *et al.*, 2013; Kaufmann *et al.*, 2014) or sometimes raise safety concerns during early human clinical trials (Diemert *et al.*, 2012). However, the use of an animal model cannot be bypassed for evaluating vaccines, and based on common ancestry and relatively high conservation of genes and their expression profiles (Mestas, Hughes, 2004), immune mechanisms that take place *in vivo* following vaccination with L3 or Mf in mice could help identify the "immune signatures" that are necessary to trigger protection, and see if there is overlap with human makers of natural protection.

Until recently studying these immune responses to infection would have been performed in a highly focused manner, by either investigating antibody responses, abundance of certain cells types, or cytokines. These methods have provided tremendous insight as to how and what type of the immune responses are being triggered, however the immune system is complex with considerable in-built redundancy, and consequently a focused approach can be limited when it comes to analysing interacting features of a complex system (Furman, Davis, 2015).

The introduction of high-throughput technology such as microarrays, has opened a broader approach to the analysis of immune responses, in which expression levels of tens of thousands of genes can be measured simultaneously (Weiner *et al.*, 2015). One of the aims of the work described in this chapter was to investigate whether markers (predictors) of immunity can be detected in the blood, and whether commonalities could be identified between mouse and human expression profiles during filarial infection. Whole blood was chosen as the sample tissue, as not only is it the least invasive sampling method (compared to tissue biopsies) and often the only practical option, but also filarial infections elicit a systemic response, which is best captured by analysing the blood during an infection.

The use of microarray technology for studying helminth infection is relatively uncommon compared to other parasites and pathogens (Kwarteng, Ahuno, 2016), with most studies investigating the host responses to helminths such as schistosomiases and *Nippostrongylus brasiliensis* (a helminth closely related to human hookworms) (Zhou *et al.*, 2016), with even fewer studies on filarial nematodes (Kwarteng, Ahuno, 2016). In most of the helminth studies, changes in gene expression were measured in tissue samples (i.e. not readily applicable as a non-invasive approach) or isolated responses in filtered cell populations, instead of whole blood. However, in other pathogen infections such as bacteria (Schoolnik, 2002), viruses (Gaucher *et al.*, 2008; Slobedman, Cheung, 2008; Querec *et al.*, 2008) and malaria (Vahey *et al.*, 2010), microarray analysis of whole blood samples has proved useful in identifying disease or vaccine associated biomarkers and underlying immune responses. Therefore, whole blood is a feasible sample to look at for changes in gene expression in filarial

infections. There are some drawbacks to using blood, as RNA isolated from whole blood has been associated with increased noise and reduced gene expression levels, but it does prevent problematic artefacts caused by cell separation which activates cells, and therefore biases the results (Feezor *et al.*, 2004).

Another issue is that microarrays leads to very high dimensional datasets (i.e. many variables are measured), often compounded by a relatively low number of samples, which are sometimes allocated to multiple treatment groups (i.e. non-immunised, control immunisation vs antigen immunised), a problem termed the "curse of dimensionality". Thus, to overcome the low signal to noise ratio and the curse of dimensionality, sophisticated analytical approaches are needed.

Classical statistical methods, typically compare fold changes in gene expression levels between genes in treated/infected samples compared to control samples, and a threshold level based on the test statistic (P-value) is used to identify genes with a significant fold change (differentially expressed genes). This approach is simple, fast, and easy to interpret, but faces several constraints when it comes to analysing high dimensional data (Table 3.1) (Butte, 2002).

In high dimensional datasets, the expression of thousands of genes are measured simultaneously, this therefore becomes a multiple testing/hypothesis problem, because of this, simply selecting a significance of $P < 0.05$, can result in many false positive discoveries. Several solutions have been proposed, but by controlling for the false discovery rate, it can cause many false negatives, and many genes that might be important can be overlooked. Therefore, the issue with this approach is that it can only

identify genes that provide a significant amount information in isolation to other genes. In reality, the regulation and activation of most functional pathways and networks is achieved through small incremental changes. The lack of sensitivity of the "fold change" approach combined with considerable variation found in blood derived samples reduces the possibility of identifying significant pathways and discriminating gene expression profiles (Huynh-Thu *et al.*, 2012).

An alternative approach such as Weighted Gene Correlation Network Analysis (WGCNA) (Langfelder, Horvath, 2008), recognises that gene expression data is more complex than a list of differentially expressed genes, and instead considers the correlation of genes expression levels across samples. This method uses gene correlation networks, which finds highly correlated genes (genes with similar expression levels) and groups them into large clusters (co-expression clusters). The gene expression patterns can be summarised within a cluster to a "module eigengene" (ME), this is a weighed summary of gene expression within a cluster (similar to a principal component). If a cluster ME has a particular behaviour, then it is likely most of the genes in that particular cluster also have a similar behaviour. This ME is then used to measure associations between the clustered genes and the sample traits.

Although this method has been used extensively and provided insight into pathogenesis of autoimmune diseases (Sundarrajan, Arumugam, 2016), neurological disorders (Mina, 2016) and cancer (Yepes *et al.*, 2016), it cannot be used to investigate the relationship of gene expression in multi-categorical data (such as multiple experimental groups) or handle large datasets, with most published studies restricted

to the analysis of the top 1000-5000 most varying genes or differentially expressed genes.

To obtain maximal information from the high dimensional (greater number of genes measured compared to samples), multi-categorical (multiple experimental groups) microarray data, which contains complex interaction between genes, biomedical research is now increasingly turning to machine learning to identify "informative" genes (genes important in infection/vaccination) (Molla *et al.*, 2004). In this study a series of publicly available machine learning methods were put together in a "pipeline" to identify genes associated with protection either in vaccinated mice or humans living in endemic areas. There is no gold-standard methodology for comparative studies in machine learning. However, WGCNA, is a popular approach in biomedical sciences (Langfelder, Horvath, 2008). Therefore, WGCNA was conducted on a subset of the data and results were used to validate results produced by the pipeline.

Table 3.1. Microarray analysis methods.

Methods	Advantages	Disadvantages
Differential fold change	Simple, fast, and their output is easy to interpret.	Accounting for multiple-testing reduces the sensitivity of the test when many genes are measured simultaneously (such as in high dimensional data) and struggles with multi-categorical data, therefore, this potentially misses out informative genes that have a small change in expression.
Gene correlation networks	More robust with complex data set and at finding important genes within noisy data.	Cannot deal with high dimensional data sets and struggles with multi-categorical data.
Machine learning	Good at analysing high dimensional, multi-labelled microarray data, which contains complex interaction between genes.	Can sometimes be complicated to implement and hard to interpret.

Chapter aims:

- To determine if changes in gene expression can be detected in whole blood, collected from mice vaccinated with infective L3 or Mf from *L. sigmodontis* using a machine learning pipeline.

- To determine if changes in gene expression associated with protection in humans exposed to *O. volvulus* can be measured in whole blood using a machine learning pipeline.

- To determine if mechanisms that are important for protection in mice and humans are the same.

3.2 Introduction to machine learning concepts

Machine learning is a branch of computer science that utilises algorithms that "learn" from data to build models capable of predicting properties of unknown or new data. These can be split into two broad categories, supervised and unsupervised methods (Libbrecht, Noble, 2015) (Figure 3.1).

- **Supervised methods** are trained on labelled data, with labels equivalent to response variables (i.e. experimental group the samples belong to), and then used to make predictions on unlabelled data (i.e. new or unknown data). For example, when using genomic data, a supervised learning algorithm may build a model on a subset of the gene expression data for which sample information is known (i.e. experimental group to which it belongs, or whether a patient is or is not infected); this is known as the "training" dataset. Then the "trained" model is subsequently used to predict the characteristics of the remaining data.

- **Unsupervised methods** find internal structure or relationships within the data without any prior knowledge of the sample labels. These methods do not try to predict the sample label.

Another categorisation of machine leaning methods which mostly applies to the supervised method is based on the type of output data. When the sample labels are categorical (i.e. different experimental groups, immunised/non immunised) this becomes a *classification* problem, however when sample labels are continuous (i.e. parasite counts) this is then *regression*. Some approaches are amenable to both classification and regression, while other algorithms can only perform one of the two.

To find the genes that best differentiate between the samples, depending on their experimental group (immunised/non-immunised control) or susceptibility to infection in humans (infected/endemic normal), several supervised algorithms exist, each with its strengths and weakness (Table 3.2)(Pirooznia *et al.*, 2008; Bolón-Canedo *et al.*, 2014).

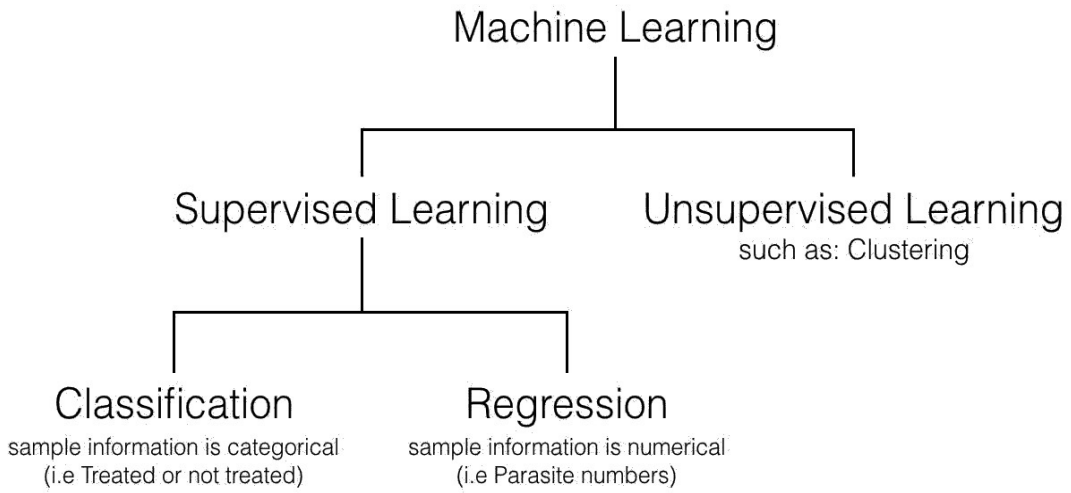

Figure 3.1. Overview of machine learning categorisation. Machine learning can be broadly split into two categories *supervised* and *unsupervised* algorithms. Supervised algorithms build models on samples were the response variable is known, and are often used to predict the response of new data. Whereas, unsupervised algorithms find internal structure or relationships within the data without any prior knowledge on the sample labels. Supervised algorithms can be further classified into *classification* or *regression* algorithms, although some supervised algorithms can do both regression and classification. Classification is done, when the responses variable in categorical, and regression is used when the response variable is continuous.

Table 3.2. Common classification machine learning methods, used to identify differentially expressed genes in microarray data.

Methods	Advantages	Disadvantages
Naïve Bayes (a type of Bayesian classifier)	Fast and easy to implement, allows for missing values; only requires a small training dataset to estimate the parameters.	It assumes variables are independent of each other; and highly dimensional data increases the computational cost.
Decision Trees	Can handle heterogeneous data; works well with complex interactions among variables (genes), easily interpretable results, that are ranked in order of importance, can handle missing data.	Algorithms are unstable, with high bias and therefore prone to overfitting.
Random Forest	Can handle heterogeneous data, works well with complex interactions among variables (genes), easily interpretable results, that are ranked in order of importance, performs well with high dimensional data, resistant to overfitting.	Does not allow missing data, loses stability when correlated variables are present.
Support Vector Machines	Robust to noise, performs well with small sample size, fast to compute.	Input variables need to be numerical and scaled, works better when there are no interactions between variables, sensitive to parameter choice, output of model is not easily interpretable.
K – nearest neighbours	Simple model to build, you do not need to have any prior knowledge of the data.	Severely affected by noisy or irrelevant variables, sensitive to parameter choice, there are large variations in prediction performance every time the model is run, it is computationally costly to run.
Artificial Neural Network	Can be used to model complex relationships between inputs and outputs and find patterns in data, can handle noisy data.	Input variables need to be numerical and scaled; output of model is not easily interpretable and can be very slow to train.

3.2.1 Generalisation error

Machine learning works by building a model on a subset of the data available, this is called the "training" dataset, and the remaining data is used to evaluate the model, and this is termed the "testing" dataset. One of the main considerations when building the "learning" algorithm is how generalisable the model is, in other words how good is the learned model at predicting unknown data (Haury *et al*., 2011).

- **Generalisation error** – is a measure of how accurately a model can predict the labels (response variable) of new data or previously unseen data, such as which treatment group the samples belong to.

For example, if gene expression is to be used to predict the vaccine efficacy in patients, a model would be built on a training dataset, in which the vaccine efficacy in patients is known (i.e. whether the vaccine successfully induced protection). The model is then used to predict vaccine efficacy in new patients (the testing set), and the difference between what the model predicts and reality, is the generalisation error. Therefore, a model with low generalisation error would be good at predicting vaccine efficacy in new patients.

Having a low generalisation error in classification means that the model is good at classifying new samples into their correct experimental group, it is simply a measure of how good the model is, and this can be used to compare different machine learning methods with one another.

3.2.2 Variance and bias trade-off

The generalisation error contains two sources of error, bias and variance. There is an inherent trade-off between bias and variance, and they prevent supervised learning methods to be used beyond their original training dataset.

- **Bias** - The error between what the model predicts and what the outcome is, therefore high bias, suggests that the model is too simplistic and is missing the relevant relationships between variables (genes) and sample outcomes (vaccine efficacy).

- **Variance** – Applies to the stability of the model in the face of "noise". If the model is built again on a different subset of the data, would the model produce the same outcome? If the model has high variance, then the model is fitting (being built) on the noise in the data and not finding the underlying relationship between the variables (genes), this is termed overfitting.

Since there is a trade-off between variance and bias, it is not always easy to minimise both simultaneously, however there are certain concepts / techniques that are utilised by machine learning methods, to minimise these errors.

3.2.3 Techniques used to improve bias and variance errors

There are some techniques that can be used to improve bias and variance, such as:

- **Dimensionality reduction** – This reduces the number of variables (genes) the model is built on, increasing the stability of the model by simplifying it (Nguyen

et al., 2015). If the number of variables (which in machine learning terminology are called features, and in context of this study are genes), are much larger than the number of samples this gives a highly dimensional dataset, and some models are better at handling this than others. If the dataset is too highly dimensional, a step can be added to identify variables of importance and discard the unwanted variables, which creates a lower dimensional space on which to run the supervised learning algorithm. Alternatively, variables can be grouped together and transformed or summarised into a new feature, therefore reducing the total number of variables (genes).

- **Ensemble of learning algorithms** – This consists of using the same learning algorithm multiple times on the same dataset or subset thereof, the principle is to combine a group of "weak learners" to form a "strong learner" (Dietterich, 2000). There are different ways of doing this, with the most common methods being:

    I. *Bagging* (portmanteau of "bootstrap aggregating"), this is when a machine learning model is built on a random subset of the data, drawn with replacement from the original dataset. This is repeated multiple times, with each time a new subset of data drawn. This produces a series of weak models that are combined to boost performance. Bagging increases the stability (variance) of the model, and is used by Random Forest.

II. *Boosting,* in the current context, uses a series of a "weak" learners, or model, which are weighted according to their performance and combined into a strong learner to boost their performance. The difference with bagging is that the subset selection is not random and depends on the performance of the previous models. Boosting tends to improve the accuracy (bias) of the model, but it is prone to overfitting[1].

- **Inbuilt parameters** – Learning algorithms also have tuneable parameters that control bias and variance

3.2.4 Characteristics of the data to consider

Different machine learning methods exist, and some are better suited to certain datasets, therefore there are several considerations to make before choosing an algorithm, such as:

- Does the data contain lots of noise (irrelevant genes)? – In the case of this study, whole blood was used to measure changes in gene expression data, and may contain a lot of noise, therefore choosing an algorithm with high accuracy and low stability is preferred.

- Does the data contain variables of the same type (count / continuous / discrete data)? – Some algorithms will work better if only one data type is present,

[1] A model which describes random error or noise, instead of the underlying relationship.

whereas some have no issue using different types. In this study, only gene expression data will be investigated so this problem does not arise, although it could become an issue if non-continuous variables/factors are added to the analysis.

- Are there interactions between the variables or are they independent? – Genes are known to interact and regulate each other, and therefore have complex interactions, therefore algorithms such as decision trees, random forest and neural networks would be the preferred analysis.

3.3 Methods in Machine Learning commonly used for microarray data analysis

For microarray data, machine learning techniques such as Bayesian classifiers, Decision Trees, Random Forest (RF), Support Vector Machines (SVM) and Artificial Neural Network (ANN) have extensively been compared to one another (Pirooznia *et al.*, 2008; Swan *et al.*, 2013; Hemphill *et al.*, 2014; Karimpour-Fard *et al.*, 2015). When considering which algorithm to use the issues mentioned above need to be considered, as microarrays are notoriously highly dimensional (large number of genes), small numbers of samples, contain uninformative genes with regards to the processes of interest, and genes not only interact with each other but their expression is often correlated.

A few different machine learning algorithms have been applied to microarray data, with the most popular being SVM and RF, which have been used or modified for a wide variety of tasks from identifying cancer subtypes (Anaissi *et al.*, 2013) to predicting patients responses to treatments (Gim *et al.*, 2016).

3.3.1 Support Vector Machine

Support Vector Machines (SVM) is a machine learning method used for both *classification* and *regression* tasks, and was first conceived by Cortes and Vapnik (Cortes, Vapnik, 1995). The simplest type of SVM, tries to draw a straight line (hyperplane) that best separates the data, so that data points belonging to one group/category are on one side of the line, and the data points in another category are on the other side (Figure 3.2). SVM assigns a weight to the variables depending on

how close they are to the line drawn, and these weights can be used to compute a variable ranking score.

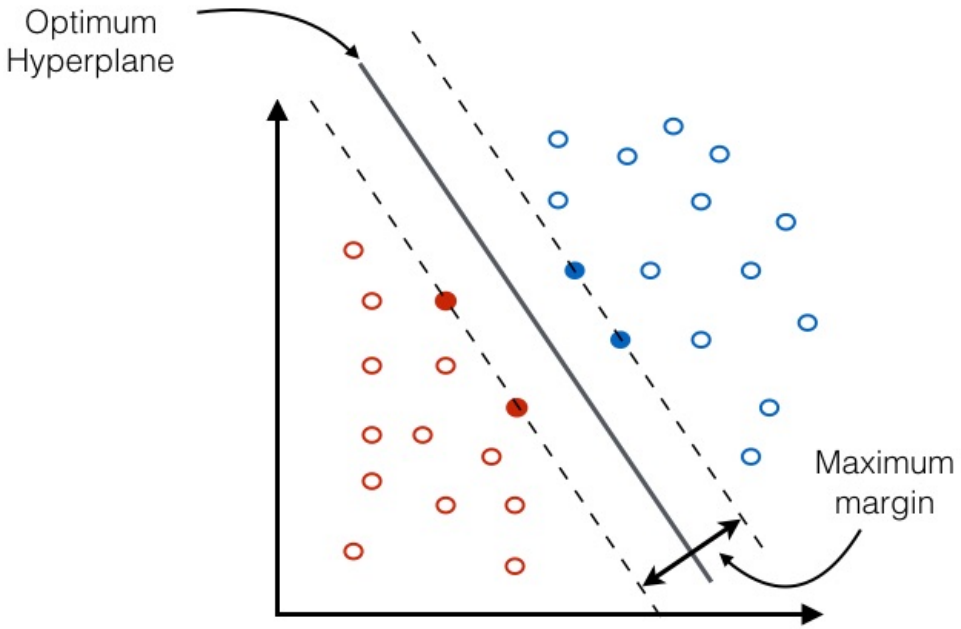

Figure 3.2. Schematic representation of Support Vector Machine in two dimensions. SVM is a supervised learning methods that draws a line (line in 2D, plane in 3D and hyperplane in higher dimensions) that separate two classes (red / blue data points). In theory, infinite numbers of lines could be drawn to separate the data, but SVM calculates the straight line that best separates the two classes of points with largest distance to the nearest data points (maximum margin). In other words, SVM draws an optimal line perfectly in the middle of two classes, so that points in the red class are on one side of the line and the points in the blue class are on the other side of the line. When the two classes of the data are not linearly separable, the points are projected into a higher dimensional space where linear separation may be possible. Support vectors are the data points that lie closest to the line (filled red and blue circles). Because these are the most difficult points to classify, these are important for classification, and each point is assigned a weight depending on its distance with the hyperplane.

3.3.1.1 SVM with Recursive Feature Elimination

Guyon et al modified SVM, by using it in a backward elimination procedure for gene selection (Guyon *et al.*, 2002). For this, a SVM is initially fitted on all the genes in the dataset, this assigns a weight to the variables (genes) which is used to compute a gene coefficient, that can be ranked so that genes with the smallest coefficients are eliminated. Then a new SVM is built on the remaining genes, and a new gene coefficient is calculated and again the genes with smallest coefficients are removed, this is repeated until a final number of genes are met. The number of genes to remove at each iteration, and the final number of gene to retain, are the set of parameters that can be optimised to get an accurate final classifier.

Removing variables (genes) each time a SVM model is built is termed *recursive feature elimination* (RFE). Support vector machine with recursive feature elimination (SVM-RFE) was initially created for cancer classification using gene expression data (Guyon *et al.*, 2002), and is continuously being improved, such as using an *ensemble* approach (multiple SVM-RFE built), which uses subsamples of the original data to create multiple SVM-RFE (Duan *et al.*, 2005).

A limitation of SVM-RFE is that it can only handle binary classification. Therefore, to extend the application of SVM-RFE to multi-class (i.e. multiple treatment groups) problems Zhang et al (2015) proposed that a multi-class dataset could be split into multiple binary problems (one versus all method). These models were applied to different DNA microarray datasets and achieved higher classification accuracy

compared to other SVM-RFE methods (Zhang, Xiaojuan Huang, 2015). However, like most of these algorithms they are not easily incorporated into statistical software.

3.3.2 Random Forest

Random forest (RF) is a robust classification algorithms (Breiman, 2001), and has many characteristics that makes it well suited for gene expression data (Chen, Ishwaran, 2012):

- it is well adapted to handle high dimensional data that can often be noisy,
- it can accommodate categorical (classification) and continuous (regression) data,
- it handles imbalanced multiclass data,
- it is less prone to overfitting compared to other machine learning techniques,
- it has a fewer parameters to fine-tune making it easier to optimise,
- it provides measures of feature (gene) importance.

For classification problems, RF has been extensively compared to other machine learning algorithms on microarray datasets, and has been found to be the most efficient method on the datasets that were tested (Lee *et al.*, 2005; Díaz-Uriarte, Alvarez de Andrés, 2006; McKinney *et al.*, 2006; Heidema *et al.*, 2006; Jeong, Soni, 2015).

RF is an ensemble of classification trees (decision trees), which uses different subsets of a dataset to generate many decision trees and then combines the results. Using the

bagging[2] technique, each tree is independently built using a bootstrap sample of the data (subsample of the dataset which is replaced after being used to fit a tree), each tree is recorded and combined at the end.

The goal of a decision tree, is to create a model that can predict the value of unknown variables (classify new samples), based on the input variables. For example, can the experimental group of samples be predicted based on their gene expression?

Decision trees start with a parent node, these are then split into two daughter nodes in a recursive way (i.e. the daughter node in turn becomes parent node, and is subsequently split into two daughter nodes) (Figure 3.3). At the parent node, the decision tree algorithm chooses the variables at that level, that best separates the data into "pure" classes, i.e. the gene that best separates the samples into either immunised or non-immunised mice. For example, if gene *A* is found by the algorithm to be the most discriminative gene, this means that if gene *A* has an expression over defined level (*x*), those samples are associated with immunised mice, whereas an expression level below *x*, that sample will be more likely to be from non-immunised mice.

At each parent node, the algorithm searches for a variable (present at that node) that best separates the data, this process is applied to each parent node, and as the tree grows less and less variables (gene) are present, till either a homogenous tree (where samples

[2] Bagging – Bootstrap aggregating, when classifiers are built on different subsets of the dataset, each subset is replaced after classifier is built, by sampling with replacement variables will be repeated in several subsets.

cannot be split anymore as they all belong to one treatment group) or non-homogenous terminal (the end of the tree is reached and no more genes are left to split) (Figure 3.3).

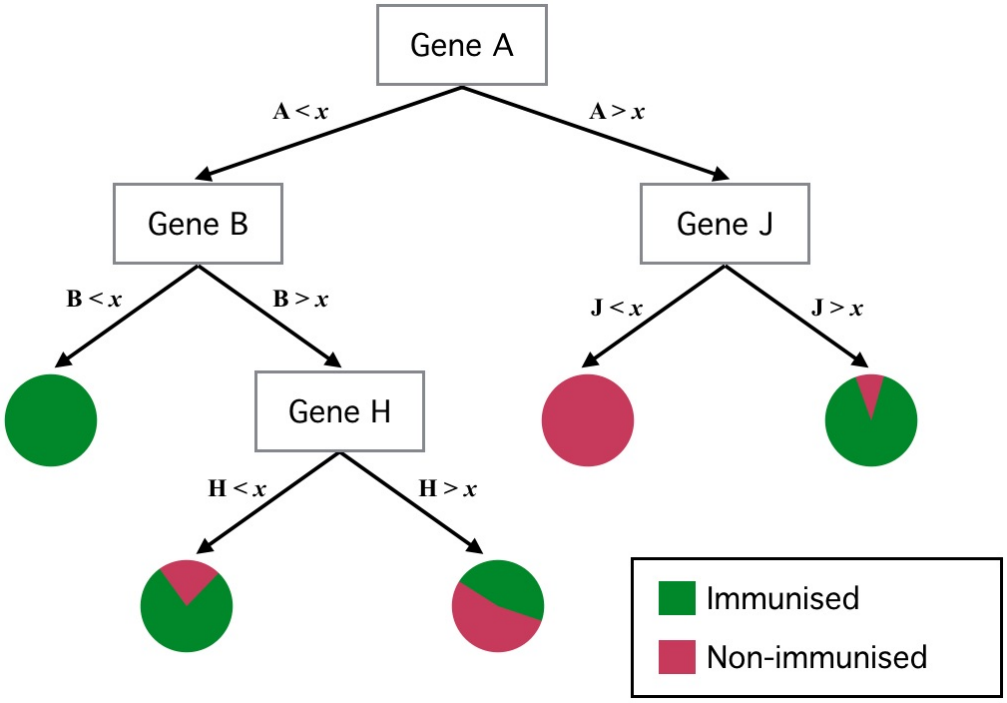

Figure 3.3. Decision Tree. The diagram represents an example of a decision tree using gene expression data. Decision trees work by finding genes that best discriminate between the experimental groups, in the example above, immunised or non-immunised. At each node of the tree (rectangles), the algorithms finds the gene which is the best at splitting samples into experimental treatment groups. The node at which the split is happening is termed the parent node, at the this point the data separated into two sister nodes depending on the expression level. For example, "gene A" is chosen by the algorithm to be best gene at splitting the data into experimental groups, and splits the samples into two daughter nodes depending on expression level of "gene A", i.e. samples, in which "gene A" has an expression level above splitting criteria (x) will go to one sister node, and the samples with expression levels of "gene A" below the splitting criteria will go to the other sister node. These sister nodes then in turn become parent nodes, and the algorithm then finds another gene that best splits the data, this happens recursively until the tree is fully grown, either till perfect classification is reach, so when all the samples in that node belong to one treatment group (full circles) or till there is no more gene to split the data (semi-circles and represent the percentage of predicted treatment). Figure adapted from (Ainali, 2013)

3.3.2.1 Optimising random forest parameters

Random forest has two main parameters to optimise:

- **The number of trees to construct** – RF uses bootstrap aggregation (sampling with replacement) to sub-sample the dataset and trees are grown from those sub-samples. This allows RF to construct multiple trees on different samples of the dataset, therefore the optimal number of trees to grow (subsamples to take from the dataset) needs to be optimised, as this can affect the model accuracy.

- **The number of variables to consider when splitting a node** – RF selects the variable that best splits the node, from a random subset of variables, compared to other decision trees that look at all variables when considering the split. Therefore, the number of random variables to consider when looking for the best split is the second parameter that needs to be optimised. This allows random forest to build many trees with low correlation (unlikely to comprise identical trees). For example, if one or a few genes are strong predictors of the response variable, these gene will be selected in many trees, causing many of the trees in a random forest to look alike (correlated), and this will affect the accuracy and stability of the model.

To achieve a RF model with the lowest generalisation error, the two parameters mentioned above need to be optimised. To do this, RF has an internal estimation of model accuracy, which is called the out-of-bag (OOB) error score (Mitchell, 2011).

- **The OOB score** – is calculated for each tree grown, by setting aside one third of the samples before building the tree, this becomes the OOB data. The OOB data is then used to measure the accuracy of the tree built. The OOB score is calculated for each tree constructed, and then averaged to become the OOB error score for the random forest.

To optimise the RF model, a procedure called *k*-fold cross-validation is used. This builds a random forest on a percentage of the data using different combinations of parameters, and for each combination the model is evaluating itself using the samples that were left out when building the model, this avoids the risk of overfitting. The best practice is to use k=10 (10-fold cross-validation), this means that 10 RF models are built on each combination of parameters, the OOB error score for the 10 RF models are calculated and then averaged out. This is repeated for each combination of parameters, and the combination of parameters that gives the lowest OOB score are then used to train a final RF classifier.

3.3.2.2 RF ability to rank variables in order of importance

Advantages of RF over other supervised methods is that it provides a measure of variable importance (gene importance in our case). For classification problems, RF computes a measure of variable importance, by calculating a mean decrease in Gini for each variable, which can then be ranked, to find which variables (genes) are the strongest predictors of the response variable.

- **Mean decrease in Gini** – A measure of how each variable (gene) contributes to the homogeneity of the nodes (if nodes contain the majority of the samples of one experimental group).

For example, if a variable is important, it will split the data, which has samples belonging to different experimental groups, into pure samples groups (all coming from the same experimental group). Therefore, each time a variable (gene) is used to split a node, the Gini coefficient for the child node is calculated (based on how pure the node is) and compared to that of the parent node, giving a decrease in Gini value for that variable.

Since random forest builds many trees and variables can be used in more than one tree, the decrease in Gini coefficient for each variable is calculated across all trees and averaged out, giving the mean decrease in Gini for that variable. The greater the mean decrease in Gini is, the more important the variable (Breiman, 2001).

3.3.2.3 RF issues with stability

Because of the intrinsic randomness of random forest, there can be instability in the variable importance ranking (Somorjai *et al.*, 2003; Michiels *et al.*, 2005; Kursa, 2014; Wang *et al.*, 2016). That is, in repeated runs of RF on the same data using the same parameters, there can be a different order of variable importance ranking. The issue of stability in importance rankings is increasingly receiving attention. Strobl et al (2008) showed that instability in variable ranking in RF was due to highly correlated variables, proving that these are being used interchangeably in a decision trees of the RF.

3.3.2.4 Dealing with correlated variables

Due to the nature of gene expression data, it is very likely that there are many correlated genes. There are a several ways to deal with correlated variables, such as: increasing the number of trees grown (Haury *et al.*, 2011); grouping correlated genes together (clustering) (Tolosi, Lengauer, 2011); or backward elimination strategies such as recursive feature elimination (removing the least important variables every time a model is build) based on the rankings of the variables in the previous RF models.

Adding a recursive feature elimination (RFE) to random forest works by first fitting an RF model to the whole training dataset; this model ranks variables in order of importance. A defined percentage of variables with the lowest rankings are removed. A new RF model is built on the remaining variables and the ranking of the new set of variables are re-calculated, and again a percentage of variables with the lowest ranking are removed. This is repeated till the set of variables that gives the smallest OOB error rate are left (Díaz-Uriarte, Alvarez de Andrés, 2006; Genuer *et al.*, 2010).

This method of recursive feature elimination combined with random forest when tested on correlated datasets was found to reduce the effect of correlation on the variable importance measure and produces stable models (same rankings were obtained if initial training set was slightly changed) (Gregorutti *et al.*, 2013).

3.3.3 Performance metrics

To measure the stability of the variable selection and to evaluate the accuracy[3] of the classification models built, several metrics can be used.

3.3.3.1 Accuracy measures

A common approach of evaluating the performance of classification models, sometimes referred to as model accuracy, is to use cross-validation (Kohavi, 1995). In cross validation, a percentage of the samples are held out and the classifier is built on the remaining data. The classifier is then used to predict the labels (experimental group/response variable) of the held-out data. If the predicted labels match the true set of labels then the accuracy is 1. This is repeated multiple times and the average accuracy score is given.

A simple way of representing this is using a confusion matrix. Figure 3.5 shows an example of a confusion matrix for a multiclass problem (more than two experimental groups). It shows the number of samples that the model accurately predicted per label. Since there is more than one label involved, the overall accuracy is the fraction of correctly classified samples over the total numbers of predictions.

If the datasets are unbalanced (not equal numbers of samples in each experimental group), it can lead to misleading results, therefore other metrics such as precision and

[3] Accuracy – How well the classification model is at predicting unknown data.

recall are used. For multiclass problems precision and recall are computed for each of the class labels and averaged out.

- **Precision** – Measures how accurate the classification model is at predicting each class (experimental groups), in a new dataset (Figure 3.4).

$$Precision = \frac{True\ positives}{True\ positive + False\ positives}$$

- **Recall** – Measures how good the classifier is at finding all the correct labels per class (experimental group). This is the same as sensitivity (Figure 3.4).

$$Recall = \frac{True\ positives}{True\ positives + False\ negatives}$$

For example, if a model produced both high precision and recall, this would mean that the model has accurate predictions (precision) and is also good at finding all the samples that belong to that class (recall). A model with high precision and low recall, means the model predicts few results per class but most of them are accurate predictions. Low precision and high recall is the opposite; the model produces lots of predictions per class but a lot of them are incorrect. Ideally both high precision and recall are wanted.

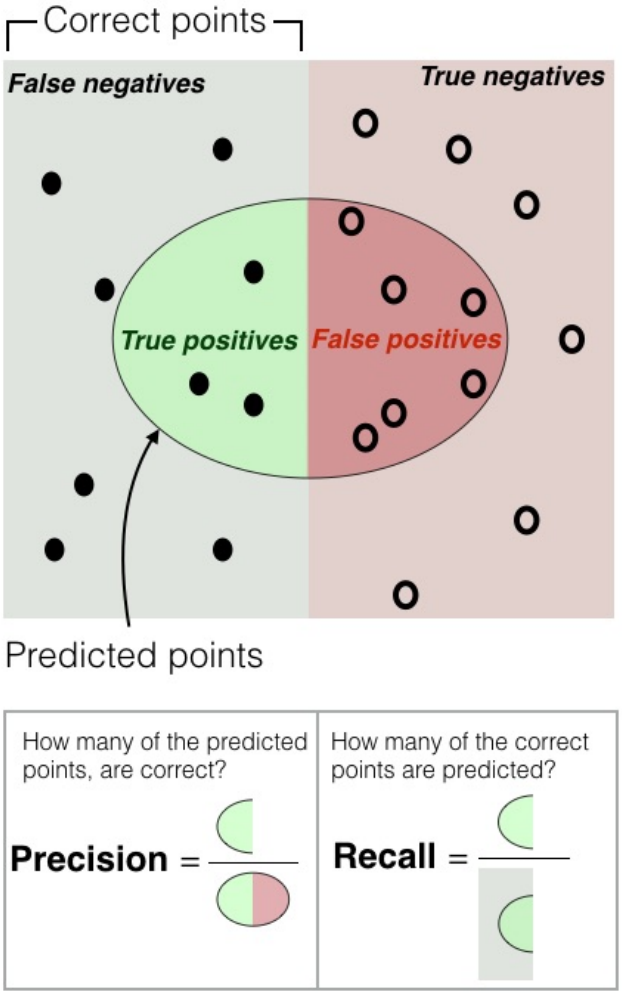

Figure 3.4. Precision and recall. Precision is the fraction of predicted points that are correct among all the predicted outcomes. Whereas recall (also called sensitivity), is the fraction of correct points that were predicted over the total number of correct points. Precision is true positives over true positives and false positives. Recall is true positives over true positives and false negatives.

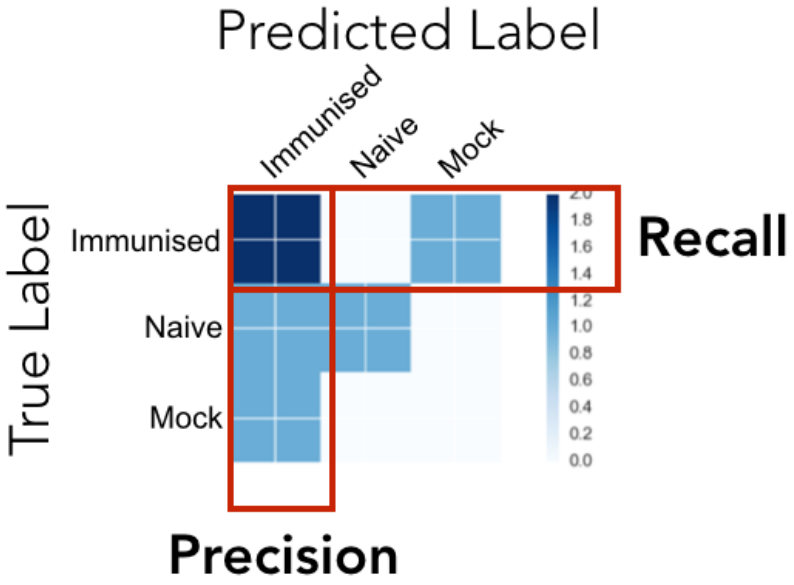

Figure 3.5 Confusion Matrix. This confusion matrix was constructed from a RF model built on a dataset from this study. An RF model was built on a subset of the data (training data), and then used to predict the samples labels from remaining data (test data - not used to build the model). In this case, there were 6 samples in the test dataset belong to the experimental groups (Labels): Immunised, Naïve or Mock. These were then compared with the labels the RF model predicted. In this example, the model was able to accurately identify two out of three samples as belonging to immunised mice (the third was misclassified as mock). The model also correctly identified a naïve mice, second sample was misclassified as immunised. The model was unable to predict any of the mock samples. Precision and recall is also calculated to evaluate Machine Learning models, and when models are built on multi-class problems (multiple experimental groups), precision and recall is calculated for each experiment group and a mean is calculated. In the confusion matrix above, for the immunised group precision was defined by the ratio of predicted samples to known immunised samples. In example above, 2 samples were correctly predicted out of 4 predictions. Recall was defined as the fraction of immunised samples it can retrieve and in this example, 2 immunised samples were identified out of a possible 3.

3.3.3.2 Stability measures

The stability of the model is measured as the robustness of variable rankings, that is produced by the classification model over repeated runs (can also be called variance). Measuring stability requires a similarity measure, for this the Jaccard index was used (Wang, 2015; Wang *et al.*, 2016) to evaluate stability of the top ranked variables over repeated run, the closer the index is to 1 the more similar the gene rankings are between each other.

3.4 Machine Learning Pipeline for Gene Selection

Despite some of the drawbacks RF faces when dealing with microarray data, it remains one of the most robust classification algorithms available. Therefore, a gene selection "pipeline" was put together using publicly available machine learning algorithms including RF, to identify genes that explain the differences seen in protection in vaccinated mice and humans living in endemic areas, from microarray data obtained from blood samples. Before the Machine Learning (ML) pipeline can be applied, the raw gene expression data from the Illumina microarrays were pre-processed to account for the variability that arises from microarray processing, such as RNA hybridisation to plates, scanning and image analysis (details in section 3.5.5). This pipeline combines a clustering algorithm with multiple RF classifications to identify genes of interest, and consists of 4 steps (Figure 3.6):

- **Step 1: Dimensionality reduction** - The first step uses a clustering method, to group genes based on the expression pattern across the samples, and then the gene

expressions within those gene clusters are summarised (mean expression). This was done for two purposes:

I. To remove noisy genes. This is on the assumption that genes which are functionally relevant and work together are more likely to have similar expression levels and therefore cluster together (Heyer *et al.*, 1999). The advantage of the clustering method used in this pipeline, over others is that it does not force all data points into a cluster and some genes remain unclustered. For this study, the unclustered genes are deemed as noise; these genes might be part of an activated pathway but the clustering algorithm is unable to detect it.

II. Clustering is used for dimensionality reduction. This is when the original data is summarised into a smaller dimension (smaller numbers of genes), by reducing all the genes within a cluster to a single variable, this results in a smaller set of uncorrelated variables (genes). This dimensionality reduction is a necessary step since microarray data is highly dimensional (more genes compared to samples), and therefore reducing the dimensionality of the data helps improve the stability and accuracy of the subsequent classifier.

- **Step 2: Random forest for feature selection (RF-FS)** – This next step of the pipeline, which is known as feature selection, involves the use of random forest (RF), as a classifier to rank the summarised gene clusters in order of importance, according to which gene cluster best classified the data into experimental groups (i.e. immunised or non-immunised mice). Random forest is repeated over 10

iterations and the gene clusters that are consistently top-ranked over the 10 iterations are deemed the most informative gene clusters.

- **Step 3: Random forest for quality control of gene selection (RF-QC)** – The top ranked gene clusters are then used to build a second set of RF models, this is to see if these gene clusters can accurately predict the experimental groups the samples came from (i.e. if they came from immunised or non-immunised mice / whether they were from naturally protected or infected humans).

- **Step 4: Functional analysis** – The genes that belong to the top gene clusters, are then subsequently used in a pathway enrichment analysis to identify what pathways the gene belongs to and how they are associated with immunity and infection.

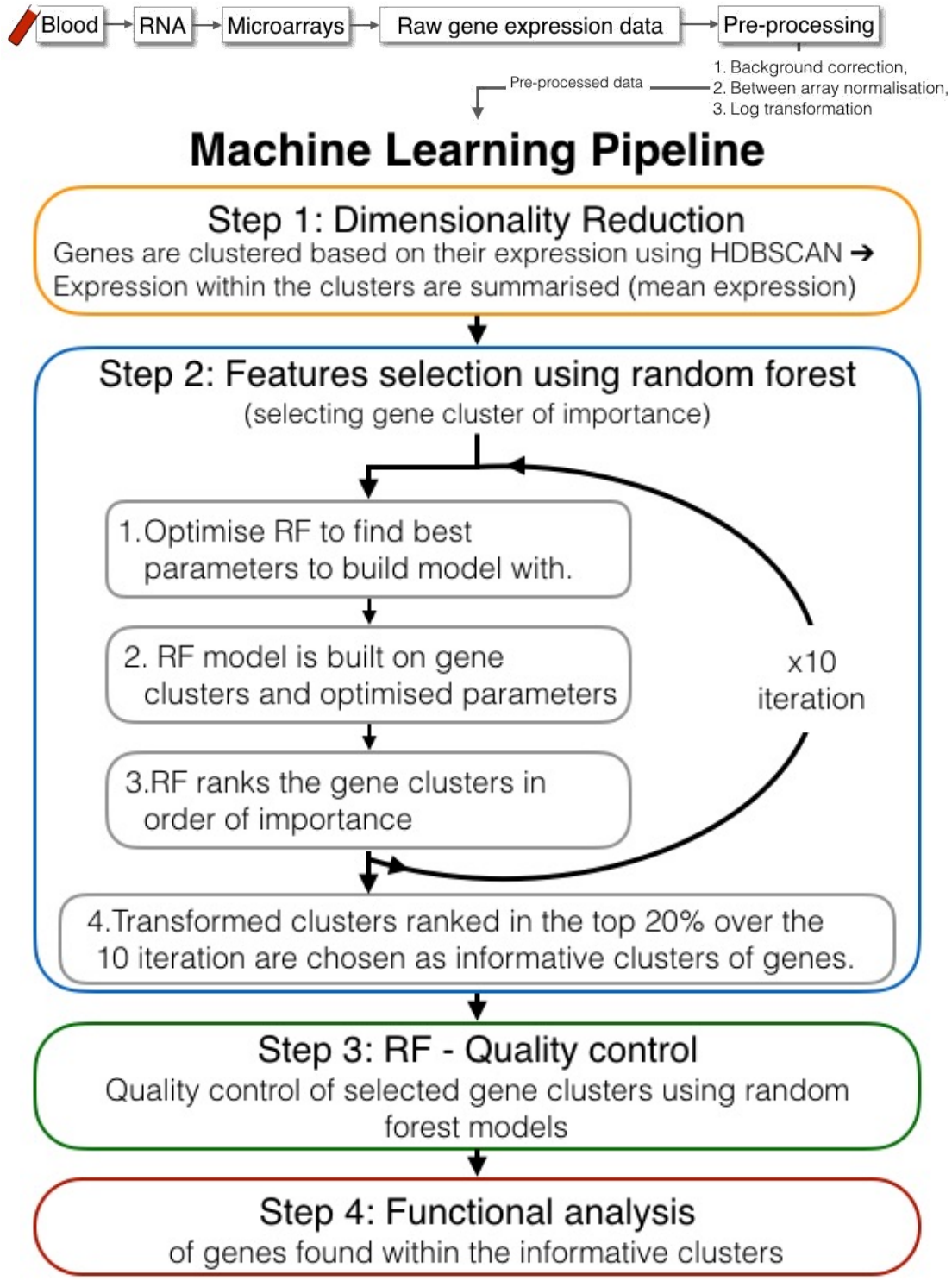

Figure 3.6 Machine Learning (ML) Pipeline for Gene Selection. The input data for the pipeline was pre-processed Illumina microarray data, the pre-processing involved background correction, between-array quantile normalisation and $\log_2$ transformation (see pre-processing section for more details). The first step in the pipeline involves *dimensionality reduction*, using

a density-based algorithm called HDBSCAN, it finds genes that have similar gene expression levels and groups them together, but does not force all points into a cluster, and therefore some genes remain unclustered. The gene clusters are summarised by calculating the average (the mean) expression of the genes within that cluster, this forms a new variable that is piped to the next step. Step 2, is termed the feature selection step, as informative features (variables and in this case gene clusters) are detected. This step uses Random Forest to rank the gene cluster, on how good they are at classifying the samples into the different experimental groups (immunised group / infection status). There are four parts to selecting the gene clusters: 1) when building a RF model two parameters need to be optimised, this was done by building models with different combinations of parameters, and the pair of parameters that gave the highest accuracy score was chosen. 2) Then a RF classifier was built with the gene clusters using the best set of parameters. 3) The optimised RF ranks the gene clusters in order of importance, and the ranking is saved. Step 1-3 are then repeated 10 times (this is because RF can be unstable in ranking variables). 4) The gene clusters that were consistently ranked in the top 20% over the 10 iteration, were chosen as the most informative gene clusters. Step 3 in the pipeline, involved doing a quality check on the gene clusters chosen in step 2 as being informative, for this a second set of RF classifier are built using only the informative gene clusters. The accuracy of the models are measured, to see if indeed they are good at classifying the samples into their experimental group. Step 4, once the gene clusters have passed the quality check, the genes within the clusters are extracted, and functional analysis performed on them to find their function or what pathways they belonged to.

3.4.1 Step 1: Dimensionality reduction

For dimensionality reduction, a density based clustering algorithm called Hierarchical Density-Based Spatial Clustering Applications with Noise (HDBSCAN) was used. HDBSCAN was developed by Campello, Moulavi, Zimek and Sander (2013) by improving the density based clustering algorithm DBSCAN, to allow the identification of clusters with different densities / size (Campello *et al*., 2013).

HDBSCAN works by finding density-connected regions (densely connected genes) in the data, which are then defined as clusters, whereas points that do not belong to any density-connected regions are labelled as noise (Figure 3.8).

Advantages of HDBSCAN are that:

- It does not force all data points (genes) into a cluster, unlike other clustering algorithms, such as K-means, which does.
- It can find clusters of varying densities, unlike the original density based clustering (DBSCAN) algorithm.
- It is more robust to parameter selection, since it only requires one input parameter to be fine-tuned, *minimum cluster size* which is the minimum number of samples in that group to be considered a cluster, any groups smaller then this will be considered noise.
- It is stable, producing the same clusters when the algorithm is used again.
- HDBSCAN was implemented into a python package by Leland McInnes, making it simple to apply to the data (McInnes, 2015).

Before clustering could be applied, the data was standardised by removing the mean and scaling to unit variance. In practice this means that each gene measured, had a mean of 0 and a variance of 1.

HDBSCAN has one parameter that needs to be selected, this is the *minimum cluster size*, which is the minimum number of samples that are needed to consider a group as

a cluster. Normally, when the *minimum cluster size* increases, the number of clusters identified by HDSCAN decreases (Figure 3.7.A). However, this was not the case in some of the datasets used in this study and no uniform decrease in cluster numbers were observed when the *minimum cluster size* is increased, possibly due to small sample size. However, when several datasets were merged (increasing the sample size), HDBSCAN produced a constant decrease in cluster numbers as the *minimum cluster size* increased. Therefore, for the purpose of this study the *minimum cluster size*, was chosen as the smallest number, which produced the largest number of clusters, for example in Figure 3.7.B, *minimum cluster size* was chosen as 4, as this produced the greatest number of clusters (65 clusters).

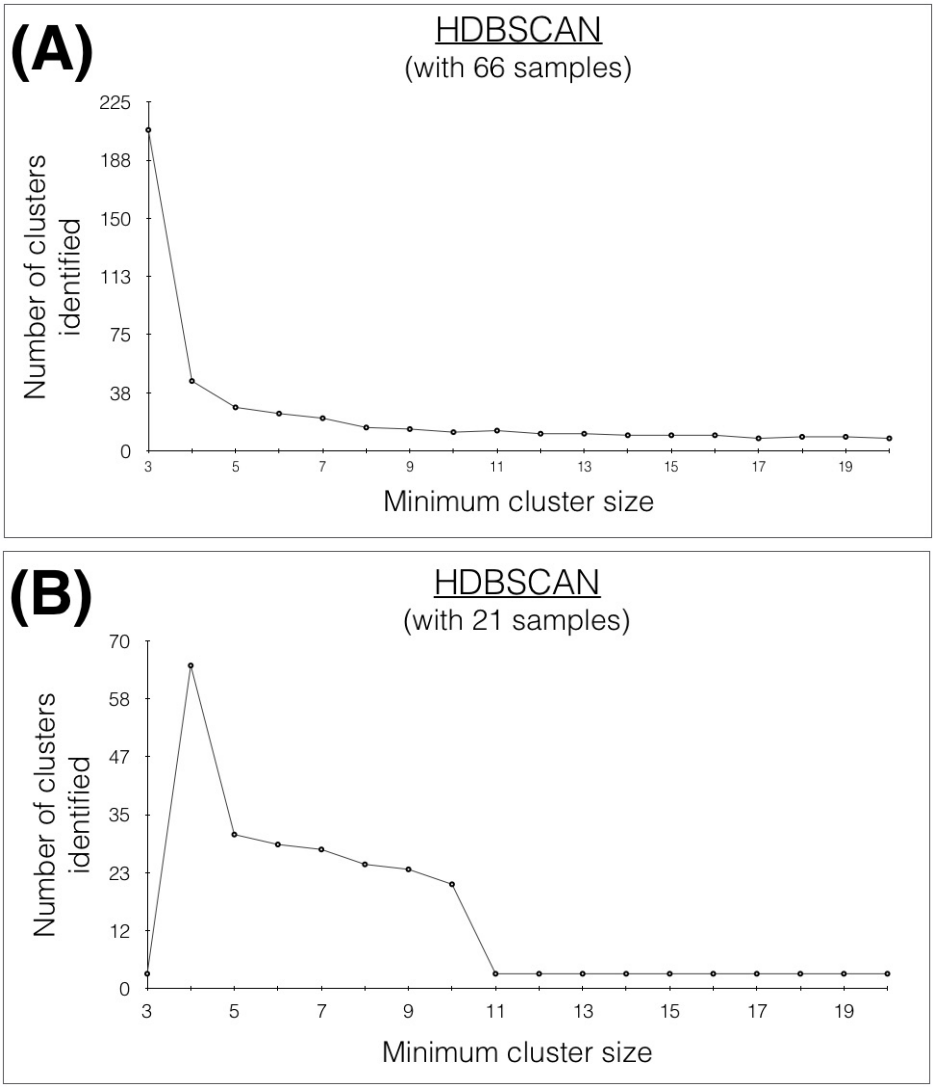

Figure 3.7 Selecting *minimum cluster size* for HDBSCAN clustering. *Minimum cluster size* is the parameter that needs to be chosen, for HDBSCAN clustering. *Minimum cluster size* is the minimum number of samples that are needed to consider a group of samples as a cluster. A) HDBSCAN clustering over a range of *minimum cluster size* (min_clust_size: 3-20), on a dataset with 66 samples (this is from merging samples from different time points in an immunisation time course), this produced a constant decrease in number of clusters produced. B) Whereas when HDBSCAN was applied to a single time point in the immunisation time course, with 21 samples (all time points in the mice immunisation time course had samples ranging from 20-24 samples), there was not a constant decrease in numbers of clusters produced. Therefore, for the purpose of this study, the *minimum cluster size* was chosen as the smallest number that produced the greatest number of clusters. For example in B), a *minimum cluster size* of 4 was chosen, as this produced the greatest number of clusters (65).

HDBSCAN assigns a number to each gene according to which cluster it belongs to, and assigns a -1 to genes that are unclustered (noise) (Figure 3.8). Once genes have been clustered, the expression of genes within a cluster can be summarised (one new data point is created for that cluster, instead of multiple genes), so that the dimensionality (number of genes) can be reduced. Therefore, for each cluster (genes labelled ≥ 0), the mean expression of the genes within that cluster was calculated, this groups correlated genes together to form a smaller set of gene clusters. Genes that were not clustered, were removed from the analysis.

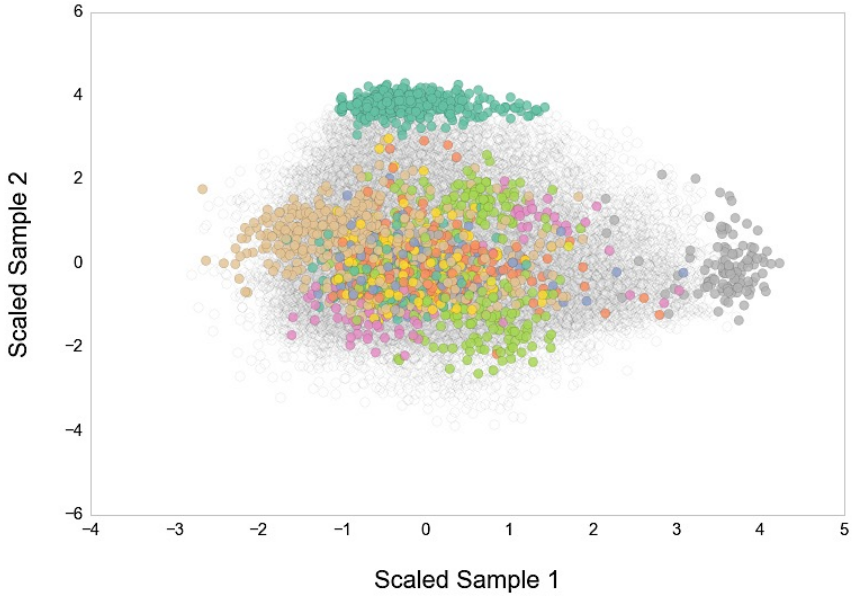

Figure 3.8. HDBSCAN clustering. A representation of HDBSCAN clustering. For simplicity only two dimensions (2 samples) are shown, but in reality there are multiple dimensions (equal to the number of samples). Each point represents a scaled gene, the colours represent different clusters and the grey points are the unclustered genes (noise) that are filtered out using HDBSCAN. As this is a multi-dimensional dataset, it is difficult to differentiate between the clusters in the centre, but looking towards to the edge clusters are apparent.

3.4.2 Step 2: Random forest for feature selection (RF-FS)

Random forest was used for feature selection, this means that random forest was used to find the gene clusters (produced by HDBSCAN), that best describes the difference between the experimental groups. For this, the summarised gene clusters were used to build a RF model, which then ranks them in order of importance, on how good they were at differentiating between the different labels (experimental groups). Because random forest struggles with stability in variable ranking, 10 random forest models were built, and the genes constantly ranked in the top 20% across the 10 iterations were chosen as being informative. The process of feature selection (gene selection) using random forest consist of five steps.

1) **Optimisation of the random forest classifier parameters** – Random forest has two parameters that needs to be optimised, the number of trees grown from a subset (bootstrapped sample) of the data (*n_estimators* parameter in Scikit-learn, Python, (Pedregosa *et al.*, 2011)) and the number of variables to consider when looking for the best split of a node (*max_features* parameter in Scikit-learn), RF does not look at all samples at a node to split the data like with decision trees. A range of values were chosen for each parameter and different combinations were tested out using a method called cross-validation, which trains a RF model on 80% of the samples and the remaining 20% are used to validate the model. In python there is a function called *Stratified_Shuffle_Split* (Scikit-learn, Python), which facilitates cross-validation and ensuring that there is at least one sample from each experimental group in the validation (test) dataset. To ensure that each sample is present in the validation set at least once, this was repeated 10 times for each combination of parameters. The

combination of parameters that gave the highest *OOB_score* scores (i.e. gives the most generalisable model), were chosen and used to train the RF classifier.

- *n_estimators* parameter – The number of trees grown were chosen from a range of 1000-2000, even though Figure 3.9 shows that the number trees grown after 500 has no effect on the accuracy of the model (OOB score), but the greater the number of trees the better the stability of feature ranking (Genuer *et al.*, 2010; Haury *et al.*, 2011).

- *max_features* parameter – The number of variables to consider when looking for the best split was either the square root of all the features (variables/ genes), 10%, or 40% of all the features.

2) **Fitting random forest model to the data using the optimised parameters** – A random forest model was built on the data using the parameters optimised in step 1. Similar to the optimisation step, the accuracy of the RF model (built with the optimised parameter), was measured using 10-fold cross-validation. For this 20% of the data was held out when building the RF model, and then used to measure how accurate the model was at predicting the held-out data.

3) **Ranking variables (genes clusters)** – The RF model ranks the gene clusters in order of importance using the Gini index. Although other metrics can be used, the Gini index was chosen because it is less affected by correlated data than other importance measures (Strobl *et al.*, 2008). The Gini index is prone to give higher ranking to variables with categorical data (Strobl *et al.*, 2007). In this case, it is not

a concern, as the data in this study are all gene expression data and are therefore continuous data.

4) **Selecting the top ranked gene clusters** – Random forest can suffer from instability when ranking variables in order of importance, especially if these variables are correlated. Therefore, step 1 to 3 were repeated 10 times, and the gene clusters that were consistently ranked in the top 20% of the 10 iterations, were considered as the informative gene clusters. Twenty percent was used as the cut-off value, for two reasons:

- The top 20% ranked gene clusters over the ten iterations gave a high stability index across the datasets.
- The number of genes that belonged to those gene clusters were enough to perform pathway analysis.

5) **Measuring accuracy and stability of the random forest models built for feature selection** – The accuracy, precision and recall of the RF models built on gene clusters, were measured for each of the 10 RF models and the averages out, the closer the values were to 1, the more accurate the models were. Stability was measured using the Jaccard index, which quantifies how similar the ranking of the gene clusters were in the top 20% rankings. The closer the Jaccard index was to 1, the more similar the rankings were, therefore a stable model.

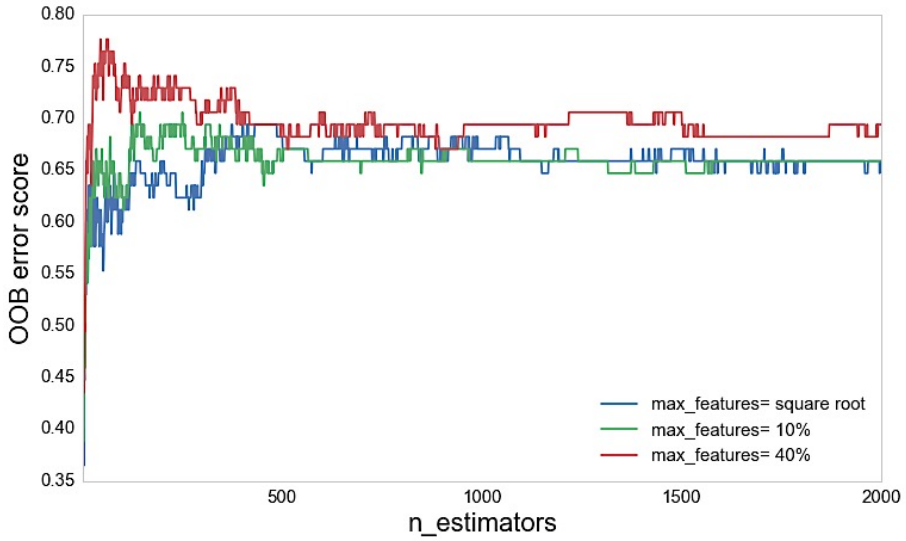

Figure 3.9 OOB_score cross-validation**.** In RF models, there are two parameters that are important to optimise. The OOB_score is used to choose the best set of parameters, and the closer the value is to 1, the more accurate the model build using that set of parameter. Above is an example of how the OOB_score changes for varying values of: *n_estimators*, which the number of trees to grown from each bootstrap subsamples of the data; and *max_features*, which is the number of variables to consider when looking for the best split of a node. Random Forest models were built using range of n_estimators (2-2000) with either the square root, 10%, or 40% of the variables as the *max_features* value. In the machine learning pipeline *n_estimators* was chosen from a range of 1000-2000, although this range is higher than required according to the graph above, a higher value for *n_estimators* provides a more stable feature ranking.

3.4.3 Step 3: Random forest for quality control of gene selection (RF-QC)

This section of the pipeline was used to evaluate the accuracy (quality) of the gene clusters selected by the random forest above. This was done by measuring the gene clusters ability to differentiate the samples by experimental group (in mice models) or parasitological presentation (in human data), by building a second round of random forest models, only using the informative gene clusters (top 20% gene clusters). This second round of RF models must also be optimised to find the best parameters before

the model can be built. The accuracy, precision and recall was measured for this second random forest used for quality control (RF-QC). This was done to verify that the gene clusters were indeed good at predicting the experimental groups from which they were derived. It is also another way of confirming that the selected genes are important in relation to immunisation (mouse studies) or infection (in human datasets).

3.4.4 Step 4: Functional and pathway analysis

Once the gene clusters were identified as "important", using the second round of RF models, the genes within these clusters were analysed. To do this functional pathway analysis was conducted, using in clusterProfiler (Yu *et al.*, 2012) in R. This finds the biological functions and pathways that are over-represented in a list of genes.

Two databases were used to identify the over-represented pathways or biological functions:

- **The Gene Ontology (GO) project** – To categorise genes based on their Biological Processes (BP) (Ashburner *et al.*, 2000).
- **The Reactome database** – To find pathways and larger processes the genes belong to (Croft *et al.*, 2011).

The degree of over-representation was calculated using cumulative hypergeometric distributions, which is the probability of finding genes belonging to a functional GO category or Reactome pathway within the gene list. A P-value was calculated for each of the pathway or processes using the Benjamini-Hochberg method to correct for multiple testing (Benjamini, Hochberg, 1995), and the significance threshold was

chosen as an adjusted P-value ≥0.05. Furthermore, at least five genes from the input list were needed for a term or pathway to be classified as over-represented, and the reference background was all the genes present on the microarray. The comparison between dataset is done by measuring the overlap of pathways over-represented.

3.5 Datasets

To investigate and compare human and murine protective immune responses to filarial infections, the machine learning pipeline was applied to four microarray datasets collected by E PIAF[4] partners.

Two of these datasets were created from the mouse model of human filarial infection, *Litomosoides sigmodontis* in BALB/c mice. Protective immunity in this model can be achieved through vaccination by inoculating mice with either irradiated infective larvae (L3) (Le Goff *et al.*, 2000) or microfilariae (Ziewer *et al.*, 2012).

The other two datasets are from human samples, these were collected in Ghana, in foci where either onchocerciasis (*O. volvulus*) or lymphatic filariasis (*Wuchereria bancrofti*) is endemic. Within these populations, a small percentage of individuals are naturally protected showing no signs of infection (no pathology, parasites nor circulating filarial antigen) even though they are constantly exposed, these individuals are termed endemic normal (EN). To make a fair comparison between mice and human responses whole blood was chosen instead of measuring responses in specific tissues.

3.5.1 L3 Vaccination dataset

To investigate the changes induced by immunisation with *L. sigmodontis* infective L3, mice were split into four groups of 6. Three of these groups were immunised with

[4] E PIAF – Enhanced Protective Immunity Against Filariasis, HEALTH-2009-4.3.1-1 Contract 242131

either: infective *L. sigmodontis* L3 (Unirradiated); irradiated *L. sigmodontis* L3 (Irradiated); or a mock dose of RPMI 1640 (Gibco) (Mock). The immunisations occurred 28, 21 and 14 days before the challenge inoculation with infective L3 (non-irradiated). The last group was left as the naïve control (no immunisation nor challenge). For each mouse whole blood samples were collected: 6 hours after the 1^{st,} immunisation, 6 hours after the challenge and 10 days after the challenge (Figure 3.10). Day 10 was chosen as the end point of this experiment because most of the parasite killing happens within two days of challenge. This vaccination targets the incoming L3 mostly at the site of inoculation, and protection is achieved because less L3 are reach the pleural cavity, therefore less maturing to adults compared to non-vaccinated controls. Therefore, to minimise cost the experiment was stopped at day 10 which is enough time to see any changes in gene expression.

The blood samples and RNA extractions were done by E PIAF partners, Dr Sabine Specht (University Hospital Bonn, Germany and Institute of Laboratory Animal Science, Vetsuisse Faculty, University of Zurich, Switzerland) and Dr Coralie Martin (Museum National d'Histoire Naturelle, France). All experimental procedures: labelling, array hybridization to Illumina MouseWG-6 BeadChip arrays (MouseWG6_V2_0_R3_11278593_A, Illumina) and array scanning was performed by Fios Genomics, Edinburgh. Fios Genomics also conducted the quality control of the arrays, for which 4 samples failed QC control. In addition, a 5th sample was found to incompletely hybridise to the array, and a 6th sample was missing altogether. As a consequence, 6 samples were removed from the analyses.

Initial analysis of the data, showed that the RF pipeline could not distinguish between irradiated and unirradiated groups, a closer look at the parasite survival at day 10 post challenge shows very little differences in parasite burden between these two groups (Figure 3.11A). Therefore, the two L3 immunised groups were merged as one and named "Immunised". The new immunised group had a lower parasite burden compared to the control immunisation (Figure 3.11), demonstrating that immunisation with L3 induces protection in these datasets. Table 3.3 summarises the characteristics of the L3 immunisation dataset taken forward for analysis.

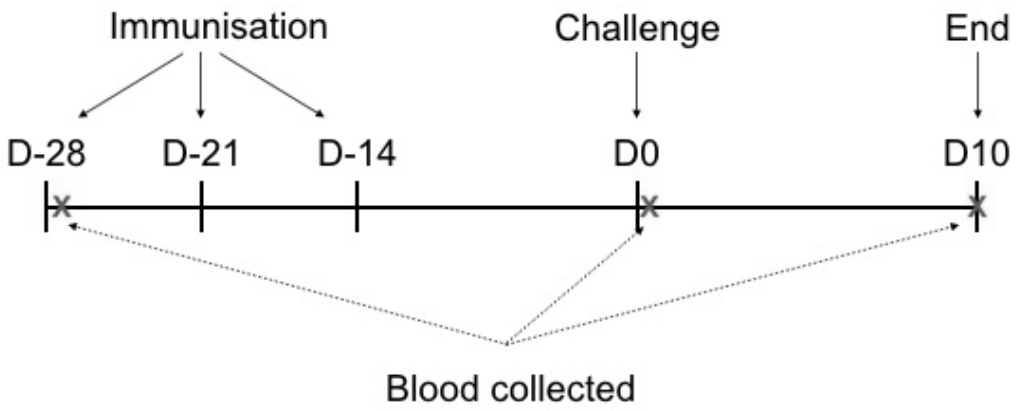

Figure 3.10 L3 immunisation time course. 24 BALB/c mice were split into four groups, three of these groups received subcutaneous immunisation of either: 40 infective *L. sigmodontis* L3 (Unirradiated); 40 irradiated *L. sigmodontis* L3 (Irradiated); or a mock dose of RPMI 1640 (Gibco) (Mock). The immunisations occurred at 28, 21 and 14 days before a challenge inoculation of 40 L3. The remaining group was the naïve control, which received no immunisation nor challenge. For every mouse in the immunisation time course whole blood was collected: 6 hours after the 1st immunisation (Day -28), 6 hours after the challenge (Day 0) and at 10 days after the challenge (Day 10).

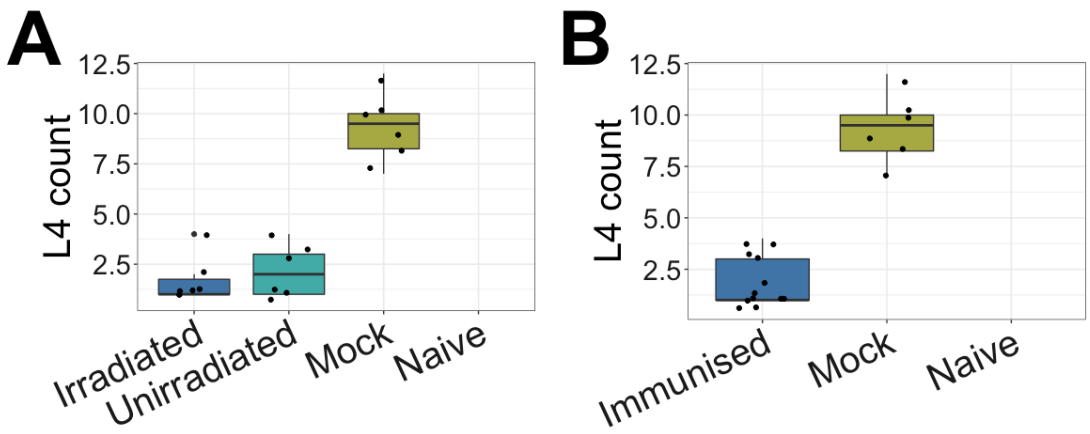

Figure 3.11 L4 Worm Counts. A) Shows *L. sigmodontis* L4 number in the pleural cavity at day 10 post challenge of mice either immunised with Irradiated L3, Unirradiated L3 or a Mock immunisation with RPMI. Naïve mice were never challenged or immunised, therefore have no parasite burden. B) Shows the *L. sigmodontis* L4 number when the Irradiated and Unirradiated treatment group are merged together.

Table 3.3. Characteristic of L3 vaccination dataset. Showing the number of samples, number of different treatment groups (classes), number of genes measured for each sample.

Time point	No of Samples	No. Classes	No. Probes
Day -28	20	3	45281
Day 0	22	3	45281
Day 10	24	3	45281

3.5.2 Mf Vaccination dataset

To investigate the changes induced by microfilariae immunisation, mice were split into four groups, three of which received immunisation of either: Microfilariae and Alum (MfA); Microfilariae (Mf); or an Alum control (Alum). The immunisation occurred 28, 21 and 14 days before the challenge, where the three immunised groups and the fourth control group (Control) were subcutaneously challenged with 40 infective L3. Whole blood samples were collected for each mouse: 6 hours after each of the three immunisations; 6 hours after the challenge; at day 49 (adults worms are present) post challenge; and day 67 post challenge (blood circulating Mf are seen) (Figure 3.12).

The whole blood samples were collected and RNA was extracted by E PIAF partners, Dr Sabine Specht (University Hospital Bonn, Germany and Institute of Laboratory Animal Science, Vetsuisse Faculty, University of Zurich, Switzerland) and Dr Coralie Martin (Museum National d'Histoire Naturelle, France). RNA hybridisation to Illumina MouseWG-6 BeadChip arrays (Mouse WG6_V2_0_R3_11278593_A, Illumina) was done by Fios Genomics, Edinburgh, who also conducted the quality control. In this time course 15 samples were missing and 7 arrays failed the quality control, therefore 22 samples were removed from the subsequent analyses, unfortunately this meant that at the day -28 (6 hours after 1st immunisation) time point, the Control group was missing and therefore day -28 was removed from analysis (Table 3.4). Parasitological readouts of worm numbers in the pleural cavity (Figure 3.13A) and microfilariae circulating in 30µl of blood (Figure 3.13B) were collected on Day 67. Table 3.4 summarises the characteristics of the Mf immunisation dataset taken forward for analysis.

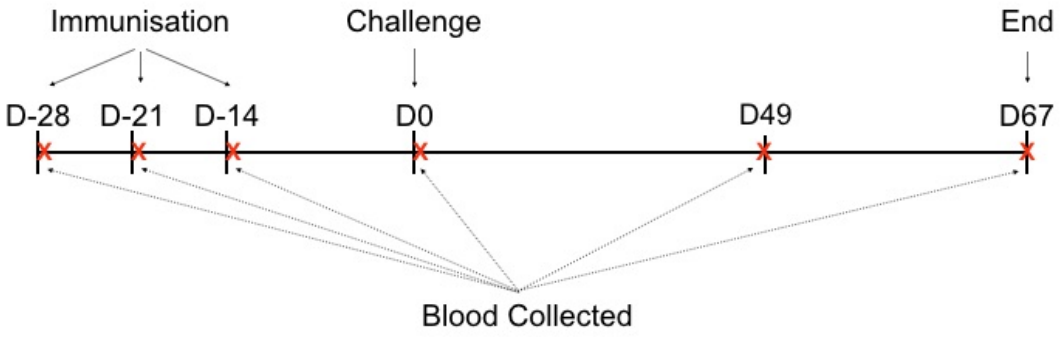

Figure 3.12 Mf vaccination time course. 24 BALB/c mice were split into four groups, three of which were immunised three times with either: 10^5 Microfilariae and Alum (MfA); 10^5 Microfilariae alone (Mf); or Alum on its own as a control (Alum). The immunisations occurred 28, 21 and 14 before the challenge with 40 infective L3, the fourth group only received a challenge (Control). For each mice whole blood was collected 6 hours after each immunisation (Day -28, Day -21 and Day -14), 6 hours after the challenge (Day 0), 49 days post challenge (Day 49) adults from the challenge should have developed, and 67 days post challenge (Day 67) when circulating microfilariae would be present.

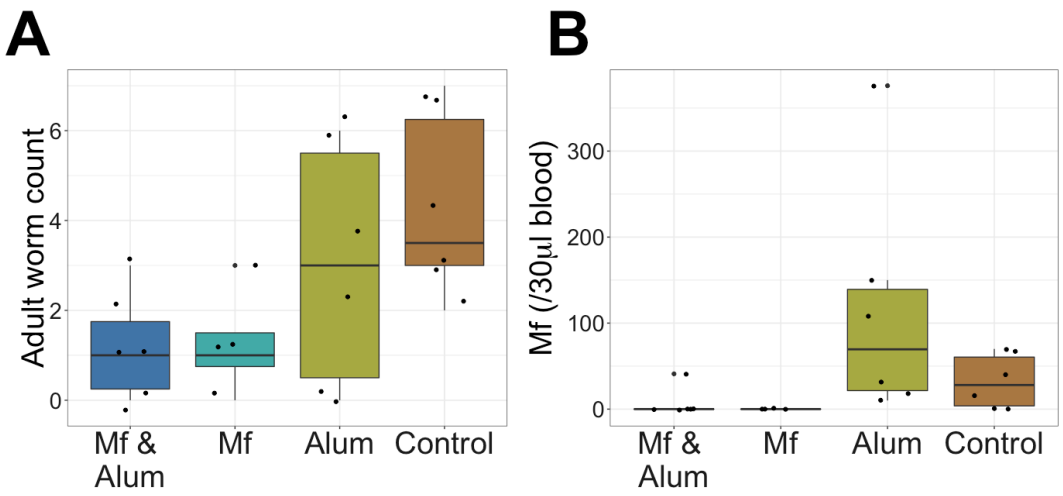

Figure 3.13. **Parasite Counts.** A) shows the number of adult worm found in the pleural cavity and B) Shows the number of microfilariae found in the 30µl of blood at day 67 post challenge in the microfilariae vaccination time course.

Table 3.4. Characteristic of Mf vaccination dataset. Showing the number of samples, number of different treatment groups (classes), number of genes measured for each sample.

Time point	No. of Samples	No. of Classes	No. of probes
Day -21	20	4	45281
Day -14	23	4	45281
Day 0	22	4	45281
Day 49	21	4	45281
Day 67	22	4	45281

3.5.3 *Wuchereria bancrofti* endemic area dataset

Whole blood samples were collected from patients living in a *Wuchereria bancrofti* endemic area in the western regions of Ghana (Ahanta West and Nzema East districts) by E PIAF partners, Dr Sabine Specht (University Hospital Bonn, Germany and Institute of Laboratory Animal Science, Vetsuisse Faculty, University of Zurich, Switzerland), Prof Achim Hoerauf (University Hospital Bonn, Germany), Dr Alex Debrah (Kwame Nkrumah University, Ghana), Dr Laura Layland (University Hospital Bonn, Germany), Gnatoulma Katataw (University Hospital Bonn, Germany, and University of Lome, Togo) and Alexander Kwarteng (Kwame Nkrumah University, Ghana).

A total of 184 whole blood human samples were collected and RNA extracted by the E PIAF collaborators from the Kwame Nkrumah University of Science and Technology (Kumasi, Ghana). The RNA samples were hybridised to Illumina HumanHT12 BeadChip arrays (HumanHT12_V4_0_R2_15002873_B) by FIOS

Genomics, Edinburgh who also conducted quality control on the samples, of which 12 arrays (samples) failed the quality control and were therefore removed from the dataset. A summary of the characteristics of the data are found in Table 3.5. All the samples also had clinical information associated with them, such as:

- Phenotype of patients: Circulating filarial antigen (CFA) and Mf positive for *W. bancrofti* was measured, with 118 patients being CFA positive with of which 48 were Mf positive. For the purpose of this analysis patients with who were CFA positive with or without Mf were classified as infected, and patients with CFA negative and no Mf were classified as endemic normals (EN).

- Coinfection with other filarial nematodes, although no coinfections with *O. volvulus* or L. *loa* were present

- Whether the patients had other helminth infections: 9 patients were infected with *Ascaris*, 9 with hookworm infections, 1 patient had both *Ascaris* and hookworms, 1 sample with schistosomiasis and 2 had trichuriasis.

- Whether they had protozoan infections: 4 patients had *Plasmodium falciparum* and 1 with *P. vivax*; 4 patients also had *Giardia*.

- How many rounds of Ivermectin treatment the patients received; although not all patients received Ivermectin, 66 patients received 1 round and 12 received 2 rounds.

- Information on the patients age, gender (43 Females, 129 males) and district they live in.

3.5.4 *Onchocerca volvulus* endemic area dataset

Whole blood samples were collected from patients living in Ghana in an onchocerciasis endemic area, by the E PIAF partners, Dr Sabine Specht (University Hospital Bonn, Germany and Institute of Laboratory Animal Science, Vetsuisse Faculty, University of Zurich, Switzerland), Prof Achim Hoerauf (University Hospital Bonn, Germany), Dr Alex Debrah (Kwame Nkrumah University, Ghana), Dr Laura Layland (University Hospital Bonn, Germany), Gnatoulma Katataw (University Hospital Bonn, Germany, and University of Lome, Togo) and Alexander Kwarteng (Kwame Nkrumah University, Ghana).

A total of 167 whole blood human samples were collected and RNA extracted by the E PIAF collaborators from the Kwame Nkrumah University of Science and Technology (Kumasi, Ghana). The RNA samples were hybridised to Illumina HumanHT12 BeadChip arrays (HumanHT12_V4_0_R2_15002873_B) by FIOS Genomics, Edinburgh who also conducted quality control on the samples, of which 5 samples failed the quality control and were therefore removed from the dataset. A summary of the characteristics of the dataset is found in Table 3.5. All the samples also had clinical information associated with them, such as:

- Whether patients were *O. volvulus* microfilariae positive; these were classified as infected. If no microfilariae were found in skin snips, then patients were classified as endemic normals (EN), in total 109 patients were classified as infected (87 presenting nodules) and 53 as endemic normal.

- Whether patients received treatment; 63 patients received one round of Ivermectin, 1 patient received 6 round of Ivermectin and no patients received doxycycline.

- Infection with other parasites; 23 patients also had *P. falciparum* infections.

- Also present is information on patients, age, weight, sex (73 Females, 89 Males) and what village they live in.

The human datasets are highly variable (in contrast to data from animal experiments) as a consequence of many confounding factors, such as whether patients have co-infections and/or received Ivermectin treatment. Therefore, as an initial analysis samples from patients that had other infections, either other helminths (i.e. hookworms or *Ascaris*), or protozoan infections (malaria or *Giardia*); and patients who had received Ivermectin treatment were removed from the analysis. Removing the confounding factors helped with the interpretability of the results, and therefore for the purpose of this study only cleaner samples were looked at.

Table 3.5 Characteristic of human dataset. Showing the number of samples, number of different treatment groups (classes), number of genes measured for each sample.

Dataset		No. of Samples	No. of Classes	No. of probes
Wuchereria bancrofti	All samples	172	2 (Infected / EN)	46698
	Pure samples	71	2 (Infected / EN)	46698
Onchocerca volvulus	All samples	162	2 (Infected / EN)	46698
	Pure samples	84	2 (Infected / EN)	46698

3.5.5 Pre-processing of Illumina microarray data

The raw microarray data was imported into R programming language (R Development Core Team,, n.d.), using the *lumi* Bioconductor package (Du *et al.*, 2008). The raw data was pre-processed using, the *neqc* function in the *limma* Bioconductor package (Ritchie *et al.*, 2015) to:

1) background correction, using Normal-exponential convolution, which uses negative controls (Ding *et al.*, 2008; Xie *et al.*, 2009),

2) between-array quantile normalisation,

3) $\log_2$ transformation.

This package was found to be the best pre-processing strategy, giving the highest precision for a given bias (Shi *et al.*, 2010).

To find the genes associated with the probes on the Illumina microarrays, the Illumina probe identifiers were matched to nuID annotation (a unique identifier) using the *lumi* package (Du *et al.*, 2008), and then corresponding Entrez Id were extracted, as these are needed for functional analysis using the *ClusterProfiler* package. All the machine learning methods used were carried out in python programming language (http://www.python.org/).

3.6 Evaluation of pipeline performance

As this area of research is getting significant attention, more alternative algorithms are emerging, however there is still a lack of gold-standard methods or datasets to compare it to. Therefore, to evaluate the pipeline's performance, its accuracy at predicting withheld data for each dataset was measured and compared to the accuracy of other machine learning methods (SVM-RFE, RF-RFE) often used on gene expression data. The machine learning results were further compared to a non-machine learning method of analysing gene expression which has been popular in biomedical sciences, WGCNA (Langfelder, Horvath, 2008).

3.6.1 Pipeline performance on the murine and human datasets.

The pipeline works by initially clustering genes, and summarising the expression of the genes within each cluster. These gene clusters are then ranked in order of importance using random forest (RF). This step uses random forest (RF) to select informative genes and is therefore named random forest for feature selection (RF-FS) for the purpose of this study. During the RF-FS step, RF models were built using the summarised gene clusters and ranks the gene clusters in order of importance. These RF models were built 10 times on the same dataset, to see how repeatable the gene rankings were (how stable it is). The gene clusters that were consistently ranked in the top 20% over the ten iterations of RF were classified as the most informative gene clusters (Table 3.6). These are termed informative as these gene clusters best describe the difference between the experimental groups.

The informative gene clusters are then used to build a second RF model, used as a quality control step, termed random forest for quality control (RF-QC). RF-QC was done to make sure that the gene clusters selected by the first RF were truly important. Different metrics were used to evaluate the model throughout the pipeline. The stability of the feature selection RF (RF-FS) model was evaluated using a Jaccard index, and then the accuracy of the two sets of random forest models (RF-FS and RF-QC) were assessed (Table 3.6).

3.6.1.1 Stability measures

In an ideal scenario, RF ranks the gene clusters in the same order of importance every time a RF model is built on the same dataset, using the same parameters. However, this is not always the case, therefore a similarity index (Jaccard index) is used to measure how similar the rankings are between the 10 RF iterations. The closer the Jaccard index is to 1, the more similar the rankings are. Across the datasets (both murine and human datasets), the pipeline's ability to rank variables in order of importance was relatively stable, with high index measures all above 0.8, except for the *W. bancrofti* dataset which gave an index of 0.77 (Table 3.6).

3.6.1.2 Accuracy measures

Another important measure is how accurate the pipeline is at classifying data, RF works by choosing variables (genes) which best classify the samples into their respective experimental groups, so that when to new data is used the RF model is able to predict the experimental group of the samples. Therefore, how well RF predicts the new data is used as measure of accuracy. In this pipeline, the accuracy of RF models

was measured in the form of precision[5], recall[6] and 10 fold cross-validation accuracy (Table 3.6). These were calculated for both:

- **The random forest used for feature selection (RF-FS)** – Accuracy of the random forest models used for feature selection (selecting the informative gene clusters) was measured. Since 10 RF models were built on the same data, the average accuracy, precision and recall were calculated.

- **The random forest used for quality control (RF-QC)** – The accuracy of the random forest models used for quality control, was also measured. This was to make sure that the gene clusters identified by the random forest feature selection step were indeed good at classifying the data into their respective experimental groups.

In all datasets the accuracy of the RF models improves when they are built on the gene clusters previously identified as being important by the first set of RF models (RF-FS). This signifies that the genes selected by the random forest models (in the feature selection step) are indeed good at classifying the data into the experimental groups.

In the L3 vaccination dataset, the ML pipeline was able to identify gene clusters that were informative for each time point, as the accuracy of the random forest for quality control at day -28, day 0 and day 10, was 0.95, 0.86 and 0.88 respectively. Figure 3.14 represents a confusion matrix for the Day -28 time point in the L3 vaccination dataset, which shows the difference between the RF ability to predict unknown data when trained on all gene clusters (Figure 3.14, RF-FS), or on gene clusters identified as important (Figure 3.14, RF-QC).

In the Mf vaccination experiment, the pipeline identified gene clusters at day -21, day -14 and day 0, that gave a high accuracy score of 0.70, 0.82 and 0.88 respectively, suggesting that those genes are good at identifying the different immunisation groups. However, by days 49 and 67 post challenge, the accuracy was lower, with an accuracy of 0.42 and 0.53, respectively.

Table 3.6. Performance of pipeline. The Jaccard index represents how similar the top 20% ranking of the gene cluster, over 10 iterations of RF. The accuracy, precision[5] and recall[6] are measured by 10-fold cross-validation (data is split into a testing and validation set 10 times, and the results are represented as averages and with standard errors in parantheses, the closer the measure is to 1 the more accurate the classifier is). Random Forest for Feature Selection (RF-FS) is the RF models used to selected informative gene clusters. Random Forest for Quality Control (RF-QC is RF models used to validate the gene clusters selected by the RF-FS.

Datasets		Jaccard index	Accuracy		Precision		Recall	
			RF-FS	RF-QC	RF-FS	RF-QC	RF-FS	RF-QC
L3 vaccination	Day -28	**0.82** (± 0.01)	0.50 (± 0.39)	**0.95** (± 0.40)	0.34 (± 0.36)	**0.81** (± 0.42)	0.47 (± 0.27)	**0.90** (± 0.24)
	Day 0	**0.87** (± 0.01)	0.62 (± 0.22)	**0.86** (± 0.18)	0.47 (± 0.30)	**0.80** (± 0.44)	0.64 (± 0.16)	**0.92** (± 0.20)
	Day 10	**0.81** (± 0.01)	0.60 (± 0.24)	**0.88** (± 0.25)	0.40 (± 0.31)	**0.70** (± 0.64)	0.62 (± 0.25)	**0.78** (± 0.42)
Mf vaccination	Day -21	**0.80** (± 0.03)	0.57 (± 0.32)	**0.70** (± 0.20)	0.34 (± 0.39)	**0.52** (± 0.32)	0.55 (± 0.20)	**0.68** (± 0.23)
	Day -14	**0.90** (± 0.01)	0.57 (± 0.45)	**0.82** (± 0.45)	0.35 (± 0.39)	**0.70** (± 0.44)	0.53 (± 0.35)	**0.69** (± 0.20)
	Day 0	**0.95** (± 0.01)	0.65 (± 0.40)	**0.88** (± 0.25)	0.42 (± 0.28)	**0.64** (± 0.31)	0.65 (± 0.24)	**0.72** (± 0.27)
	Day 49	**0.93** (± 0.02)	0.12 (± 0.25)	**0.42** (± 0.39)	0.03 (± 0.09)	**0.35** (± 0.39	0.15 (± 0.33)	**0.44** (± 0.37)
	Day 67	**1.000**	0.23 (± 0.27)	**0.53** (± 0.35)	0.14 (± 0.19)	**0.33** (± 0.36)	0.28 (± 0.35)	**0.53** (± 0.27)
W. bancrofti immunity		**0.77** (± 0.01)	0.68 (± 0.05)	**0.79** (± 0.13)	0.44 (± 0.00)	**0.85** (± 0.68)	0.68 (± 0.80)	**0.80** (± 0.15)
O. volvulus immunity		**0.89** (± 0.01)	0.71 (± 0.12)	**0.85** (± 0.11)	0.66 (± 0.27)	**0.78** (± 0.18)	0.69 (± 0.09)	**0.84** (± 0.12)

[5] Precision – The fraction of true positives identified out of all the predictions. (P = Tp/(Tp +Fp))

[6] Recall – The fraction of true positives predicted, which would have been identified (R= Tp/(Tp+Fn))

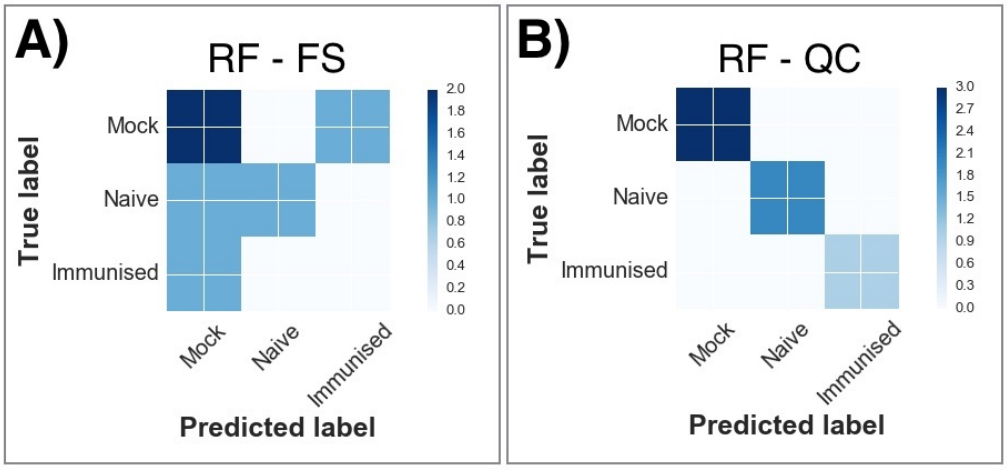

Figure 3.14. Confusion Matrix. Confusion matrix illustrating the classification accuracy between A) the random forest built on all gene clusters, i.e. random forest for feature selection (RF-FS), and B) the random forest built on gene clusters selected as being important by RF-FS, this is the random forest used for the quality control of the gene clusters selected (RF-QC). The confusion matrix above are based on random forest models built on the data from the Day -28 time point in the L3 immunity dataset, this corresponds to 6 hours following the first immunisation. Labels are the experimental groups the samples belong to: Immunised group signifies mice received a dose of either irradiated or non-irradiated L3; mice in the mock group received a dose of RPMI media as a mock immunisation; and naïve mice received no immunisation. The confusion matrix in B) shows that the RF built on informative gene, was able to predict the experimental group the samples belong to, whereas in A), the RF model built on all the gene, shows that there are some samples that are misclassified. Suggesting that genes selected are indeed good at differentiating between experimental group the samples belong, and are therefore informative on what is happening in vivo in blood following immunisation.

3.6.1.3 Stability and accuracy of unclustered genes

The importance of the clustering step (dimensionality reduction) in the machine learning pipeline was evaluated by building RF models of un-clustered genes and measuring their stability and accuracy. For this RF models were built on the L3

vaccination dataset using the pre-processed gene expression data (without the clustering step). 10 RF models were built at each time point (Day -28, day 0 and day 10), and the stability and accuracy of the RF models were measured (Table 3.7). The clustering step (dimensionality reduction) was necessary to increase the accuracy of RF, which is much lower without clustering (Table 3.7).

Clustering also increases the stability of the RF models. Thus, when RF models are built on all the genes (no clustering), a Jaccard index of 0.14-0.26 was recorded, but when clustered, Jaccard index of between 0.95 and 0.86 was obtained (Table 3.7). Using gene clusters, the RF model is able to consistently rank the clusters in a similar order of importance.

Table 3.7. Performance of pipeline without the clustering step on L3 vaccination dataset. The machine learning pipeline was applied to the L3 vaccination without the clustering step and compared to the pipeline with clustering step in terms of model accuracy and stability. The Jaccard index represents how similar the gene rankings were in the top 20% ranked (closer to 1 the more similar the ranking, so more stable the pipeline), the accuracy was measured by 10-fold cross-validation.

	Jaccard Index (Stability)		**Accuracy**	
Time points	**Without clustering**	**With clustering**	**Without clustering**	**With clustering**
Day -28	0.26 (± 0.01)	**0.82** (± 0.01)	0.61 (± 0.24)	**0.95** (± 0.40)
Day 0	0.13 (± 0.01)	**0.87** (± 0.01)	0.20 (± 0.20)	**0.86** (± 0.18)
Day 10	0.14 (± 0.01)	**0.81** (± 0.01)	0.30 (± 0.30)	**0.88** (± 0.25)

3.6.2 Comparison with alternative machine learning methods

The pipelines accuracy was compared to two other machine learning algorithms, Support Vector Machine with Recursive Feature Elimination (SVM-RFE, (Zhang, Xiaojuan Huang, 2015) and Random Forest with Recursive Feature Elimination (RF-RFE, (Díaz-Uriarte, Alvarez de Andrés, 2006). These alternative machine learning methods were applied to the L3 vaccination dataset. Using a 10 fold cross-validation, these alternative machine learning methods were optimised to output the number of variables (genes) that gave the highest accuracy.

The machine learning (ML) pipeline used in this study gave a higher classification accuracy scores compared to SVM-RFE and RF-RFE (Table 3.8). However, only a few genes were found to overlap between the gene list produced by SVM-RFE and RF-RFE and the ML pipeline (Table 3.8). Furthermore, the running time of the ML pipeline is much faster than both SVM-RFE and RF-RFE, and ML pipeline in this study is easier to optimise and allows for more flexibility in choosing the number of features (genes/clusters) to take forward for further analysis. SVM-RFE and RF-RFE, identifies the genes that are important, but the number of genes that one is left with at the end of the analysis, that are classified as important cannot be chosen. Whereas, with the ML pipeline, the percentage of the top ranked genes can be changed depending the subsequent analysis.

Table 3.8.Accuracy of SVM-RFE and RF-RFE, on the L3 vaccination dataset, measured by 10-fold cross-validation (splits the data into a testing and validation set 10 times, the results are averages and standard error of each classification), the genes classified as important by the SVM-RFE and RF-RFE methods were compared to the genes extracted from the clustered classified as important the pipeline in this study, by measuring the overlap in gene (% overlap).

Time points	ML Pipeline		SVM-RFE			RF-RFE		
	Accuracy	No. genes	Accuracy	No. genes	Overlap	Accuracy	No. genes	Overlap
Day -28	**0.95** (± 0.20)	421	**0.80** (± 0.37)	9	0 %	**0.60** (± 0.24)	31	30%
Day 0	**0.82** (± 0.22)	218	**0.34** (± 0.18)	205	6.8 %	**0.68** (± 0.20)	31	3.2%
Day 10	**0.80** (± 0.44)	268	**0.30** (± 0.20)	66	0.37 %	**0.50** (± 0.00)	31	0%

3.6.3 Comparison with a non-machine learning method used in the microarray literature

A popular non-machine learning method for the analysis of gene expression data is Weighted Gene Correlation Network Analysis (WGCNA), which works by finding highly correlated groups of co-expressed genes (these have similar gene expressions across the samples). In this study, WGCNA was used to identify groups of genes that were associated with immunisation. The genes identified by WGCNA were then compared to the genes identified by the ML pipeline.

However, WGCNA cannot analyse multiclass data and struggles with large dimensional data (i.e would not be able to analyse the 46,000 genes measured by microarrays). For this reason, comparison was restricted to data from the Mf vaccinated mice and genes involved in immune responses identified using the Reactome database (R-MMU-168256.1) and Gene Ontology database (GO:0002376), a total of 2,568 genes. The Mf immunisation dataset has a measure of Mf counts at Day 67 and this provides the simplest way to split mice into a binary outcome of presence or absence of Mf.

WGCNA essentially clusters genes based on their expression, and a principal component is calculated for each cluster. The first component of the clusters are then correlated with presence or absence of Mf, to see if there is an association between the gene clusters and presence or absence of the parasite. WGCNA was conducted in R using the parameters recommended the WGCNA guidelines (Langfelder, Horvath, 2008): *deepSplit* =2, cut height = 0.99 and minimum module size = 10.

At each time points a different number of clusters were produced, with varying numbers of genes in each (Table 3.9). The cluster first principal component was calculated (a way of summarising the cluster) and correlated with a binary outcome of presence or absence of Mf seen at Day 67. The clusters found to be significantly associated with the presence or absence of Mf were further investigated using pathway analysis to determine any association with parasite numbers. Pathway analysis was the same as that used by the machine learning pipeline, except genes associated with immune responses where used as the reference background.

Table 3.9. WGCNA per time point: The number of clusters (groups of genes) identified at each time point; the average number genes per cluster; and number of clusters significantly associated with protection.

Time point	Number of clusters	Average size of clusters (genes)	Number of clusters associated with protection*
Day -21	127	20	6 (263 genes)
Day -14	66	39	1 (158 genes)
Day 0	24	108	4 (232 genes)
Day 49	90	29	5 (125 genes)
Day 67	105	25	3 (122 genes)

*Clusters significantly correlated with presence/absence of Mf at day 67, with a P-value <0.05 (T-test)

The pathways identified by WGCNA were compared to the pathways identified by the machine learning pipeline. For the post immunisation time points (Day -21 and Day -14) both methods identified similar pathways being triggered by the Mf immunisation (Table 3.10) such as: negative regulation of T cell receptor signalling (TCR and PD-1 signalling); interferon signalling (IFN-γ and IFN-β); antigen processing and cross-presentation. However, WGCNA also identified genes associated with regulation of

MAP Kinases involved in the activation Toll Like Receptor (TLR) cascades (Table 3.10).

Pathways identified by WGCNA following the challenge infection (Day 0) or 49 days later (Day 49) are not particularly informative (Table 3.10). At Day 0 genes involved in nucleus organisation were identified. At Day 49 genes associated with response to external stimulus (GO:0009605) were identified: this is a broad category of processes that involves any change of activity of a cell, in terms of movement, secretion, enzyme production, gene expression due to an external stimulus.

Interestingly at Day 67 WGCNA successfully identified pathways associated with interferon signalling (IFN-γ and IFN-β), antigen presentation and regulation of immune responses (Table 3.10), the same as after immunisation. The genes involved with these processes are more highly expressed in mice with microfilariae circulating in their blood, suggesting that WGCNA can detect the immune responses to Mf being produced by the adult worms and which is not possible using ML pipeline. However, this observation may also be explained by the fact, that for the WGCNA the data was split into presence and absence of Mf, which was not done in the ML pipeline as the focus was on the effect of immunisations.

Although there may not be complete overlap of results from the new ML pipeline and the popular WGCNA method, the ML pipeline successfully identifies the same biologically relevant pathways as WGCNA. Moreover, since it is not restricted to candidate genes (immune genes) and two groups, it is able to identify greater number of genes and processes that are being triggered following immunisation.

Table 3.10. Summary of processes/pathways found as being important by WGCNA and the ML pipeline, using the GO database and Reactome database for each time point. (Full list Supplementary Table S2 and Table S3)

Time Point	Summarised processes / pathways identified by WGCNA	Overlap with pathways identified by the ML pipeline
Day -21	TCR signalling	✓
	PD-1 signalling	✓
	Antigen processing and presentation	✓
	Antigen cross-presentation	✓
	Interferon signalling (IFN-γ and IFN-β)	✓
	Regulation of immune responses	✓
	Regulation of ERK (a MAPK) involved in TLR cascades	
Day -14	Interferon signalling (IFN-γ and IFN-β)	✓
Day 0	Nucleus organization	
Day 49	Response to external stimulus	
Day 67	Defence response to other organisms	
	Interferon signalling (IFN-γ and IFN-β)	
	Regulation of immune responses	
	Antigen processing and presentation (MHC-I)	

3.7 Biological relevance of results from machine learning pipeline

3.7.1 Changes in gene expression after vaccination in murine models

Functional analysis was performed on the genes selected by the machine learning pipeline to investigate the biological significance of the genes; overlap between time points within a dataset; and to uncover any overlap in protective immunity between the *L. sigmodontis* mouse models and human filarial infections (*O. volvulus* and *W. bancrofti*).

It was determined that concentrating on pathways or processes with which specific genes are associated to may give more insight into the overall responses being triggered during the vaccination time course, and determine if there is any overlap with human immune responses. As little overlap between the different time points was observed when comparing the genes selected by the ML Pipeline throughout a vaccination time course (Figure 3.15A). However, when comparing the functional pathways those genes belonged to between the time points, these showed greater overlap (Figure 3.15B). This implies that similar pathways are being triggered throughout the vaccination time course, but the genes are not always the same.

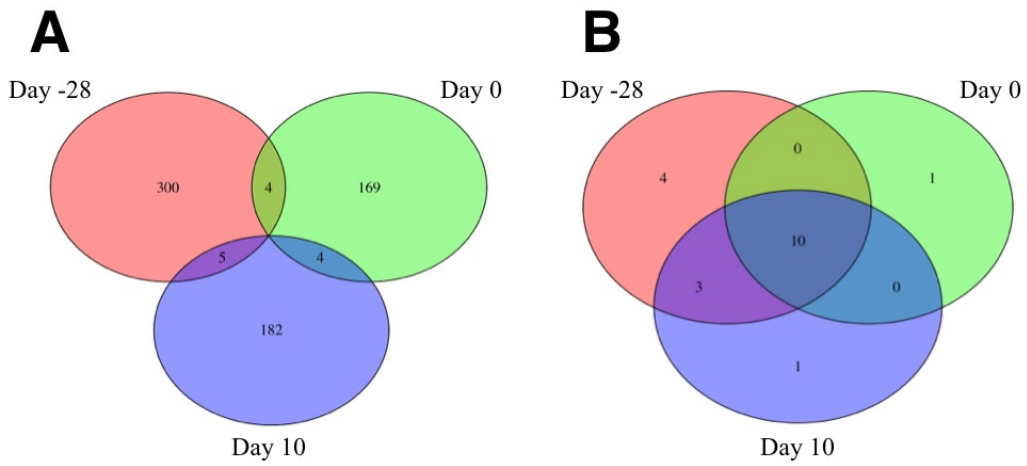

Figure 3.15 Venn diagrams of genes and their corresponding pathways selected by the pipeline for the L3 vaccination dataset. Venn diagrams showing overlap between time points in L3 vaccination dataset of A) genes from the clusters selected by the pipeline and B) "immune system process" (GO:0045087) those genes are involved in.

Over-represented pathways were identified in each of the gene lists produced by the ML pipeline. The degree of over-representation was calculated using cumulative hypergeometric distributions, which is the probability of finding genes belonging to a functional Gene Ontology (GO) category or Reactome pathway within the gene list. A P-value was calculated for each of the pathway or processes using the Benjamini-Hochberg method to correct for multiple testing (Benjamini, Hochberg, 1995).

Different numbers of GO terms associated with Biological Process (BP) were found at each time point, and different pathways were identified using the Reactome database (Table 3.11). Because there are some discrepancies between the pathways found between the two databases, both databases were used in order to obtain the best overview of what is biologically happening at each time point.

Overall the list of genes identified by the ML pipeline at the time points after immunisation had the greatest number of over-represented GO terms and pathways.

This analysis suggests, that the early events in infection are important in determining the outcome to infection and this fits with what is known from the *L. sigmodontis* literature (Babayan *et al.*, 2003), especially in terms of initial CD4+ T cells responses, which are crucial in determining immunity to later stages of infection (Taylor *et al.*, 2009).

The conclusion is that a vaccine would need to skew the immune system to a protective phenotype from the onset of infection, as once the infection is established it is hard to reverse the immune-regulatory pathways triggered, and hence why in these vaccination experiments more pathways were found over-represented at the earlier time points and not once infection has established itself.

Table 3.11. Over-represented terms and pathways across the murine datasets. The number of genes selected by the pipeline for each time point; the number of genes annotated; number of over-represented pathways using the Gene Ontology (GO) database looking at Biological Processes (BP) and Reactome database

Datasets & Time points		**No of Genes selected by the pipeline**	**No of genes annotated**	**Terms/Pathways using:**	
				GO BP Terms	**Reactome**
L3 Vaccination	Day -28	421	309	3	0
	Day 0	218	177	0	0
	Day 10	268	191	0	2
Mf Vaccination	Day -21	216	146	59	41
	Day -14	128	92	35	14
	Day 0	218	158	0	0
	Day 49	144	113	2	0
	Day 67	145	125	0	0

3.7.1.1 Informative genes and pathways in L3 immunised mice

Among the genes identified from the L3 vaccination dataset by the pipeline, biological processes were significantly over-represented at day -28 and were involved with neutrophil movement (Chemotaxis: GO:0030593; Migration: GO:1990266) (Figure 3.16A). The genes involved in these processes have increased expression in the immunised group after the 1st immunisation (day -28), followed by increased expression in the mock group after challenge (day 0), and by day 10 expression across the groups is relatively similar (Figure 3.17). This suggests that the machine learning pipeline is detecting the innate response to incoming L3 and that such responses have been demonstrated to have a major role in controlling the early stages of filarial infection in the skin (Pionnier *et al.*, 2016).

Interestingly, over-represented pathways associated with regulation of interferon (IFN) α signalling (Reactome: 5992081) were found at Day 10 (Figure 3.16). Both IFN-α and IFN-β are type I interferons, which exert multiple functions in the immune system and have various stimulatory and suppressive effects on dendritic cells (DC), macrophages, natural killer (NK) cells, T and B lymphocytes. While mostly associated with viral infections, they have recently been shown to be induced by bacteria or bacterial products, protozoa such as *Leishmania spp* and *Plasmodium spp*, and in helminths by *Schistosoma mansoni* eggs (Bogdan *et al.*, 2004). Both IFN-α and IFN-β have been shown to have immunomodulatory properties and are therefore used as a potential therapeutic for human hepatic alveolar echinococcosis (Godot *et al.*, 2003).

Four genes from the IFN-α regulation pathway were selected by the ML pipeline; two genes encoding IFN-α, Ifna5 and Ifna9; and two genes involved in the regulation of cytokine signalling Socs1 (negative feedback loops) and Ptpn11 (Sh2). However, the precise role of IFN-α and its regulation in *L. sigmodontis* infection remains unclear (Figure 3.18).

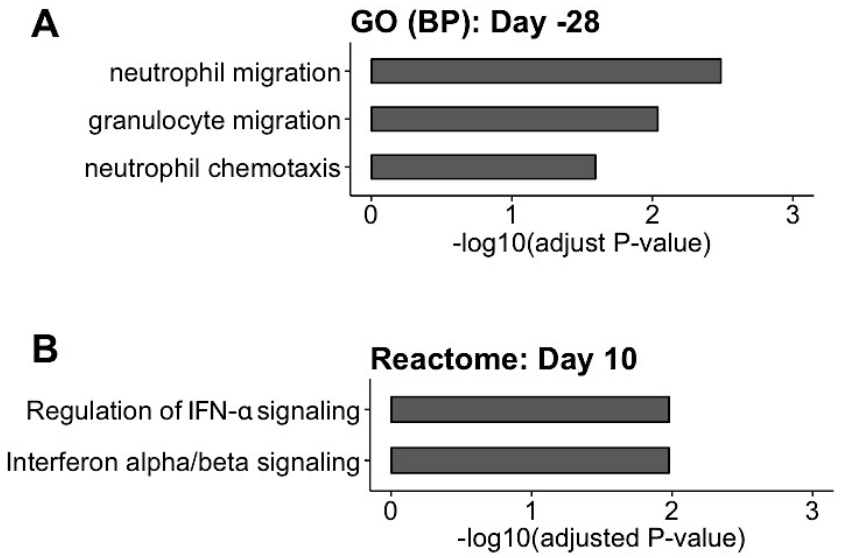

Figure 3.16 Over-represented terms found in the genes identified by the ML pipeline, in mice immunised with L3. A) Biological processes using the Gene Ontology database were found over-represented at Day -28 (6 hours post 1st immunisation) in the L3 vaccination time course, these were associated with neutrophil movement (Chemotaxis: GO:0030593; Migration: GO:1990266). B) Using the Reactome database, pathways involved in IFN-α signalling were found over-represented at Day 10. Each process or pathway is represented as an over-representation score, the –log10 of the adjusted P-value (which accounts for multiple testing using the Benjamini-Hochberg method) and denotes the significance of the terms over-represented, the higher the value the more significant the GO term was. An adjusted P-value < 0.05 and Q-value < 0.2 was used as the significance cut-off.

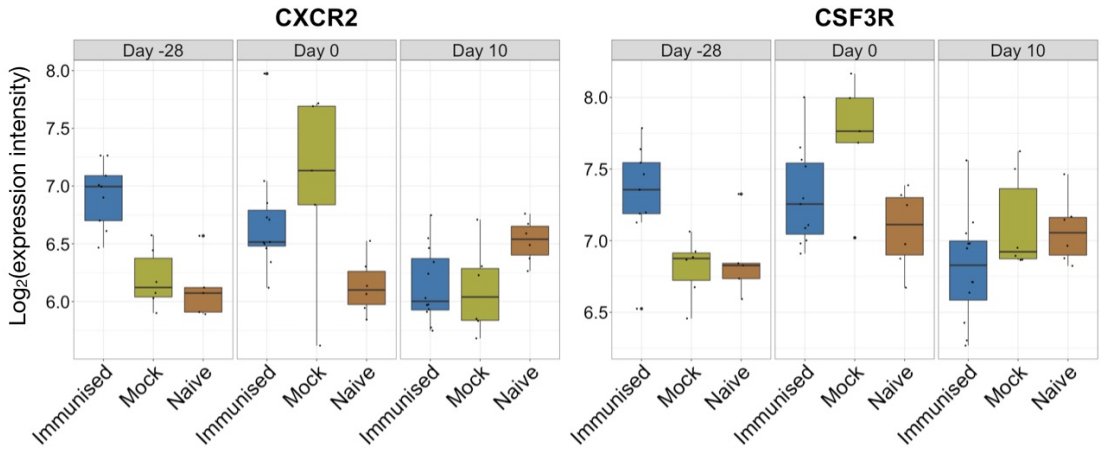

Figure 3.17 Gene expression of genes involved in neutrophil pathways across the different time points. Log_2 of expression intensity of Cxcr2 and Csf3r on day -28 (6hr after 1st immunisation), day 0 (6hrs after challenged) and day 10 time points (*L. sigmodontis* worms from challenge, that have survived would have matured to L4 stage and some will have reached the immature adult stage). Cxcr2 is a receptor for IL-8 and mediates neutrophil migration. Csf3r is a receptor for Csf3 which is a cytokine that controls the production, differentiation and function of granulocytes. Both of these genes are involved in pathways associated with neutrophil migration (GO:1990266), neutrophil chemotaxis (GO:0030593).

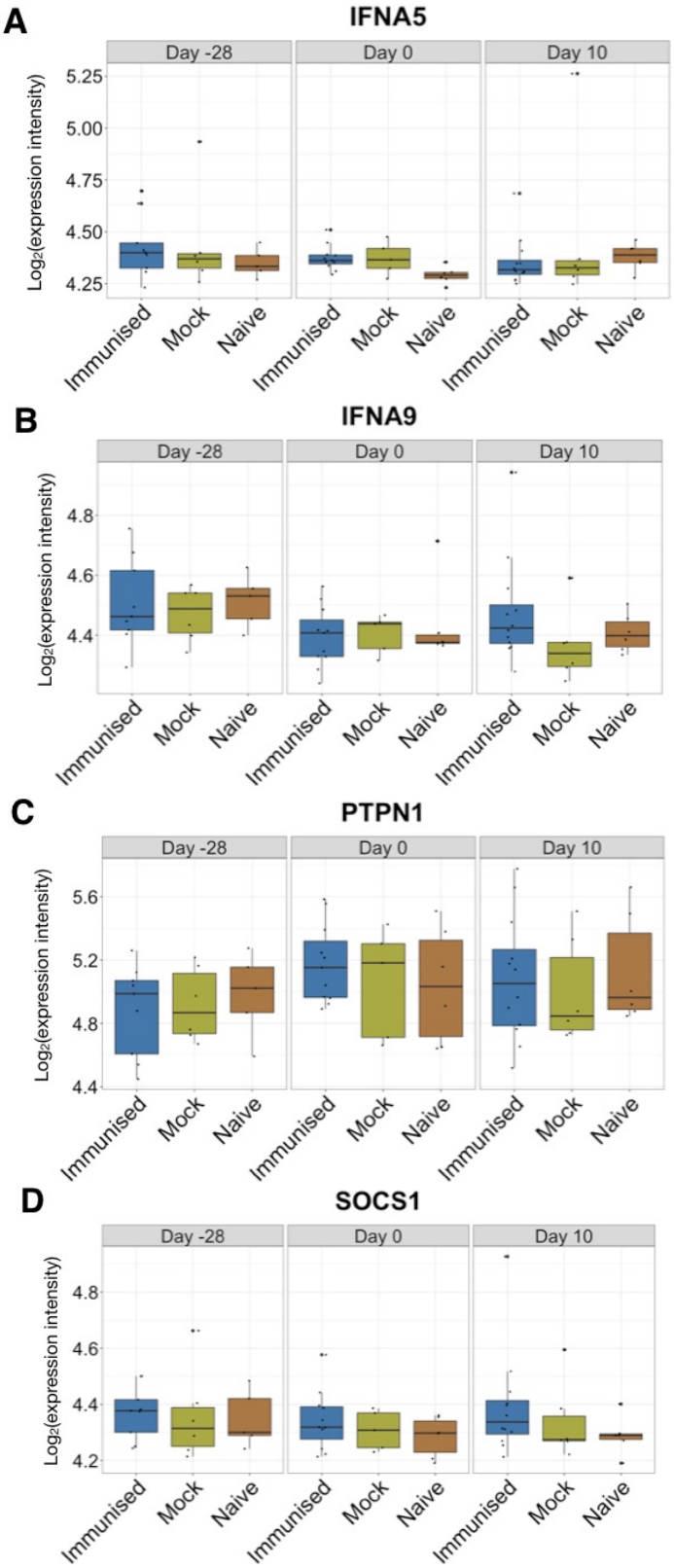

Figure 3.18. Gene expression of genes involved in IFN-α pathway across the different time points. Log_2 expression intensity of IFN-α encoding genes (A) Ifna5, (B) Ifna9; genes

involved in the regulation of cytokine signalling (C) Ptpn11 and (D) Socs1, on Day -28 (6hr after 1st immunisation), Day 0 (6hrs after challenged) and Day 10 time points. Ifna5 is a gene that encodes IFN-α5, and Ifna9 is a gene that encodes IFN-α9. Ptpn1 is a member of the protein tyrosine phosphate family, these catalyze the hydrolysis of the phosphate monoesters specifically on tyrosine residues, and are known to be signalling molecules that regulate a variety of cellular processes. Socs1 is a member of the STAT-induces STAT inhibitor and are known to suppress cytokine signalling. Altogether these genes are associated with pathways involved in the regulation of IFN-α signalling and were selected by the ML pipeline at day 10 as being informative.

3.7.1.2 Informative genes and pathways in Mf immunised mice

Similar to that seen following L3 vaccination, time points 6 hours after Mf vaccination were the most informative with over-represented biological processes found at day-21 and day -14 (Table 3.11).

Since large numbers of GO terms were found over-represented at day -21 and day -14, only the top 20 significant BP GO terms are represented in Figure 3.19, with some overlap seen between the time points (Figure 3.20). At day 0 (6 hours post L3 challenge) the ML pipeline did not identify any biological process as being differentially expressed between the treatment groups. This could be that there is little difference in response to incoming L3 between Mf immunised and the controls, and that the changes in gene expression that determine the outcome of infection happen early on after immunisation, and again once the worms have matured, since it's known that the Mf vaccine affects adult worms fertility and not the incoming L3 (Ziewer *et al.*, 2012). The full list of significantly over-represented GO categories for Biological Process are summarised in Appendix B Table S4.

The main biological processes found by the pipeline for day -21 were processes involved in the response to IFN with an emphasis in IFN-β and IFN-γ; antigen processing and presentation with processes triggering lymphocyte activation and adaptive immune responses; homotypic cell to cell adhesion of immune cells such as T cells including their activation and aggregation; and defence responses to single cell organisms such as bacteria and protozoa.

At day -14 processes such as response to IFN-β and IFN-γ; antigen processing and presentation, and defence responses to protozoan and bacteria were found over-represented. The difference between the two immunisation time points is that day -14 also had over-representation of processes involved in regulation of gene expression (epigenetics).

In addition, Day 49 had GO terms over-represented in processes involved in regulation of gene expression through epigenetic changes and chromatin silencing. Day 49 corresponds to a time when adults are found in the pleural cavity but no Mf are yet present.

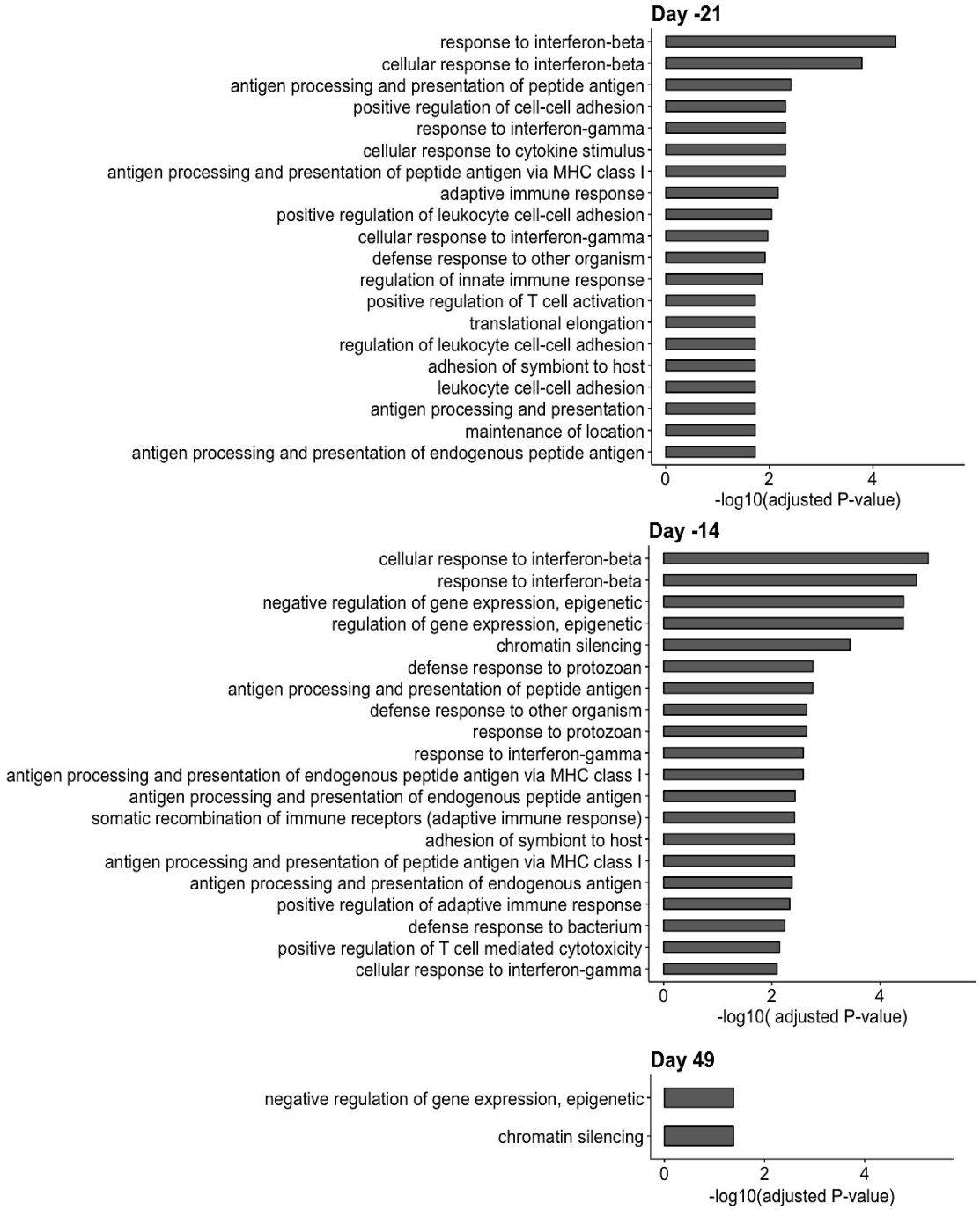

Figure 3.19. Gene Ontology Terms of biological processes in the Mf immunity dataset. Time point Day -21, Day -14 and Day 49 in the Mf immunity dataset had significantly over-represented GO terms. The over-representation score is the –log10 of the adjusted P-value (which accounts for multiple testing using the Benjamini-Hochberg method) and denotes the significance of the GO terms, the higher the value the more significant the GO term is. The significance cut-off is an adjusted P-value < 0.05 and a Q-value < 0.2.

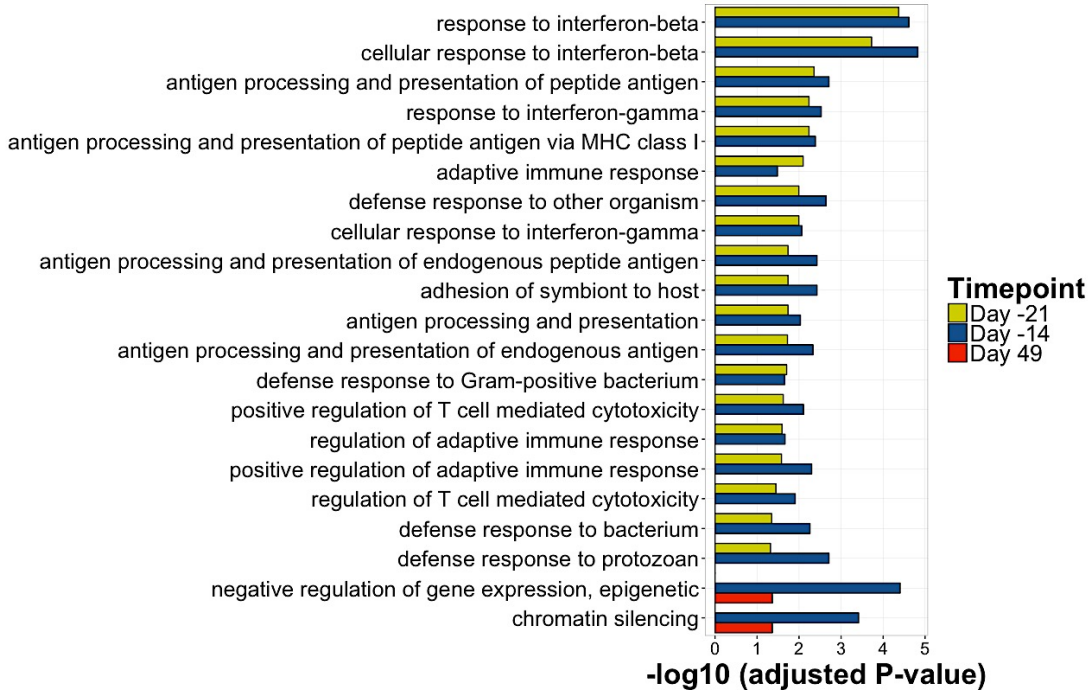

Figure 3.20. Overlap of over-represented Gene Ontology Terms for biological processes in the Mf immunity dataset. There is overlap between terms at Day -28 and Day -14; and then between Day -14 and Day 49. The data is presented as enrichment scores, which is the –log10 of the adjusted P-value (accounting for multiple testing using the Benjamini-Hochberg method) and denotes the significance of the GO terms within a gene list, the higher the value the more significant the GO term is. A GO term was classified as significantly over-represented, if the term had an adjusted P-value < 0.05 and a Q-value < 0.2.

Pathway analysis based on the Reactome database also revealed significantly over represented pathways at time points directly after immunisation, the majority of which were related to the immune system. The over-represented pathways following immunisation were mostly associated with genes involved in the initiation of adaptive immune response, in terms of antigen processing and antigen presentation (MHC I mediated); and interferon signalling, with an emphasis on IFN-γ. Furthermore, at day -21, pathways involved in T cell signalling such as T cell receptor (TCR) signalling; T

cell co-stimulation signalling through receptors of the CD28 family (including PD-1 which is a negative regulator of TCR signalling and leads to the dephosphorylation of CD3 zeta chains) were over-represented (Figure 3.21).

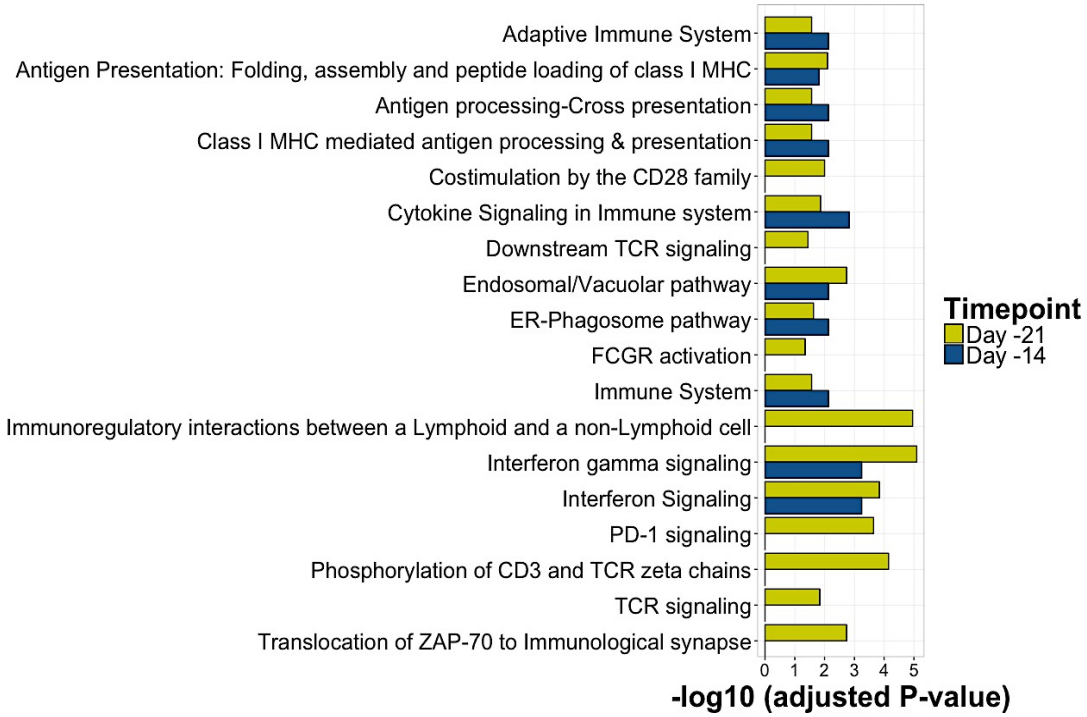

Figure 3.21. Pathway over-represented using Reactome database. Using the Reactome database over-represented pathways were found in the gene lists extracted for day -21 and day -14 of the Mf immune dataset. The data is presented as enrichment scores, which is the –log10 of the adjusted P-value (accounting for multiple testing using the Benjamini-Hochberg method) and denotes the significance of the GO terms within a gene list, the higher the value the more significant the GO term is. The analysis accounted for multiple testing using the Benjamini-Hochberg method, and each process was over-represented with a P-adjusted-value < 0.05 and a q-value < 0.2.

All together the GO and Reactome pathways found over-represented highlight the fact that time points after immunisations are most informative of immunisation and

consequently to the outcome of infection. Therefore, measuring changes in gene expression at the end of a vaccination time course would not be as informative as was seen in chapter 2. Despite there being some discrepancies between the pathways identified between the two databases (GO and Reactome), there are common themes.

Responses to IFN seemingly play an important role at the earlier time points, as genes involved in these processes have a similar expression across the time point, showing a response to Mf, with higher expression after immunisation with Mf and again at Day 67 in groups that have higher Mf burdens (Figure 3.23). This has been demonstrated *in vivo*, where single stage infection with Mf in mice induces the production of IFN-ɣ (Lawrence *et al.*, 2000). Genes involved in IFN-β pathways were also detected; although commonly associated with antiviral immunity IFN-β has been shown to have immunomodulatory properties in bacterial infections, whereby DC produce IFN-β upon activation by LPS, leading to induction of T cell tolerance by enhancing PD-1 binding and therefore facilitating Treg generation (Wang *et al.*, 2014), or where *Schistosoma mansoni* eggs induce myeloid DC to produce IFN-β as a potential immune evasion strategy (Trottein *et al.*, 2004), but there is no documented role of IFN-β in filarial infections. Another possible explanation is that there is a lot of overlap between the genes involved in interferon signalling pathways (IFN-β, IFN-γ and IFN-α), and because the genes selected by the pipeline are involved in both and are not specific to either pathways, they are shown as over-represented. Comparing the gene assigned to each pathway shows that indeed they are very similar genes and are shared by both IFN-γ and IFN-β pathways. What can be concluded is that Mf trigger IFN

responses, with a possible novel role of IFN-β but further *in vivo* experiments would need to be done to confirm this.

Pathways involved in initiation of the adaptive immune responses were also upregulated after immunisation. These include antigen processing and presentation, and T cell recruitment and activation. Although most of the genes are associated with antigen presentation are genes that form the MHC class I molecules (H2 class I genes, orthologue for HLA-A genes in humans), all had greater expression when Mf were present (Figure 3.24), suggesting that Mf antigens are being cross-presented to $CD8^+$ T cells, and that these might play a larger role in innate (initial immune) response to Mf, which has been suggest to be the case for modulation of infection in chronically infected *Loa loa* (Steel *et al.*, 2012).

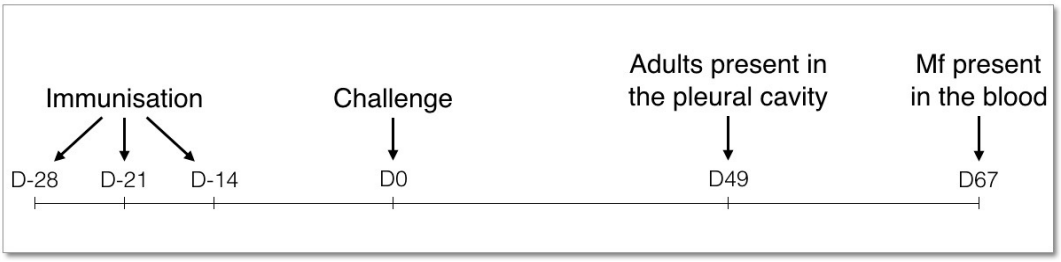

Figure 3.22. Timeline of Mf immunisation time course. Day -28, -21 and -14 before the challenge infection, mice received immunisation, with either Mf on their own, Mf in alum, alum alone as a control, or media as a second control. At day 0, mice received a challenge infection with L3. By day 49 adult worms should be present in the pleural cavity of the mice and at day 67, Mf should be found circulating the blood, this also corresponds to the end of vaccination experiments.

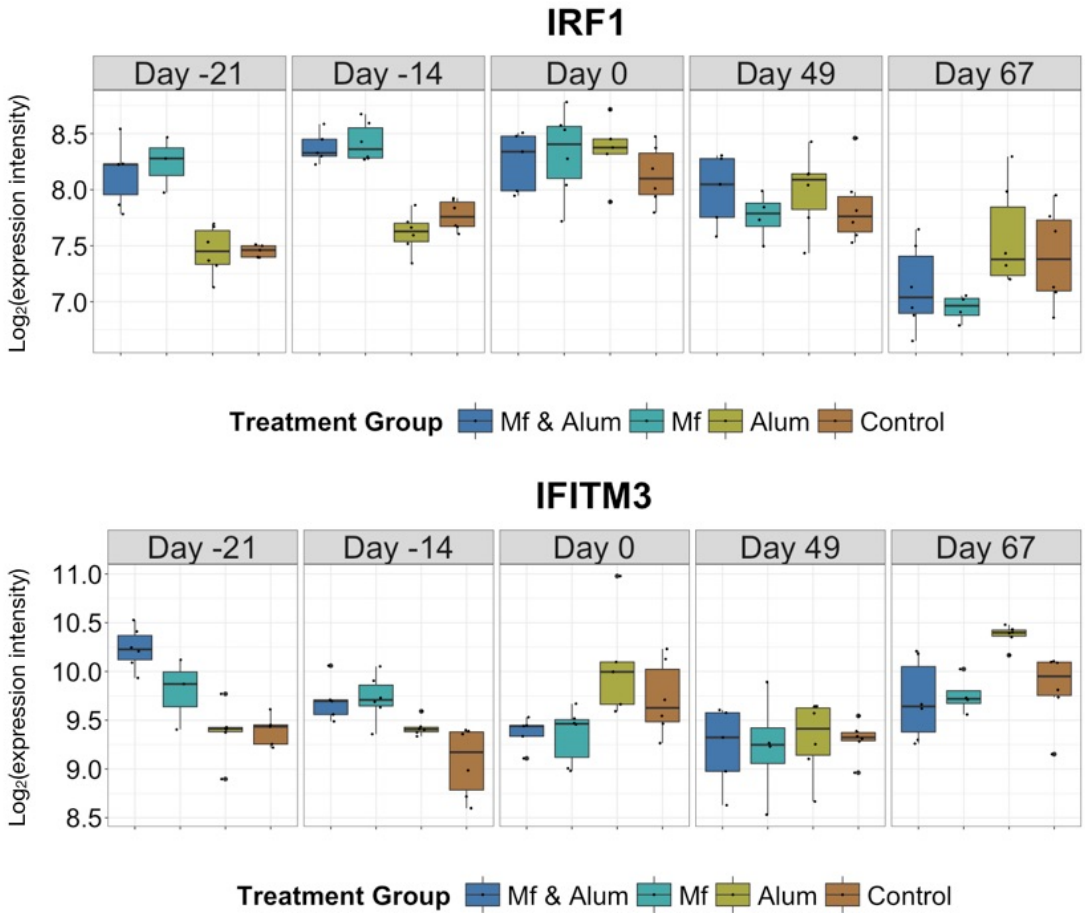

Figure 3.23. Expression of Ifitm3 and Irf1 genes throughout the Mf vaccination time course. Interferon regulatory factor 1 (Irf1) is a gene that encodes a transcription factor and interferon-induces transmembrane protein 3 (Ifitm3) encodes a gene that is an interferon induced membrane protein that is most known to be involved in immunity to virus. Both genes are involved in pathways responsible for responses to IFN-γ and IFN-β (IFN-γ: Gene Ontology GO:0034341, Reactome: R-MMU-913531.1; and IFN-β GO:0035456). Other genes were selected by ML pipeline that were involved in these pathways (B2m, Cdc37, Gbp2b, Gbp2, Gbp3, H2-D1, H2-Q8, H2-T23, Ifitm3, Irf1, Irf7, Sp100) and all follow a similar pattern of expression, with higher expression when Mf are present such as after immunisation and again at day 67 in groups with higher Mf burdens.

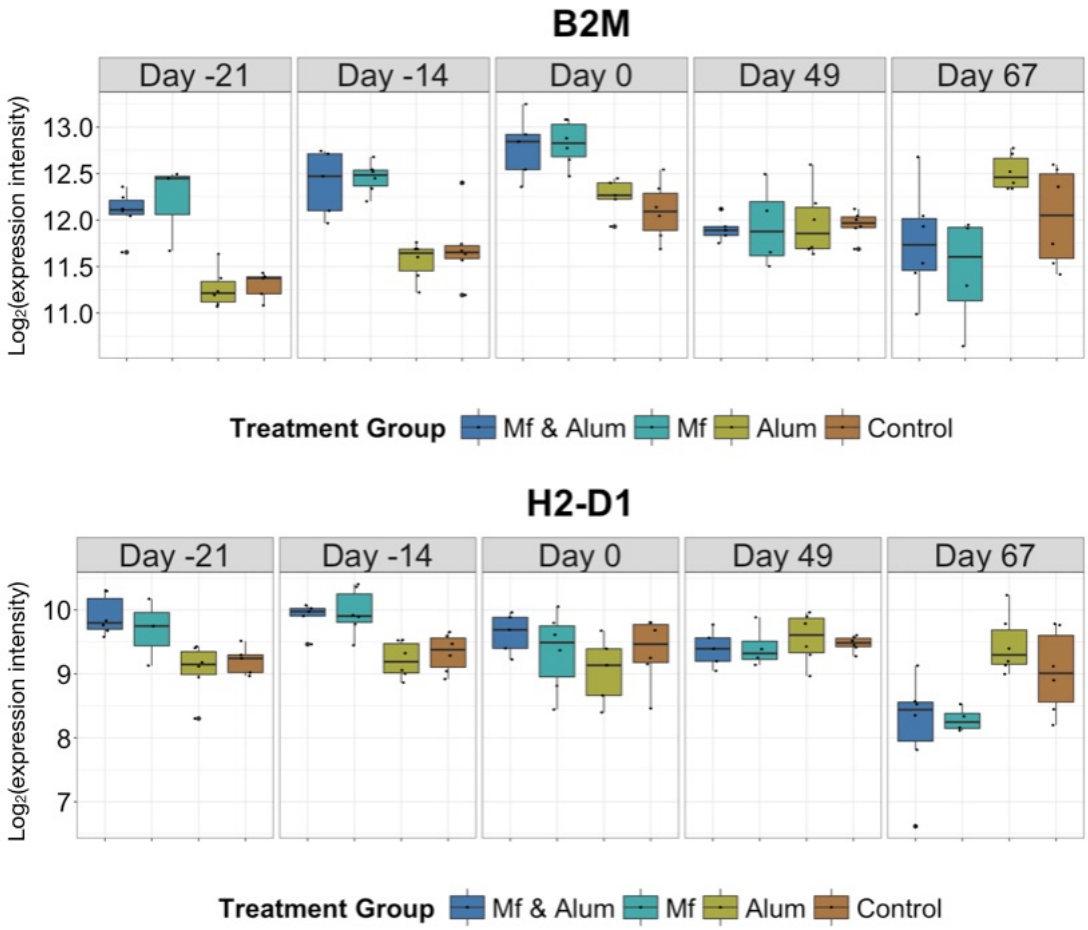

Figure 3.24. Expression of B2m and H2-D1 genes throughout the Mf vaccination time course. Beta-2 microglobulin (B2m) is a gene encoding a component of MHC class I molecules and histocompatibility 2, D region locus 1 (H2-D1) is a gene encoding is a component of MHC class I molecules. Both genes are involved in pathways associated with antigen processing and presentation, and T cell activation, this is because these genes form part of the MHC class I structure, other genes were also selected by the pipeline that form this complex (B2m, H2-K1, H2-Q6, H2-Q8, H3-T23), and all follow a similar pattern of expression, with higher expression when Mf are present such as after immunisation and again at day 67 in groups with higher Mf burdens.

3.7.2 Human Filariasis

To find pathways associated with protection in human onchocerciasis and lymphatic filariasis, the ML pipeline was used initially on all 162 and 172 samples from the *O. volvulus and W. bancrofti* datasets respectively, however no significantly over-represented pathways where found in *W. bancrofti* dataset, and only pathways associated with cell cycle were identified in *O. volvulus*.

Clinical data from human samples is inherently noisier than data collected from controlled experiments (e.g. mice vaccination experiments). Samples from patient with co-infections, either from other helminths (i.e. hookworms or *Ascaris*), or protozoa infections (malaria or *Giardia*); and patients who had received Ivermectin treatment were removed from the analysis. This left 85 and 71 samples for the *O. volvulus* and *W. bancrofti* dataset respectively (Table 3.12). When the pipeline was applied to these "cleaner" samples, a higher classification accuracy and stability was achieved. Over-represented pathways were found in the genes identified as being associated with protection by the ML Pipeline for both *O. volvulus* and *W. bancrofti* individuals (Table 3.12).

Table 3.12. Number of genes identified as important in distinguishing infected and protected individuals in *O. volvulus* and *W. bancrofti* endemic areas. The number of genes selected by the pipeline for each human datasets; number of annotated genes; number of over-represented pathways/process in the gene list produced by the ML pipeline, using Gene Ontology Biological Process and Reactome database.

Human Datasets	No Genes selected by ML pipeline	No genes annotated	Over-represented Terms/Pathways using:	
			BP GO Terms	Reactome
O. volvulus	275	198	53	12
W. bancrofti	195	128	0	4

Onchocerca volvulus endemic area dataset

Overall 53 GO terms were found over-represented in the gene list produced by the pipeline from *O. volvulus* dataset, and 12 using the reactome database (Figure 3.25). These terms were summarised based on their function to get an overview of processes important in onchocerciasis infections:

- Killing of cells of other organisms (GO:0031640):

  _ Defence responses to fungus (GO: 0050832) and Gram-positive bacterium (GO:0050830)

  _ Modification of morphology or physiology of other organism (GO:0035821), these are process that are involved in killing a variety of organisms

  _ Innate immune responses in mucosa (GO: 0002227)

  _ Acute inflammatory responses (GO:0002526)

_ Negative regulation of growth of symbiont involved in interaction with host (GO:0044146)

- Humoral immune response (GO:0006959):

_ Organ or tissue specific immune response (GO:0002251) – an immune response happening in an organ or tissue, such as liver, brain, mucosa or nervous system

_ Leukocyte mediated immunity (GO:0002443) -including DC

- DNA-dependent DNA replication (GO:0006261):

_ Anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process (GO: 0031145) - chemical reactions and pathways resulting in the breakdown of a protein or peptide by hydrolysis of its peptide bond.

- DNA conformational change (GO:0071103)
- Chromosome segregation (GO:0007059)
- G1/S transition of mitotic cell cycle (GO:0000082)
- Cell killing (GO:0001906)
- Negative regulation of protein binding

Wuchereria bancrofti endemic area dataset

In the lymphatic filariasis dataset, no GO terms and 4 pathways using Reactome database were found significantly over-represented, all associated with T cell receptor signalling (Figure 3.25). The genes involved in these processes (HLA-DP and HLA-DR) had higher expression in infected individuals compared to endemic normal (Figure 3.26). These genes are part of the MHC class II cell surface receptors, which is the ligand for CD4 T cell receptors and therefore associated with T cell receptor signalling, which can either trigger T cell activation or suppression.

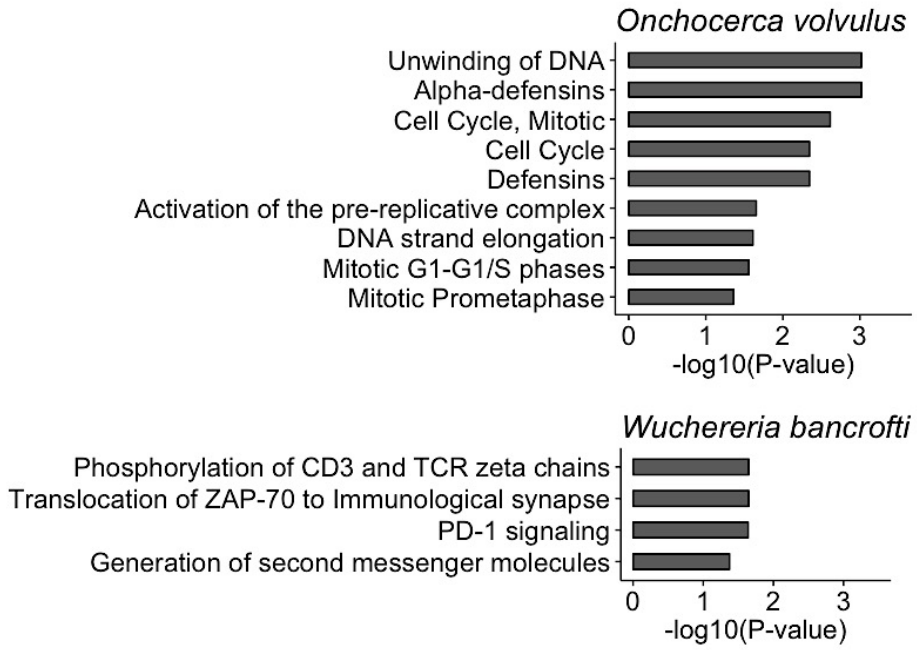

Figure 3.25. Over-represented pathways were detected in both Onchocerciasis and Lymphatic Filariasis datasets. Using the reactome database over-represented pathways were detected in the gene list produced by the ML pipeline for both the onchocerciasis and lymphatic filariasis datasets, (these were gene expression for individuals living in *O. volvulus* or *W. bancrofti* endemic areas respectively). The data is presented as over-representation scores, the –log10 of the adjusted P-value (accounting for multiple testing using the Benjamini-Hochberg method) and denotes the significance of the GO terms within a gene list, the higher the value the more significant the GO term is. The analysis accounted for multiple testing using the Benjamini-Hochberg method, and significance defined as a P-adjusted-value < 0.05 and a q-value < 0.2.

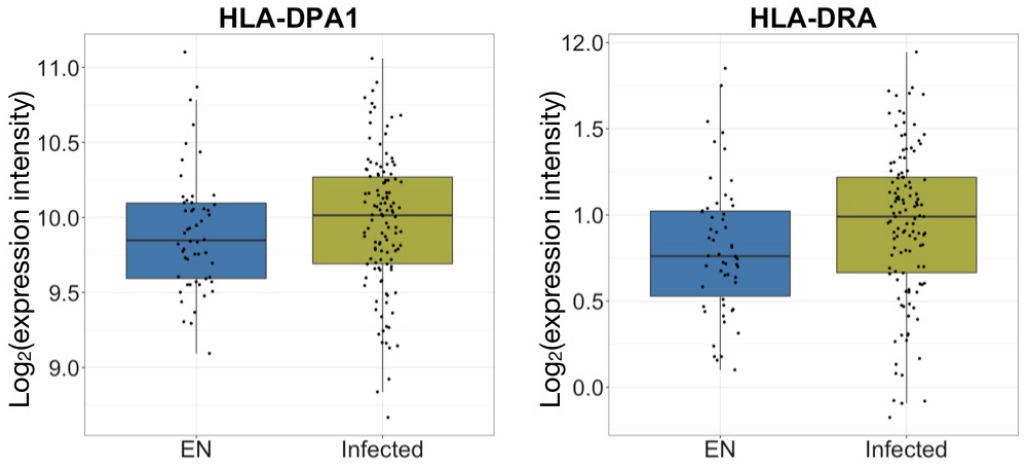

Figure 3.26. Gene expression of HLA-DPA1 and HLA-DRA in individuals living in *W. bancrofti* endemic area of Ghana. Both of these genes are associated with T cell receptor signalling and are in fact part of the MHC class II molecule. Both HLA-DPA1 and HLA-DRA have increased expression in individuals infected with *W. bancrofti* compared to endemic normal. GLM's were used to fins statistical differences between the groups, but no statistical difference was found.

3.7.3 Overlap between mice and human protective immunity

As one of the aims of this study was to investigate if there was any overlap in protection in humans living in endemic areas and immunised mice, pathways associated with protection were compared between both systems. Pathways associated with Mf immunity, such as defence against bacteria were found to be associated with protection in individuals living in *O. volvulus* endemic areas, these pathways could be being triggered by *Wolbachia* found in *L. sigmodontis* Mf and *O. volvulus*. By contrast T cell receptor signalling pathways were found to be important for protection in both individuals living in *W. bancrofti* endemic areas and in *L. sigmodontis* Mf immunisations (Figure 3.27).

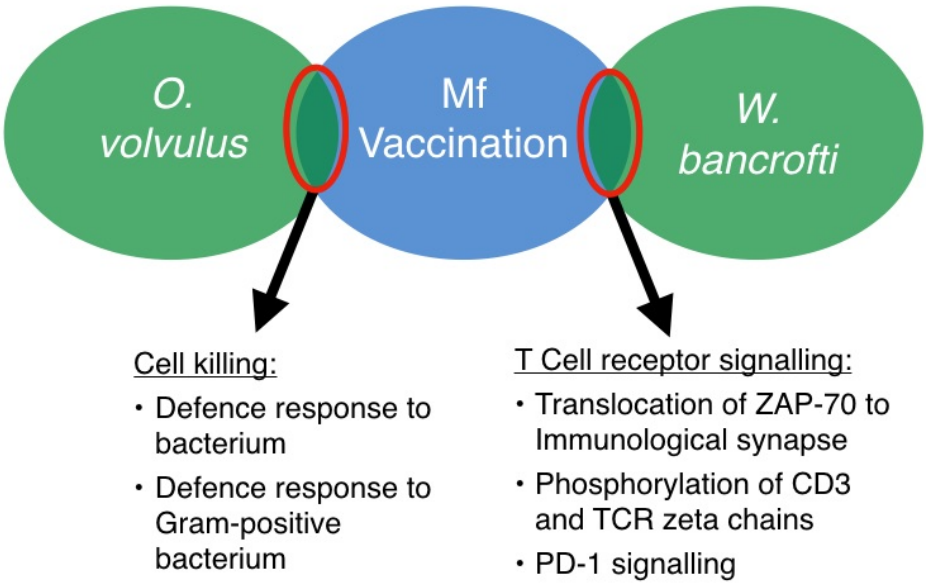

Figure 3.27. Representation of overlap in pathways between human datasets (*W. bancrofti* and *O. volvulus* endemic areas) and Mf vaccination dataset. Overlap in GO terms and reactome pathways that are identical matches between human datasets and mice datasets, only the Mf vaccination out of both murine models had an identical match. Genes involved in cell killing mostly associated with bacteria killing were over-represented in both *O. volvulus* and the immunization time points of the Mf vaccination dataset, and genes involved in T cell receptor signalling were found over-represented in both *W. bancrofti* and the immunization time points of the Mf vaccination dataset.

3.8 Conclusions and future uses

Microarray array technology has become an important tool to dissect the complex mechanisms involved in diseases, in this study microarrays were used to identify the underlying pathways associated with protection, induced in vaccination experiments in *L. sigmodontis* a murine model of filariasis, and in humans living in either an onchocerciasis or lymphatic filariasis endemic area. Gene expression was measured in whole blood as this is the most convenient sampling method for both humans and murine studies. The disadvantage of starting with whole blood is that the "background noise" is relatively high. Lower backgrounds can be obtained if isolated cell populations are used (Bondar *et al.*, 2014) however, this is not a realistic proposition in the context of either the murine or human studies described here, and systemic responses would be overlooked.

To analyse the microarray data, machine learning techniques were used to create a pipeline to identify informative gene expression in murine and human datasets that may differentiate between protected or infected individuals. The advantages of a machine learning pipeline over traditional and popular methods such as differential fold changes or gene co-expression networks, is that the pipeline can incorporate samples with multiple experimental groups and is not restricted to comparing two groups or disease states. In addition, this approach can handle large multi-dimensional data sets and is robust at detecting small changes in gene expression.

The ML Pipeline uses publicly available algorithms including a density based clustering algorithm to overcome the problem of dimensionality that afflicts microarray data. This algorithm successfully cluster genes by their expression pattern,

and identifies genes that do not cluster. The genes within a cluster were "summarised" so that a new variable can represent the expression of many genes within a cluster. The summarised gene clusters were then used to build a random forest model that ranks the gene clusters in order of importance, i.e. by how well they discriminate between the experimental groups in mice (immunised, mock or naïve) or infection status in humans (infected or EN). Random forest was repeated over several iterations and the gene clusters that routinely fell into the top 20% at each iteration were selected as being "informative" gene clusters. The function of the individual genes within the informative clusters were determined using pathway analysis. Compared to other machine learning, the ML Pipeline provides greater accuracy and is more flexible than recursive methods such as RF-RFE and SVM-RFE.

Biologically relevant genes were successfully identified from whole blood gene expression data from mice vaccinated with either live attenuated infective L3 larval stage or microfilariae. Furthermore, it was shown that these different vaccination protocols elicited protection through different mechanisms. Thus, in the case of L3 vaccination, protection was associated with neutrophil recruitment and migration. Such a response has also been demonstrated by Pionnier et al (2016) who showed that subcutaneous delivery of L3 resulted in an increase in neutrophils in the skin and which is essential for the early control of infection (Pionnier *et al.*, 2016). In contrast, protection associated with Mf immunisation involved IFN signalling and IFN responses which corresponds to what is seen *in vivo* where mice immunised with Mf had stronger IFN-γ responses compared to controls (Ziewer *et al.*, 2012) and IFN-γ

RNA levels in re-stimulated splenocytes were increased after the onset of patency in *L. sigmodontis* infected BALB/c (Taubert, Zahner, 2001).

The results presented in this study clearly show that genes identified by the ML pipeline are biologically relevant to filarial infections and the use of whole blood provides a convenient and workable starting point for additional studies.

The ML pipeline also demonstrated that adaptive immune responses are being triggered 6-hours post immunisation. Following the final (3^{rd}) immunisation (Day -14), antibody production by B cells is triggered (GO:0002822, somatic recombination of receptors). The results also may suggest that Th1 responses such as IFN-α/β and MHC I antigen cross-presentation might play a greater role in filarial immunity than has been reported in the literature, at least during the initial stages of exposure and infection.

A comparison of human and murine datasets showed that protective responses evoked in mice by immunisation with microfilariae most closely resembled protective pathways induced in human through infection with either *O. volvulus* or *W. bancrofti* infections. This is not to say that protective responses evoked by L3 vaccination in mice may not be mirrored in humans. To investigate this possibility would however be difficult because it would require collecting blood shortly after an initial exposure and in practice, this would mean taking samples from very young children. This is a critical gap in our knowledge of human immune responses to filarial infections and one crucial for the successful development of a vaccine. It is a gap that may be closed by analysis

of blood samples taken from cattle infected with *O ochengi* under conditions of known exposure (Makepeace, Tanya, 2016).

The ML pipeline does have certain limitations, one of them being sample size. The dataset used have relatively large sample sizes compared to other microarray studies using traditional statistical methods, which usually contain 2 or 3 samples per treatment group (Zhou *et al.*, 2016), but our data sets are smaller than many studies using machine learning techniques where data analyses rarely contains less than 100 samples. Increasing sample size will increase accuracy and stability (Kim, 2009) but while sample sizes of 500-800 are suggested, such numbers would be prohibitively expensive and unfeasible in the case of experimental mouse studies.

To mitigate against the relatively small numbers, murine studies could be better designed to maximise difference between control and test groups. For example, in the case of filarial infections, this would mean comparing mice presenting with specific and divergent parasite (Mf) loads. Furthermore, now that it is known that time points after immunisation are the most informative, gene expression could ideally be measured after each after immunisation time point and then after challenge.

One of the major issues that concerns vaccine development is adverse reaction to the vaccination, since filarial infections can present with severe Th2 driven pathology (Brattig, 2004; Babu, Nutman, 2012). Therefore, detecting genes or pathways that determine the outcome of pathology will be important in developing a safe vaccine that would lead to protection without any pathology.

The work presented in this chapter demonstrates that machine learning techniques can be applied to highly dimensional microarray data from whole blood to understand the molecular events that happen *in vivo* to filarial infections. Such methods could be used to identify biomarkers of protection to vaccination, which would benefit vaccine trials (Nakaya *et al.*, 2011a) particularly if such markers could predict outcome of vaccination within a relatively short period (e.g. within 3-4 weeks of final immunisation). Such characteristics would be particularly helpful in the case of onchocerciasis infections because of the long incubation period (18 months). The ability to accurately predict the immunogenicity of a vaccine within the first few days or weeks of a clinical trial would help determine the endpoint any clinical trial and ultimately save money and time. However, any biomarkers of protection would have to be well defined, as correlates of efficacy in healthy young adults may not be necessarily the same as biomarkers of protection in the elderly, or children under 5, which are the primary targets for most vaccination campaigns.

Chapter 4. Structural modification of the CPI immunomodulator rescues DC function.

4.1 Introduction

Traditional vaccine development has largely been based on the use of inactivated or attenuated vaccines without knowing the vaccine's exact mechanisms of protection. Despite attenuated vaccines (irradiated L3) for filarial parasites showing protection in animal models, it would not only be unethical but also unfeasible for use in humans. Thus, vaccine development has turned to the use of subunit vaccines, based on parasite antigens, however filarial nematodes are complex parasites, and the chances of finding effective antigens which confer protective immunity are low. Fortunately, knowledge of the immune responses induced by the filarial parasites can be used for a more rational approach to vaccine design. It is recognised that protection against filarial parasites is dominated by Th2 immune responses with some Th1 responses induced by *Wolbachia* (present within the nematodes); however filarial parasites induce a state of hypo-responsiveness that predisposes towards a chronic infection. Therefore, a vaccine strategy for filarial parasites has been to target the parasite's immunomodulators that induce this state of hypo-responsiveness, as well as evoking a Th2 response.

Vaccine candidates were selected based on their role in immunomodulation and their potential to induce protective immunity in several animal models of filariasis. With this strategy, a cysteine protease inhibitor (CPI) was identified (Table 4.1). In the *L. sigmodontis* model, *Ls*-CPI vaccine had no effect on protection nor was it able to generate strong specific immune response (Babayan *et al.*, 2012). However, when structurally modified through a single point mutation (Murray *et al.*, 2005), CPIm has

consistently proven to be more protective than the native form of CPI. CPIm induces protection and increased immune stimulation (Table 4.1), in immunisation experiments with: a DNA vaccine in BALB/c mice against *L. sigmodontis* (Babayan *et al.*, 2012)*;* a recombinant protein vaccine in gerbils against *B. malayi* (Arumugam *et al.*, 2014b); or in chamber model in BALB/c mice against *O. volvulus* (Hess *et al.*, 2014).

Table 4.1. Immunisation experiments with CPI and mutated CPI (CPIm).

Parasite	Model	Protection following CPI vaccination	Protection following CPIm vaccination
B. malayi	Jirds (permissive host)	No protection but did alter adult worm migration and final niche location (Arumugam *et al.*, 2014a)	~48 reduction in worm burden, as well as reduced female fecundity (Arumugam *et al.*, 2014b)
L. sigmodontis	BALB/c mice (permissive host)	No change in circulating Mf and failed to generate strong specific immune responses. (Babayan *et al.*, 2012)	Some reduction in circulating Mf on its own and 90% decrease in Mf when combined with other antigen. (Babayan *et al.*, 2012)
O. ochengi	Cattle (permissive host)	In combination with other antigens, 42% decrease in Mf numbers in dermis. (Makepeace *et al.*, 2009)	
O. volvulus	BALB/c mice – Diffusion chamber	~49% reduction in L3 survival (Hess *et al.*, 2014)	

Elucidating the mechanism that underlies vaccine induced protection can not only provide insight into the effectiveness of the vaccine candidate, but will also have implications for formulation of other and more efficacious vaccines. CPI is known to interfere with antigen processing in the histocompatibility complex class II (MHC-II) antigen pathway *in vitro* (Manoury *et al.*, 2001). In different nematodes such as *Heligmosomoides polygyrus* and *Nippostrongylus brasiliensis*, CPI is able to modulate

the differentiation and activation of bone-marrow-derived dendritic cells (BMDC); interferes with antigen and MHC-II molecule processing; and interferes with Toll-like receptor (TLR) signalling pathways, resulting in functionally deficient dendritic cells (DC) (Sun *et al.*, 2013). This may render DC ineffective at initiating strong adaptive immune responses, similar to effect of *B. malayi* L3 on DC *in vivo* (Sharma *et al.*, 2016). Therefore, it was proposed that CPIm induces greater protection in immunisation experiment, by rescuing DC activity, which leads to increased protective immune responses, seen as a decrease in parasite burden at the end of a vaccine trial.

This would fit with what was previously observed, that mechanisms induced 6 hours after immunisation are the most predictive of the outcome of infection (Chapter 3). These observations also agree with earlier work with *L. sigmodontis*, which identified early changes in immune responses to be determinants of the outcome of infection, with either the induction of protective or regulatory phenotypes (Taylor *et al.*, 2009; Babayan *et al.*, 2010). Since DC are professional antigen presenting cells and are key for the development of an adaptive immune response, it was suggested that DC are involved in the induction of either adaptive Th2 responses (Balic *et al.*, 2004; Smith *et al.*, 2012; Guigas, 2014; Cook *et al.*, 2015) or modulatory Treg responses (Carvalho *et al.*, 2009; Everts *et al.*, 2010). Therefore, an intervention, such as a vaccine, that can boost DC responses at the early stages of infection, could provide protection and reduce pathology.

The present study was designed to examine how *Ls*-CPIm vaccine, previously shown to be protective, affects the early stages of the adaptive immune response, and therefore compare the effects of *Ls*-CPI and *Ls*-CPIm on antigen presenting cell function both

in vivo and *in vitro*. Consistent with the hypothesis that mutation of CPI provides an early advantage in activation of DC, *in vitro* increased expression of cell surface MHC-II and CD86 were detected, as well as an increase in IL-6 and IL-12p40 production when DC were loaded with Ls-CPIm as compared to *Ls*-CPI.

4.2 Methods & Materials

4.2.1 Ethics statement

All procedures involving animals were approved by the University of Edinburgh and the University of Glasgow ethical review committee, and performed under license from the UK Home Office in accordance with the Animals (Scientific Procedures) Act 1986.

4.2.2 Mice

All mice used to prepare bone marrow-derived dendritic cells (BMDC), harvest naïve T cells and in the vaccination experiment were female BALB/c mice that were obtained from either the Anne Walker Animal unit of the University of Edinburgh or purchased from Charles River (UK). Mice were housed in individually ventilated cages (IVC) at either the University of Edinburgh or the University of Glasgow. In each vaccination experiment the treatment groups were randomly allocated to avoid any cage effects. All mice were between 6-8 weeks of age before the start of any procedure.

4.2.3 Generation of bone marrow-derived DC

Bone marrow-derived dendritic cells (BMDC) were collected by flushing out the bone marrow of the femurs and tibias of BALB/c mice with PBS using a 1ml syringe fitted with 23G needle. Bone marrow was then suspended in 5ml of complete-RPMI (cRPMI): RPMI-1640 (Sigma-Aldrich); 10% heat-inactivated foetal bovine serum (FBS) (Sigma-Aldrich); 100 U/ml of penicillin-streptomycin (Gibco); and 1% of 1 x

L-glutamine (Gibco). Red blood cells in the suspended bone-marrow were lysed by adding 2ml RBC lysis buffer (Sigma-Aldrich) for 5 minutes and lysis stopped using 10ml of cRPMI, and cells were then collected by centrifuging at 400 x g for 5mins at 4°C. Cells were re-suspended in 5ml cRPMI which was further supplemented with 20 ng/ml of granulocyte-macrophage colony-stimulating factor (GM-CSF, PeproTech), and the cell concentration was determined using a haemocytometer (VWR).

The cells were suspended to a concentration of $2x10^5$ cells/ml in cRPMI with GM-CSF, and cultured in batches of 10ml for 10 days at 37°C with 5% CO2 by: feeding the cells with 10ml of fresh cRPMI and GM-CSF on day 3; and then changing the media on days 6 and 8, by removing 9ml of media (carefully not to aspirate any cells) and adding 10ml of fresh cRPMI and GM-CSF. On day 10, non-adherent cells were harvested by aspirating all the media without removing any cells stuck to the plates, adherent cells were washed by centrifuging at 400 x g for 5mins at 4°C, and re-suspended in cPRMI containing 5ng/ml of GM-CSF; these cells were used as immature DC.

4.2.4 *In vitro* BMDC stimulation assays

The immature DC harvested on day 10 were re-suspended in a 96-well U-bottom plate (ThermoFisher) at $1.1x10^5$ cells/well and stimulated for 24hrs (day 11) with either: a) 50 μg/ml of recombinant proteins CPI or CPIm treated with 10μg/ml polymyxin B (PmB, Sigma-Aldrich) to ensure no residual contamination with liposaccharides (LPS); b) 50 μg/ml of ovalbumin (OVA) also treated with PmB as a Th2 responses control; c) 0.1μg/ml of LPS (Sigma-Aldrich) on its own as Th1 response control; or d)

left unstimulated by adding cRPMI. After 24 hours of DC stimulation, cells were centrifuged at 400 x g for 5mins at 4°C and half of the supernatant fluid was collected for cytokine quantification, and the remaining supernatant was used to re-suspend the cells for T-cell co-cultures.

All recombinant proteins used were produced by Creative Biomart (Shirley, NY, USA) by cloning CPI and CPIm into pET24a and pET28a respectively. These plasmids were used to transform *E. coli* BL21 (DE3) cells, and expression of encoded proteins was induced by incubation in the presence of IPTG for 12 hours at 16°C. Expressed proteins were detected by SDS-PAGE and purified from bacterial lysates by affinity over a polyhistidine column.

4.2.5 T cell isolation and co-culture with BMDC

Spleens were harvested from BALB/c mice and placed in 5ml cRPMI. Naïve T cells were isolated from the spleens by first homogenising spleen cells between two pieces of 70µm nylon mesh (Fisher Scientific) using forceps to obtain a single cell suspension. Red blood cells were then removed from the single cell suspension by adding 2ml of RBC lysis buffer for 5mins; the reaction was stopped by adding 10ml of cRPMI and centrifuged at 400 x g for 5mins at 4ºC. The supernatant was poured off and cells re-suspended in 2ml of cRPMI to determine their concentration using a haemocytometer.

The Pan T cell isolation kit II (MACS) was used to isolate T cells from the spleen cell suspension by depleting non-target cells. The first step was to label non-target cells with a cocktail of biotin-conjugated monoclonal antibodies against CD11b, CD11c,

CD19, CD45R, CD49b, CD105, MHC class II and Ter-119. Once the cells were labelled with biotin-conjugated antibodies, anti-biotin monoclonal antibodies conjugated to MicroBeads were added. The cells were then washed through a magnetic MACS® Column in the magnetic field of a MACS Separator. The labelled cells are retained within the column and the wash through was collected as unlabelled T cells. The concentration of T cells was determined using a haemocytometer and re-suspended to a concentration of $1x10^4$cell/µl in T cell buffer. T cell buffer: sterile 1xPBS (Gibco™) containing 0.5% bovine serum albumin (BSA, Sigma-Aldrich), 2mM EDTA (UltraPure™, ThermoFisher), pH 7.2, and degassed using a sterile bottle top filter with a 0.45µm pore size (Nalgene™, ThermoFisher).

For DC and T cell co-cultures, naïve T cells were added to the stimulated BMDC ($1x10^6$ cells/well) in 96 well plates, at a ratio of 1:10 of DC to T cells ($1x10^7$ cells/well). The T cells and DC were left in co-cultures for either 24 hours, 48 hours or 96 hours, and after each time point the cells were centrifuged at 400 x g for 5mins at 4°C, and the supernatant fluid was collected for cytokine quantification by ELISA and cells harvested for FACS analysis.

4.2.6 Immunisation protocol

All DNA plasmids for immunisation were cloned following the recommendations of the pcDNA 3.1 Directional TOPO Expression Kit (Invitrogen). CPI (AF229173.1) was amplified from *L. sigmodontis* cDNA, and the mutation in CPIm was obtained by site directed mutagenesis, as previously describe (Babayan *et al.*, 2012). The pIL-4 and pMIP1α DNA vaccines used as adjuvants were also created by Babayan et al.

Table 4.2. Plasmids used in vaccination

Plasmid Name	Plasmid vector backbone	Insert	Role
pCPI	pcDNA3.1	Accession: AF229173.1	Vaccine antigen
pCPIm	pcDNA3.1	Mutated form of *L. sigmodontis* CPI-2. (Accession: AF229173.1) Point mutation of asparagine at position 66 to lysine	Vaccine antigen
pIL-4	pcDNA3.1	Mus musculus IL-4 (NCBI Gene ID: 16189)	Adjuvant
pMIP1α	pcDNA3.1	Mus musculus MIP1α (Accession: M73061.1])	Adjuvant
pEmpty	pcDNA3.1	-	Non-coding plasmid control

In immunisation experiments, 24 BALB/c mice were split into four groups (n=6), with three groups receiving DNA vaccines or pEmpty control, whereas the fourth group received no DNA vaccine (Naïve). Since DNA vaccines (purified plasmids) were stored at -20°C, the day before vaccination the plasmids were placed in a 4°C fridge to allow them to thaw out slowly. The day of immunisation the concentration of the plasmids were quantified and vaccine cocktails were made up to total of 80µg of DNA, with equal quantities of each plasmid, this meant that the quantity of each individual plasmid decreased as the number of different plasmids were added to a vaccine cocktail (Table.4.3). The plasmids were delivered in 50µl doses and elution buffer (QIAGEN: 10mM Tris-Cl, pH 8.5) was used to make up the volume.

DNA vaccine cocktails were administered to the tibialis anterior muscle of the left leg with a 27G needle, immediately followed by electroporation with an ECM 830 generator and Tweezertrodes (BTX Harvard Apparatus) using a setting of 8 pulses,

200 V/cm, 40ms duration with 460ms intervals. During the vaccination process mice were under anaesthesia using gas inhalation of isoflurane and placed on heat pads. To wake mice, the gas inhalation was removed and mice recovered naturally on heat pads. The treatments were randomised per cage so that each cage had at least one of each of the treatment groups. The mediastinal and parathymic lymph nodes and spleens were harvested 24 hours following vaccination and placed in 5ml of cRPMI.

Table 4.3. Amount of plasmid added to the different vaccine cocktails in a DNA vaccine experiment.

	DNA plasmid					
Vaccines	**pCPI**	**pCPIm**	**pIL-4**	**pMIP1α**	**pEmpty**	**Total DNA**
Ls-CPI (Adj)	40µg	-	20µg	20µg	-	80µg
Ls-CPIm (Adj)	-	40µg	20µg	20µg	-	80µg
pEmpty (Adj)	-	-	20µg	20µg	40µg	80µg
Naive	-	-	-	-	-	-

4.2.7 Processing of spleens and lymph nodes

Lymph nodes and spleens collected following immunisation were dissociated to obtain a single cell suspension by grinding them through a 70µm nylon mesh (Fisher Scientific) using forceps in a petri-dish. The single cell suspensions were given a red blood lysis treatment, by centrifuging the cells at 400 x g for 5mins at 4ºC, removing the supernatants and suspending the cells in the remaining cRPMI and treating these with 2ml of RBC lysis buffer (Sigma-Aldrich) for 4 mins. The RBC lysis buffer was stopped by adding 10ml of cRPMI and cells centrifuged at 400 x g for 5mins at 4ºC, the supernatant poured off and cells re-suspended in 2ml of cRPMI. The cell concentration of each sample was determined using a hemocytometer, and the cells

were made up to a concentration of 10^6 cells/ml using cRPMI, which were then used for FACS analysis.

4.2.8 Cytokine quantification by ELISA

Concentrations of IL-12p40, IL-6, IL-10, IL-4 and IFNγ were measured by sandwich ELISA, in the supernatant fluids of DC cell cultures 24 hours after DC stimulation, and in DC and T cell co-cultures either after 24 hours, 48 hours, or 96 hours after the addition of T cells. Flat bottom NUNC plates (Thermo Scientific) were coated with 50µl/well of coating antibody in their respective buffer (Table 4.4), and left in the dark at 4°C overnight. The following day plates were washed four times in wash buffer: 1xPBS containing 0.05% Tween-20 (Sigma-Aldrich). Coating antibodies were blocked by incubating the plates for 1hr at 37°C in the dark with 200µl/well of 4% BSA in 1xPBS. The plates were then washed four times with the wash buffer before adding 50µl/well of either the samples (DC supernatants) in triplicate and the standards for each cytokines which were serially diluted (1 in 2) in their appropriate buffer (Table 4.4) these were used to determine the concentration of the samples. The plates were then left overnight at 4°C in the dark.

The next day the plates were washed four times in wash buffer, and 50µl/well of biotinylated detection antibodies were added to the plates at the recommended concentration and buffer (Table 4.4). These were then incubated at 37°C for 1hr in the dark, and washed four times using wash buffer. Then AMDEX streptavidin-peroxidase (Sigma-Aldrich) was added to each well (50µl) in a 1/6000 dilution in the same buffer as for the detection antibodies, then plates were incubated in the dark for 30 minutes

at 37ºC. Once the final incubation was done the plates were washed five times in wash buffer, and 50µl/well of TMB-H_2O_2 (Merck Chemicals) was added to the plate and left to develop in the dark at room temperature for 5-10mins until a gradient of blue is seen, the reaction was then stopped by adding 100µl of H_2SO_4 (1M). Absorbance was measured at 405nm and the concentration of cytokine levels in the samples were determined using the standard curve made using the O.D. values of the standards against their known concentrations (Figure 4.1).

Table 4.4. Concentrations (conc) and buffers used for the antibodies in the sandwich ELISA for IL-12p40, IL-6, IL-10, IL-4 and IFNγ quantification.

Cytokine		Concentration (µg/ml)	Buffer	Clone/ Isotype
IL-12p40	Coating	2	0.2M Na_2HPO_4, pH6.5	C15.6/Rat IgG1 [1]
	Standard	start conc: 0.005	10% NCS* in PBS	N/A [2]
	Detection	0.5	10% NCS* in PBS	C17.8/Rat IgG2aκ [3]
IL-6	Coating	1	0.1M Na_2HPO_4, pH12	MP5-20F3/Rat IgG1 [1]
	Standard	start conc: 0.025	10% NCS* in PBS	N/A [2]
	Detection	0.5	10% NCS* in PBS	MP5-32C11/Rat IgG2a [1]
IL-10	Coating	4	0.2M $NaHPO_4$, pH6.5	JES5-16E3/Rat IgG2bκ [4]
	Standard	start conc: 0.05	10% NCS* in PBS	N/A [2]
	Detection	0.25	10% NCS* in PBS	JES5-2A5/Rat IgG2bκ[4]
IL-4	Coating	1.2	1xPBS	11B11/Rat IgG [5]
	Standard	start conc: 0.008	10% NCS* in PBS	N/A [2]
	Detection	0.25	10% NCS* in PBS	BVD6-24G2/Rat IgG1[3]
IFNγ	Coating	1.5	0.1M $NaHCO_3$, pH9.6	AN-18/Rat IgG1[3]
	Standard	start conc: 0.05	10% NCS* in PBS	N/A [2]
	Detection	1	10% NCS* in PBS	R4-6A2/Rat IgG1[3]

*NCS, newborn calf serum (ThermoFisher); [1]BD Bioscience, [2]Peprotech, [3]BioLegend, [4]eBioscience, [5]BD Pharmingen

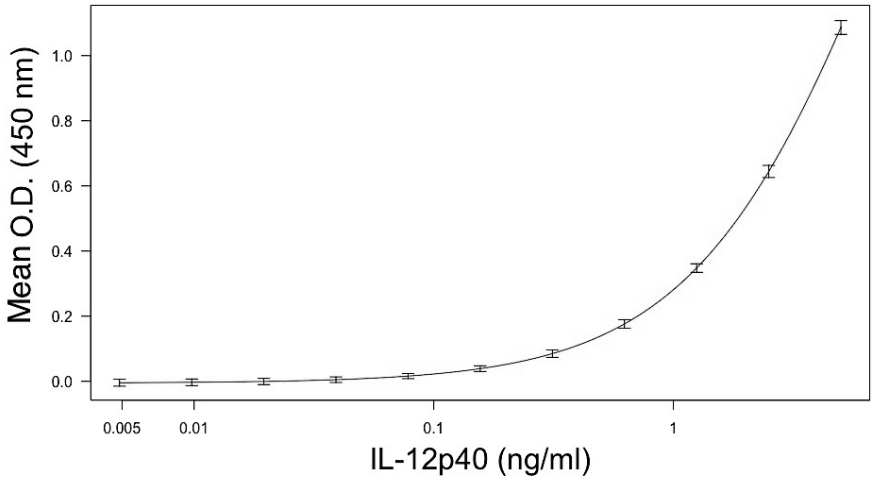

Figure 4.1. Cytokine ELISA standard curve. An example of a standard curve used to calculate cytokine concentration in samples, this was the standard curve for IL-12p40, and was used to determine the IL-12p40 levels in the wells.

4.2.9 Flow cytometry analysis

Flow cytometry was used to analyse cell surface markers on DC and T cell from co-cultures, and DC from spleens and lymph nodes following immunisation. From the single cell suspension, 10^6 cells of cells were added to FACS tubes and non-specific binding was blocked using anti-CD16/32 (BD Pharmingen) at 5µg/ml in 1/20 mouse serum for 30mins in the dark at 4°C. These cells were then washed in FACS buffer (PBS with 2% FBS), centrifuged at 400 x g for 5mins at 4°C, supernatant discarded and re-suspended before adding the staining antibodies for the surface markers, all diluted 1/400 (Table 4.5.). Stained cells were left for 30mins in the dark at 4°C followed by a wash in FACS buffer, centrifuged at 400 x g for 5mins at 4°C, supernatant discarded and cells were resuspended in 170µl FACS buffer. Cells were analysed using an LSR II (BD Biosciences), running FACSDiva software (BD Biosciences). Some cells were left unstained as negative controls and compensation

beads (Affymetrix, eBioscience) were stained with each individual staining antibody in a 1/400 dilution as the compensation control. Analysis of the flow cytometry was performed using Flowjo (Tree star).

Table 4.5. Multicolour panel for flow cytometry of DC and T cell from co-cultures and lymph nodes (LN) and spleens harvested from mice 24hrs following immunisation.

For *in vitro* DC and T cell co-cultures		For LN and spleen cell suspensions
DC stain	**T cell stain**	**DC stain**
CD11c – APC [4]	CD3–PeCy7 [3]	CD3–APC/Cy7[1]
CD40 – PE[1]	CD4–FITC [3]	CD11c–BrilliantViolet421[1]
CD80 – PeCy7[1]	CD19- PE[4]	CD11b–BrilliantViolet711[1]
CD86–BrilliantViolet605 [4]	CD8 – AF700 [1]	CD19 –AlexaFluor700[4]
F4/80 – AF700 [1]		CD40 – FITC[1]
MHCII–Violet500 [1]		CD80 –APC[1]
		CD86–PE [4]
		F4/80–PE/Cy7 [1]
		Gr1–BrilliantViolet605 [1]
		MHCII–PerCP [1]

[1]BioLegend; [2]BD Horizon; [3]BD Pharmingen; [4]eBiosciences

4.2.10 Microarray datasets

Human microarray datasets are the same as in Chapter 3. These reveal gene expression patterns of individuals living in two different foci in Ghana: one endemic for *O. volvulus;* and the other *W. bancrofti*. Whole blood human samples were collected and RNA extracted by the E PIAF collaborators Dr Sabine Specht (University Hospital Bonn, Germany), Prof Achim Hoerauf (University Hospital Bonn, Germany), Dr Alex Debrah (Kwame Nkrumah University, Ghana), Dr Laura Layland (University Hospital Bonn, Germany) and Gnatoulma Katataw (University Hospital Bonn, Germany, and University of Lome, Togo), and processed at the Kwame Nkrumah University of

Science and Technology (Kumasi, Ghana). The RNA samples were hybridised to Illumina HumanHT12 BeadChip arrays (HumanHT12_V4_0_R2_15002873_B) by FIOS Genomics, Edinburgh who also conducted quality control on the samples. All samples that failed the quality control (QC) were removed from the dataset (Table 4.6).

Individuals were split into infected and endemic normal (EN) groups; EN are individuals that show no clinical or parasitological signs of infection. Therefore in *O. volvulus* foci individuals with no detectable Mf in skin snips or nodules were classified as EN, and in *W. bancrofti* foci individuals with no detectable Mf in their blood and no circulating filarial antigen (CFA) to *W. bancrofti* were classified as EN.

The raw microarray files were processed in R programming language (R Development Core Team,, n.d.) using the *lumi* Bioconductor package (Du *et al.*, 2008), in the same manner as chapter 3. Briefly, the *neqc* function in the *limma* Bioconductor package (Ritchie *et al.*, 2015) was used for background correction using negative controls (*normexp*) (Ding *et al.*, 2008; Xie *et al.*, 2009), followed by between-array quantile normalisation and $\log_2$ transformation, as this was found to be the best pre-processing strategy giving the best precision for a given bias (Shi *et al.*, 2010).

To find the genes associated with the probes on the Illumina microarrays, the probe identifiers were converted to nuID annotation using the *lumi* package (Du *et al.*, 2008). Genes associated with the MHC class II (R-HSA-2132295.3), CD28 co-stimulatory (R-HSA-388841.2) and TCR signalling (R-HSA-388841.2) pathways were extracted from the Reactome database (www.reactome.org), and used to subset the datasets, so only those genes are analysed.

Table 4.6.Information on samples used in human microarray.

		***O. volvulus* endemic foci**	***W. bancrofti* endemic foci**
Number of samples which passed QC		166	162
Number of samples classified as:	• Infected	118	109
	• EN	48	53
Other parameters recorded:		• Age • Co-infection with helminths (none were co-infected with other filarial nematodes) • Co-infection with protozoa • Rounds of Ivermectin	• Age • Co-infection with protozoa • Rounds of Ivermectin (none received doxycycline) • Presence of nodules

4.2.11 Statistical analysis

All statistical analyses were performed in R (Ihaka, Gentleman, 1996). Generalised linear models (GLM) followed by the Shapiro-Wilk normality test to assess the normality of the residuals, were used to compare the effects of the different vaccines on DC *in vivo*, and the different treatments on DC stimulation *in vitro*. For in the *in vitro* data, means± SE are shown. For microarray analysis, GLMs were also used to detected differences between infected and endemic normal individuals, since there were 137 genes for MHC-II signalling, 89 genes for CD28 co-stimulation signalling and 138 genes for TCR signalling, GLMs were looped through and multiple testing was accounted for using the *qvalue* package in Bioconductor (Storey, 2015). All graphs were produced with the package *ggplot2* (Wickham, n.d.).

4.3 Results

4.3.1 Modifying CPI rescues pro-inflammatory cytokine production by DC *in vitro*

To assess whether modifying the immunomodulator CPI effects DC activation, immature BMDC were stimulated with either recombinant *Ls*-CPI or its modified counterpart which has a mutation, asn66 to lys66 (*Ls*-CPIm). The immature BMDC were stimulated for 24 hours with either *Ls*-CPI, *Ls*-CPIm, OVA a Th2 control, LPS as Th1 control, or left unstimulated (Media). Overall immature BMDC stimulated with the different antigens showed an increased in activation over unstimulated cells, seen as an increase in cytokine production. Treatment with *Ls*-CPIm increased (non-significant) the production of the pro-inflammatory cytokines IL-12p40 (0.23 ± 0.19) and IL-6 (0.52 ± 0.42) compared to its native form *Ls*-CPI (Figure 4.2), whereas treatment with *Ls*-CPI showed increased levels of IL-10 cytokine production compared to *Ls*-CPIm and OVA (Figure 4.2).

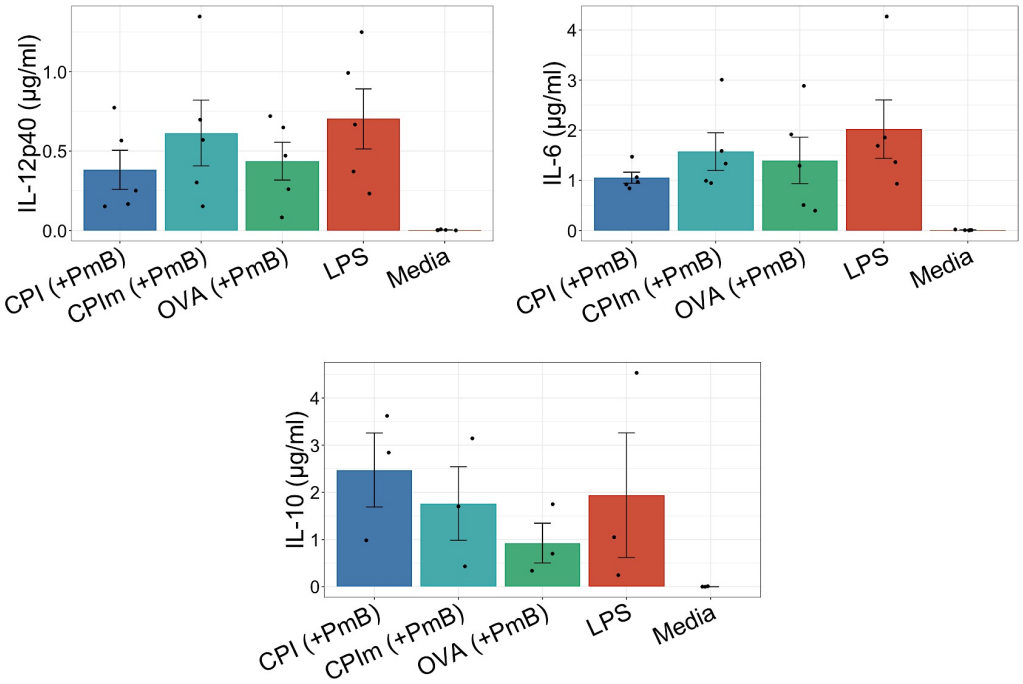

Figure 4.2. Cytokine production by BMDC in response to stimulation with either *Ls*-CPI or *Ls*-CPIm. Bone marrow cells were cultured for 10 days with GM-CSF, to obtain immature DC. Immature DC were harvested and cultured for 24 hours with either: *Ls*-CPI (50 µg/ml) with polymixin B (PmB) to deplete lipopolysaccharides (LPS); *Ls*-CPIm (50 µg/ml) with PmB; Ovalbumin (OVA) (50 µg/ml) with PmB as a Th2 control; LPS as Th1 control; or media (complete RPMI) as an unstimulated control. Supernatants were harvested and cytokines IL-12p40, IL-6 and IL10 concentration were measured using sandwich ELISA. Data represented as means (± SE) of 5 independent experiments for IL-12p40 and IL-6, whereas IL-10 could only be measured in 3 independent experiments. GLM were used to compare statistical significant differences taking into account the independent experiments, however no significant difference was found. The first DC culture experiment was done by Marjorie Besençon.

4.3.2 Increased DC antigen presentation with *Ls*-CPIm stimulation

For the adaptive immune responses to be triggered, DC must deliver a series of defined signals to naive T cells, including the process of antigen presentation via MHC class II to naive T cells with the appropriate co-stimulation signals (CD86, CD80 and CD40). CPI's immunomodulatory property is its ability to inhibit antigen presentation by MHC class II molecules, therefore to assess if mutating CPI restores DC ability to present antigens to naive T cells, immature BMDC were stimulated with either *Ls*-CPI, *Ls*-CPIm, OVA, LPS or left unstimulated (media control) for 24hrs to obtain mature DC, which were then cultured with naive T cells for either 24, 48 and 96 hours. The BMDC and T cells co-cultures were harvested, and stained for flow cytometry analysis. CD11c was used to select for DC populations of cells, and the expression of MHC-II and co-stimulatory molecules were analysed.

The number of $CD11c^+$ dendritic cells in cultures increased over time following addition of T cells (Figure 4.3.A). Since there was only one experiment for each time point, and similar trends in expression of MHC-II and CD86 were seen across the time points, expression was calculated as a fold change over unstimulated (media) controls per time point, as to combine experiments and analyse the difference between treatment groups. Treatment with *Ls*-CPIm significantly increased the expression of MHC-II on $CD11c^+$ cells, with a 0.5 fold increase in expression (Figure 4.3.B) and a 0.4 fold increase in CD86 expression compared to BMDC treated with *Ls*-CPI (Figure 4.3.B), signifying an increase in antigen presentation and co-stimulation.

The expression of CD40 and CD80 was only measured at the 24-hour time point, but only CD40 showed a change in expression between unstimulated and stimulated cells, with BMDC stimulated with *Ls*-CPIm having increased expression over *Ls*-CPI stimulated cells (Figure 4.3. C), since CD40 expression was only measured once it is not possible to conclude that it is increased.

Changes in naive T cell differentiation were also assessed, by measuring the fold change in $CD4^+$ T cells between the treatment groups, although there was an increase in $CD4^+$ T cell differentiation between stimulated and unstimulated (media) cells, no difference between *Ls*-CPI and *Ls*-CPIm treated cells was detected (Figure 4.4).

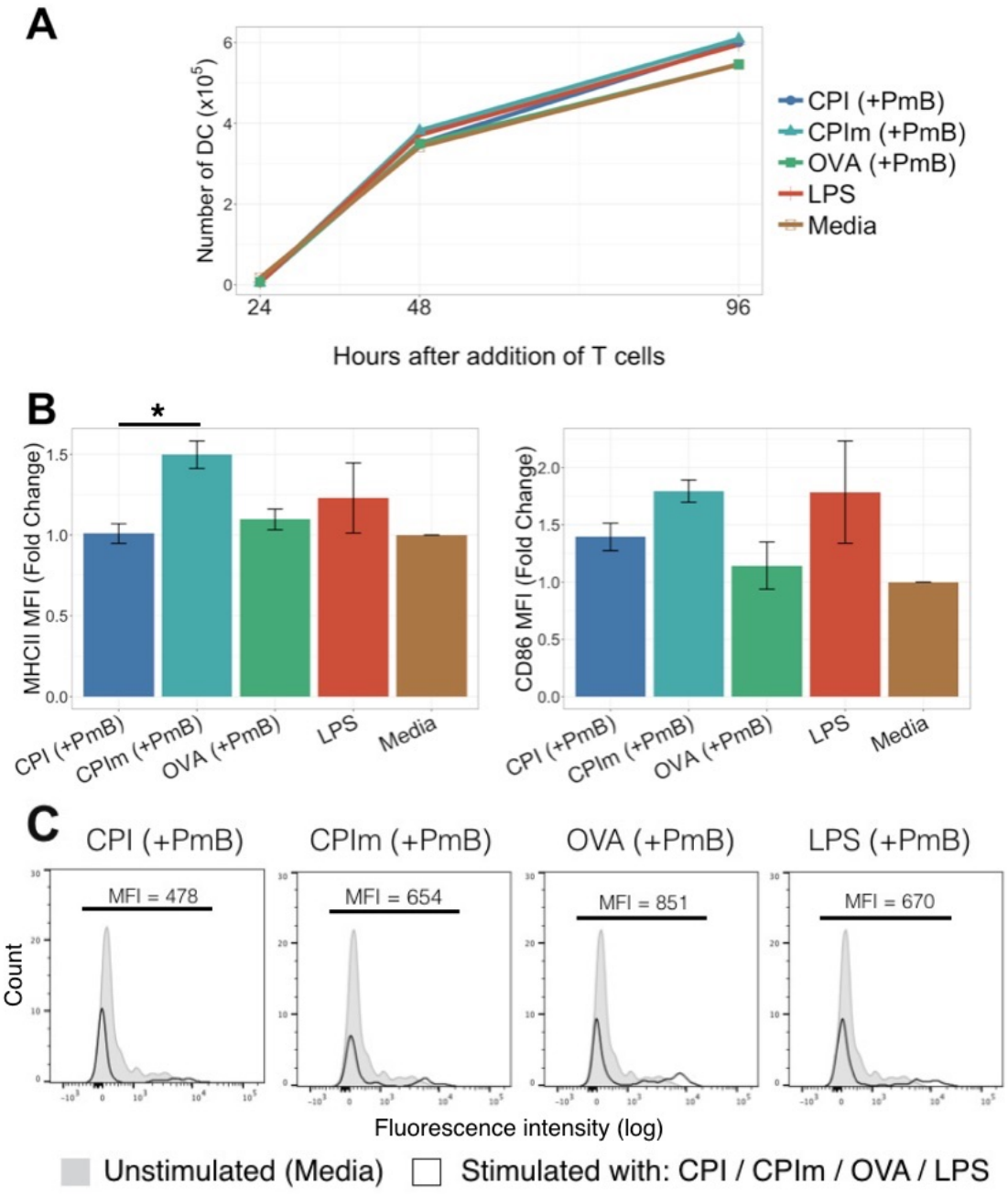

Figure 4.3. Effects of recombinant *Ls*-CPI and *Ls*-CPIm on expression of MHC-II and co-stimulatory molecules on BMDC following co-culture with naïve T cells. Immature BMDC obtained from 10-day old bone marrow cell cultures, were stimulated for 24 hours with either: *Ls*-CPI (50 µg/ml) with polymixin B (PmB) to deplete lipopolysaccharides (LPS); *Ls*-CPIm (50 µg/ml) with PmB; Ovalbumin (OVA) (50 µg/ml) with PmB as a Th2 control; LPS as Th1 control; or media (complete RPMI) as an unstimulated control. Following stimulation, BMDC were harvested and co-cultured with naïve T cells (1:10 of DC to T cells) for 24 hours,

48 hours or 96 hours. Co-cultured cells were harvested and stained for DC using CD11c, MHC-II and co-stimulatory molecule CD86, and analysed by flow cytometry. A) Number of DC ($CD11c^+$ cells) harvested at 24, 48, and 96 hours of co-culture. B) Fold change of mean fluorescence intensity (MFI) of CD86 and MHC-II expression by $CD11c^+$, between stimulated and media control (unstimulated) BMDC. Data shown are means (± SE) of the three BMDC and T cell co-cultures ending at 24, 48 or 96 hours of co-culture. C) FACS histograms of expression of CD40 on CD11+c cells from BMDC and T cell co-cultures, at the 24-hour time point, expression shown as MFI between unstimulated (shaded grey) and either *Ls*-CPI (+PmB), *Ls*-CPIm (+PmB), OVA and LPS (black line). Statistical significant differences were measured using GLM with experiment end points as a random effect, * P-value < 0.05. The 48 and 96 hour time points experiments were done by Marjorie Besençon.

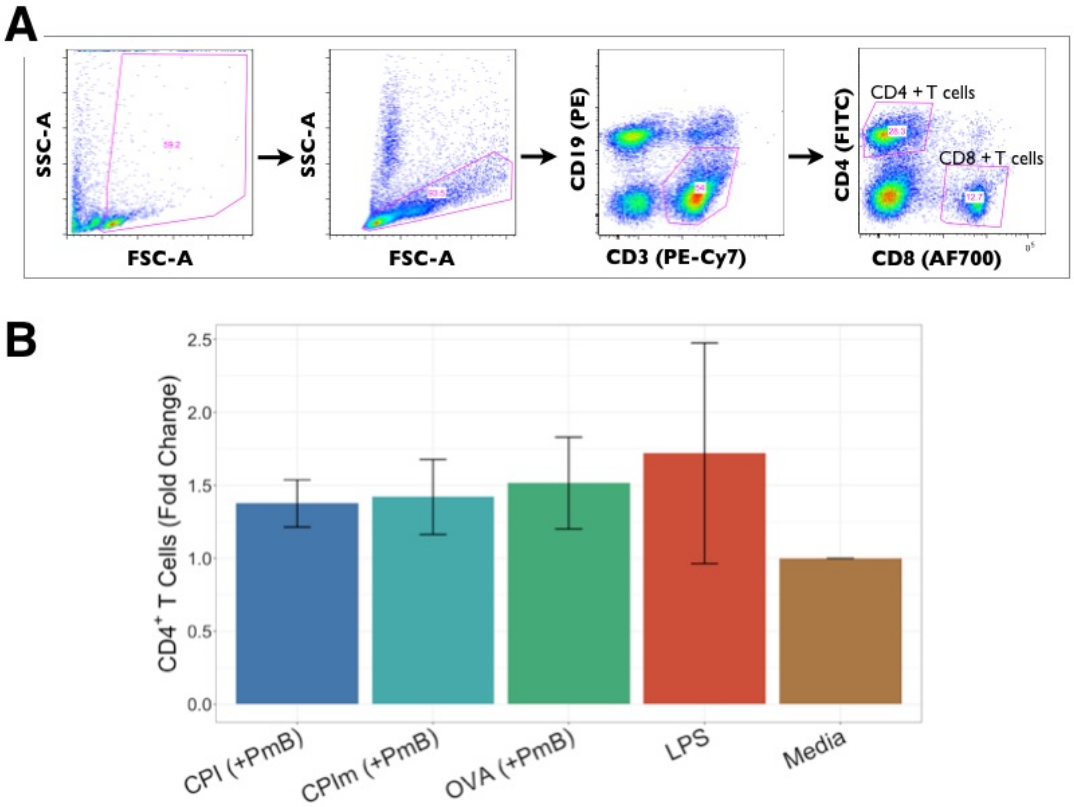

Figure 4.4. Expansion of CD4$^+$ T-cells, induced by BMDC stimulated with either *Ls*-CPI or *Ls*-CPIm. Immature bone-marrow derived DC (BMDC) were stimulated with either: *Ls*-CPI (50 µg/ml) with polymixin B (PmB) to deplete lipopolysaccharides (LPS); *Ls*-CPIm (50 µg/ml) with PmB; ovalbumin (OVA) (50 µg/ml) with PmB as a Th2 control; LPS as Th1 control; or media (complete RPMI) as an unstimulated control, for 24 hours. The stimulated BMDC were then co-cultured with naïve T cells for 24-hours, 48-hours or 96-hours. A) Gating strategy for the detection of CD4+ T cells. B) Differences in CD4$^+$ T cells (CD3$^+$CD4$^+$ cells) harvested from co-cultures with BMDC, were measured as fold changes over unstimulated cells (Media) at each time point to compare all three experiments together. Data is presented as mean (± SE) fold change of the three independent experiments (time points). The 48 and 96 hour time points experiments were done by Marjorie Besençon.

4.3.3 Change in cytokine profiles in DC and T cell co-cultures

Changes in cytokine production after the addition of naïve T cells, were measured as fold changes over unstimulated cells, as there was variation between the independent experiments. Although numbers of $CD11c^+$ DC in cultures increased over time (Figure 4.5. A), production of IL-12p40 and IL-6 peaked at 48 hours after the addition of T cells, furthermore IL-10 could only be detected at 48 hours. A slight increase in IL-12p40 production by *Ls*-CPIm treated cells was detected, similar to what was seen before the addition of naive T cells (Figure 4.5.B), however no difference in IL-6 production nor IL-10 production after the addition of T cells was detected (Figure 4.5.B,C). Furthermore, levels of IL-4 and IFN-γ were undetectable, therefore it is hard to say whether there was a change in $CD4^+$ T cell activation.

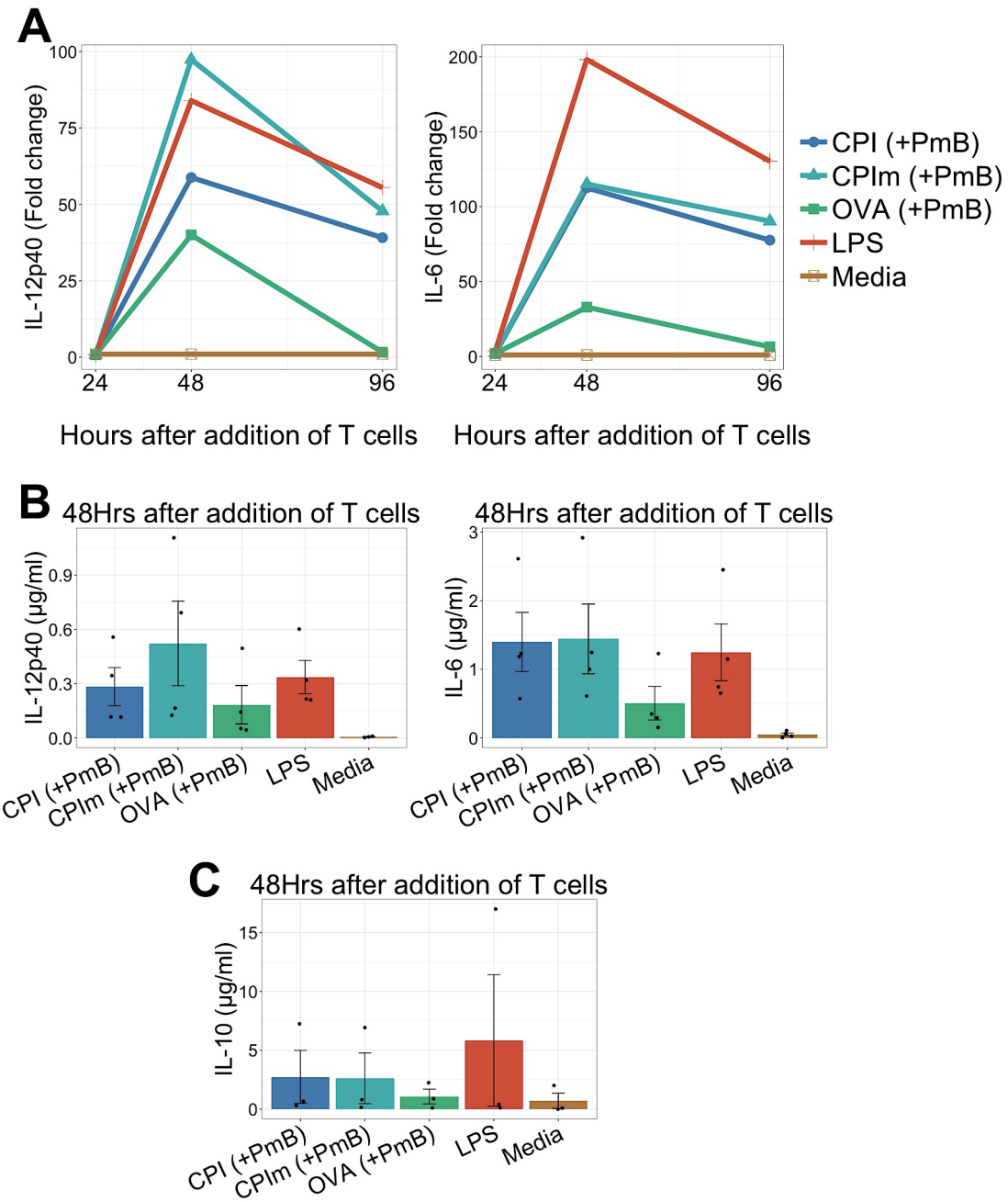

Figure 4.5.Effects of recombinant *Ls*-CPI and *Ls*-CPIm on cytokine production by BMDC following co-culture with naïve T cells. Immature BMDC obtained from 10-day old bone marrow cell cultures, were stimulated for 24 hours with either: *Ls*-CPI (50 µg/ml) with polymixin B (PmB) to deplete lipopolysaccharides (LPS); *Ls*-CPIm (50 µg/ml) with PmB; Ovalbumin (OVA) (50 µg/ml) with PmB as a Th2 control; LPS as Th1 control; or media (complete RPMI) as an unstimulated control. Following stimulation, BMDC were harvested and co-cultured with naïve T cells (1:10 of DC to T cells) for 24-hours, 48-hours or 96-hours. A) Dynamics of cytokines IL-12p40 and IL-6 production, at the different end points. Cytokine

levels are represented as fold change over media cultured BMDC (unstimulated BMDC) to compare cytokine production between the different co-culture end points. Since, 48 hours of BMDC and T cell co-cultures produced the greatest levels of cytokines, this time point was repeated. B) Production of IL-12p40 (μg/ml) and IL-6 (μg/ml) at 48 hour of BMDC and T cell co-cultures, presented as means (± SE) of four independent experiments; and C) production of IL-10 (μg/ml) at 48 hour of BMDC and T cell co-cultures, presented as means (± SE) of three independent experiments. The first 48 and 96 hour time points experiments were done by Marjorie Besençon.

4.3.4 Undetectable changes in DC activation following immunisation *in vivo*

The effects of the native and modified CPI on DC maturation were investigated *in vivo* by immunising BALB/c mice with a DNA vaccine containing ether the *Ls*-CPI or *Ls*-CPIm antigen in combination with adjuvant plasmids for IL-4 and MIP1α. Spleens and lymph nodes were harvested 24 hour following immunisation, flow cytometry was used to measure the expression of MHC class II and the co-stimulatory molecules CD86, CD80 and CD40 on $CD11c^+$ DC present in the spleen and lymph nodes. However, no change in expression of MHC-II nor the co-stimulatory molecules could be detected between the immunisation groups, on $CD11c^+$ cells from the spleen (Figure 4.6) or lymph nodes (data not shown). This could be due the complex nature and poorly controlled factors inherent to *in vivo* settings (e.g. baseline variation, timing of stimulation from DNA incorporation in exogenous protein expression).

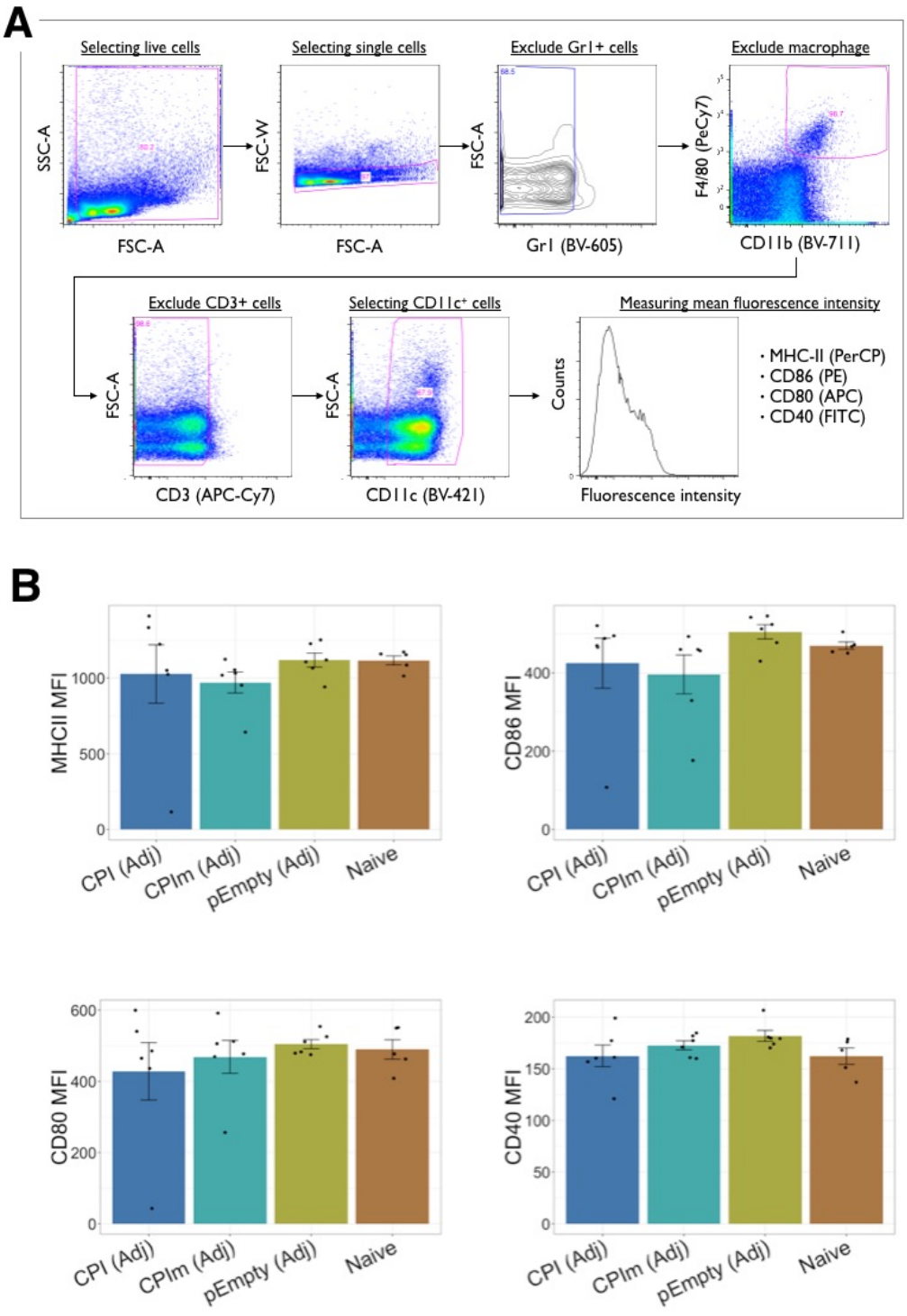

Figure 4.6.*In vivo* expression of MHC-II and co-stimulatory molecules on CD11c$^+$ DC following immunisation with either *Ls*-CPI or *Ls*-CPIm. Female BALB/c mice were immunised with DNA vaccines intramuscularly, and spleens were harvested for flow cytometry to measure expression of MHC-II, CD86, CD80 and CD40 on DC, classified as CD11c$^+$ cells. Mice received a DNA plasmids containing antigen inserts for *Ls*-CPI with

adjuvant, *Ls*-CPIm with adjuvant or an empty pcDNA3.1 plasmid (pEmpty) with adjuvant, or were left non-immunised termed as Naïve. Adjuvant plasmids contained inserts for IL-4 and MIP1α. A) Represents the gating strategy used to isolate DC population and measure the fluorescence intensity of MHC-II, CD86, CD80 and CD40. B) Expression was measured as geometric mean fluorescence intensity (MFI). Data presented are medians (horizontal line), interquartile (bars), extreme values (vertical lines), and individual values/mice (points), with an n=5-6 per group.

4.3.5 Changes in gene expression associated with DC and T cell activation in humans

Since *in vivo* immunisation experiments in mice failed to detect differences in DC activation, more likely due to difficulties in isolating the correct DC population. It was then suggested that examination of human gene expression data, might reveal differences in DC activation, considering that using the machine learning pipeline (Chapter 3), pathways involved in T cell activation were found to be important at differentiating between infected and endemic normal individuals living in areas endemic for onchocerciasis and lymphatic filariasis.

The microarray gene expressions data available was obtained from whole blood, therefore it is impossible to select for specific cells and measure expression of genes within them. Instead expression of genes involved in MHC-II signalling (R-HSA-2132295.3) and CD28 co-stimulation (R-HSA-388841.2) according to the reactome database (Croft *et al.*, 2011) were investigated, by comparing expression in naturally protected (EN) and infected individuals using generalised linear models (GLMs). To verify that GLMs could detect differences in expression, genes involved in TCR signalling (R-HSA-202403.4) were also investigated to confirm what the machine

learning pipeline detected. Two datasets were available, one from individuals living in onchocerciasis (*O. volvulus*) endemic areas in Ghana with Mf numbers measured per mg of skin, the other dataset was from individuals living in lymphatic filariasis (*W. bancrofti*) endemic areas in Ghana with Mf numbers and detection of circulating *W. bancrofti* antigens measured in the blood. Neither dataset had co-infections with filarial nematodes.

In individuals living in onchocerciasis endemic areas, no differences in expression of genes specific for MHC-II and CD28 co-stimulation signalling, such as HLA-DRA and CD86, were detected (Figure 4.7). There were differences in expression of genes associated with T cell receptor signalling, such as increased expression of ICOS and CD3E in EN, although not significant when accounting for multiple testing (Figure 4.7). ICOS is an inducible T-cell co-stimulator found expressed on activated T cells and CD3E, is a gene encoding the CD3ε which forms part of the CD3 T cell co-receptor.

In individuals living in lymphatic filariasis endemic areas, expression of HLA-DRA, ICOS and CD74 were significantly increased in infected compared to naturally protected EN (Figure 4.8). Although ICOS expression was higher in infected individuals, it was negatively associated with Mf burden (Figure 4.9), HLA-DRA and CD74 also have a negative correlation but not as strong. HLA-DRA gene encodes for the DR alpha chain of MHC class II and CD74 encodes the invariant chain involved in the formation and transport of MHC class II protein.

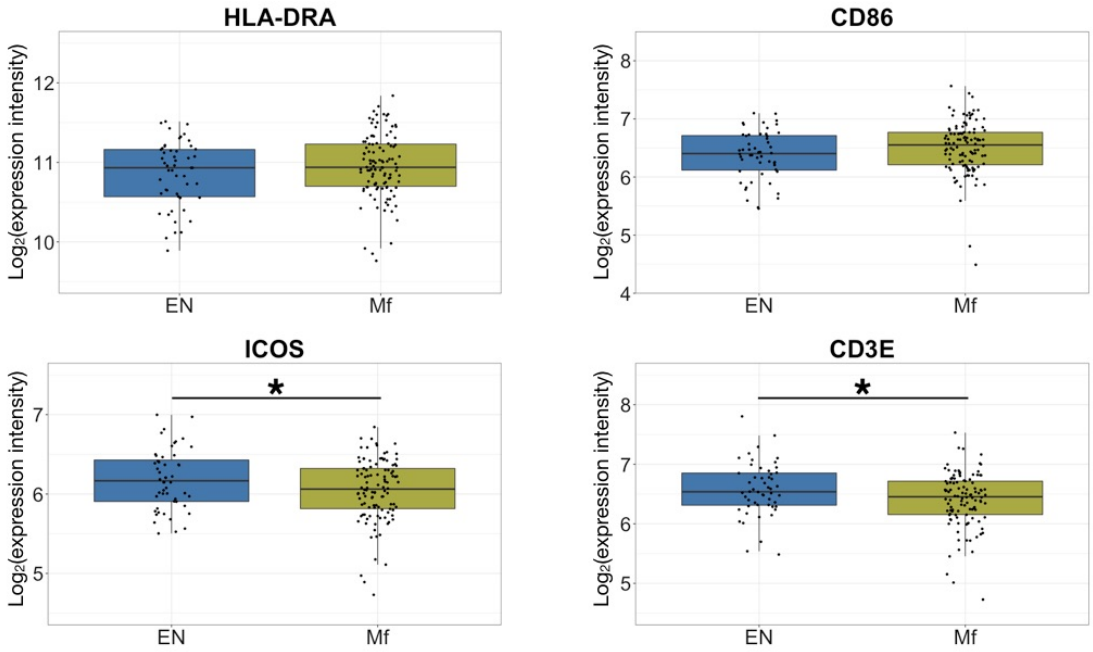

Figure 4.7. Expression of HLA-DRA, CD86, ICOS and CD3ε in whole blood from individuals living in *O. volvulus* endemic areas. Gene expression were measured using Illumina microarray plates, and differences in expression between infected (Mf) and naturally protected endemic normal (EN) were analysed using GLMs. HLA-DRA is a component of the MHC-class II receptor; CD86 is a receptor found on dendritic cells which induces the proliferation and activation of T cells; ICOS is an inducible T-cell co-stimulator; and CD3ε is a component of the T cell receptor. Individuals were classified as EN if no Mf were present in skin snip. * P-value <0.05, not accounting for multiple testing.

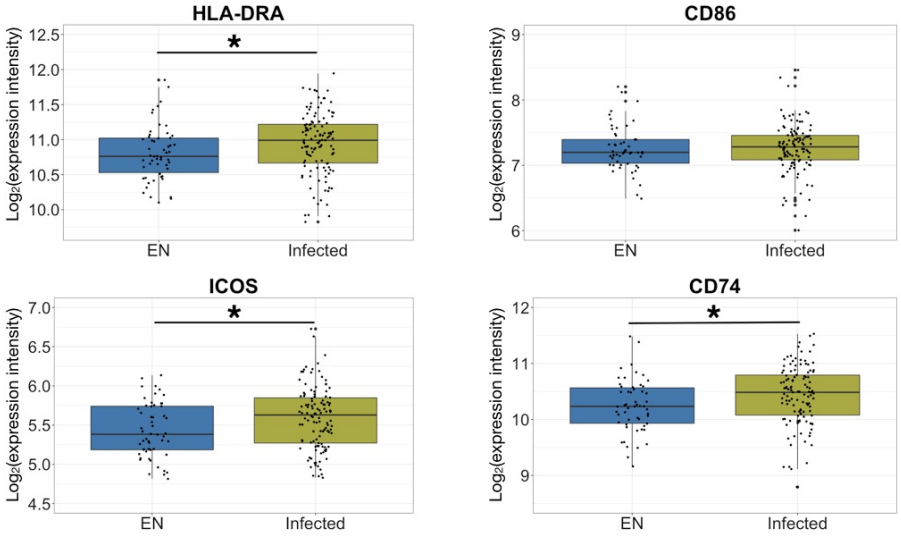

Figure 4.8.Expression of HLA-DRA, CD86, ICOS and CD3ε in individuals living in *W. bancrofti* endemic area of Ghana. Gene expressions were measured using Illumina microarray plates, and differences in expression between infected and naturally protected endemic normal were analysed using GLMs. HLA-DRA is a component of the MHC-class II receptor; CD86 is a receptor found on dendritic cells which induces the proliferation and activation of T cells; ICOS is an inducible T-cell co-stimulator; and CD74 encodes the invariant chain protein which is associated with the formation and transport MHC-II. Individuals were classified as EN if no Mf were present in their blood and no circulating antigens were detected. * P-value <0.05, not accounting for multiple testing.

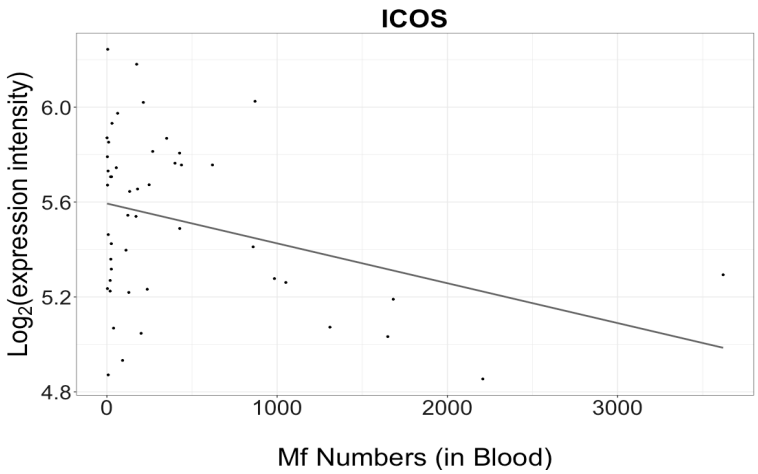

Figure 4.9. Expression of ICOS compared to microfilariae counts found in blood of individuals living in lymphatic filariasis endemic areas. Gene expression were measured using Illumina microarray plates, and GLM were used to measure the association between ICOS and Mf numbers present in the blood (samples without Mf present were removed). ICOS is an inducible T-cell co-stimulator and is negatively associated with Mf numbers found in the blood of individuals living in foci endemic to *W. bancrofti*. P-value <0.021.

4.4 Discussion

Vaccination experiments with the immunomodulator *L. sigmodontis* cysteine protease inhibitor (*Ls*-CPI), which has been shown to inhibit antigen presentation in MHC class II molecules (Manoury *et al.*, 2001), induced little protection against challenge infection. However, when genetically modified to remove its immunomodulatory site, CPIm showed a greater efficacy, with a reduction in circulating microfilariae compared to CPI and non-vaccinated controls (Babayan *et al.*, 2012). As it has been demonstrated that the early immune responses in infection or following immunisation determine the outcome of infection (Chapter 3) (Taylor *et al.*, 2009; Babayan *et al.*, 2010), and that CPI affects antigen presentation on cells such as dendritic cells (DC), changes in DC activation following stimulation/immunisation with CPI or CPIm were investigated *in vitro* and *in vivo*. This demonstrates how antigen (CPI) modification can affect the activation of dendritic cells and priming of T cells.

In vitro experiments with bone-marrow derived DC (BMDC), showed an increase in DC activation following stimulation with CPIm, with increased levels of IL-6 and IL-12p40 production. Both cytokines are frequently associated with Th1 immunity but since helminths generally fail to induce DC production of pro-inflammatory cytokines, often with minimal changes in DC messenger RNA expression, in this section IL-12p40 and IL-6 are simply used to measure the change in DC activation (Cook *et al.*, 2015).

Due to the lack of cytokine production by DC, it was proposed that Th2 responses in helminths infections (and allergens) were induced as the default pathway when Th1 DC responses failed to be activated (MacDonald, Maizels, 2008). However, a range of

molecules have been associated with DC ability to induce Th2 immunity (Cook *et al.*, 2015), such as CD40 (MacDonald *et al.*, 2002), CD80/86 (Whelan *et al.*, 2000), OX40L (Jenkins *et al.*, 2007), CCL17 (Medoff *et al.*, 2008), RELMα (Cook *et al.*, 2012) and NF-κB (Artis *et al.*, 2005). This demonstrates that DC can prime a Th2 response to helminth and allergens, however the specific mechanisms are unknown.

In filarial infections IL-10 has been linked to parasite-driven immunomodulation (Satoguina *et al.*, 2002; Simons *et al.*, 2010; Specht *et al.*, 2011b), and in the experiments described above an increase in IL-10 production by DC stimulated with CPI was also detected. In other nematode infections IL-10 has been directly linked to DC, such as in *Ascaris suum* infections where their ES products have the ability to modulate the antigen presenting ability of DC via an IL-10 mediated mechanism (Silva *et al.*, 2006) and in *H. polygyrus* with ES-treated BMDC generate regulatory T cells that produced high levels of IL-10 (Segura *et al.*, 2007). This demonstrates that CPI could be pushing DC to a more regulatory phenotype whereas CPIm causes an increase in DC maturation.

DC are critical for the development of an effective adaptive immune responses, as they are one of the main antigen-presenting cells, delivering a series of defined signals to naive T cells (Liu, 2001b). In DC and T cell co-cultures, there were significant increases in MHC class II antigen presentation and an increase in CD86 expression when DC were stimulated with CPIm. Furthermore, a small increase in CD40 expression was detected on BMDC treated with CPIm, and together with MHC II and CD86, these molecules have been shown to be required for DC mediated Th2 induction (MacDonald *et al.*, 2002; Straw *et al.*, 2003)

In attempts to identify similar changes in DC activation following immunisation *in vivo*, mice were immunised with either *L. sigmodontis* CPI (*Ls*-CPI) or the *Ls*-CPIm. However, no differences between the immunised groups could be detected. DC are a complex population of cells, contrary to *in vitro* bone-marrow derived DC which are homogenous. *In vivo*, mice have several subpopulations of migratory and lymphoid-resident DC, each with different phenotypes and functional properties (Segura, Villadangos, 2009). Therefore, the fact that no change in expression were detected *in vivo*, in this instance could be because specific DC subsets were not being isolated before the expression of activation markers were measured. In studies with *B. malayi* L3 infection, it was observed that lymphoid DC and $CD8^+$ plasmacytoid DC peaked at day 7 post infection in spleens and mesenteric lymph nodes of infected mice (Sharma *et al.*, 2016), therefore it could also be possible that looking at 24hrs post immunization is too short to detect any changes in DC activation or looking at mesenteric is too far away for DC to travel on such a small immune response induced by plasmid vaccine.

The role of CPI in immunomodulation has previously been demonstrated, in *in vitro* experiments, spleens cells exposed to CPI (cystatin) from *L. sigmodontis,* showed decreased antigen specific responsiveness (Pfaff *et al.*, 2002), and in experiments using human PBMC, stimulation with CPI (onchocystatin) from *O. volvulus*, suppressed antigen driven proliferation, with increased production of IL-10, decrease in IL-12p40, reduced expression of HLA-DR and costimulatory molecule CD86, corroborating the *in vitro* experiments with *Ls*-CPI and BMDC (Schönemeyer *et al.*, 2001).

Although CPI shows similar effect on human cells as with mice cells, early modulation of the immune responses by filarial parasites in humans cannot be investigated, since

it is impossible to identify individuals that have just been infected, as infected individuals are detected by the presence of Mf or adult parasites. Humans however do exhibit naturally induced protection (Hoerauf, Brattig, 2002; Brattig, 2004), and in chapter 3 pathways involved in T cell activation were found to be important at differentiating between infected and EN individuals living in areas endemic for onchocerciasis and lymphatic filariasis. Therefore, taking a candidate gene approach, genes involved in MHC-II, CD28-costimulatory and TCR signalling were investigated, however only genes involved in TCR signalling were detected as being important, supporting the results from machine learning approach taken in chapter 3. Using a system biology approach, the adaptive T-cell immune response induced by vaccination was successfully predicted using transcriptomic data obtained from DC (sorted spleen DC), 6 hours following vaccination (Dérian *et al.*, 2016). Therefore, it is possible to use early changes in DC gene expression to predict adaptive immune responses and protection, but using a more focused approach such as investigating individual cell population.

The work presented in this chapter shows that by modifying the immunomodulatory CPI antigen, DC responses are rescued, seen as an increase in DC maturation and activation, leading to increased protection seen at later stage of infection. This may suggest that DC can indeed prime the immune system towards Th2 pathways at the very early stages on infection, and that by boosting the immune response to DC following immunisation, would provide protection, such circumstances have implications for the formulation of any future vaccines.

Summary

- Structural modifications to CPI induces an increase in DC activation *in vitro.*
- Early changes in DC activation could explain increased protection seen late on infection, following immunisation with CPIm.

Chapter 5. General discussion

Current control of onchocerciasis relies on annual mass drug administration (MDA) with Ivermectin, (Mectizan donation project, (Thylefors, 2008)), whose widespread distribution in mass treatment campaigns has reduced the incidence of skin and eye disease in some foci in Latin America (World Health Organization, 2016b) and West Africa (Mali, Senegal (Diawara *et al.*, 2009) and Nigeria (Tekle *et al.*, 2012)). Such success led to the speculation that Ivermectin treatment alone could eliminate (and possibly eradicate) onchocerciasis. However, after more than 25 years of distribution of Ivermectin, onchocerciasis remains endemic across 27 countries in Africa, including in areas where Ivermectin treatment has been given continuously for over 15 years (Katabarwa *et al.*, 2013; Lamberton *et al.*, 2015; Wanji, 2015; Kamga *et al.*, 2016).

Mathematical modelling has demonstrated that relying on MDA alone for the elimination of onchocerciasis in Africa will not be sufficient (Turner *et al.*, 2013; 2014b). Several factors contribute to the persistence of onchocerciasis:

- Ivermectin only kills the Mf and any interruption of their production is temporary (Basáñez *et al.*, 2008),
- In areas with high prevalence, such as forest areas, Ivermectin cannot completely interrupt transmission (Cupp *et al.*, 2011; Cheke, 2017),
- There is growing evidence for the emergence of Ivermectin resistance (Osei-Atweneboana *et al.*, 2011; Pion *et al.*, 2013),
- Ivermectin cannot be used in areas with *O. volvulus* and *L. loa* co-endemicity. Individuals infected with both parasites are at risk of severe and possible fatal adverse reaction if treated with Ivermectin and this situation prevents

implementation of control programmes using MDA with this drug (Chesnais *et al.*, 2017),

- The drug cannot be given to children under 5 or pregnant women, therefore this leaves a reservoir of the infection, even in areas where MDA has been implemented (Stolk *et al.*, 2015b).

If onchocerciasis is to be eliminated, and perhaps eradicated, new interventions are required, such as a novel drug that kills adult worms or a vaccine. Mathematical modelling has shown that a vaccine against onchocerciasis would complement MDA efforts and decrease the chances of re-emergence of *O. volvulus* infections, especially in foci where onchocerciasis has been successfully controlled (Turner *et al.*, 2015). It would also protect the considerable financial investment in onchocerciasis control that has been made over the decades.

Filarial nematodes are multi-cellular organisms that induce an immunomodulatory milieu, allowing them to establish chronic infections that may last up to 20 years (Doetze *et al.*, 2000; Hoerauf *et al.*, 2005). However, it is also apparent that a small proportion of individuals develop protective immunity that keeps them free of parasites and disease despite a life-time of exposure (Hoerauf, Brattig, 2002; Brattig, 2004). This protective immunity is associated with a Th2 response, which is also responsible for the protective immunity induced by vaccination in murine models of filarial infections (Allen *et al.*, 2008; Katawa *et al.*, 2015; Kwarteng, Ahuno, 2017).

A strategy has been proposed to target parasite-driven immunomodulators and allow the host to mount a protective Th2 responses (Babayan *et al.*, 2012; Nisbet *et al.*, 2013). However, simply reducing the parasites immunomodulation may not be sufficient to

induce full protection, thus it has been suggested that the inclusion of a second antigen could induce greater protection.

Over the years of *O. volvulus* research, several potential vaccines candidates have been identified and tested in animal models (Morris *et al.*, 2013). The most promising of these are: *Ov*-RAL-2 (Lustigman *et al.*, 1992b; Bradley *et al.*, 1993; Hess *et al.*, 2014), *Ov*-103 (Lustigman *et al.*, 1992b; Hess *et al.*, 2014) and mutated *Ov*-CPI-2 (CPIm (Babayan *et al.*, 2012; Arumugam *et al.*, 2014b)). Each have consistently conferred high levels of protection (Babayan *et al.*, 2012; Arumugam *et al.*, 2014b; Hess *et al.*, 2014; Arumugam *et al.*, 2016).

CPIm, is derived from CPI-2, a known immunomodulator (Manoury *et al.*, 2001; Vray *et al.*, 2002; Gregory, Maizels, 2008), which has been modified to remove its modulatory residues (Murray *et al.*, 2005). In immunisation experiments in the *L. sigmodontis* model, using *Ls*-CPIm DNA vaccines reduced the numbers of circulating Mf (Babayan *et al.*, 2012), and in *B. malayi* model using *Bm*-CPIm recombinant protein vaccine, induced a 48% reduction in worm burden (Arumugam *et al.*, 2014b), providing evidence that targeting the immunomodulatory molecules of the parasites is a feasible vaccine strategy.

With this result in mind, the excretory and secretory (E/S) molecules of adult female worms were investigated to identify other potential immunomodulators (Armstrong *et al.*, 2014), as adult females have been shown to be responsible for the survival of Mf (Hoffmann *et al.*, 2001). From these analyses, a protein containing six ShK domains was particularly distinctive, as it was found in the ES of all mammalian life stages of

L. sigmodontis (Armstrong *et al.*, 2014) and homologues of *Ls*-ShK have been hypothesised to be able to modulate memory T cells (Beeton *et al.*, 2011; Chhabra *et al.*, 2014).

Initial DNA vaccine experiments with *Ls*-ShK in the *L. sigmodontis* model, showed promise by inducing protection in an initial experiment, as no Mf were detected the blood of *Ls*-ShK immunised mice (Duprez. J, MRes Thesis, 2013, University of Edinburgh). However, in repeat immunisation experiments, *Ls*-ShK failed to induce any protection, perhaps because of its immunomodulator properties (Armstrong *et al.*, 2014; Chhabra *et al.*, 2014). It was demonstrated that removing suppressive functions of immunomodulator vaccine candidates (e.g. CPI) could induce significant protection (Babayan *et al.*, 2012; Arumugam *et al.*, 2014b). Therefore, in subsequent vaccination experiments *Ls*-ShK was modified, in the hope that this would increase specific immune response and therefore induce protection.

DNA vaccines were chosen to screen vaccine candidates, in part because they are relatively easy and cheap to produce. However, it must be noted that DNA vaccines do not always induce strong protective immunity despite being able to generate specific immune responses (Donnelly *et al.*, 2005). Therefore, peptide vaccines were used to further investigate the candidate antigens. Peptides are an attractive alternative, as they encode specific epitopes of an antigen, and as consequence they are able to induce a highly targeted immune responses while eliminating possible allergic or hyper-reactive responses (Li *et al.*, 2014).

Peptides derived from *Ls*-RAL and *Ls*-103 were grouped together as these are highly immunogenic candidates, and peptides from known or hypothesised immunomodulators, *Ls*-CPI-2, *Ls*-ShK and *Ls*-Tgh2 were grouped together. This was done because individual peptides would take too long to test and immunising with a single antigenic epitope might not be enough to elicit an appropriate immune response.

Immunisation with the immunogenic or immunomodulatory peptide vaccine resulted in decreased numbers of circulating Mf in the blood. In contrast, the numbers of adult worms appeared unaffected by any vaccination protocol. This suggests that the peptides vaccines are affecting the Mf life stages only. Investigation of adult female fecundity revealed that the immunomodulatory and immunogenic groups of peptides are affecting either the development of Mf *in uteri* or the number (density) of Mf produced. This reduced adult female fecundity can also be seen in other immunisation experiments, such as in *L. sigmodontis* model with either DNA vaccination with *Ls*-CPIm (Babayan *et al.*, 2012) or the Mf life stage (Ziewer *et al.*, 2012); in *B. malayi* model with *Bm*-CPIm protein recombinants (Arumugam *et al.*, 2014b); as well as with immunisation with *B. malayi* ES in gerbils (Zipperer *et al.*, 2013). Suggesting that the vaccine candidates, are inducing protection by reducing female fecundity, instead of affecting adult worm survival.

These early results show promise and prompted questions about the use of a combination of antigens in a single inoculum. The groups which received all peptides in an immunisation dose showed no protection, but we know that combination of *Ov*-Ral-2 and *Ov*-103 does increase protective efficacy over individual antigens (Hess *et al.*, 2014; 2016). Therefore, selecting the most effective combination of antigens is

essential, and different combinations of immunogenic and immunomodulatory antigens will have to be investigated in the future.

The vaccine candidates were selected based on their protective efficacy in several animal models (measured as a decrease in parasite burden), however the mechanisms by which they induce protection and correlates of vaccine-induced immunity remain unknown. Identifying correlates of vaccine-induced protective immunity that are predictive of vaccine efficacy would aid in the screening of vaccine candidates, because currently vaccine efficacy is determined as decrease in parasite burden, measured late in a vaccination time course. Therefore, if vaccine efficacy could be predicted in the early stages of infection or even before infection, time and money could be saved (Mastelic *et al.*, 2013).

Several studies have taken a "systems biology/vaccinology" approach to identify correlates of vaccine induced immunity. Systems vaccinology consists of using high-throughput technologies including DNA microarrays, protein arrays and deep sequencing to enable systems-wide measurements, combined with predictive modelling. The advantage of a systems-wide approach versus conventional immunological methods which only analyse a single or small numbers of components of the immune system at a given time, is that by measuring changes across a whole system the full complexity and dynamics of the human immune system can be analysed.

To identify correlates of protection following immunisation in the *L. sigmodontis* model, a systems wide approach was taken. Blood samples were collected throughout

an immunisation experiment, with either irradiated L3 or microfilariae immunisations (Le Goff *et al.*, 2000; Ziewer *et al.*, 2012), and RNA from these samples were hybridised to Illumina microarrays, allowing the expression of tens of thousands of genes to be measured.

Machine learning methods were used to analyse the gene expression datasets over more commonly used methods such differential fold change or weighted gene correlation networks (WGCNA) (Langfelder, Horvath, 2008), because the microarray datasets produced were highly dimensional (many gene measured compared to sample numbers), with complex interaction between genes working within networks/pathways, for which the other methods were not well suited or powerful enough to detect changes (Butte, 2002; Huynh-Thu *et al.*, 2012; Sundarrajan, Arumugam, 2016). Machine learning has been used for a variety of tasks in biomedical sciences, such as classifying cancer subtypes (Anaissi *et al.*, 2013), to predicting outcomes of treatment (Gim *et al.*, 2016), with a newer role in vaccinology, such as predicting antibody responses to potential HIV vaccines (Choi *et al.*, 2015).

Using machine learning (ML), biologically relevant gene signatures were successfully detected following immunisation with either irradiated L3 and Mf. Gene signatures identified after immunisation with irradiated L3 were associated with neutrophil migration and chemotaxis. In *in vivo* studies, neutrophils recruitment to site of L3 infection have also been associated with protection (Pionnier *et al.*, 2016).

Immunisation with Mf elicited changes in gene expression, associated with antigen presentation by MHC class I, interferon signalling and T cell receptor signalling.

Immune responses to Mf are known to be linked to IFN-γ (Lawrence *et al.*, 2000; Saeftel *et al.*, 2003; Ziewer *et al.*, 2012), and has also been successfully identified by the ML pipeline. The ML pipeline also further identified IFN-α and IFN-β as being involved with Mf immunity, to date, activity of these cytokines has not been linked to filarial infection. However, in schistosomiasis IFN-β was produced by DC in response to *Schistosoma mansoni* eggs and is thought to be an immune evasion strategy (Trottein *et al.*, 2004), and IFN-β produced by DC stimulated with LPS was suggested to generate regulatory T cells (Wang *et al.*, 2014).

This study has shown that it is feasible to use a system wide approach to detect correlates of protection and information on the mechanism triggered by the vaccines in blood sampled during filarial infections. Furthermore, it demonstrates that it is the time points after immunisation that are the most predictive of vaccine efficacy. This approach has also been applied to commercially available vaccines such as: live attenuated vaccine YF-17D against yellow fever (Gaucher *et al.*, 2008; Querec *et al.*, 2008); the trivalent inactivated influenza vaccine (TIV) (Bucasas *et al.*, 2011; Nakaya *et al.*, 2011b); and a two vaccines against meningitis (MPSV4 and MCV4) (Li *et al.*, 2013b), and similar to this study, they found that the early gene signatures were predictive of vaccine immunogenicity, although mechanism of action did differ.

The results presented revealed that the best time point to identify immune correlates of protection was immediately after immunisation, and not when infection has been established. This can also be demonstrated when measuring antibody levels, which are often cumulative as infection progresses and can be positively associated with parasite

numbers. Similarly changes in gene expression measured late in infection, using either qPCR arrays (Chapter 2) or high-throughput techniques such as microarrays (Chapter 3), is too late to detect any changes associated with protection, and detectable changes in expression of immune genes were positively association with parasite burden.

Most investigations of immune responses against murine and human filarial infection have been carried out during chronic infections (once infection has been established), as often this is the only point infection can be detected. However, this could mean that interpretation of results and identification of protective responses may be difficult, since we have demonstrated that it is the early time point in infection that are the most informative. This is consistent with Taylor et al who showed that early priming of the immune system was important in determining the outcome of infection (Taylor *et al.*, 2009).

A wide range of pathways were found associated with protection using the ML pipeline on human gene expression data. Gene activity involved in the defence against other pathogens, humoral immune responses, DNA dependent replication and cell cycle pathways were detected in blood taken from onchocerciasis patients. In *W. bancrofti* infections, gene activity associated with protection appeared to be primarily involved with T cell receptor signalling. One of the reasons for which gene signatures associated with protection in humans is more varied, is because humans can be re-infected, have co-infections with other diseases, vary in age, unlike mice experiments, which are more controlled.

Despite the human dataset being noisier, the ML analysis detected common activity/pathways between the murine and human filarial infections. These results further increases confidence in the use of the *L. sigmodontis* mouse model for vaccine studies. Being able to bridge the gap between mice and humans will be tremendously helpful for vaccine development and indeed in the interpretation of any processes where model systems must be used. Mouse models have been useful and are necessary at identifying immune responses to filarial infections and evaluating potential vaccine candidates. Murine models exhibit much less variation in gene responses to infection than found in human subjects and hence results of murine studies provide a convenient starting point for analysis of human data.

However, there are significant differences in both the innate and adaptive arm of the immune system between humans and mice (Mestas, Hughes, 2004) and these should be carefully considered when interpretation gene expression data. Translating vaccine research from animal models to human trials is not always as straightforward. Some vaccines that have showed promise in mouse models, lacked efficacy in human trials (Gray *et al.*, 2011; Tameris *et al.*, 2013; Kaufmann *et al.*, 2014) or raised safety concerns during clinical trials (Diemert *et al.*, 2012).

Now that it has been shown that whole blood is a feasible sample to measure correlates of protection, these methods could be applied during vaccine trials. As filarial parasites have long life cycles, the ability to predict the immunogenicity of a vaccine within the first few days or weeks of a clinical trial, would help predict protection levels at the endpoint and ultimately save money and time if the vaccine candidate is deemed unlikely to be efficacious. However, these biomarkers of protection would have to be

well defined, as correlates of efficacy in healthy young adults may not be necessarily the same as biomarkers of protection in the elderly or children under 5 which are the targeted population (Nakaya *et al.*, 2011a).

Identifying biomarkers or signatures of safety will also be critical for successful vaccine development and to avoid adverse reactions in clinical trials. The human data set used in the studies describe here lacked information on pathology and therefore identification of potential signatures of adverse reactions was limited.

Since the microarray analysis identified the early time points following immunisation as being the most predictive of vaccine efficacy, the effect of *Ls*-CPIm (one of the vaccine candidates) on dendritic cells (DC) was investigated. Dendritic cells are a population of professional antigen presenting cells that have the ability to sense foreign pathogens and initiate a type 1, type 2 or regulatory immune response (Motran *et al.*, 2016). Pathogens have evolved strategies to modulate DC, such as the excretory/secretory products produced by filarial nematodes, which have shown to impair DC function, and suppress both Th1 and Th2 adaptive immune response as well as inducing regulatory T cells (Segura *et al.*, 2007; Carvalho *et al.*, 2009; Terrazas *et al.*, 2013). *Ls*-CPIm is derived from CPI-2, which is known to affect antigen presentation by MHC II molecules on antigen presenting cells such as DC (Manoury *et al.*, 2001). Immunisation with the native CPI shows little protection in *L. sigmodontis* models (Babayan *et al.*, 2012), and in *B. malayi* model, immunisation with CPI changes the migratory patterns of adult worms and but confers no protection (Arumugam *et al.*, 2014a). However, when mutated (Murray *et al.*, 2005), there is an

increase in protection seen following immunisation (Babayan *et al.*, 2012; Arumugam *et al.*, 2014b).

In vitro bone marrow derived DC cultures were used to show that the mutation of CPI rescues DC activation and maturation. Since DC are a key link between innate and adaptive immunity (Liu, 2001a), it was suggested that CPIm works by increasing the activation of DC and hence increasing adaptive immunity, which eventually leads to protective immunity. Although this could not be demonstrated *in vivo*, this was most likely due to lack of sensitivity associated with a presumed small population of DC that have taken up the CPI or CPIm. Measuring changes in gene expression in whole blood is ideal to detect systemic changes, however if one wants to measure a more specific cellular response to immunisation, such as ones by DC, then cells must first be isolated. Therefore, to determine if CPIm vaccine is indeed rescuing DC activation *in vivo*, then changes in gene expression would need to be investigated within isolated DC population.

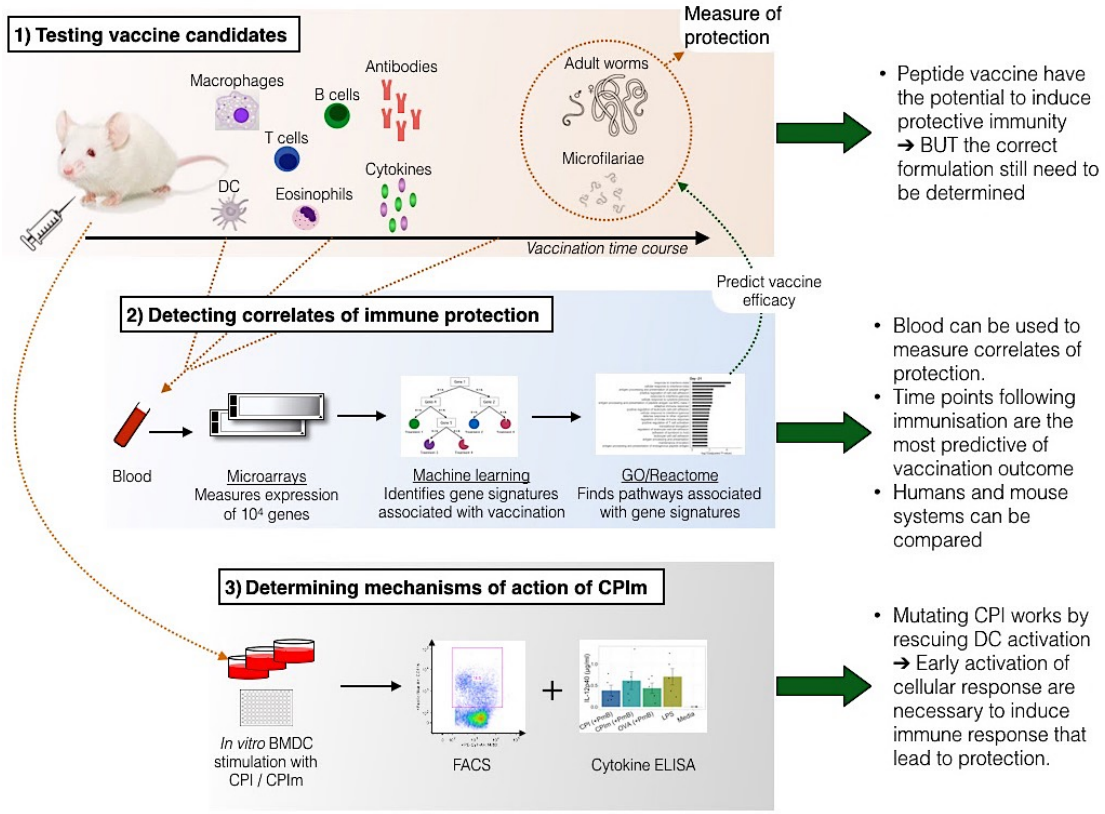

Figure 5.1. Overview of thesis. 1) Vaccine candidates were tested in the *L. sigmodontis* mouse model, either using DNA or peptide vaccines. Vaccine efficacy or protection is measured as a change in adult worm and microfilariae numbers, at the end of a vaccination time course, in *L. sigmodontis* model this is around 65-days post challenge infection. Immunological readouts such as percentage of cell types, cytokine levels and antibody levels were also measured at the end of the time course, although these are not always predictive of vaccine efficacy. 2) Determine correlates of vaccine induced protection can be beneficial for vaccine trials, such as identifying early predictors of vaccine efficacy could help speed up clinical trials. The simplest method to identify correlates of protection would be to identify gene signatures in blood using high-throughput techniques (microarrays), but whether these could be measured in blood during filarial infection remained unknown, and if possible at what point during infection, since at the end it is too late to detect any changes. Therefore, blood was collected at several time points in a vaccination time course, RNA was extracted and hybridised to Illumina microarrays to measure the expression of tens of thousands of genes. Since the microarrays generated complex datasets, a machine learning pipeline was created to detect changes in gene signatures between immunised and non-immunised mice. This pipeline successfully detected changes in gene expression that were known to be induced by these vaccines, showing that blood is indeed a good tissue to use and that machine learning

techniques can detect biologically significant changes in gene expression. Furthermore, time points following immunisation showed to be the most informative, therefore in the future these can be used to predict vaccine efficacy. Parallel to mice studies, blood samples were collected from naturally protected and infected individuals living in foci endemic for *O. volvulus* and *W. bancrofti*, using microarrays and the machine learning pipeline, correlates of protection in human were investigated and some overlap was found between vaccine induced immunity in mice and natural protection in humans. 3) Most vaccine candidates were chosen without knowing their mechanism of action, since one of the vaccine candidates, *Ls*-CPIm, was derived from CPI-2, an immunomodulatory known to affect antigen presentation on cells such as DC. Therefore, the effect of *Ls*-CPI and mutated *Ls*-CPI (CPIm) on bone-marrow derived DC (BMDC) were compared, finding that *Ls*-CPIm rescues DC activation. Showing that increased protection induced by *Ls*-CPIm is due to increased DC activity, although this was not shown *in vivo*. With the microarrays analysis that shows that is the early time points following immunisation that determines the outcome of vaccination, it could be suggested that it is the early activation of cellular response, such as DC that are necessary to induce immune response that lead to protection.

5.1 Conclusion

The work presented in this thesis shows that targeting immunomodulatory molecules is a feasible strategy for vaccination against onchocerciasis. Experience suggest that it might be necessary to modify the native molecule to remove their immunomodulatory properties as part of the formulation of any vaccine. This was done to one of the most promising vaccine candidate, CPI. A mutated form of CPI that lacks immunomodulatory activity evoked greater protection which was associated with increased DC activation. The correct formulation of any vaccine will be a critical component for successful development of a vaccine.

The use of a systems approach and machine learning has demonstrated that changes in gene expression can be detected in whole blood following exposure to and immunisation against filarial infections. This opens many possibilities for vaccine development, by more rapidly understanding the mechanisms associated with successful immunisation and formulation to optimise the immunogenicity and efficacy of vaccines, to achieve stronger and longer protection. The potential to recognise and predict possible adverse reactions to specific molecules will improve safety of any vaccine. This approach also brings the possibility of identifying early predictors of vaccine efficacy and this knowledge could greatly speed up vaccine trials.

It has been demonstrated that early changes in the immune system immediately following exposure to infection or immunisation are critical in determining the outcome of infection. These observation questions the value of many earlier investigations of immune response to filariae that have used material collected from patients presenting with chronic infections. The results also suggest that to determine

if an immunisation is successful, a detailed investigation of responses immediately after exposure to infection is needed. This can be done in murine models but for humans studies the focus must now turn to the examination of very young children living in endemic areas.

These studies have, by necessity, focussed on the capacity of experimental vaccines to reduce worm burden. A vaccine that reduces the number of microfilariae in the skin will reduce morbidity. However, if onchocerciasis is to be eliminated from any region, or even eradicated. The reduction in microfilariae loads must be such as to block transmission. The aim is to vaccinate individuals currently omitted from MDA, such as pre-school children, and by including this cohort in an integrated control programme with Ivermectin distribution could be a major step towards elimination of onchocerciasis.

References

Abegunde, A. T., Ahuja, R. M., Okafor, N. J. (2016). Doxycycline plus ivermectin versus ivermectin alone for treatment of patients with onchocerciasis. Abegunde, A. T. (ed.). The Cochrane database of systematic reviews, 197 (1), Chichester, UK: John Wiley & Sons, Ltd, p.CD011146. [Online]. Available at: doi:10.1002/14651858.CD011146.pub2.

Abraham, D., Grieve, R. B., Holy, J. M., Christensen, B. M. (1989). Immunity to larval *Brugia malayi* in BALB/c mice: protective immunity and inhibition of larval development. American Journal of Tropical Medicine and Hygiene, 40 (6), p.598–604.

Abraham, D., Leon, O., Schnyder-Candrian, S., Wang, C. C., Galioto, A. M., Kerepesi, L. A., Lee, J. J., Lustigman, S. (2004). Immunoglobulin E and Eosinophil-Dependent Protective Immunity to Larval *Onchocerca volvulus* in Mice Immunized with Irradiated Larvae. Infection and Immunity, 72 (2), p.810–817. [Online]. Available at: doi:10.1128/IAI.72.2.810-817.2004.

Abraham, D., Lucius, R., Trees, A. J. (2002). Immunity to *Onchocerca spp*. in animal hosts. Trends in Parasitology, 18 (4), p.164–171.

Achukwi, M. D., Harnett, W., Enyong, P., Renz, A. (2007). Successful vaccination against *Onchocerca ochengi* infestation in cattle using live *Onchocerca volvulus* infective larvae. Parasite Immunology, 29 (3), p.113–116. [Online]. Available at: doi:10.1111/j.1365-3024.2006.00917.x.

Ainali, C. (2013). Machine learning for translational medicine. PhD Thesis. King's College London, University of London.

Ajendra, J., Specht, S., Neumann, A.-L., Gondorf, F., Schmidt, D., Gentil, K., Hoffmann, W. H., Taylor, M. J., Hoerauf, A., H bner, M. P. (2014). ST2 Deficiency Does Not Impair Type 2 Immune Responses during Chronic Filarial Infection but Leads to an Increased Microfilaremia Due to an Impaired Splenic Microfilarial Clearance. Stager, S. (ed.). PLoS ONE, 9 (3), p.e93072–13. [Online]. Available at: doi:10.1371/journal.pone.0093072.

Ajendra, J., Specht, S., Ziewer, S., Schiefer, A., Pfarr, K., Parčina, M., Kufer, T. A., Hoerauf, A., Hübner, M. P. (2016). NOD2 dependent neutrophil recruitment is required for early protective immune responses against infectious *Litomosoides sigmodontis* L3 larvae. Nature Publishing Group, 6, Nature Publishing Group, p.1–16. [Online]. Available at: doi:10.1038/srep39648.

Al-Qaoud, K. M., Fleischer, B., Hoerauf, A. (1998). The Xid defect imparts susceptibility to experimental murine filariosis--association with a lack of antibody and IL-10 production by B cells in response to phosphorylcholine. International Immunology, 10 (1), Oxford University Press, p.17–25. [Online].

Available at: doi:10.1093/intimm/10.1.17.

Al-Qaoud, K. M., Pearlman, E., Hartung, T., Klukowski, J., Fleischer, B., Hoerauf, A. (2000). A new mechanism for IL-5-dependent helminth control: neutrophil accumulation and neutrophil-mediated worm encapsulation in murine filariasis are abolished in the absence of IL-5. International Immunology, 12 (6), p.899–908.

Al-Qaoud, K. M., Taubert, A., Zahner, H., Fleischer, B., Hoerauf, A. (1997). Infection of BALB/c mice with the filarial nematode *Litomosoides sigmodontis*: role of CD4+ T cells in controlling larval development. Infection and Immunity, 65 (6), American Society for Microbiology (ASM), p.2457–2461.

Albiez, E. J., Newland, H. S., White, A. T., Kaiser, A., Greene, B. M., Taylor, H. R., Buttner, D. W. (1988). Chemotherapy of onchocerciasis with high doses of diethylcarbamazine or a single dose of ivermectin: microfilaria levels and side effects. Tropical medicine and parasitology: official organ of Deutsche Tropenmedizinische Gesellschaft and of Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ), 39 (1), p.19–24.

Allen, J. E., Adjei, O., Bain, O., Hoerauf, A., Hoffmann, W. H., Makepeace, B. L., Schulz-Key, H., Tanya, V. N., Trees, A. J., Wanji, S., Taylor, D. W. (2008). Of Mice, Cattle, and Humans: The Immunology and Treatment of River Blindness. Lustigman, S. (ed.). 2 (4), p.e217. [Online]. Available at: doi:10.1371/journal.pntd.0000217.t001.

Allen, J. E., Daub, J., Guiliano, D., McDonnell, A., Lizotte-Waniewski, M., TayloR, D. W., Blaxter, M. (2000). Analysis of Genes Expressed at the Infective Larval Stage Validates Utility of *Litomosoides sigmodontis* as a Murine Model for Filarial Vaccine Development. Infection and Immunity, 68 (9), p.5454–5458. [Online]. Available at: doi:10.1128/IAI.68.9.5454-5458.2000.

Anaissi, A., Kennedy, P. J., Goyal, M., Catchpoole, D. R. (2013). A balanced iterative random forest for gene selection from microarray data. BMC Bioinformatics, 14 (1), BioMed Central, p.261. [Online]. Available at: doi:10.1186/1471-2105-14-261.

Anand, S. B., Murugan, V., Prabhu, P. R., Anandharaman, V., Reddy, M. V. R., Kaliraj, P. (2008). Comparison of immunogenicity, protective efficacy of single and cocktail DNA vaccine of *Brugia malayi* abundant larval transcript (ALT-2) and thioredoxin peroxidase (TPX) in mice. Acta Tropica, 107 (2), p.106–112. [Online]. Available at: doi:10.1016/j.actatropica.2008.04.018.

Anand, S. B., Rajagopal, V., Kaliraj, P. (2012). *Brugia malayi* Thioredoxin Peroxidase as a Potential Vaccine Candidate Antigen for Lymphatic Filariasis. Applied Biochemistry and Biotechnology, 167 (5), p.1351–1364. [Online]. Available at: doi:10.1007/s12010-012-9643-6.

APOC. (2013). The World Health Organization Year 2013 Progress Report. p.1–64.

Arlot, S., Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4 (0), p.40–79. [Online]. Available at: doi:10.1214/09-SS054.

Armstrong, S. D., Babayan, S. A., Lhermitte-Vallarino, N., Gray, N., Xia, D., Martin, C., Kumar, S., Taylor, D. W., Blaxter, M. L., Wastling, J. M., Makepeace, B. L. (2014). Comparative Analysis of the Secretome from a Model Filarial Nematode (*Litomosoides sigmodontis*) reveals Maximal Diversity in Gravid Female Parasites. Molecular & Cellular Proteomics, p.1–58.

Armstrong, S. D., Xia, D., Bah, G. S., Krishna, R., Ngangyung, H. F., LaCourse, E. J., McSorley, H. J., Kengne-Ouafo, J. A., Chounna Ndongmo, P. W., Wanji, S., Enyong, P. A., Taylor, D. W., Blaxter, M. L., Wastling, J. M., Tanya, V. N., Makepeace, B. L. (2016). Stage-specific Proteomes from *Onchocerca ochengi*, Sister Species of the Human River Blindness Parasite, Uncover Adaptations to a Nodular Lifestyle. Molecular & cellular proteomics: MCP, 15 (8), American Society for Biochemistry and Molecular Biology, p.2554–2575. [Online]. Available at: doi:10.1074/mcp.M115.055640.

Arndts, K., Specht, S., Debrah, A. Y., Tamarozzi, F., Klarmann Schulz, U., Mand, S., Batsa, L., Kwarteng, A., Taylor, M., Adjei, O., Martin, C., Layland, L. E., Hoerauf, A. (2014). Immunoepidemiological profiling of onchocerciasis patients reveals associations with microfilaria loads and ivermectin intake on both individual and community levels. 8 (2), p.e2679. [Online]. Available at: doi:10.1371/journal.pntd.0002679.

Artis, D., Kane, C. M., Fiore, J., Zaph, C., Shapira, S., Joyce, K., Macdonald, A., Hunter, C., Scott, P., Pearce, E. J. (2005). Dendritic cell-intrinsic expression of NF-kappa B1 is required to promote optimal Th2 cell differentiation. The Journal of Immunology, 174 (11), p.7154–7159.

Arumugam, S., Bin Zhan, Abraham, D., Ward, D., Lustigman, S., Klei, T. R. (2014a). Vaccination with recombinant *Brugia malayi* cystatin proteins alters worm migration, homing and final niche selection following a subcutaneous challenge of Mongolian gerbils (*Meriones unguiculatus*) with *B. malayi* infective larvae. Parasites & Vectors, 7 (1), Parasites & Vectors, p.1–7. [Online]. Available at: doi:10.1186/1756-3305-7-43.

Arumugam, S., Wei, J., Liu, Z., Abraham, D., Bell, A., Bottazzi, M. E., Hotez, P. J., Zhan, B., Lustigman, S., Klei, T. R. (2016). Vaccination of Gerbils with *Bm*-103 and *Bm*-RAL-2 Concurrently or as a Fusion Protein Confers Consistent and Improved Protection against *Brugia malayi* Infection. Mitre, E. (ed.). PLOS Neglected Tropical Diseases, 10 (4), p.e0004586–26. [Online]. Available at: doi:10.1371/journal.pntd.0004586.

Arumugam, S., Wei, J., Ward, D., Abraham, D., Lustigman, S., Bin Zhan, Klei, T. R.

(2014b). Vaccination with a genetically modified *Brugia malayi* cysteine protease inhibitor-2 reduces adult parasite numbers and affects the fertility of female worms following a subcutaneous challenge of Mongolian gerbils (*Meriones unguiculatus*) with *B. malayi* infective larvae. International Journal for Parasitology, 44 (10), Australian Society for Parasitology Inc., p.675–679. [Online]. Available at: doi:10.1016/j.ijpara.2014.05.003.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. Nature Genetics, 25 (1), p.25–29. [Online]. Available at: doi:10.1038/75556.

Attout, T., Martin, C., Babayan, S. A., Kozek, W. J., Bazzocchi, C., Oudet, F., Gallagher, I. J., Specht, S., Bain, O. (2008). Pleural cellular reaction to the filarial infection *Litomosoides sigmodontis* is determined by the moulting process, the worm alteration, and the host strain. Parasitology International, 57 (2), p.201–211. [Online]. Available at: doi:10.1016/j.parint.2008.01.001.

Awadzi, K., Edwards, G., Duke, B. O. L., Opoku, N. O., Attah, S. K., Addy, E. T., Ardrey, A. E., Quartey, B. T. (2013). The co-administration of ivermectin and albendazole - safety, pharmacokinetics and efficacy against *Onchocerca volvulus*. Annals of Tropical Medicine and Parasitology, 97 (2), Taylor & Francis, p.165–178. [Online]. Available at: doi:10.1179/000349803235001697.

Babayan, S. A., Luo, H., Gray, N., Taylor, D. W., Allen, J. E. (2012). Deletion of Parasite Immune Modulatory Sequences Combined with Immune Activating Signals Enhances Vaccine Mediated Protection against Filarial Nematodes. Davies, S. J. (ed.). 6 (12), p.e1968. [Online]. Available at: doi:10.1371/journal.pntd.0001968.t001.

Babayan, S. A., Read, A. F., Lawrence, R. A., Bain, O., Allen, J. E. (2010). Filarial parasites develop faster and reproduce earlier in response to host immune effectors that determine filarial life expectancy. Babayan, S. A., Read, A. F., Lawrence, R. A., Bain, O., Allen, J. E. (eds.). PLoS Biology, 8 (10), p.e1000525. [Online]. Available at: doi:10.1371/journal.pbio.1000525.t001.

Babayan, S., Attout, T., Harris, A., Taylor, M., Le Goff, L., Vuong, P., Renia, L., Allen, J., Bain, O. (2006). Vaccination against filarial nematodes with irradiated larvae provides long-term protection against the third larval stage but not against subsequent life cycle stages. International Journal for Parasitology, 36 (8), p.903–914. [Online]. Available at: doi:10.1016/j.ijpara.2006.04.013.

Babayan, S., Ungeheuer, M.-N. E., Martin, C., Attout, T., Belnoue, E., Snounou, G., enia, L. R., Korenaga, M., Bain, O. (2003). Resistance and Susceptibility to Filarial Infection with *Litomosoides sigmodontis* Are Associated with Early Differences in Parasite Development and in Localized Immune Reactions. 12 ed.

Infection and Immunity, 71, p.6820–6829. [Online]. Available at: doi:10.1128/IAI.71.12.6820–6829.2003.

Babu, S., Ganley, L. M., Klei, T. R., Shultz, L. D., Rajan, T. V. (2000). Role of gamma interferon and interleukin-4 in host defense against the human filarial parasite *Brugia malayi*. Infection and Immunity, 68 (5), p.3034–3035.

Babu, S., Nutman, T. B. (2012). Immunopathogenesis of lymphatic filarial disease. Seminars in Immunopathology, 34 (6), p.847–861. [Online]. Available at: doi:10.1007/s00281-012-0346-4.

Bain, R. K. (1999). Irradiated vaccines for helminth control in livestock. International Journal for Parasitology, 29 (1), p.185–191.

Balic, A., Harcus, Y., Holland, M. J., Maizels, R. M. (2004). Selective maturation of dendritic cells by *Nippostrongylus brasiliensis*-secreted proteins drives Th2 immune responses. European Journal of Immunology, 34 (11), p.3047–3059. [Online]. Available at: doi:10.1002/eji.200425167.

Basano, S. de A., Fontes, G., Medeiros, J. F., Aranha Camargo, J. S. de A., Souza Vera, L. J., Parente Araújo, M. P., Pires Parente, M. S., Mattos Ferreira, R. de G., Barreto Crispim, P. D. T., Aranha Camargo, L. M. (2014). Sustained clearance of *Mansonella ozzardi* infection after treatment with ivermectin in the Brazilian Amazon. The American journal of Tropical Medicine and Hygiene, 90 (6), p.1170–1175. [Online]. Available at: doi:10.4269/ajtmh.13-0410.

Basáñez, M. G., Walker, M., Turner, H. C., Coffeng, L. E., de Vlas, S. J., Stolk, W. A. (2016). River Blindness: Mathematical Models for Control and Elimination. Advances in parasitology, 94, Elsevier, p.247–341. [Online]. Available at: doi:10.1016/bs.apar.2016.08.003.

Basáñez, M.-G., Pion, S. D. S., Boakes, E., Filipe, J. A. N., Churcher, T. S., Boussinesq, M. (2008). Effect of single-dose ivermectin on *Onchocerca volvulus*: a systematic review and meta-analysis. The Lancet Infectious Diseases, 8 (5), p.310–322. [Online]. Available at: doi:10.1016/S1473-3099(08)70099-9.

Beck, L., Spiegelberg, H. L. (1989). The polyclonal and antigen-specific IgE and IgG subclass response of mice injected with ovalbumin in alum or complete Freund's adjuvant. Cellular immunology, 123 (1), p.1–8.

Beekman, N. J., Schaaper, W. M., Tesser, G. I., Dalsgaard, K., Kamstrup, S., Langeveld, J. P., Boshuizen, R. S., Meloen, R. H. (1997). Synthetic peptide vaccines: palmitoylation of peptide antigens by a thioester bond increases immunogenicity. The journal of peptide research: official journal of the American Peptide Society, 50 (5), p.357–364.

Beeton, C., Pennington, M. W., Norton, R. S. (2011). Analogs of the sea anemone potassium channel blocker ShK for the treatment of autoimmune diseases.

Inflammation & allergy drug targets, 10 (5), p.313–321.

Benjamini, Y., Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society. Series B (Methodological), 57 (1), [Royal Statistical Society, Wiley], p.289–300.

Bennuru, S., Cotton, J. A., Ribeiro, J. M. C., Grote, A., Harsha, B., Holroyd, N., Mhashilkar, A., Molina, D. M., Randall, A. Z., Shandling, A. D., Unnasch, T. R., Ghedin, E., Berriman, M., Lustigman, S., Nutman, T. B. (2016). Stage-Specific Transcriptome and Proteome Analyses of the Filarial Parasite *Onchocerca volvulus* and Its *Wolbachia* Endosymbiont. mBio, 7 (6), p.e02028–16–11. [Online]. Available at: doi:10.1128/mBio.02028-16.

Bennuru, S., Meng, Z., Ribeiro, J. M. C., Semnani, R. T., Ghedin, E., Chan, K., Lucas, D. A., Veenstra, T. D., Nutman, T. B. (2011). Stage-specific proteomic expression patterns of the human filarial parasite *Brugia malayi* and its endosymbiont *Wolbachia*. Proceedings of the National Academy of Sciences of the United States of America, 108 (23), National Acad Sciences, p.9649–9654. [Online]. Available at: doi:10.1073/pnas.1011481108.

Bennuru, S., Pion, S. D. S., Kamgno, J., Wanji, S., Nutman, T. B. (2014). Repurposed Automated Handheld Counter as a Point-of-Care Tool to Identify Individuals 'At Risk' of Serious Post-Ivermectin Encephalopathy. Sakanari, J. (ed.). PLOS Neglected Tropical Diseases, 8 (9), p.e3180–e3187. [Online]. Available at: doi:10.1371/journal.pntd.0003180.

Bennuru, S., Semnani, R., Meng, Z., Ribeiro, J. M. C., Veenstra, T. D., Nutman, T. B. (2009). *Brugia malayi* Excreted/Secreted Proteins at the Host/Parasite Interface: Stage- and Gender-Specific Proteomic Profiling. Ghedin, E. (ed.). PLOS Neglected Tropical Diseases, 3 (4), p.e410–e412. [Online]. Available at: doi:10.1371/journal.pntd.0000410.

Blacklock, D. B. (2016). The Development of *Onchocerca Volvulus* in *Simulium Damnosum*. Annals of Tropical Medicine and Parasitology, 20 (1), Taylor & Francis, p.1–48. [Online]. Available at: doi:10.1080/00034983.1926.11684476.

Bogdan, C., Mattner, J., Schleicher, U. (2004). The role of type I interferons in non-viral infections. Immunological reviews, 202 (1), Munksgaard International Publishers, p.33–48. [Online]. Available at: doi:10.1111/j.0105-2896.2004.00207.x.

Bolón-Canedo, V., Sánchez-Maroño, N., Alonso-Betanzos, A., Benítez, J. M., Herrera, F. (2014). A review of microarray datasets and applied feature selection methods. Information Sciences, 282 (C), Elsevier Inc., p.111–135. [Online]. Available at: doi:10.1016/j.ins.2014.05.042.

Bondar, G., Cadeiras, M., Wisniewski, N., Maque, J., Chittoor, J., Chang, E., Bakir,

M., Starling, C., Shahzad, K., Ping, P., Reed, E., Deng, M. (2014). Comparison of Whole Blood and Peripheral Blood Mononuclear Cell Gene Expression for Evaluation of the Perioperative Inflammatory Response in Patients with Advanced Heart Failure. Sussman, M. A. (ed.). PLoS ONE, 9 (12), p.e115097–22. [Online]. Available at: doi:10.1371/journal.pone.0115097.

Bouchery, T., Dénécé, G., Attout, T., Ehrhardt, K., Lhermitte-Vallarino, N., Hachet-Haas, M., Galzi, J. L., Brotin, E., Bachelerie, F., Gavotte, L., Moulia, C., Bain, O., Martin, C. (2012a). The Chemokine CXCL12 Is Essential for the Clearance of the Filaria *Litomosoides sigmodontis* in Resistant Mice. Rénia, L. (ed.). PLoS ONE, 7 (4), p.e34971. [Online]. Available at: doi:10.1371/journal.pone.0034971.t001.

Bouchery, T., Ehrhardt, K., Lefoulon, E., Hoffmann, W., Bain, O., Martin, C. (2012b). Differential tissular distribution of *Litomosoides sigmodontis* microfilariae between microfilaremic and amicrofilaremic mice following experimental infection. Parasite, 19 (4), p.351–358. [Online]. Available at: doi:10.1051/parasite/2012194351.

Boussinesq, M. (2013). Loiasis. Annals of Tropical Medicine and Parasitology, 100 (8), p.715–731. [Online]. Available at: doi:10.1179/136485906X112194.

Boussinesq, M., Gardon, J., Gardon-Wendel, N., Chippaux, J.-P. (2003). Clinical picture, epidemiology and outcome of Loa-associated serious adverse events related to mass ivermectin treatment of onchocerciasis in Cameroon. Filaria Journal, 2 Suppl 1, p.S4. [Online]. Available at: doi:10.1186/1475-2883-2-S1-S4.

Boyd, A., Killoran, K., Mitre, E., Nutman, T. B. (2015). Pleural cavity type 2 innate lymphoid cells precede Th2 expansion in murine *Litomosoides sigmodontis* infection. EXPERIMENTAL PARASITOLOGY, 159 (c), Elsevier Inc, p.118–126. [Online]. Available at: doi:10.1016/j.exppara.2015.09.006.

Bradley, J. E., Tuan, R. S., Shepley, K. J., Tree, T. I., Maizels, R. M., Helm, R., Gregory, W. F., Unnasch, T. R. (1993). Onchocerca volvulus: characterization of an immunodominant hypodermal antigen present in adult and larval parasites. Experimental Parasitology, 77 (4), p.414–424. [Online]. Available at: doi:10.1006/expr.1993.1101.

Brattig, N. W. (2004). Pathogenesis and host responses in human onchocerciasis: impact of *Onchocerca* filariae and *Wolbachia* endobacteria. Microbes and Infection, 6 (1), p.113–128. [Online]. Available at: doi:http://dx.doi.org/10.1016/j.micinf.2003.11.003.

Brattig, N. W., Lepping, B., Timmann, C., Buttner, D. W., Marfo, Y., Hamelmann, C., Horstmann, R. D. (2002). *Onchocerca volvulus*-exposed persons fail to produce interferon-gamma in response to *O. volvulus* antigen but mount proliferative responses with interleukin-5 and IL-13 production that decrease with increasing

microfilarial density. The Journal of infectious diseases, 185 (8), p.1148–1154. [Online]. Available at: doi:10.1086/339820.

Brattig, N. W., Tischendorf, F. W., Strote, G., Medina-de la Garza, C. E. (1991). Eosinophil-larval-interaction in onchocerciasis: heterogeneity of *in vitro* adherence of eosinophils to infective third and fourth stage larvae and microfilariae of *Onchocerca volvulus*. Parasite Immunology, 13 (1), p.13–22.

Breiman, L. (2001). Random Forests. Machine Learning, 45 (1), Kluwer Academic Publishers, p.5–32. [Online]. Available at: doi:10.1023/A:1010933404324.

Bucasas, K. L., Franco, L. M., Shaw, C. A., Bray, M. S., Wells, J. M., Nino, D., Arden, N., Quarles, J. M., Couch, R. B., Belmont, J. W. (2011). Early Patterns of Gene Expression Correlate With the Humoral Immune Response to Influenza Vaccination in Humans. Journal of Infectious Diseases, 203 (7), p.921–929. [Online]. Available at: doi:10.1093/infdis/jiq156.

Butte, A. (2002). The use and analysis of microarray data. Nature Reviews Drug Discovery, 1 (12), p.951–960. [Online]. Available at: doi:10.1038/nrd961.

Campello, R. J. G. B., Moulavi, D., Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In: Advances in Knowledge Discovery and Data Mining, Lecture Notes in Computer Science, 7819 (Chapter 14), Berlin, Heidelberg: Springer, Berlin, Heidelberg, p.160–172. [Online]. Available at: doi:10.1007/978-3-642-37456-2_14.

Cano, J., Rebollo, M. P., Golding, N., Pullan, R. L., Crellen, T., Soler, A., Kelly- Hope, L. A., Lindsay, S. W., Hay, S. I., Bockarie, M. J., Brooker, S. J. (2014). The global distribution and transmission limits of lymphatic filariasis: past and present. Parasites & Vectors, 7 (1), p.1–19. [Online]. Available at: doi:10.1186/s13071-014-0466-x.

Carter, T., Sumiya, M., Reilly, K., Ahmed, R., Sobieszczuk, P., Summerfield, J. A., Lawrence, R. A. (2007). Mannose-Binding Lectin A-Deficient Mice Have Abrogated Antigen-Specific IgM Responses and Increased Susceptibility to a Nematode Infection. The Journal of Immunology, 178 (8), p.5116–5123. [Online]. Available at: doi:10.4049/jimmunol.178.8.5116.

Carvalho, L., Sun, J., Kane, C., Marshall, F., Krawczyk, C., Pearce, E. J. (2009). Review series on helminths, immune modulation and the hygiene hypothesis: Mechanisms underlying helminth modulation of dendritic cell function. Immunology, 126 (1), p.28–34. [Online]. Available at: doi:10.1111/j.1365-2567.2008.03008.x.

Chandrashekar, R., Curtis, K. C., Lu, W., Weil, G. J. (1998). Molecular cloning of an enzymatically active thioredoxin peroxidase from *Onchocerca volvulus*. Molecular and Biochemical Parasitology, 93 (2), p.309–312. [Online]. Available at: doi:http://dx.doi.org/10.1016/S0166-6851(98)00041-3.

Cheke, R. A. (2017). Factors affecting onchocerciasis transmission: lessons for infection control. Expert Review of Anti-infective Therapy, 15 (4), Taylor & Francis, p.1–11. [Online]. Available at: doi:10.1080/14787210.2017.1286980.

Chen, X., Ishwaran, H. (2012). Random forests for genomic data analysis. Genomics, 99 (6), Elsevier Inc., p.323–329. [Online]. Available at: doi:10.1016/j.ygeno.2012.04.003.

Chesnais, C. B., Takougang, I., Paguélé, M., Pion, S. D., Boussinesq, M. (2017). Excess mortality associated with loiasis: a retrospective population-based cohort study. The Lancet Infectious Diseases, 17 (1), p.108–116. [Online]. Available at: doi:10.1016/S1473-3099(16)30405-4.

Chhabra, S., Chang, S. C., Nguyen, H. M., Huq, R., Tanner, M. R., Londono, L. M., Estrada, R., Dhawan, V., Chauhan, S., Upadhyay, S. K., Gindin, M., Hotez, P. J., Valenzuela, J. G., Mohanty, B., Swarbrick, J. D., Wulff, H., Iadonato, S. P., Gutman, G. A., Beeton, C., Pennington, M. W., Norton, R. S., Chandy, K. G. (2014). Kv1.3 channel-blocking immunomodulatory peptides from parasitic worms: implications for autoimmune diseases. FASEB journal: official publication of the Federation of American Societies for Experimental Biology. [Online]. Available at: doi:10.1096/fj.14-251967.

Cho-Ngwa, F., Liu, J., Lustigman, S. (2010). The *Onchocerca volvulus* Cysteine Proteinase Inhibitor, *Ov*-CPI-2, Is a Target of Protective Antibody Response That Increases with Age. Mackenzie, C. D. (ed.). 4 (8), p.e800. [Online]. Available at: doi:10.1371/journal.pntd.0000800.t003.

Choi, I., Chung, A. W., Suscovich, T. J., Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., O'Connell, R. J., Francis, D., Robb, M. L., Michael, N. L., Kim, J. H., Alter, G., Ackerman, M. E., Bailey-Kellogg, C. (2015). Machine learning methods enable predictive modeling of antibody feature: function relationships in RV144 vaccinees. PLoS Computational Biology, 11 (4), p.e1004185. [Online]. Available at: doi:10.1371/journal.pcbi.1004185.

Coffeng, L. E., Stolk, W. A., Hoerauf, A., Habbema, D., Bakker, R., Hopkins, A. D., de Vlas, S. J. (2014). Elimination of African Onchocerciasis: Modeling the Impact of Increasing the Frequency of Ivermectin Mass Treatment. McCaw, J. M. (ed.). PLoS ONE, 9 (12), p.e115886–25. [Online]. Available at: doi:10.1371/journal.pone.0115886.

Coffeng, L. E., Stolk, W. A., Zouré, H. G. M., Veerman, J. L., Agblewonu, K. B., Murdoch, M. E., Noma, M., Fobi, G., Richardus, J. H., Bundy, D. A. P., Habbema, D., de Vlas, S. J., Amazigo, U. V. (2013). African Programme for Onchocerciasis Control 1995–2015: Model-Estimated Health Impact and Cost. Basáñez, M.-G. (ed.). 7 (1), p.e2032. [Online]. Available at: doi:10.1371/journal.pntd.0002032.s001.

Cook, P. C., Jones, L. H., Jenkins, S. J., Wynn, T. A., Allen, J. E., MacDonald, A. S. (2012). Alternatively activated dendritic cells regulate CD4+ T-cell polarization in vitro and in vivo. Proceedings of the National Academy of Sciences of the United States of America, 109 (25), National Acad Sciences, p.9977–9982. [Online]. Available at: doi:10.1073/pnas.1121231109.

Cook, P. C., Owen, H., Deaton, A. E. E. M., Borger, J. G., Brown, S. L., Clouaire, T., Jones, G.-R., Jones, L. H., Lundie, R. J., Marley, A. K., Morrison, V. L., Phythian-Adams, A. T., Wachter, E., Webb, L. M., Sutherland, T. E., Thomas, G. D., Grainger, J. R., Selfridge, J., McKenzie, A. N. J., Allen, J. E., Fagerholm, S. C., Maizels, R. M., Ivens, A. C., Bird, A., MacDonald, A. S. (2015). A dominant role for the methyl-CpG-binding protein Mbd2 in controlling Th2 induction by dendritic cells. Nature Communications, 6, Nature Publishing Group, p.1–11. [Online]. Available at: doi:10.1038/ncomms7920.

Cortes, C., Vapnik, V. (1995). Support-Vector Networks. Machine Learning, 20 (3), Kluwer Academic Publishers-Plenum Publishers, p.273–297. [Online]. Available at: doi:10.1023/A:1022627411411.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., Stein, L. (2011). Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Research, 39 (Database issue), Oxford University Press, p.D691–D697. [Online]. Available at: doi:10.1093/nar/gkq1018.

Cupp, E. W., Sauerbrey, M., Richards, F. (2011). Elimination of human onchocerciasis: History of progress and current feasibility using ivermectin (Mectizan®) monotherapy. Acta Tropica, 120, Elsevier B.V., p.S100–S108. [Online]. Available at: doi:10.1016/j.actatropica.2010.08.009.

D'Ambrosio, M. V., Bakalar, M., Bennuru, S., Reber, C., Skandarajah, A., Nilsson, L., Switz, N., Kamgno, J., Pion, S., Boussinesq, M., Nutman, T. B., Fletcher, D. A. (2015). Point-of-care quantification of blood-borne filarial parasites with a mobile phone microscope. Science translational medicine, 7 (286), American Association for the Advancement of Science, p.286re4–286re4. [Online]. Available at: doi:10.1126/scitranslmed.aaa3480.

Dalton, J. P., Mulcahy, G. (2001). Parasite vaccines--a reality? Veterinary Parasitology, 98 (1-3), p.149–167.

De Andres, B., Rakasz, E., Hagen, M., McCormik, M. L., Mueller, A. L., Elliot, D., Metwali, A., Sandor, M., Britigan, B. E., Weinstock, J. V., Lynch, R. G. (1997). Lack of Fc-epsilon receptors on murine eosinophils: implications for the functional significance of elevated IgE and eosinophils in parasitic infections. Blood, 89 (10), p.3826–3836.

Dérian, N., Bellier, B., Pham, H. P., Tsitoura, E., Kazazi, D., Huret, C., Mavromara, P., Klatzmann, D., Six, A. (2016). Early Transcriptome Signatures from Immunized Mouse Dendritic Cells Predict Late Vaccine-Induced T-Cell Responses. Altan-Bonnet, G. (ed.). PLoS Computational Biology, 12 (3), Public Library of Science, p.e1004801. [Online]. Available at: doi:10.1371/journal.pcbi.1004801.

Diawara, L., Traoré, M. O., Badji, A., Bissan, Y., Doumbia, K., Goita, S. F., Konaté, L., Mounkoro, K., Sarr, M. D., Seck, A. F., Toé, L., Tourée, S., Remme, J. H. F. (2009). Feasibility of Onchocerciasis Elimination with Ivermectin Treatment in Endemic Foci in Africa: First Evidence from Studies in Mali and Senegal. Basáñez, M.-G. (ed.). PLOS Neglected Tropical Diseases, 3 (7), p.e497–15. [Online]. Available at: doi:10.1371/journal.pntd.0000497.

Diemert, D. J., Pinto, A. G., Freire, J., Jariwala, A., Santiago, H., Hamilton, R. G., Periago, M. V., Loukas, A., Tribolet, L., Mulvenna, J., Correa Oliveira, R., Hotez, P. J., Bethony, J. M. (2012). Generalized urticaria induced by the *Na*-ASP-2 hookworm vaccine: implications for the development of vaccines against helminths. The Journal of allergy and clinical immunology, 130 (1), p.169–76.e6. [Online]. Available at: doi:10.1016/j.jaci.2012.04.027.

Dietterich, T. G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. Machine Learning, 40 (2), Kluwer Academic Publishers, p.139–157. [Online]. Available at: doi:10.1023/A:1007607513941.

Ding, L. H., Xie, Y., Park, S., Xiao, G., Story, M. D. (2008). Enhanced identification and biological validation of differential gene expression via Illumina whole-genome expression arrays through the use of the model-based background correction methodology. Nucleic Acids Research, 36 (10), p.e58–e58. [Online]. Available at: doi:10.1093/nar/gkn234.

Díaz-Uriarte, R., Alvarez de Andrés, S. (2006). Gene selection and classification of microarray data using random forest. BMC Bioinformatics, 7 (1), p.3. [Online]. Available at: doi:10.1186/1471-2105-7-3.

Doetze, A., Satoguina, J., G, B., Rau, T., Loliger, C., Fleischer, B., Hoerau, A. (2000). Antigen-specific cellular hyporesponsiveness in a chronic human helminth infcetion is mediated by Th3/Tr1- type cytokines IL-10 and transforming growth factor-B but not by a Th1 to Th2 shift. The Japanese Society of Immunlogy, 12 (5), p.623–630.

Donnelly, J. J., Wahren, B., Liu, M. A. (2005). DNA Vaccines: Progress and Challenges. The Journal of Immunology, 175 (2), p.633–639. [Online]. Available at: doi:10.4049/jimmunol.175.2.633.

Drame, P. M., Fink, D. L., Kamgno, J., Herrick, J. A., Nutman, T. B. (2014). Loop-Mediated Isothermal Amplification for Rapid and Semiquantitative Detection of

Loa loa Infection. Journal of Clinical Microbiology, 52 (6), p.2071–2077. [Online]. Available at: doi:10.1128/JCM.00525-14.

Drame, P. M., Meng, Z., Bennuru, S., Herrick, J. A., Veenstra, T. D., Nutman, T. B. (2016). Identification and Validation of *Loa loa* Microfilaria-Specific Biomarkers: a Rational Design Approach Using Proteomics and Novel Immunoassays. mBio, 7 (1), p.e02132–15–8. [Online]. Available at: doi:10.1128/mBio.02132-15.

Du, P., Kibbe, W. A., Lin, S. M. (2008). lumi: a pipeline for processing Illumina microarray. Bioinformatics, 24 (13), p.1547–1548. [Online]. Available at: doi:10.1093/bioinformatics/btn224.

Duan, K. B., Rajapakse, J. C., Wang, H., Azuaje, F. (2005). Multiple SVM-RFE for Gene Selection in Cancer Classification With Expression Data. 4 (3), IEEE, p.228–234. [Online]. Available at: doi:10.1109/TNB.2005.853657.

Duerr, H. P., Dietz, K., Schulz-Key, H., Buttner, D. W., Eichner, M. (2004). The relationships between the burden of adult parasites, host age and the microfilarial density in human onchocerciasis. International Journal for Parasitology, 34 (4), p.463–473. [Online]. Available at: doi:10.1016/j.ijpara.2003.11.008.

Duprez, J. (2013). Using immunomodulatory molecules as vaccine targets for filarial nematodes. Taylor, D. W., Babayan, S. A. MRes Thesis. University of Edinburgh. p.1–139.

Eberhard, M. L., Cupp, E. W., Katholi, C. R., Richards, F. O., Unnasch, T. R. (2017). Skin snips have no role in programmatic evaluations for onchocerciasis elimination: a reply to Bottomley et al. Parasites & Vectors, 10 (1), Parasites & Vectors, p.1–3. [Online]. Available at: doi:10.1186/s13071-017-2090-z.

Eberhard, M. L., Dickerson, J. W., Boyer, A. E., Tsang, V. C., Zea-Flores, R., Walker, E. M., Richards, F. O., Zea-Flores, G., Strobert, E. (1991). Experimental *Onchocerca volvulus* infections in mangabey monkeys (*Cercocebus atys*) compared to infections in humans and chimpanzees (*Pan troglodytes)*. American Journal of Tropical Medicine and Hygiene, 44 (2), p.151–160.

Elson, L. H., H, M. C., Y, W. P., N, E. A., Bradley, J. E., Guderian, R. H., Nutman, T. B. (2008). Immunityto Onchocerciasis: Putative Immune Persons Produce a Thl-like Response to *Onchocerca volvulus*. p.1–8.

Everts, B., Smits, H. H., Hokke, C. H., Yazdanbakhsh, M. (2010). Helminths and dendritic cells: Sensing and regulating via pattern recognition receptors, Th2 and Treg responses. European Journal of Immunology, 40 (6), p.1525–1537. [Online]. Available at: doi:10.1002/eji.200940109.

Feezor, R. J., Baker, H. V., Mindrinos, M., Hayden, D., Tannahill, C. L., Brownstein, B. H., Fay, A., MacMillan, S., Laramie, J., Xiao, W., Moldawer, L. L., Cobb, J.

P., Laudanski, K., Miller-Graziano, C. L., Maier, R. V., Schoenfeld, D., Davis, R. W., Tompkins, R. G. (2004). Whole blood and leukocyte RNA isolation for gene expression analyses. Physiological genomics, 19 (3), p.247–254. [Online]. Available at: doi:10.1152/physiolgenomics.00020.2004.

Filipe, J. A. N., Boussinesq, M., Renz, A., Collins, R. C., Vivas-Martinez, S., Grillet, M.-E., Little, M. P., Basáñez, M.-G. (2005). Human infection patterns and heterogeneous exposure in river blindness. Proceedings of the National Academy of Sciences, 102 (42), National Acad Sciences, p.15265–15270. [Online]. Available at: doi:10.1073/pnas.0502659102.

Fischer, P., Bamuhiiga, J., Buttner, D. W. (1997). Treatment of human Mansonella streptocerca infection with ivermectin. Tropical medicine & international health : TM & IH, 2 (2), p.191–199.

Francis, H., Awadzi, K., Ottesen, E. A. (1985). The Mazzotti reaction following treatment of onchocerciasis with diethylcarbamazine: clinical severity as a function of infection intensity. American Journal of Tropical Medicine and Hygiene, 34 (3), p.529–536.

Fraser, C. C., Altreuter, D. H., Ilyinskii, P., Pittet, L., LaMothe, R. A., Keegan, M., Johnston, L., Kishimoto, T. K. (2014). Generation of a universal CD4 memory T cell recall peptide effective in humans, mice and non-human primates. Vaccine, 32 (24), Elsevier Ltd, p.2896–2903. [Online]. Available at: doi:10.1016/j.vaccine.2014.02.024.

Fresnay, S., McArthur, M. A., Magder, L., Darton, T. C., Jones, C., Waddington, C. S., Blohmke, C. J., Angus, B., Levine, M. M., Pollard, A. J., Sztein, M. B. (2016). Salmonella Typhi-specific multifunctional CD8+ T cells play a dominant role in protection from typhoid fever in humans. Journal of Translational Medicine, 14 (1), BioMed Central, p.1–14. [Online]. Available at: doi:10.1186/s12967-016-0819-7.

Furman, D., Davis, M. M. (2015). New approaches to understanding the immune response to vaccination and infection. Vaccine, 33 (40), Elsevier Ltd, p.5271–5281. [Online]. Available at: doi:10.1016/j.vaccine.2015.06.117.

Gallin, M. Y., Tan, M., Kron, M. A., Rechnitzer, D., Greene, B. M., Newland, H. S., White, A. T., Taylor, H. R., Unnasch, T. R. (1989). *Onchocerca volvulus* recombinant antigen: physical characterization and clinical correlates with serum reactivity. The Journal of infectious diseases, 160 (3), p.521–529.

Gao, Y., Chen, L., Hou, M., Chen, Y., Ji, M., Wu, H., Wu, G. (2013). TLR2 Directing PD-L2 Expression Inhibit T Cells Response in *Schistosoma japonicum* Infection. Gobert, G. N. (ed.). PLoS ONE, 8 (12), p.e82480–12. [Online]. Available at: doi:10.1371/journal.pone.0082480.

Gardon, J., Gardon-Wendel, N., Demanga-Ngangue, Kamgno, J., Chippaux, J.-P.,

Boussinesq, M. (1997). Serious reactions after mass treatment of onchocerciasis with ivermectin in an area endemic for *Loa loa* infection. The Lancet, 350 (9070), p.18–22. [Online]. Available at: doi:10.1016/S0140-6736(96)11094-1.

Gaucher, D., Therrien, R., Kettaf, N., Angermann, B. R., Boucher, G., Filali-Mouhim, A., Moser, J. M., Mehta, R. S., Drake, D. R., Castro, E., Akondy, R., Rinfret, A., Yassine-Diab, B., Said, E. A., Chouikh, Y., Cameron, M. J., Clum, R., Kelvin, D., Somogyi, R., Greller, L. D., Balderas, R. S., Wilkinson, P., Pantaleo, G., Tartaglia, J., Haddad, E. K., Sekaly, R. P. (2008). Yellow fever vaccine induces integrated multilineage and polyfunctional immune responses. Journal of Experimental Medicine, 205 (13), p.3119–3131. [Online]. Available at: doi:10.1084/jem.20031598.

Genuer, R., Poggi, J.-M., Tuleau-Malot, C. (2010). Pattern Recognition Letters. 31 (14), Elsevier B.V., p.2225–2236. [Online]. Available at: doi:10.1016/j.patrec.2010.03.014.

Gerdts, V., Littel-van den Hurk, S. V. D., Griebel, P. J., Babiuk, L. A. (2007). Use of animal models in the development of human vaccines. Future Microbiology, 2 (6), p.667–675. [Online]. Available at: doi:10.2217/17460913.2.6.667.

Gilbert, J., Nfon, C. K., Makepeace, B. L., Njongmeta, L. M., Hastings, I. M., Pfarr, K. M., Renz, A., Tanya, V. N., Trees, A. J. (2005). Antibiotic chemotherapy of onchocerciasis: in a bovine model, killing of adult parasites requires a sustained depletion of endosymbiotic bacteria (*Wolbachia* species). The Journal of infectious diseases, 192 (8), p.1483–1493. [Online]. Available at: doi:10.1086/462426.

Gim, J., Cho, Y. B., Hong, H. K., Kim, H. C., Yun, S. H., Wu, H.-G., Jeong, S.-Y., Joung, J.-G., Park, T., Park, W.-Y., Lee, W. Y. (2016). Predicting multi-class responses to preoperative chemoradiotherapy in rectal cancer patients. Radiation Oncology, 11 (1), Radiation Oncology, p.1–8. [Online]. Available at: doi:10.1186/s13014-016-0623-9.

Godot, V., Harraga, S., Podoprigora, G., Liance, M., Bardonnet, K., Vuitton, D. A. (2003). IFNα-2a protects mice against a helminth infection of the liver and modulates immune responses. Gastroenterology, 124 (5), p.1441–1450. [Online]. Available at: doi:10.1016/S0016-5085(03)00273-7.

Golden, A., Stevens, E. J., Yokobe, L., Faulx, D., Kalnoky, M., Peck, R., Valdez, M., Steel, C., Karabou, P., Banla, M., Soboslay, P. T., Adade, K., Tekle, A. H., Cama, V. A., Fischer, P. U., Nutman, T. B., Unnasch, T. R., de los Santos, T., Domingo, G. J. (2016). A Recombinant Positive Control for Serology Diagnostic Tests Supporting Elimination of *Onchocerca volvulus*. Mackenzie, C. D. (ed.). PLOS Neglected Tropical Diseases, 10 (1), p.e0004292. [Online]. Available at: doi:10.1371/journal.pntd.0004292.

Gomez-Escobar, N., Gregory, W. F., Maizels, R. M. (2000). Identification of tgh-2, a

filarial nematode homolog of *Caenorhabditis elegans* daf-7 and human transforming growth factor beta, expressed in microfilarial and adult stages of *Brugia malayi*. Infection and Immunity, 68 (11), p.6402–6410.

Gomez-Escobar, N., Lewis, E., Maizels, R. M. (1998). A novel member of the transforming growth factor-beta (TGF-beta) superfamily from the filarial nematodes *Brugia malayi* and *B. pahangi*. Experimental Parasitology, 88 (3), p.200–209. [Online]. Available at: doi:10.1006/expr.1998.4248.

Gray, C. A., Lawrence, R. A. (2002). A role for antibody and Fc receptor in the clearance of Brugia malayi microfilariae. European Journal of Immunology, 32, p.1114–1120.

Gray, G. E., Allen, M., Moodie, Z., Churchyard, G., Bekker, L.-G., Nchabeleng, M., Mlisana, K., Metch, B., de Bruyn, G., Latka, M. H., Roux, S., Mathebula, M., Naicker, N., Ducar, C., Carter, D. K., Puren, A., Eaton, N., McElrath, M. J., Robertson, M., Corey, L., Kublin, J. G. (2011). Safety and efficacy of the HVTN 503/Phambili Study of a clade-B-based HIV-1 vaccine in South Africa: a double-blind, randomised, placebo-controlled test-of-concept phase 2b study. The Lancet Infectious Diseases, 11 (7), p.507–515. [Online]. Available at: doi:10.1016/S1473-3099(11)70098-6.

Greene, B. M., Taylor, H. R., Cupp, E. W., Murphy, R. P., White, A. T., Aziz, M. A., Schulz-Key, H., D'Anna, S. A., Newland, H. S., Goldschmidt, L. P. (1985). Comparison of ivermectin and diethylcarbamazine in the treatment of onchocerciasis. The New England journal of medicine, 313 (3), p.133–138. [Online]. Available at: doi:10.1056/NEJM198507183130301.

Gregorutti, B., Michel, B., Saint-Pierre, P. (2013). Correlation and variable importance in random forests. arXiv.org. [Online]. Available at: doi:10.1007/s11222-016-9646-1.

Gregory, W. F., Maizels, R. M. (2008). Cystatins from filarial parasites: Evolution, adaptation and function in the host–parasite relationship. The International Journal of Biochemistry & Cell Biology, 40 (6-7), p.1389–1398. [Online]. Available at: doi:10.1016/j.biocel.2007.11.012.

Grote, A., Lustigman, S., Ghedin, E. (2017). Lessons from the genomes and transcriptomes of filarial nematodes. Molecular and Biochemical Parasitology, Elsevier B.V., p.1–7. [Online]. Available at: doi:10.1016/j.molbiopara.2017.01.004.

Guigas, B. (2014). Priming dendritic cells forTh2 polarization: lessons learned from helminths and implications for metabolic disorders. p.1–9. [Online]. Available at: doi:10.3389/fimmu.2014.00499/abstract.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46 (1-3),

Kluwer Academic Publishers, p.389–422. [Online]. Available at: doi:10.1023/A:1012487302797.

Gyapong, J. O., Kumaraswami, V., Biswas, G., Ottesen, E. A. (2005). Treatment strategies underpinning the global programme to eliminate lymphatic filariasis. Expert opinion on pharmacotherapy, 6 (2), p.179–200. [Online]. Available at: doi:10.1517/14656566.6.2.179.

Haben, I., Hartmann, W., Specht, S., Hoerauf, A., Roers, A., Muller, W., Breloer, M. (2013). T-cell-derived, but not B-cell-derived, IL-10 suppresses antigen-specific T-cell responses in Litomosoides sigmodontis-infected mice. European Journal of Immunology, 43 (7), p.1799–1805. [Online]. Available at: doi:10.1002/eji.201242929.

Hagan, T., Nakaya, H. I., Subramaniam, S., Pulendran, B. (2015). Systems vaccinology: Enabling rational vaccine design with systems biological approaches. Vaccine, 33 (40), Elsevier Ltd, p.5294–5301. [Online]. Available at: doi:10.1016/j.vaccine.2015.03.072.

Hansen, R. D. E., Trees, A. J., Bah, G. S., Hetzel, U., Martin, C., Bain, O., Tanya, V. N., Makepeace, B. L. (2011). A worm's best friend: recruitment of neutrophils by *Wolbachia* confounds eosinophil degranulation against the filarial nematode *Onchocerca ochengi*. Proceedings. Biological sciences, 278 (1716), The Royal Society, p.2293–2302. [Online]. Available at: doi:10.1098/rspb.2010.2367.

Harnett, W. (2014). Article in press. Molecular and Biochemical Parasitology, Elsevier B.V., p.1–7. [Online]. Available at: doi:10.1016/j.molbiopara.2014.03.007.

Harrison, R. A., Wu, Y., Egerton, G., Bianco, A. E. (2000). DNA immunisation with *Onchocerca volvulus* chitinase induces partial protection against challenge infection with L3 larvae in mice. Vaccine, p.647–655.

Hartmann, W., Schramm, C., Breloer, M. (2015). *Litomosoides sigmodontis* induces TGF-β receptor responsive, IL-10-producing T cells that suppress bystander T-cell proliferation in mice. European Journal of Immunology, 45 (9), p.2568–2581. [Online]. Available at: doi:10.1002/eji.201545503.

Haury, A.-C., Gestraud, P., Vert, J.-P. (2011). The Influence of Feature Selection Methods on Accuracy, Stability and Interpretability of Molecular Signatures. Teh, M.-T. (ed.). PLoS ONE, 6 (12), p.e28210–e28212. [Online]. Available at: doi:10.1371/journal.pone.0028210.

Hayashi, Y., Noda, K., Shirasaka, A., Nogami, S., Nakamura, M. (1984). Vaccination of BALB/c mice against *Brugia malayi* and *B. pahangi* with larvae attenuated by gamma irradiation. The Japanese journal of experimental medicine, 54 (4), p.177–181.

Heidema, A. G., Boer, J. M. A., Nagelkerke, N., Mariman, E. C. M., van der A, D. L.,

Feskens, E. J. M. (2006). The challenge for genetic epidemiologists: how to analyze large numbers of SNPs in relation to complex diseases. BMC Genetics, 7, p.23. [Online]. Available at: doi:10.1186/1471-2156-7-23.

Hemphill, E., Lindsay, J., Lee, C., Măndoiu, I. I., Nelson, C. E. (2014). Feature selection and classifier performance on diverse biological datasets. BMC Bioinformatics, 15 (Suppl 13), BioMed Central Ltd, p.S4. [Online]. Available at: doi:10.1186/1471-2105-15-S13-S4.

Hess, J. A., Bin Zhan, Bonne-Année, S., Deckman, J. M., Bottazzi, M. E., Hotez, P. J., Klei, T. R., Lustigman, S., Abraham, D. (2014). International Journal for Parasitology. International Journal for Parasitology, 44 (9), Australian Society for Parasitology Inc., p.637–646. [Online]. Available at: doi:10.1016/j.ijpara.2014.04.006.

Hess, J. A., Zhan, B., Torigian, A. R., Patton, J. B., Petrovsky, N., Zhan, T., Bottazzi, M. E., Hotez, P. J., Klei, T. R., Lustigman, S., Abraham, D. (2016). The Immunomodulatory Role of Adjuvants in Vaccines Formulated with the Recombinant Antigens *Ov*-103 and *Ov*-RAL-2 against *Onchocerca volvulus* in Mice. McSorley, H. J. (ed.). PLOS Neglected Tropical Diseases, 10 (7), p.e0004797–23. [Online]. Available at: doi:10.1371/journal.pntd.0004797.

Hewitson, J. P., Harcus, Y. M., Curwen, R. S., Dowle, A. A., Atmadja, A. K., Ashton, P. D., Wilson, A., Maizels, R. M. (2008). The secretome of the filarial parasite, *Brugia malayi*: Proteomic profile of adult excretory–secretory products. Molecular and Biochemical Parasitology, 160 (1), p.8–21. [Online]. Available at: doi:10.1016/j.molbiopara.2008.02.007.

Hewitson, J. P., Maizels, R. M. (2014). Vaccination against helminth parasite infections. Expert Review of Vaccines, 13 (4), p.473–487. [Online]. Available at: doi:10.1586/14760584.2014.893195.

Heyer, L. J., Kruglyak, S., Yooseph, S. (1999). Exploring expression data: identification and analysis of coexpressed genes. Genome Research, 9 (11), p.1106–1115.

Hise, A. G., Gillette-Ferguson, I., Pearlman, E. (2003). Immunopathogenesis of *Onchocerca volvulus* keratitis (river blindness): a novel role for TLR4 and endosymbiotic *Wolbachia* bacteria. Journal of Endotoxin Research, 9 (6), p.390–394. [Online]. Available at: doi:10.1179/096805103225002746.

Hoerauf, A. (2008). Filariasis: new drugs and new opportunities for lymphatic filariasis and onchocerciasis. Current Opinion in Infectious Diseases, 21 (6), p.673–681. [Online]. Available at: doi:10.1097/QCO.0b013e328315cde7.

Hoerauf, A., Brattig, N. (2002). Resistance and susceptibility in human onchocerciasis – beyond Th1 vs Th2. Trends in Parasitology, 18, p.25–31.

Hoerauf, A., Kruse, S., Brattig, N. W., Heinzmann, A., Mueller-Myhsok, B., Deichmann, K. A. (2002). The variant Arg110Gln of human IL-13 is associated with an immunologically hyper-reactive form of onchocerciasis (sowda). Microbes and Infection, 4 (1), p.37–42.

Hoerauf, A., Satoguina, J., Saeftel, M., Specht, S. (2005). Immunomodulation by filarial nematodes. Parasite Immunology, 27 (10-11), p.417–429. [Online]. Available at: doi:10.1111/j.1365-3024.2005.00792.x.

Hoerauf, A., Volkmann, L., Hamelmann, C., Adjei, O., Autenrieth, I. B., Fleischer, B., Buttner, D. W. (2000). Endosymbiotic bacteria in worms as targets for a novel chemotherapy in filariasis. The Lancet, 355 (9211), p.1242–1243. [Online]. Available at: doi:10.1016/S0140-6736(00)02095-X.

Hoffmann, W., Pfaff, A. W., Schulz-Key, H., Soboslay, P. T. (2001). Determinants for resistance and susceptibility to microfilaraemia in *Litomosoides sigmodontis* filariasis. p.1–9.

Honglin. (2011). Experimental DNA vaccine against filariasis. Taylor, D. W. (ed.). p.1–252.

Hotez, P. J., Diemert, D., Bacon, K. M., Beaumier, C., Bethony, J. M., Bottazzi, M. E., Brooker, S., Couto, A. R., Freire, M. D. S., Homma, A., Lee, B. Y., Loukas, A., Loblack, M., Morel, C. M., Oliveira, R. C., Russell, P. K. (2013). The Human Hookworm Vaccine. Vaccine, 31 Suppl 2, p.B227–B232. [Online]. Available at: doi:10.1016/j.vaccine.2012.11.034.

Hotez, P. J., Kamath, A. (2009). Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden. Cappello, M. (ed.). PLOS Neglected Tropical Diseases, 3 (8), p.e412. [Online]. Available at: doi:10.1371/journal.pntd.0000412.

Houweling, T. A. J., Karim-Kos, H. E., Kulik, M. C., Stolk, W. A., Haagsma, J. A., Lenk, E. J., Richardus, J. H., de Vlas, S. J. (2016). Socioeconomic Inequalities in Neglected Tropical Diseases: A Systematic Review. Knopp, S. (ed.). PLOS Neglected Tropical Diseases, 10 (5), Public Library of Science, p.e0004546–28. [Online]. Available at: doi:10.1371/journal.pntd.0004546.

Huynh-Thu, V. A., Saeys, Y., Wehenkel, L., Geurts, P. (2012). Statistical interpretation of machine learning-based feature importance scores for biomarker discovery. Bioinformatics, 28 (13), p.1766–1774. [Online]. Available at: doi:10.1093/bioinformatics/bts238.

Hübner, M. P., Torrero, M. N., Mitre, E. (2010). Type 2 immune-inducing helminth vaccination maintains protective efficacy in the setting of repeated parasite exposures. Vaccine, 28 (7), p.1746–1757. [Online]. Available at: doi:10.1016/j.vaccine.2009.12.016.

Ihaka, R., Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics, 5 (3), p.299–314. [Online]. Available at: doi:10.1080/10618600.1996.10474713.

Jackson-Jones, L. H., Duncan, S. M., Magalhaes, M. E. N. S., Campbell, S. M., Maizels, R. M., McSorley, H. J., Allen, J. E., zech, C. E. C. B. E. N. E. (2016). Fat-associated lymphoid clusters control local IgM secretion during pleural infection and lung inflammation. Nature Communications, 7, Nature Publishing Group, p.1–14. [Online]. Available at: doi:10.1038/ncomms12651.

Jenkins, S. J., Perona-Wright, G., Worsley, A. G. F., Ishii, N., MacDonald, A. S. (2007). Dendritic Cell Expression of OX40 Ligand Acts as a Costimulatory, Not Polarizing, Signal for Optimal Th2 Priming and Memory Induction In Vivo. The Journal of Immunology, 179 (6), p.3515–3523. [Online]. Available at: doi:10.4049/jimmunol.179.6.3515.

Jeong, R., Soni, A. (2015). A comprehensive analysis of classification algorithms for cancer prediction from gene expression. In: 2015, New York, New York, USA: ACM Press, p.525–526. [Online]. Available at: doi:10.1145/2808719.2811443.

Johnson, E. H., Lustigman, S., Kass, P. H., Irvine, M., Browne, J., Prince, A. M. (1995). *Onchocerca volvulus*: a comparative study of in vitro neutrophil killing of microfilariae and humoral responses in infected and endemic normals. Experimental Parasitology, 81 (1), p.9–19. [Online]. Available at: doi:10.1006/expr.1995.1087.

Johnson, E. H., Schynder-candrian, S., Rajan, T. V., Nelson, F. K., Lustigman, S., Abraham, D. (1998). Immune responses to third stage larvae of *Onchocerca volvulus* in interferon-gamma and interleukin-4 knockout mice. Parasite Immunology, 20 (7), Blackwell Publishing Ltd, p.319–324. [Online]. Available at: doi:10.1046/j.1365-3024.1998.00148.x.

Joseph, S. K., Sambanthamoorthy, S., Dakshinamoorthy, G., Munirathinam, G., Ramaswamy, K. (2012). Protective immune responses to biolistic DNA vaccination of *Brugia malayi* abundant larval transcript-2. Vaccine, 30 (45), Elsevier Ltd, p.6477–6482. [Online]. Available at: doi:10.1016/j.vaccine.2012.07.084.

Kamga, G.-R., Dissak-Delon, F. N., Nana-Djeunga, H. C., Biholong, B. D., Mbigha-Ghogomu, S., Souopgui, J., Zouré, H. G. M., Boussinesq, M., Kamgno, J., Robert, A. (2016). Still mesoendemic onchocerciasis in two Cameroonian community-directed treatment with ivermectin projects despite more than 15 years of mass treatment. 3rd ed. Parasites & Vectors, 9 (1), Parasites & Vectors, p.1–12. [Online]. Available at: doi:10.1186/s13071-016-1868-8.

Karadjian, G., Fercoq, F., Pionnier, N., Vallarino-Lhermitte, N., Lefoulon, E., Nieguitsila, A., Specht, S., Carlin, L. M., Martin, C. (2017). Migratory phase of *Litomosoides sigmodontis* filarial infective larvae is associated with pathology

and transient increase of S100A9 expressing neutrophils in the lung. Brehm, K. (ed.). PLOS Neglected Tropical Diseases, 11 (5), p.e0005596–25. [Online]. Available at: doi:10.1371/journal.pntd.0005596.

Karimpour-Fard, A., Epperson, L. E., Hunter, L. E. (2015). A survey of computational tools for downstream analysis of proteomic and other omic datasets. Human Genomics, Human Genomics, p.1–11. [Online]. Available at: doi:10.1186/s40246-015-0050-2.

Katabarwa, M. N., Lakwo, T., Habomugisha, P., Agunyo, S., Byamukama, E., Oguttu, D., Tukesiga, E., Unoba, D., Dramuke, P., Onapa, A., Tukahebwa, E. M., Lwamafa, D., Walsh, F., Unnasch, T. R. (2013). Transmission of *Onchocerca volvulus* continues in Nyagak-Bondo focus of northwestern Uganda after 18 years of a single dose of annual treatment with ivermectin. The American journal of tropical medicine and hygiene, 89 (2), p.293–300. [Online]. Available at: doi:10.4269/ajtmh.13-0037.

Katawa, G., Layland, L. E., Debrah, A. Y., Horn, von, C., Batsa, L., Kwarteng, A., Arriens, S., W Taylor, D., Specht, S., Hoerauf, A., Adjobimey, T. (2015). Hyperreactive Onchocerciasis is Characterized by a Combination of Th17-Th2 Immune Responses and Reduced Regulatory T Cells. Makepeace, B. L. (ed.). PLOS Neglected Tropical Diseases, 9 (1), p.e3414–11. [Online]. Available at: doi:10.1371/journal.pntd.0003414.

Kaufmann, S. H. E., McElrath, M. J., Lewis, D. J. M., Del Giudice, G. (2014). Challenges and responses in human vaccine development. Current Opinion in Immunology, 28, p.18–26. [Online]. Available at: doi:10.1016/j.coi.2014.01.009.

Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG (2010) Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. PLOS Biology 8(6): e1000412. [Online]. Available at: doi.org/10.1371/journal.pbio.1000412

Kim, S.-Y. (2009). Effects of sample size on robustness and prediction accuracy of a prognostic gene signature. BMC Bioinformatics, 10 (1), p.147–10. [Online]. Available at: doi:10.1186/1471-2105-10-147.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. 14 (2), p.1137–1145.

Kolaskar, A. S., Tongaonkar, P. C. (1990). A semi-empirical method for prediction of antigenic determinants on protein antigens. FEBS Letters, 276 (1-2), p.172–174.

Korten, S., Badusche, M., Buttner, D. W., Hoerauf, A., Brattig, N., Fleischer, B. (2008). Natural death of adult *Onchocerca volvulus* and filaricidal effects of doxycycline induce local Foxp3+/CD4+ regulatory T cells and granzyme expression. Microbes and Infection, 10 (3), p.313–324. [Online]. Available at:

doi:10.1016/j.micinf.2007.12.004.

Korten, S., Buttner, D. W., Schmetz, C., Hoerauf, A., Mand, S., Brattig, N. (2009). The nematode parasite *Onchocerca volvulus* generates the transforming growth factor-beta (TGF-beta). Parasitology Research, 105 (3), p.731–741. [Online]. Available at: doi:10.1007/s00436-009-1450-9.

Korten, S., Hoerauf, A., Kaifi, J. T., Buttner, D. W. (2010). Low levels of transforming growth factor-beta (TGF-beta) and reduced suppression of Th2-mediated inflammation in hyperreactive human onchocerciasis. Parasite Immunology, 138 (01), p.35–45. [Online]. Available at: doi:10.1017/S0031182010000922.

Korten, S., Wildenburg, G., Darge, K., Buttner, D. W. (1998). Mast cells in onchocercomas from patients with hyperreactive onchocerciasis (sowda). Acta Tropica, 70 (2), p.217–231.

Kurath, G. (2008). Biotechnology and DNA vaccines for aquatic animals. Revue scientifique et technique (International Office of Epizootics), 27 (1), p.175–196.

Kursa, M. B. (2014). Robustness of Random Forest-based gene selection methods. BMC Bioinformatics, 15 (1), BioMed Central, p.8. [Online]. Available at: doi:10.1186/1471-2105-15-8.

Kwarteng, A., Ahuno, S. (2016). The Potentials and Pitfalls of Microarrays in Neglected Tropical Diseases: A Focus on Human Filarial Infections. Microarrays, 5 (3), p.20–13. [Online]. Available at: doi:10.3390/microarrays5030020.

Kwarteng, A., Ahuno, S. T. (2017). Immunity in filarial infections: lessons from animal models and human studies. Scandinavian journal of immunology, 85 (4), p.251–257. [Online]. Available at: doi:10.1111/sji.12533.

Lamberton, P. H. L., Cheke, R. A., Winskill, P., Tirados, I., Walker, M., Osei-Atweneboana, M. Y., Biritwum, N.-K., Tetteh-Kumah, A., Boakye, D. A., Wilson, M. D., Post, R. J., Basáñez, M.-G. (2015). Onchocerciasis Transmission in Ghana: Persistence under Different Control Strategies and the Role of the Simuliid Vectors. Brockhouse, C. (ed.). 9 (4), p.e0003688. [Online]. Available at: doi:10.1371/journal.pntd.0003688.s002.

Lange, A. M., Yutanawiboonchai, W., Lok, J. B., Trpis, M., Abraham, D. (1993). Induction of protective immunity against larval *Onchocerca volvulus* in a mouse model. American Journal of Tropical Medicine and Hygiene, 49 (6), p.783–788.

Lange, A. M., Yutanawiboonchai, W., Scott, P., Abraham, D. (1994). IL-4- and IL-5-dependent protective immunity to *Onchocerca volvulus* infective larvae in BALB/cBYJ mice. Journal of immunology (Baltimore, Md. : 1950), 153 (1), p.205–211.

Langfelder, P., Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics, 9 (1), p.559–13. [Online]. Available at: doi:10.1186/1471-2105-9-559.

Lariviere, M., Vingtain, P., Aziz, M., Beauvais, B., Weimann, D., Derouin, F., Ginoux, J., Schulz-Key, H., Gaxotte, P., Basset, D. (1985). Double-blind study of ivermectin and diethylcarbamazine in African onchocerciasis patients with ocular involvement. The Lancet, 2 (8448), p.174–177.

Lawrence, R. A., Allen, J. E., Gray, C. A. (2000). Requirements for in vivo IFN-gamma induction by live microfilariae of the parasitic nematode, *Brugia malayi*. Parasite Immunology, 120 (Pt 6), p.631–640.

Le Goff, L., Lamb, T. J., Graham, A. L., Harcus, Y., Allen, J. E. (2002). IL-4 is required to prevent filarial nematode development in resistant but not susceptible strains of mice. International Journal for Parasitology, 32 (10), p.1277–1284.

Le Goff, L., Martin, C., Oswald, I. P., Vuong, P. N., Petit, G., Ungeheuer, M. N., Bain, O. (2000). Parasitology and immunology of mice vaccinated with irradiated Litomosoides sigmodontis larvae. Parasite Immunology, 120 (3), p.271–280.

Le Goff, L., Petit, G., TayloR, D. W., Hoffmann, W., Bain, O. (1997). Early reduction of the challenge recovery rate following immunization with irradiated infective larvae in a filaria mouse system. Tropical medicine and International Health, 2 (12), p.1170–1174.

Lechner, C. J., Gantin, R. G., Seeger, T., Sarnecka, A., Portillo, J., Schulz-Key, H., Karabou, P. K., Helling-Giese, G., Heuschkel, C., Banla, M., Soboslay, P. T. (2012). Chemokines and cytokines in patients with an occult *Onchocerca volvulus* infection. Microbes and Infection, 14 (5), Elsevier Masson SAS, p.438–446. [Online]. Available at: doi:10.1016/j.micinf.2011.12.002.

Lee, J. W., Lee, J. B., Park, M., Song, S. H. (2005). An extensive comparison of recent classification tools applied to microarray data. Computational Statistics and Data Analysis, 48 (4), p.869–885. [Online]. Available at: doi:10.1016/j.csda.2004.03.017.

Li, B.-W., Rush, A., Zhang, S. R., Curtis, K. C., Weil, G. J. (2004). Antibody responses to *Brugia malayi* antigens induced by DNA vaccination. Filaria Journal, 3 (1), p.1. [Online]. Available at: doi:10.1186/1475-2883-3-1.

Li, S., Nakaya, H. I., Kazmin, D. A., Oh, J. Z., Pulendran, B. (2013a). Systems biological approaches to measure and understand vaccine immunity in humans. 25 (3), Elsevier Ltd, p.209–218. [Online]. Available at: doi:10.1016/j.smim.2013.05.003.

Li, S., Rouphael, N., Duraisingham, S., Romero-Steiner, S., Presnell, S., Davis, C., Schmidt, D. S., Johnson, S. E., Milton, A., Rajam, G., Kasturi, S., Carlone, G.

M., Quinn, C., Chaussabel, D., Palucka, A. K., Mulligan, M. J., Ahmed, R., Stephens, D. S., Nakaya, H. I., Pulendran, B. (2013b). Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. Nature immunology, 15 (2), p.195–204. [Online]. Available at: doi:10.1038/ni.2789.

Li, W., Joshi, M., Singhania, S., Ramsey, K., Murthy, A. (2014). Peptide Vaccine: Progress and Challenges. Vaccines, 2 (3), p.515–536. [Online]. Available at: doi:10.3390/vaccines2030515.

Libbrecht, M. W., Noble, W. S. (2015). Machine learning applications in genetics and genomics. Nature Publishing Group, 16 (6), Nature Publishing Group, p.321–332. [Online]. Available at: doi:10.1038/nrg3920.

Lightowlers, M. W., Donadeu, M., Elaiyaraja, M., Maithal, K., Kumar, K. A., Gauci, C. G., Firestone, S. M., Sarasola, P., Rowan, T. G. (2016). Anamnestic responses in pigs to the *Taenia solium* TSOL18 vaccine and implications for control strategies. Parasite Immunology, 143 (4), Cambridge University Press, p.416–420. [Online]. Available at: doi:10.1017/S0031182016000202.

Linhart, B., Narayanan, M., Focke-Tejkl, M., Wrba, F., Vrtala, S., Valenta, R. (2014). Prophylactic and therapeutic vaccination with carrier-bound Bet v 1 peptides lacking allergen-specific T cell epitopes reduces Bet v 1-specific T cell responses via blocking antibodies in a murine model for birch pollen allergy. Clinical & Experimental Allergy, 44 (2), p.278–287. [Online]. Available at: doi:10.1111/cea.12216.

Lipner, E. M., Dembele, N., Souleymane, S., Alley, W. S., Prevots, D. R., Toé, L., Boatin, B., Weil, G. J., Nutman, T. B. (2006). Field applicability of a rapid-format anti-*Ov*-16 antibody test for the assessment of onchocerciasis control measures in regions of endemicity. The Journal of infectious diseases, 194 (2), p.216–221. [Online]. Available at: doi:10.1086/505081.

Liu, Y. J. (2001a). Dendritic cell subsets and lineages, and their functions in innate and adaptive immunity. Cell, 106 (3), p.259–262.

Liu, Y. J. (2001b). Dendritic cell subsets and lineages, and their functions in innate and adaptive immunity. Cell, 106 (3), p.259–262. [Online]. Available at: doi:10.1016/S0092-8674(01)00456-1.

Lustigman, S., Brotman, B., Huima, T., Prince, A. M., McKerrow, J. H. (1992a). Molecular cloning and characterization of onchocystatin, a cysteine proteinase inhibitor of *Onchocerca volvulus*. The Journal of biological chemistry, 267 (24), ASBMB, p.17339–17346.

Lustigman, S., Brotman, B., Johnson, E. H., Smith, A. B., Huima, T., Prince, A. M. (1992b). Identification and characterization of an *Onchocerca volvulus* cDNA clone encoding a microfilarial surface-associated antigen. Molecular and

Biochemical Parasitology, 50 (1), p.79–93.

Lustigman, S., James, E. R., Tawe, W., Abraham, D. (2002). Towards a recombinant antigen vaccine against *Onchocerca volvulus*. Trends in Parasitology, 18 (3), p.135–141.

MacDonald, A. S., Maizels, R. M. (2008). Alarming dendritic cells for Th2 induction. Journal of Experimental Medicine, 205 (1), Rockefeller Univ Press, p.13–17. [Online]. Available at: doi:10.1084/jem.20072665.

MacDonald, A. S., Straw, A. D., Dalton, N. M., Pearce, E. J. (2002). Cutting edge: Th2 response induction by dendritic cells: a role for CD40. Journal of immunology (Baltimore, Md. : 1950), 168 (2), p.537–540.

Madhumathi, J., Prince, P. R., Anugraha, G., Kiran, P., Rao, D. N., Reddy, M. V. R., Kaliraj, P. (2010). Identification and characterization of nematode specific protective epitopes of *Brugia malayi* TRX towards development of synthetic vaccine construct for lymphatic filariasis. Vaccine, 28 (31), Elsevier Ltd, p.5038–5048. [Online]. Available at: doi:10.1016/j.vaccine.2010.05.012.

Maizels, R. M., Blaxter, M. L., Scott, A. L. (2001a). Immunological genomics of *Brugia malayi*: filarial genes implicated in immune evasion and protective immunity. Parasite Immunology, 23 (7), p.327–344. [Online]. Available at: doi:10.1046/j.1365-3024.2001.00397.x.

Maizels, R. M., Gomez-Escobar, N., Gregory, W. F., Murray, J., Zang, X. (2001b). Immune evasion genes from filarial nematodes. International Journal for Parasitology, 31, p.889–898.

Makepeace, B. L., Gill, A. C. (2016). *Wolbachia*. In: Rickettsiales, Rickettsiales: Biology, Molecular Biology, Epidemiology, and Vaccine Development, (Chapter 21), Cham: Springer International Publishing, p.465–512. [Online]. Available at: doi:10.1007/978-3-319-46859-4_21.

Makepeace, B. L., Jensen, S. A., Laney, S. J., Nfon, C. K., Njongmeta, L. M., Tanya, V. N., Williams, S. A., Bianco, A. E., Trees, A. J. (2009). Immunisation with a Multivalent, Subunit Vaccine Reduces Patent Infection in a Natural Bovine Model of Onchocerciasis during Intense Field Exposure. Oliveira, S. C. (ed.). PLOS Neglected Tropical Diseases, 3 (11), Public Library of Science, p.e544. [Online]. Available at: doi:10.1371/journal.pntd.0000544.

Makepeace, B. L., Tanya, V. N. (2016). 25 Years of the *Onchocerca ochengi* Model. Trends in Parasitology, 32 (12), Elsevier Ltd, p.966–978. [Online]. Available at: doi:10.1016/j.pt.2016.08.013.

Manchang, T. K., Ajonina-Ekoti, I., Ndjonka, D., Eisenbarth, A., Achukwi, M. D., Renz, A., Brattig, N. W., Liebau, E., Breloer, M. (2014). Immune recognition of *Onchocerca volvulus* proteins in the human host and animal models of

onchocerciasis. Journal of Helminthology, p.1–12. [Online]. Available at: doi:10.1017/S0022149X14000224.

Manoury, B., Gregory, W. F., Maizels, R. M., Watts, C. (2001). Bm-CPI-2, a cystatin homolog secreted by the filarial parasite *Brugia malayi*, inhibits class II MHC-restricted antigen processing. Current Biology, 11 (6), p.447–451.

Marechal, P., Le Goff, L., Petit, G., Diagne, M., TayloR, D. W., Bain, O. (1996). The fate of the filaria *Litomosoides sigmodontis* in susceptible and naturally resistant mice. Parasite, 3 (1), p.25–31. [Online]. Available at: doi:10.1051/parasite/1996031025.

Marino, S., Gideon, H. P., Gong, C., Mankad, S., McCrone, J. T., Lin, P. L., Linderman, J. J., Flynn, J. L., Kirschner, D. E. (2016). Computational and Empirical Studies Predict Mycobacterium tuberculosis-Specific T Cells as a Biomarker for Infection Outcome. Beauchemin, C. A. A. (ed.). PLoS Computational Biology, 12 (4), p.e1004804–e1004830. [Online]. Available at: doi:10.1371/journal.pcbi.1004804.

Martin, C., Al-Qaoud, K. M., Ungeheuer, M. N., Paehle, K., Vuong, P. N., Bain, O., Fleischer, B., Hoerau, A. (2000a). IL-5 is essential for vaccine-induced protection and for resolution of primary infection in murine filariasis. Medical Microbiology and Immunology, 189, p.67–74.

Martin, C., Le Goff, L., Ungeheuer, M.-N., Vuong, P. N., Bain, A. O. (2000b). Drastic Reduction of a Filarial Infection in Eosinophilic Interleukin-5 Transgenic Mice. Infection and Immunity, 68, p.3651–3656.

Martin, C., Saeftel, M., Vuong, P. N., Babayan, S., Fischer, K., Bain, O., Hoerauf, A. (2001). B-Cell Deficiency Suppresses Vaccine-Induced Protection against Murine Filariasis but Does Not Increase the Recovery Rate for Primary Infection. Infection and Immunity, 69 (11), p.7067–7073. [Online]. Available at: doi:10.1128/IAI.69.11.7067-7073.2001.

Mastelic, B., Lewis, D. J. M., Golding, H., Gust, I., Sheets, R., Lambert, P.-H. (2013). Potential use of inflammation and early immunological event biomarkers in assessing vaccine safety. Biologicals, 41 (2), Elsevier Ltd, p.115–124. [Online]. Available at: doi:10.1016/j.biologicals.2012.10.005.

McInnes, L. (2015). lmcinnes/hdbscan. 0 ed. GitHub. [Online]. Available at: doi:https://github.com/lmcinnes/hdbscan.git.

McKinney, B. A., Reif, D. M., Ritchie, M. D., Moore, D. J. H. (2006). Machine Learning for Detecting Gene-Gene Interactions. Applied bioinformatics, 5 (2), Springer International Publishing, p.77–88. [Online]. Available at: doi:10.2165/00822942-200605020-00002.

McSorley, H. J., Grainger, J. R., Harcus, Y., Murray, J., Nisbet, A. J., Knox, D. P.,

Maizels, R. M. (2009). daf-7-related TGF-β homologues from Trichostrongyloid nematodes show contrasting life-cycle expression patterns. Parasite Immunology, 137 (01), p.159. [Online]. Available at: doi:10.1017/S0031182009990321.

Medeiros, F., Baldwin, C. I., Denham, D. A. (1996). Brugia pahangi in cats: the passive transfer of anti-microfilarial immunity from immune to non-immune cats. Parasite Immunology, 18 (2), p.79–86.

Medoff, B. D., Seung, E., Hong, S., Thomas, S. Y., Sandall, B. P., Duffield, J. S., Kuperman, D. A., Erle, D. J., Luster, A. D. (2008). CD11b+ Myeloid Cells Are the Key Mediators of Th2 Cell Homing into the Airway in Allergic Inflammation. The Journal of Immunology, 182 (1), p.623–635. [Online]. Available at: doi:10.4049/jimmunol.182.1.623.

Mestas, J., Hughes, C. C. W. (2004). Of Mice and Not Men: Differences between Mouse and Human Immunology. Journal of immunology (Baltimore, Md. : 1950), 172 (5), p.2731–2738. [Online]. Available at: doi:10.4049/jimmunol.172.5.2731.

Metenou, S., Nutman, T. B. (2013). Regulatory T cell subsets in filarial infection and their function. Frontiers in immunology, 4, p.305. [Online]. Available at: doi:10.3389/fimmu.2013.00305.

Méndez-Samperio, P. (2016). Molecular events by which dendritic cells promote Th2 immune protection in helmith infection. Infectious Diseases, p.1–6. [Online]. Available at: doi:10.1080/23744235.2016.1194529.

Michiels, S., Koscielny, S., Hill, C. (2005). Prediction of cancer outcome with microarrays: a multiple random validation strategy. The Lancet, 365 (9458), p.488–492. [Online]. Available at: doi:10.1016/S0140-6736(05)17866-0.

Mina, E. (2016). Common disease signatures from gene expression analysis in Huntington's disease human blood and brain. Orphanet Journal of Rare Diseases, 11 (1), Orphanet Journal of Rare Diseases, p.1–13. [Online]. Available at: doi:10.1186/s13023-016-0475-2.

Mishra, R., Sahoo, P. K., Mishra, S., Achary, K. G., Dwibedi, B., Kar, S. K., Satapathy, A. K. (2014). Bancroftian filariasis: circulating B-1 cells decreased in microfilaria carriers and correlate with immunoglobulin M levels. Parasite Immunology, 36 (5), p.207–217. [Online]. Available at: doi:10.1111/pim.12105.

Mitchell, M. W. (2011). Bias of the Random Forest Out-of-Bag (OOB) Error for Certain Input Parameters. Open Journal of Statistics, 01 (03), p.205–211. [Online]. Available at: doi:10.4236/ojs.2011.13024.

Mittrücker, H.-W., Steinhoff, U., Köhler, A., Krause, M., Lazar, D., Mex, P., Miekley, D., Kaufmann, S. H. E. (2007). Poor correlation between BCG vaccination-

induced T cell responses and protection against tuberculosis. Proceedings of the National Academy of Sciences, 104 (30), National Acad Sciences, p.12434–12439. [Online]. Available at: doi:10.1073/pnas.0703510104.

Molla, M., Waddell, M., Page, D., Shavlik, J. (2004). Using Machine Learning to Design and Interpret Gene-Expression Microarrays. AI Magazine, 25 (1), p.23. [Online]. Available at: doi:10.1609/aimag.v25i1.1745.

Morales-Hojas, R., Cheke, R. A., Post, R. J. (2006). Molecular systematics of five Onchocerca species (Nematoda: *Filarioidea*) including the human parasite, *O. volvulus*, suggest sympatric speciation. Journal of Helminthology, 80 (3), p.281–290. [Online]. Available at: doi:10.1079/JOH2006331.

Moreno, Y., Geary, T. G. (2008). Stage- and Gender-Specific Proteomic Analysis of *Brugia malayi* Excretory-Secretory Products. Ghedin, E. (ed.). PLOS Neglected Tropical Diseases, 2 (10), p.e326–12. [Online]. Available at: doi:10.1371/journal.pntd.0000326.

Morris, C. P., Evans, H., Larsen, S. E., Mitre, E. (2013). A comprehensive, model-based review of vaccine and repeat infection trials for filariasis. Clinical Microbiology Reviews, 26 (3), American Society for Microbiology, p.381–421. [Online]. Available at: doi:10.1128/CMR.00002-13.

Motran, C. C., Ambrosio, L. F., Volpini, X., Celias, D. P., Cervi, L. (2016). Dendritic cells and parasites: from recognition and activation to immune response instruction. Seminars in Immunopathology, Seminars in Immunopathology, p.1–15. [Online]. Available at: doi:10.1007/s00281-016-0588-7.

Murdoch, M. E., Asuzu, M. C., Hagan, M., Makunde, W. H., Ngoumou, P., Ogbuagu, K. F., Okello, D., Ozoh, G., Remme, J. (2002). Onchocerciasis: the clinical and epidemiological burden of skin disease in Africa. Annals of Tropical Medicine and Parasitology, 96 (3), p.283–296. [Online]. Available at: doi:10.1179/000349802125000826.

Murdoch, M. E., Murdoch, I. E. (2016). Onchocerciasis. In: Intraocular Inflammation, Intraocular Inflammation, (Chapter 134), Berlin, Heidelberg: Springer Berlin Heidelberg, p.1349–1359. [Online]. Available at: doi:10.1007/978-3-540-75387-2_134.

Murdoch, M. E., Payton, A., Abiose, A., Thomson, W., Panicker, V. K., Dyer, P. A., Jones, B. R., Maizels, R. M., Ollier, W. E. (1997). HLA-DQ alleles associate with cutaneous features of onchocerciasis. The Kaduna-London-Manchester Collaboration for Research on Onchocerciasis. Human Immunology, 55 (1), p.46–52.

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., Aboyans, V., Abraham, J., Ackerman, I., Aggarwal, R., Ahn, S. Y., Ali, M. K., Alvarado, M., Anderson, H.

R., Anderson, L. M., Andrews, K. G., Atkinson, C., Baddour, L. M., Bahalim, A. N., Barker-Collo, S., Barrero, L. H., Bartels, D. H., Basáñez, M.-G., Baxter, A., Bell, M. L., Benjamin, E. J., Bennett, D., Bernabé, E., Bhalla, K., Bhandari, B., Bikbov, B., Bin Abdulhak, A., Birbeck, G., Black, J. A., Blencowe, H., Blore, J. D., Blyth, F., Bolliger, I., Bonaventure, A., Boufous, S., Bourne, R., Boussinesq, M., Braithwaite, T., Brayne, C., Bridgett, L., Brooker, S., Brooks, P., Brugha, T. S., Bryan-Hancock, C., Bucello, C., Buchbinder, R., Buckle, G., Budke, C. M., Burch, M., Burney, P., Burstein, R., Calabria, B., Campbell, B., Canter, C. E., Carabin, H., Carapetis, J., Carmona, L., Cella, C., Charlson, F., Chen, H., Cheng, A. T.-A., Chou, D., Chugh, S. S., Coffeng, L. E., Colan, S. D., Colquhoun, S., Colson, K. E., Condon, J., Connor, M. D., Cooper, L. T., Corriere, M., Cortinovis, M., de Vaccaro, K. C., Couser, W., Cowie, B. C., Criqui, M. H., Cross, M., Dabhadkar, K. C., Dahiya, M., Dahodwala, N., Damsere-Derry, J., Danaei, G., Davis, A., De Leo, D., Degenhardt, L., Dellavalle, R., Delossantos, A., Denenberg, J., Derrett, S., Jarlais, Des, D. C., Dharmaratne, S. D., Dherani, M., Diaz-Torne, C., Dolk, H., Dorsey, E. R., Driscoll, T., Duber, H., Ebel, B., Edmond, K., Elbaz, A., Ali, S. E., Erskine, H., Erwin, P. J., Espindola, P., Ewoigbokhan, S. E., Farzadfar, F., Feigin, V., Felson, D. T., Ferrari, A., Ferri, C. P., Fèvre, E. M., Finucane, M. M., Flaxman, S., Flood, L., Foreman, K., Forouzanfar, M. H., Fowkes, F. G. R., Fransen, M., Freeman, M. K., Gabbe, B. J., Gabriel, S. E., Gakidou, E., Ganatra, H. A., Garcia, B., Gaspari, F., Gillum, R. F., Gmel, G., Gonzalez-Medina, D., Gosselin, R., Grainger, R., Grant, B., Groeger, J., Guillemin, F., Gunnell, D., Gupta, R., Haagsma, J., Hagan, H., Halasa, Y. A., Hall, W., Haring, D., Haro, J. M., Harrison, J. E., Havmoeller, R., Hay, R. J., Higashi, H., Hill, C., Hoen, B., Hoffman, H., Hotez, P. J., Hoy, D., Huang, J. J., Ibeanusi, S. E., Jacobsen, K. H., James, S. L., Jarvis, D., Jasrasaria, R., Jayaraman, S., Johns, N., Jonas, J. B., Karthikeyan, G., Kassebaum, N., Kawakami, N., Keren, A., Khoo, J.-P., King, C. H., Knowlton, L. M., Kobusingye, O., Koranteng, A., Krishnamurthi, R., Laden, F., Lalloo, R., Laslett, L. L., Lathlean, T., Leasher, J. L., Lee, Y. Y., Leigh, J., Levinson, D., Lim, S. S., Limb, E., Lin, J. K., Lipnick, M., Lipshultz, S. E., Liu, W., Loane, M., Ohno, S. L., Lyons, R., Mabweijano, J., MacIntyre, M. F., Malekzadeh, R., Mallinger, L., Manivannan, S., Marcenes, W., March, L., Margolis, D. J., Marks, G. B., Marks, R., Matsumori, A., Matzopoulos, R., Mayosi, B. M., McAnulty, J. H., McDermott, M. M., McGill, N., McGrath, J., Medina-Mora, M. E., Meltzer, M., Mensah, G. A., Merriman, T. R., Meyer, A.-C., Miglioli, V., Miller, M., Miller, T. R., Mitchell, P. B., Mock, C., Mocumbi, A. O., Moffitt, T. E., Mokdad, A. A., Monasta, L., Montico, M., Moradi-Lakeh, M., Moran, A., Morawska, L., Mori, R., Murdoch, M. E., Mwaniki, M. K., Naidoo, K., Nair, M. N., Naldi, L., Narayan, K. M. V., Nelson, P. K., Nelson, R. G., Nevitt, M. C., Newton, C. R., Nolte, S., Norman, P., Norman, R., O'Donnell, M., O'Hanlon, S., Olives, C., Omer, S. B., Ortblad, K., Osborne, R., Ozgediz, D., Page, A., Pahari, B., Pandian, J. D., Rivero, A. P., Patten, S. B., Pearce, N., Padilla, R. P., Perez-Ruiz, F., Perico, N., Pesudovs, K., Phillips, D., Phillips, M. R., Pierce, K., Pion, S., Polanczyk, G. V., Polinder, S., Pope, C. A., Popova, S., Porrini, E., Pourmalek, F., Prince, M., Pullan, R. L., Ramaiah, K. D., Ranganathan, D., Razavi, H., Regan, M., Rehm, J. T., Rein, D. B., Remuzzi, G.,

Richardson, K., Rivara, F. P., Roberts, T., Robinson, C., De Leòn, F. R., Ronfani, L., Room, R., Rosenfeld, L. C., Rushton, L., Sacco, R. L., Saha, S., Sampson, U., Sanchez-Riera, L., Sanman, E., Schwebel, D. C., Scott, J. G., Segui-Gomez, M., Shahraz, S., Shepard, D. S., Shin, H., Shivakoti, R., Singh, D., Singh, G. M., Singh, J. A., Singleton, J., Sleet, D. A., Sliwa, K., Smith, E., Smith, J. L., Stapelberg, N. J. C., Steer, A., Steiner, T., Stolk, W. A., Stovner, L. J., Sudfeld, C., Syed, S., Tamburlini, G., Tavakkoli, M., Taylor, H. R., Taylor, J. A., Taylor, W. J., Thomas, B., Thomson, W. M., Thurston, G. D., Tleyjeh, I. M., Tonelli, M., Towbin, J. A., Truelsen, T., Tsilimbaris, M. K., Ubeda, C., Undurraga, E. A., van der Werf, M. J., van Os, J., Vavilala, M. S., Venketasubramanian, N., Wang, M., Wang, W., Watt, K., Weatherall, D. J., Weinstock, M. A., Weintraub, R., Weisskopf, M. G., Weissman, M. M., White, R. A., Whiteford, H., Wiebe, N., Wiersma, S. T., Wilkinson, J. D., Williams, H. C., Williams, S. R. M., Witt, E., Wolfe, F., Woolf, A. D., Wulf, S., Yeh, P.-H., Zaidi, A. K. M., Zheng, Z.-J., Zonies, D., Lopez, A. D., AlMazroa, M. A., Memish, Z. A. (2012). Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet (London, England), 380 (9859), p.2197–2223. [Online]. Available at: doi:10.1016/S0140-6736(12)61689-4.

Murray, J., Manoury, B., Balic, A., Watts, C., Maizels, R. M. (2005). *Bm*-CPI-2, a cystatin from *Brugia malayi* nematode parasites, differs from *Caenorhabditis elegans* cystatins in a specific site mediating inhibition of the antigen-processing enzyme AEP. Molecular and Biochemical Parasitology, 139 (2), p.197–203. [Online]. Available at: doi:10.1016/j.molbiopara.2004.11.008.

Na, H., Cho, M., Chung, Y. (2016). Regulation of Th2 Cell Immunity by Dendritic Cells. Immune Network, 16 (1), p.1–12. [Online]. Available at: doi:10.4110/in.2016.16.1.1.

Nakaya, H. I., Li, S., Pulendran, B. (2011a). Systems vaccinology: learning to compute the behavior of vaccine induced immunity. Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 4 (2), p.193–205. [Online]. Available at: doi:10.1002/wsbm.163.

Nakaya, H. I., Wrammert, J., Lee, E. K., Racioppi, L., Marie-Kunze, S., Haining, W. N., Means, A. R., Kasturi, S. P., Khan, N., Li, G.-M., McCausland, M., Kanchan, V., Kokko, K. E., Li, S., Elbein, R., Mehta, A. K., Aderem, A., Subbarao, K., Ahmed, R., Pulendran, B. (2011b). Systems biology of vaccination for seasonal influenza in humans. Nature immunology, 12 (8), p.786–795. [Online]. Available at: doi:10.1038/ni.2067.

Nfon, C. K., Makepeace, B. L., Njongmeta, L. M., Tanya, V. N., Bain, O., Trees, A. J. (2006). Eosinophils contribute to killing of adult *Onchocerca ochengi* within onchocercomata following elimination of *Wolbachia*. Microbes and Infection, 8 (12-13), p.2698–2705. [Online]. Available at: doi:10.1016/j.micinf.2006.07.017.

Nguyen, T.-T., Huang, J. Z., Nguyen, T. T. (2015). Unbiased Feature Selection in Learning Random Forests for High-Dimensional Data. The Scientific World Journal, 2015 (2), p.1–18. [Online]. Available at: doi:10.1162/jocn.1991.3.1.71.

Nisbet, A. J., McNeilly, T. N., Wildblood, L. A., Morrison, A. A., Bartley, D. J., Bartley, Y., Longhi, C., McKendrick, I. J., Palarea-Albaladejo, J., Matthews, J. B. (2013). Successful immunization against a parasitic nematode by vaccination with recombinant proteins. Vaccine, 31 (37), Elsevier Ltd, p.4017–4023. [Online]. Available at: doi:10.1016/j.vaccine.2013.05.026.

Nmorsi, O. P. G., Ukwandu, N. C. D., Alabi-Eric, O. J., Popoola, W., Osita-Emina, M. (2007). CD4+, CD8+, immunoglobulin status and ocular lesions among some onchocerciasis-infected rural Nigerians. Parasitology Research, 100 (6), p.1261–1266. [Online]. Available at: doi:10.1007/s00436-006-0421-7.

Ochiai, R. L., Khan, M. I., Soofi, S. B., Sur, D., Kanungo, S., You, Y. A., Habib, M. A., Sahito, S. M., Manna, B., Dutta, S., Acosta, C. J., Ali, M., Bhattacharya, S. K., Bhutta, Z. A., Clemens, J. D. (2014). Immune responses to Vi capsular polysaccharide typhoid vaccine in children 2 to 16 years old in Karachi, Pakistan, and Kolkata, India. Clinical and vaccine immunology : CVI, 21 (5), American Society for Microbiology, p.661–666. [Online]. Available at: doi:10.1128/CVI.00791-13.

Osei-Atweneboana, M. Y., Awadzi, K., Attah, S. K., Boakye, D. A., Gyapong, J. O., Prichard, R. K. (2011). Phenotypic Evidence of Emerging Ivermectin Resistance in *Onchocerca volvulus*. Lustigman, S. (ed.). 5 (3), p.e998. [Online]. Available at: doi:10.1371/journal.pntd.0000998.s001.

Otabil, K. B., Tenkorang, S. B. (2015). Filarial hydrocele: a neglected condition of a neglected tropical disease. The Journal of Infection in Developing Countries, 9 (05). [Online]. Available at: doi:10.3855/jidc.5346.

Ottesen, E. A., Hooper, P. J., Bradley, M., Biswas, G. (2008). The global programme to eliminate lymphatic filariasis: health impact after 8 years. de Silva, N. (ed.). PLOS Neglected Tropical Diseases, 2 (10), p.e317. [Online]. Available at: doi:10.1371/journal.pntd.0000317.

Pan, K.-H., Lih, C.-J., Cohen, S. N. (2005). Effects of threshold choice on biological conclusions reached during analysis of gene expression by DNA microarrays. Proceedings of the National Academy of Sciences, 102 (25), National Acad Sciences, p.8961–8965. [Online]. Available at: doi:10.1073/pnas.0502674102.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, MIT Press. [Online]. Available at: http://scikit-learn.org.

Petit, G., Diagne, M., Marechal, P., Owen, D., Taylor, D., Bain, O. (1992). Maturation of the filaria *Litomosoides sigmodontis* in BALB/c mice; comparative susceptibility of nine other inbred strains. Annales de parasitologie humaine et comparee, 67 (5), p.144–150.

Pfaff, A. W., Schulz-Key, H., Soboslay, P. T., Geiger, S. M., Hoffmann, W. H. (2000). Litomosoides sigmodontis: dynamics of the survival of microfilariae in resistant and susceptible strains of mice. Experimental Parasitology, 94 (2), p.67–74. [Online]. Available at: doi:10.1006/expr.1999.4475.

Pfaff, A. W., Schulz-Key, H., Soboslay, P. T., Taylor, D. W., MacLennan, K., Hoffmann, W. H. (2002). *Litomosoides sigmodontis* cystatin acts as an immunomodulator during experimental filariasis. International Journal for Parasitology, 32 (2), p.171–178.

Pion, S. D. S., Filipe, J. A. N., Kamgno, J., Gardon, J., Basáñez, M. G., Boussinesq, M. (2006). Microfilarial distribution of Loa loa in the human host: population dynamics and epidemiological implications. Parasite Immunology, 133 (Pt 1), p.101–109. [Online]. Available at: doi:10.1017/S0031182006000035.

Pion, S. D. S., Nana-Djeunga, H. C., Kamgno, J., Tendongfor, N., Wanji, S., Njiokou, F., Prichard, R. K., Boussinesq, M. (2013). Dynamics of *Onchocerca volvulus* Microfilarial Densities after Ivermectin Treatment in an Ivermectin-naïve and a Multiply Treated Population from Cameroon. Lammie, P. J. (ed.). 7 (2), p.e2084. [Online]. Available at: doi:10.1371/journal.pntd.0002084.s001.

Pionnier, N., Brotin, E., Karadjian, G., Hemon, P., Gaudin-Nomé, F., Vallarino-Lhermitte, N., Nieguitsila, A., Fercoq, F., Aknin, M.-L., Marin-Esteban, V., Chollet-Martin, S., Schlecht-Louf, G., Bachelerie, F., Martin, C. (2016). Neutropenic Mice Provide Insight into the Role of Skin-Infiltrating Neutrophils in the Host Protective Immunity against Filarial Infective Larvae. MacDonald, A. S. (ed.). PLOS Neglected Tropical Diseases, 10 (4), p.e0004605–e0004624. [Online]. Available at: doi:10.1371/journal.pntd.0004605.

Pirooznia, M., Yang, J. Y., Yang, M. Q., Deng, Y. (2008). A comparative study of different machine learning methods on microarray gene expression data. BMC Genomics, 9 (Suppl 1), p.S13–S13. [Online]. Available at: doi:10.1186/1471-2164-9-S1-S13.

Plotkin, S. A. (2010). Correlates of protection induced by vaccination. Clinical and vaccine immunology : CVI, 17 (7), American Society for Microbiology, p.1055–1065. [Online]. Available at: doi:10.1128/CVI.00131-10.

Pulendran, B., Li, S., Nakaya, H. I. (2010). Systems vaccinology. Immunity, 33 (4), p.516–529. [Online]. Available at: doi:10.1016/j.immuni.2010.10.006.

Querec, T. D., Akondy, R. S., Lee, E. K., Cao, W., Nakaya, H. I., Teuwen, D., Pirani, A., Gernert, K., Deng, J., Marzolf, B., Kennedy, K., Wu, H., Bennouna, S.,

Oluoch, H., Miller, J., Vencio, R. Z., Mulligan, M., Aderem, A., Ahmed, R., Pulendran, B. (2008). Systems biology approach predicts immunogenicity of the yellow fever vaccine in humans. Nature immunology, 10 (1), p.116–125. [Online]. Available at: doi:10.1038/ni.1688.

R Development Core Team. (2012). R: A language and environment for statistical computing. Vienna, Austria. [Online]. Available at: http://www.R-project.org/.

Ramaiah, K. D., Ottesen, E. A. (2014). Progress and Impact of 13 Years of the Global Programme to Eliminate Lymphatic Filariasis on Reducing the Burden of Filarial Disease. Bockarie, M. (ed.). PLOS Neglected Tropical Diseases, 8 (11), Public Library of Science, p.e3319–10. [Online]. Available at: doi:10.1371/journal.pntd.0003319.

Reif, D. M., Motsinger-Reif, A. A., McKinney, B. A., Rock, M. T., Crowe, J. E., Moore, J. H. (2008). Integrated analysis of genetic and proteomic data identifies biomarkers associated with adverse events following smallpox vaccination. Genes and Immunity, 10 (2), p.112–119. [Online]. Available at: doi:10.1038/gene.2008.80.

Renz, A., Enyong, P., Wahl, G. (1994). Cattle, worms and zooprophylaxis. Parasite, 1 (1S), p.S4–S6. [Online]. Available at: doi:10.1051/parasite/199401s1004.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research, 43 (7), Oxford University Press, p.e47–e47. [Online]. Available at: doi:10.1093/nar/gkv007.

Romanowski, B. (2014). Long term protection against cervical infection with the human papillomavirus: Review of currently available vaccines. 7 (2), p.161–169. [Online]. Available at: doi:10.4161/hv.7.2.13690.

Roncarolo, M. G., Gregori, S., Battaglia, M., Bacchetta, R., Fleischhauer, K., Levings, M. K. (2006). Interleukin-10-secreting type 1 regulatory T cells in rodents and humans. Immunological reviews, 212, p.28–50. [Online]. Available at: doi:10.1111/j.0105-2896.2006.00420.x.

Rubin, R. L., Tang, F. L., Lucas, A. H., Spiegelberg, H. L., Tan, E. M. (1986). IgG subclasses of anti-tetanus toxoid antibodies in adult and newborn normal subjects and in patients with systemic lupus erythematosus, Sjogren's syndrome, and drug-induced autoimmunity. The Journal of Immunology, 137 (8), p.2522–2527.

Rückerl, D., Allen, J. E. (2014). Macrophage proliferation, provenance, and plasticity in macroparasite infection. Immunological reviews, 262 (1), p.113–133. [Online]. Available at: doi:10.1111/imr.12221.

Saeftel, M., Arndt, M., Specht, S., Volkmann, L., Hoerauf, A. (2003). Synergism of

Gamma Interferon and Interleukin-5 in the Control of Murine Filariasis. Infection and Immunity, 71 (12), p.6978–6985. [Online]. Available at: doi:10.1128/IAI.71.12.6978-6985.2003.

Saeftel, M., Volkmann, L., Korten, S., Brattig, N., Al-Qaoud, K., Fleischer, B., Hoerauf, A. (2001). Lack of interferon-gamma confers impaired neutrophil granulocyte function and imparts prolonged survival of adult filarial worms in murine filariasis. Microbes and Infection, 3 (3), p.203–213.

Satoguina, J. S., Adjobimey, T., Arndts, K., Hoch, J., Oldenburg, J., Layland, L. E., Hoerauf, A. (2008). Tr1 and naturally occurring regulatory T cells induce IgG4 in B cells through GITR/GITR-L interaction, IL-10 and TGF-beta. European Journal of Immunology, 38 (11), p.3101–3113. [Online]. Available at: doi:10.1002/eji.200838193.

Satoguina, J. S., Weyand, E., Larbi, J., Hoerauf, A. (2005). T regulatory-1 cells induce IgG4 production by B cells: role of IL-10. Journal of immunology (Baltimore, Md. : 1950), 174 (8), p.4718–4726. [Online]. Available at: doi:10.4049/jimmunol.174.8.4718.

Satoguina, J., Mempel, M., Larbi, J., Badusche, M., Loliger, C., Adjei, O., Gachelin, G., Fleischer, B., Hoerau, A. (2002). Antigen-specific T regulatory-1 cells are associated with immunosuppression in a chronic helminth infection (onchocerciasis). Microbes and Infection, 4, p.1291–1300.

Schoolnik, G. K. (2002). Microarray analysis of bacterial pathogenicity. Advances in microbial physiology, 46, p.1–45.

Schönemeyer, A., Lucius, R., Sonnenburg, B., Brattig, N., Sabat, R., Schilling, K., Bradley, J., Hartmann, S. (2001). Modulation of human T cell responses and macrophage functions by onchocystatin, a secreted protein of the filarial nematode *Onchocerca volvulus*. Journal of immunology (Baltimore, Md. : 1950), 167 (6), p.3207–3215.

SEARO, W. (2010). The Regional Strategic Plan for Elimination of Lymphatic Filariasis - 2010-2015. p.1–27.

Segura, E., Villadangos, J. A. (2009). Antigen presentation by dendritic cells *in vivo*. Current Opinion in Immunology, 21 (1), p.105–110. [Online]. Available at: doi:10.1016/j.coi.2009.03.011.

Segura, M., Su, Z., Piccirillo, C., Stevenson, M. M. (2007). Impairment of dendritic cell function by excretory-secretory products: A potential mechanism for nematode-induced immunosuppression. European Journal of Immunology, 37 (7), p.1887–1904. [Online]. Available at: doi:10.1002/eji.200636553.

Seib, K. L., Zhao, X., Rappuoli, R. (2012). Developing vaccines in the era of genomics: a decade of reverse vaccinology. Clinical Microbiology and Infection, 18,

European Society of Clinical Microbiology and Infectious Diseases, p.109–116. [Online]. Available at: doi:10.1111/j.1469-0691.2012.03939.x.

Semnani, R. T., Liu, A. Y., Sabzevari, H., Kubofcik, J., Zhou, J., Gilden, J. K., Nutman, T. B. (2003). *Brugia malayi* Microfilariae Induce Cell Death in Human Dendritic Cells, Inhibit Their Ability to Make IL-12 and IL-10, and Reduce Their Capacity to Activate CD4+ T Cells. The Journal of Immunology, 171 (4), p.1950–1960. [Online]. Available at: doi:10.4049/jimmunol.171.4.1950.

Seok, J., Warren, H. S., Cuenca, A. G. (2013). Genomic responses in mouse models poorly mimic human inflammatory diseases. In: 2013. [Online]. Available at: doi:10.1073/pnas.1222878110/-/DCSupplemental/sapp.pdf.

Sharma, A., Sharma, P., Vishwakarma, A. L., Srivastava, M. (2016). Functional Impairment of Murine Dendritic Cell Subsets following Infection with Infective Larval Stage 3 of Brugia malayi. Adams, J. H. (ed.). Infection and Immunity, 85 (1), p.e00818–16–46. [Online]. Available at: doi:10.1128/IAI.00818-16.

Shedlock, D. J., Weiner, D. B. (2000). DNA vaccination: antigen presentation and the induction of immunity. Journal of leukocyte biology, 68 (6), p.793–806.

Shi, W., Oshlack, A., Smyth, G. K. (2010). Optimizing the noise versus bias trade-off for Illumina whole genome expression BeadChips. Nucleic Acids Research, 38 (22), p.e204–e204. [Online]. Available at: doi:10.1093/nar/gkq871.

Silva, S. R., Jacysyn, J. F., Macedo, M. S., Faquim-Mauro, E. L. (2006). Immunosuppressive components ofAscaris suumdown-regulate expression of costimulatory molecules and function of antigen-presenting cellsvia an IL-10-mediated mechanism. European Journal of Immunology, 36 (12), p.3227–3237. [Online]. Available at: doi:10.1002/eji.200636110.

Simons, J. E., Gray, C. A., Lawrence, R. A. (2010). Absence of regulatory IL-10 enhances innate protection against filarial parasites by a neutrophil-independent mechanism. Parasite Immunology, 32 (7), p.473–478. [Online]. Available at: doi:10.1111/j.1365-3024.2010.01210.x.

Simons, J. E., Rothenberg, M. E., Lawrence, R. A. (2005). Eotaxin-1-regulated eosinophils have a critical role in innate immunity against experimental *Brugia malayi* infection. European Journal of Immunology, 35 (1), p.189–197. [Online]. Available at: doi:10.1002/eji.200425541.

Simonsen, P. E., Onapa, A. W., Asio, S. M. (2011). *Mansonella perstans* filariasis in Africa. Elsevier B.V., p.1–12. [Online]. Available at: doi:10.1016/j.actatropica.2010.01.014.

Skwarczynski, M., Toth, I. (2016). Peptide-based synthetic vaccines. Chemical Science, 7, Royal Society of Chemistry, p.842–854. [Online]. Available at: doi:10.1039/C5SC03892H.

Slatko, B. E., Luck, A. N., Dobson, S. L., Foster, J. M. (2014). *Wolbachia* endosymbionts and human disease control. Molecular and Biochemical Parasitology, 195 (2), Elsevier B.V., p.88–95. [Online]. Available at: doi:10.1016/j.molbiopara.2014.07.004.

Slobedman, B., Cheung, A. K. L. (2008). Microarrays for the study of viral gene expression during human cytomegalovirus latent infection. Methods in molecular medicine, 141, p.153–175.

Smith, K. A., Harcus, Y., Garbi, N., Hämmerling, G. J., MacDonald, A. S., Maizels, R. M. (2012). Type 2 innate immunity in helminth infection is induced redundantly and acts autonomously following CD11c(+) cell depletion. Infection and Immunity, 80 (10), American Society for Microbiology, p.3481–3489. [Online]. Available at: doi:10.1128/IAI.00436-12.

Smith, K. A., Hochweller, K., Hämmerling, G. J., Boon, L., MacDonald, A. S., Maizels, R. M. (2011). Chronic Helminth Infection Promotes Immune Regulation *In Vivo* through Dominance of $CD11c^{lo}CD103^{-}$ Dendritic Cells. The Journal of Immunology, 186 (12), p.7098–7109.

Somorjai, R. L., Dolenko, B., Baumgartner, R. (2003). Class prediction and discovery using gene microarray and proteomics mass spectroscopy data: curses, caveats, cautions. Bioinformatics, 19 (12), p.1484–1491. [Online]. Available at: doi:10.1093/bioinformatics/btg182.

Specht, S., Frank, J. K., Alferink, J., Dubben, B., Layland, L. E., Denece, G., Bain, O., Forster, I., Kirschning, C. J., Martin, C., Hoerauf, A. (2011a). CCL17 Controls Mast Cells for the Defense against Filarial Larval Entry. The Journal of Immunology, 186 (8), p.4845–4852. [Online]. Available at: doi:10.4049/jimmunol.1000612.

Specht, S., Taylor, M. D., Hoeve, M., Allen, J. E., Lang, R., Hoerauf, A. (2011b). Over expression of IL-10 by macrophages overcomes resistance to murine filariasis. Experimental Parasitology, 132 (1), Elsevier Inc., p.1–7. [Online]. Available at: doi:10.1016/j.exppara.2011.09.003.

Specht, S., Volkmann, L., Wynn, T., Hoerauf, A. (2004). Interleukin-10 (IL-10) counterregulates IL-4-dependent effector mechanisms in Murine Filariasis. Infection and Immunity, 72 (11), p.6287–6293. [Online]. Available at: doi:10.1128/IAI.72.11.6287-6293.2004.

Steel, C., Nutman, T. B. (2003). CTLA-4 in filarial infections: implications for a role in diminished T cell reactivity. The Journal of Immunology, 170 (4), American Association of Immunologists, p.1930–1938. [Online]. Available at: doi:10.4049/jimmunol.170.4.1930.

Steel, C., Varma, S., Nutman, T. B. (2012). Regulation of Global Gene Expression in Human *Loa loa* Infection Is a Function of Chronicity. Unnasch, T. R. (ed.).

PLOS Neglected Tropical Diseases, 6 (2), p.e1527–13. [Online]. Available at: doi:10.1371/journal.pntd.0001527.

Steisslinger, V., Korten, S., Brattig, N. W., Erttmann, K. D. (2015). DNA vaccine encoding the moonlighting protein *Onchocerca volvulus* glyceraldehyde-3-phosphate dehydrogenase (*Ov*-GAPDH) leads to partial protection in a mouse model of human filariasis. Vaccine, 33 (43), Elsevier Ltd, p.5861–5867. [Online]. Available at: doi:10.1016/j.vaccine.2015.07.110.

Stolk, W. A., Stone, C., de Vlas, S. J. (2015a). Modelling lymphatic filariasis transmission and control: modelling frameworks, lessons learned and future directions. Advances in parasitology, 87, Elsevier, p.249–291. [Online]. Available at: doi:10.1016/bs.apar.2014.12.005.

Stolk, W. A., Walker, M., Coffeng, L. E., Basáñez, M.-G., de Vlas, S. J. (2015b). Required duration of mass ivermectin treatment for onchocerciasis elimination in Africa: a comparative modelling analysis. Parasites & Vectors, Parasites & Vectors, p.1–16. [Online]. Available at: doi:10.1186/s13071-015-1159-9.

Straw, A. D., MacDonald, A. S., Denkers, E. Y., Pearce, E. J. (2003). CD154 plays a central role in regulating dendritic cell activation during infections that induce Th1 or Th2 responses. Journal of immunology (Baltimore, Md. : 1950), 170 (2), p.727–734.

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., Zeileis, A. (2008). Conditional variable importance for random forests. BMC Bioinformatics, 9 (1), BioMed Central Ltd, p.307. [Online]. Available at: doi:10.1186/1471-2105-9-307.

Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics, 8, p.1–3. [Online]. Available at: doi:10.1186/1471-2105-8-25.

Sun, Y., Liu, G., Li, Z., Chen, Y., Liu, Y., Liu, B., Su, Z. (2013). Modulation of dendritic cell function and immune response by cysteine protease inhibitor from murine nematode parasite Heligmosomoides polygyrus. Immunology, 138 (4), p.370–381. [Online]. Available at: doi:10.1111/imm.12049.

Sundarrajan, S., Arumugam, M. (2016). Weighted gene co-expression based biomarker discovery for psoriasis detection. Gene, 593 (1), Elsevier B.V., p.225–234. [Online]. Available at: doi:10.1016/j.gene.2016.08.021.

Swan, A. L., Mobasheri, A., Allaway, D., Liddell, S., Bacardit, J. (2013). Application of Machine Learning to Proteomics Data: Classification and Biomarker Identification in Postgenomics Biology. OMICS: A Journal of Integrative Biology, 17 (12), p.595–610. [Online]. Available at: doi:10.1089/omi.2013.0017.

Tamarozzi, F., Halliday, A., Gentil, K., Hoerauf, A., Pearlman, E., Taylor, M. J.

(2011). Onchocerciasis: the Role of *Wolbachia* Bacterial Endosymbionts in Parasite Biology, Disease Pathogenesis, and Treatment. Clinical Microbiology Reviews, 24 (3), p.459–468. [Online]. Available at: doi:10.1128/CMR.00057-10.

Tamarozzi, F., Turner, J. D., Pionnier, N., Midgley, A., Guimaraes, A. F., Johnston, K. L., Edwards, S. W., Taylor, M. J. (2016). *Wolbachia* endosymbionts induce neutrophil extracellular trap formation in human onchocerciasis. Nature Publishing Group, 6 (1), Nature Publishing Group, p.1–13. [Online]. Available at: doi:10.1038/srep35559.

Tamarozzi, F., Wright, H. L., Johnston, K. L., Edwards, S. W., Turner, J. D., Taylor, M. J. (2014). Human filarial Wolbachialipopeptide directly activates human neutrophils *in vitro*. Parasite Immunology, 36 (10), p.494–502. [Online]. Available at: doi:10.1111/pim.12122.

Tameris, M. D., Hatherill, M., Landry, B. S., Scriba, T. J., Snowden, M. A., Lockhart, S., Shea, J. E., McClain, J. B., Hussey, G. D., Hanekom, W. A., Mahomed, H., McShane, H., (null). (2013). Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. Lancet (London, England), 381 (9871), p.1021–1028. [Online]. Available at: doi:10.1016/S0140-6736(13)60177-4.

Taubert, A., Zahner, H. (2001). Cellular immune responses of filaria (*Litomosoides sigmodontis*) infected BALB/c mice detected on the level of cytokine transcription. Parasite Immunology, 23 (8), p.453–462.

Taylor, M. D., Harris, A., Babayan, S. A., Bain, O., Culshaw, A., Allen, J. E., Maizels, R. M. (2007). CTLA-4 and CD4+CD25+ Regulatory T Cells Inhibit Protective Immunity to Filarial Parasites *In Vivo*. journal of immunology, 179, p.4626–4636.

Taylor, M. D., Harris, A., Nair, M. G., Maizels, R. M., Allen, J. E. (2006). F4/80+ alternatively activated macrophages control CD4+ T cell hyporesponsiveness at sites peripheral to filarial infection. Journal of immunology (Baltimore, Md. : 1950), 176 (11), p.6918–6927.

Taylor, M. D., LeGoff, L., Harris, A., Malone, E., Allen, J. E., Maizels, R. M. (2005). Removal of regulatory T cell activity reverses hyporesponsiveness and leads to filarial parasite clearance *in vivo*. Journal of immunology (Baltimore, Md. : 1950), 174 (8), p.4924–4933.

Taylor, M. D., van der Werf, N., Harris, A., Graham, A. L., Bain, O., Allen, J. E., Maizels, R. M. (2009). Early recruitment of natural CD4+Foxp3+ Treg cells by infective larvae determines the outcome of filarial infection. European Journal of Immunology, 39 (1), p.192–206. [Online]. Available at: doi:10.1002/eji.200838727.

Taylor, M. J., Hoerauf, A., Bockarie, M. (2010). Lymphatic filariasis and onchocerciasis. The Lancet, 376 (9747), Elsevier Ltd, p.1175–1185. [Online]. Available at: doi:10.1016/S0140-6736(10)60586-7.

Tchakouté, V. L., Graham, S. P., Jensen, S. A., Makepeace, B. L., Nfon, C. K., Njongmeta, L. M., Lustigman, S., Enyong, P. A., Tanya, V. N., Bianco, A. E. (2006). In a bovine model of onchocerciasis, protective immunity exists naturally, is absent in drug-cured hosts, and is induced by vaccination. Proceedings of the National Academy of Sciences, 103 (15), National Acad Sciences, p.5971–5976.

Tekle, A. H., Elhassan, E., Isiyaku, S., Amazigo, U. V., Bush, S., Noma, M., Cousens, S., Abiose, A., Remme, J. H. (2012). Impact of long-term treatment of onchocerciasis with ivermectin in Kaduna State, Nigeria: first evidence of the potential for elimination in the operational area of the African Programme for Onchocerciasis Control. Parasites & Vectors, 5 (1), BioMed Central, p.28. [Online]. Available at: doi:10.1186/1756-3305-5-28.

Tendongfor, N., Wanji, S., Ngwa, J. C., Esum, M. E., Specht, S., Enyong, P., Matthaei, K. I., Hoerauf, A. (2012). The human parasite *Loa loa* in cytokine and cytokine receptor gene knock out BALB/c mice: survival, development and localization. Parasites & Vectors, 5 (1), BioMed Central, p.43. [Online]. Available at: doi:10.1186/1756-3305-5-43.

Terrazas, C. A., Alcántara-Hernández, M., Bonifaz, L., Terrazas, L. I., Satoskar, A. R. (2013). Helminth-excreted/secreted products are recognized by multiple receptors on DCs to block the TLR response and bias Th2 polarization in a cRAF dependent pathway. FASEB journal: official publication of the Federation of American Societies for Experimental Biology, 27 (11), p.4547–4560. [Online]. Available at: doi:10.1096/fj.13-228932.

Thylefors, B. (2008). The Mectizan Donation Program (MDP). Annals of Tropical Medicine and Parasitology, 102 Suppl 1, p.39–44. [Online]. Available at: doi:10.1179/136485908X337481.

Timmann, C., Fuchs, S., Thoma, C., Lepping, B., Brattig, N. W., Sievertsen, J., Thye, T., Müller-Myhsok, B., Horstmann, R. D. (2004). Promoter haplotypes of the interleukin-10 gene influence proliferation of peripheral blood cells in response to helminth antigen. Genes and Immunity, 5 (4), p.256–260. [Online]. Available at: doi:10.1038/sj.gene.6364094.

Tolosi, L., Lengauer, T. (2011). Classification with correlated features: unreliability of feature ranking and solutions. Bioinformatics, 27 (14), p.1986–1994. [Online]. Available at: doi:10.1093/bioinformatics/btr300.

Torrero, M. N., Hubner, M. P., Larson, D., Karasuyama, H., Mitre, E. (2010). Basophils Amplify Type 2 Immune Responses, but Do Not Serve a Protective Role, during Chronic Infection of Mice with the Filarial Nematode *Litomosoides*

sigmodontis. The Journal of Immunology, 185 (12), p.7426–7434. [Online]. Available at: doi:10.4049/jimmunol.0903864.

Torrero, M. N., Morris, C. P., Mitre, B. K., Hübner, M. P., Fox, E. M., Karasuyama, H., Mitre, E. (2013). Basophils help establish protective immunity induced by irradiated larval vaccination for filariasis. Vaccine, 31 (36), p.3675–3682. [Online]. Available at: doi:10.1016/j.vaccine.2013.06.010.

Trottein, F., Pavelka, N., Vizzardelli, C., Angeli, V., Zouain, C. S., Pelizzola, M., Capozzoli, M., Urbano, M., Capron, M., Belardelli, F., Granucci, F., Ricciardi-Castagnoli, P. (2004). A Type I IFN-Dependent Pathway Induced by Schistosoma mansoni Eggs in Mouse Myeloid Dendritic Cells Generates an Inflammatory Signature. The Journal of Immunology, 172 (5), p.3011–3017. [Online]. Available at: doi:10.4049/jimmunol.172.5.3011.

Turaga, P. S. D., Tierney, T. J., Bennett, K. E., McCarthy, M. C., Simonek, S. C., Enyong, P. A., Moukatte, D. W., Lustigman, S. (2000). Immunity to Onchocerciasis: Cells from Putatively Immune Individuals Produce Enhanced Levels of Interleukin-5, Gamma Interferon, and Granulocyte-Macrophage Colony-Stimulating Factor in Response to Onchocerca volvulus Larval and Male Worm Antigens. Infection and Immunity, 68 (4), p.1905–1911. [Online]. Available at: doi:10.1128/IAI.68.4.1905-1911.2000.

Turner, H. C., Bettis, A. A., Chu, B. K., McFarland, D. A., Hooper, P. J., Ottesen, E. A., Bradley, M. H. (2016). The health and economic benefits of the global programme to eliminate lymphatic filariasis (2000–2014). Infectious Diseases of Poverty, 5 (1), Infectious Diseases of Poverty, p.1–19. [Online]. Available at: doi:10.1186/s40249-016-0147-4.

Turner, H. C., Churcher, T. S., Walker, M., Osei-Atweneboana, M. Y., Prichard, R. K., Basáñez, M.-G. (2013). Uncertainty Surrounding Projections of the Long-Term Impact of Ivermectin Treatment on Human Onchocerciasis. Lustigman, S. (ed.). 7 (4), p.e2169. [Online]. Available at: doi:10.1371/journal.pntd.0002169.s003.

Turner, H. C., Walker, M., Churcher, T. S., Bas ez, M. A.-G. (2014a). Modelling the impact of ivermectin on River Blindness and its burden of morbidity and mortality in African Savannah: EpiOncho projections. Parasites & Vectors, 7 (1), p.241–15. [Online]. Available at: doi:10.1186/1756-3305-7-241.

Turner, H. C., Walker, M., Churcher, T. S., Osei-Atweneboana, M. Y., Biritwum, N.-K., Hopkins, A., Prichard, R. K., Basáñez, M.-G. (2014b). Reaching the london declaration on neglected tropical diseases goals for onchocerciasis: an economic evaluation of increasing the frequency of ivermectin treatment in Africa. Clinical infectious diseases: an official publication of the Infectious Diseases Society of America, 59 (7), p.923–932. [Online]. Available at: doi:10.1093/cid/ciu467.

Turner, H. C., Walker, M., Lustigman, S., Taylor, D. W., Basáñez, M.-G. (2015).

Human Onchocerciasis: Modelling the Potential Long-term Consequences of a Vaccination Programme. Fenton, A. (ed.). PLOS Neglected Tropical Diseases, 9 (7), Public Library of Science, p.e0003938–19. [Online]. Available at: doi:10.1371/journal.pntd.0003938.

Turner, J. D., Langley, R. S., Johnston, K. L., Gentil, K., Ford, L., Wu, B., Graham, M., Sharpley, F., Slatko, B., Pearlman, E., Taylor, M. J. (2009). *Wolbachia* lipoprotein stimulates innate and adaptive immunity through Toll-like receptors 2 and 6 to induce disease manifestations of filariasis. The Journal of biological chemistry, 284 (33), p.22364–22378. [Online]. Available at: doi:10.1074/jbc.M901528200.

Twum-Danso, N. A. (2003). *Loa loa* encephalopathy temporally related to ivermectin administration reported from onchocerciasis mass treatment programs from 1989 to 2001: implications for the future. Filaria Journal, 2 Suppl 1, p.S7. [Online]. Available at: doi:10.1186/1475-2883-2-S1-S7.

Uniting to Combat NTDs. (2012). The London Declaration on Neglected Tropicial Diseases. London: Uniting to Combat NTDs.

Vahey, M. T., Wang, Z., Kester, K. E., Cummings, J., Heppner, D. G., Nau, M. E., Ofori Anyinam, O., Cohen, J., Coche, T., Ballou, W. R., Ockenhouse, C. F. (2010). Expression of genes associated with immunoproteasome processing of major histocompatibility complex peptides is indicative of protection with adjuvanted RTS,S malaria vaccine. Journal of Infectious Diseases, 201 (4), Oxford University Press, p.580–589. [Online]. Available at: doi:10.1086/650310.

Van der Werf, N., Redpath, S. A., Azuma, M., Yagita, H., Taylor, M. D. (2013). Th2 Cell-Intrinsic Hypo-Responsiveness Determines Susceptibility to Helminth Infection. Loke, P. (ed.). PLoS Pathogens, 9 (3), p.e1003215. [Online]. Available at: doi:10.1371/journal.ppat.1003215.s002.

Vlaminck, J., Fischer, P. U., Weil, G. J. (2015). Diagnostic Tools for Onchocerciasis Elimination Programs. Trends in Parasitology, 31 (11), Elsevier Ltd, p.571–582. [Online]. Available at: doi:10.1016/j.pt.2015.06.007.

Volkmann, L., Bain, O., Saeftel, M., Specht, S., Fischer, K., Hoerauf, A. (2003). Murine filariasis: interleukin 4 and interleukin 5 lead to containment of different worm developmental stages. p.1–11. [Online]. Available at: doi:10.1007/s00430-002-0155-9.

Volkmann, L., Saeftel, M., Bain, O., Fischer, K., Fleischer, B., Hoerauf, A. (2001). Interleukin-4 Is Essential for the Control of Microfilariae in Murine Infection with the Filaria *Litomosoides sigmodontis*. Infection and Immunity, 69 (5), p.2950–2956. [Online]. Available at: doi:10.1128/IAI.69.5.2950-2956.2001.

Vray, B., Hartmann, S., Hoebeke, J. (2002). Immunomodulatory properties of

cystatins. Cellular and molecular life sciences: CMLS, 59 (9), p.1503–1512.

Wang, F., Wang, Y. Y., Li, J., You, X., Qiu, X. H., Wang, Y. N., Gao, F. G. (2014). Increased Antigen Presentation but Impaired T Cells Priming after Upregulation of Interferon-Beta Induced by Lipopolysaccharides Is Mediated by Upregulation of B7H1 and GITRL. Gabriele, L. (ed.). PLoS ONE, 9 (8), p.e105636–11. [Online]. Available at: doi:10.1371/journal.pone.0105636.

Wang, H. (2015). Improved Variable Importance Measure of Random Forest via Combining of Proximity Measure and Support Vector Machine for Stable Feature Selection. Journal of Information and Computational Science, 12 (8), p.3241–3252. [Online]. Available at: doi:10.12733/jics20105854.

Wang, H., Yang, F., Luo, Z. (2016). An experimental study of the intrinsic stability of random forest variable importance measures. BMC Bioinformatics, 17 (1), BioMed Central, p.60. [Online]. Available at: doi:10.1186/s12859-016-0900-5.

Wanji, S. (2015). Situation analysis of parasitological and entomological indices of onchocerciasis transmission in three drainage basins of the rain forest of South West Cameroon after a decade of ivermectin treatment. 2nd ed. Parasites & Vectors, 8 (1), p.1–21. [Online]. Available at: doi:10.1186/s13071-015-0817-2.

Wanji, S., Amvongo-Adjia, N., Koudou, B., Njouendou, A. J., Chounna Ndongmo, P. W., Kengne-Ouafo, J. A., Datchoua-Poutcheu, F. R., Fovennso, B. A., Tayong, D. B., Fombad, F. F., Fischer, P. U., Enyong, P. I., Bockarie, M. (2015a). Cross-Reactivity of Filariais ICT Cards in Areas of Contrasting Endemicity of *Loa loa* and *Mansonella perstans* in Cameroon: Implications for Shrinking of the Lymphatic Filariasis Map in the Central African Region. Bottomley, C. (ed.). PLOS Neglected Tropical Diseases, 9 (11), Public Library of Science, p.e0004184–20. [Online]. Available at: doi:10.1371/journal.pntd.0004184.

Wanji, S., Eyong, E.-E., Tendongfor, N., Ngwa, C., Esuka, E., Kengne-Ouafo, A., Datchoua-Poutcheu, F., Enyong, P., Hopkins, A., Mackenzie, C. D. (2015b). Parasitological, Hematological and Biochemical Characteristics of a Model of Hyper-microfilariaemic Loiasis (Loa loa) in the Baboon (Papio anubis). Makepeace, B. L. (ed.). PLOS Neglected Tropical Diseases, 9 (11), p.e0004202–e0004224. [Online]. Available at: doi:10.1371/journal.pntd.0004202.

Wanji, S., Tendongfor, N., Nji, T., Esum, M., Che, J. N., Nkwescheu, A., Alassa, F., Kamnang, G., Enyong, P. A., Taylor, M. J., Hoerauf, A., Taylor, D. W. (2009). Community-directed delivery of doxycycline for the treatment of onchocerciasis in areas of co-endemicity with loiasis in Cameroon. Parasites & Vectors, 2 (1), p.39–10. [Online]. Available at: doi:10.1186/1756-3305-2-39.

Weiner, J., Kaufmann, S. H. E., Maertzdorf, J. (2015). High-throughput data analysis and data integration for vaccine trials. Vaccine, 33 (40), Elsevier Ltd, p.5249–5255. [Online]. Available at: doi:10.1016/j.vaccine.2015.04.096.

Whelan, M., Harnett, M. M., Houston, K. M., Patel, V., Harnett, W., Rigley, K. P. (2000). A Filarial Nematode-Secreted Product Signals Dendritic Cells to Acquire a Phenotype That Drives Development of Th2 Cells. The Journal of Immunology, 164 (12), p.6453–6460. [Online]. Available at: doi:10.4049/jimmunol.164.12.6453.

WHO. (2016). Guidelines for Stopping Mass Drug Administration and Verifying Elimination of Human Onchocerciasis: Criteria and Procedures. Geneva: World Health Organization.

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer New York. [Online]. Available at: http://had.co.nz/ggplot2/book.

World Health Organization. (2012). Accelerating work to overcome the global impact of neglected tropical diseases–A roadmap for implementation. p.1–42.

World Health Organization. (2013). Global programme to eliminate lymphatic filariasis: progress report for 2012. Weekly epidemiological record, (88), p.389–400. [Online]. Available at: http://www.who.int/wer.

World Health Organization. (2015). Global programme to eliminate lymphatic filariasis: progress report, 2014. Weekly epidemiological record, 90, p.489–504. [Online]. Available at: http://www.who.int/wer.

World Health Organization. (2016a). Global programme to eliminate lymphatic lariasis: progress report, 2015. Weekly epidemiological record, (91), p.441–460. [Online]. Available at: http://www.who.int/wer.

World Health Organization. (2016b). Progress towards eliminating onchocerciasis in the WHO Region of the Americas. Weekly epidemiological record, (91), p.501–516. [Online]. Available at: http://www.who.int/wer.

Wu, M. T., Fang, H., Hwang, S. T. (2001). Cutting Edge: CCR4 Mediates Antigen-Primed T Cell Binding to Activated Dendritic Cells. The Journal of Immunology, 167 (9), p.4791–4795. [Online]. Available at: doi:10.4049/jimmunol.167.9.4791.

Xie, Y., Wang, X., Story, M. (2009). Statistical methods of background correction for Illumina BeadArray data. Bioinformatics, 25 (6), p.751–757. [Online]. Available at: doi:10.1093/bioinformatics/btp040.

Yamazaki, S., Inaba, K., Tarbell, K. V., Steinman, R. M. (2006). Dendritic cells expand antigen-specific Foxp3+ CD25+ CD4+ regulatory T cells including suppressors of alloreactivity. Immunological reviews, 212 (1), Blackwell Publishing Ltd, p.314–329. [Online]. Available at: doi:10.1111/j.0105-2896.2006.00422.x.

Yates, J. A., Higashi, G. I. (1985). *Brugia malayi*: vaccination of jirds with 60cobalt-

attenuated infective stage larvae protects against homologous challenge. American Journal of Tropical Medicine and Hygiene, 34 (6), p.1132–1137.

Yepes, S., López, R., Andrade, R. E., Rodriguez-Urrego, P. A., López-Kleine, L., Torres, M. M. (2016). Co-expressed miRNAs in gastric adenocarcinoma. Genomics, 108 (2), Elsevier Inc., p.93–101. [Online]. Available at: doi:10.1016/j.ygeno.2016.07.002.

Yu, G., Wang, L.-G., Han, Y., He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS: A Journal of Integrative Biology, 16 (5), p.284–287. [Online]. Available at: doi:10.1089/omi.2011.0118.

Zhang, L., Xiaojuan Huang. (2015). Multiple SVM-RFE for multi-class gene selection on DNA Microarray data. In: 21 September 2015, IEEE, p.1–6. [Online]. Available at: doi:10.1109/IJCNN.2015.7280417.

Zhou, G., Stevenson, M. M., Geary, T. G., Xia, J. (2016). Comprehensive Transcriptome Meta-analysis to Characterize Host Immune Responses in Helminth Infections. Yazdanbakhsh, M. (ed.). PLOS Neglected Tropical Diseases, 10 (4), p.e0004624–20. [Online]. Available at: doi:10.1371/journal.pntd.0004624.

Ziewer, S., Hübner, M. P., Dubben, B., Hoffmann, W. H., Bain, O., Martin, C., Hoerauf, A., Specht, S. (2012). Immunization with *L. sigmodontis* Microfilariae Reduces Peripheral Microfilaraemia after Challenge Infection by Inhibition of Filarial Embryogenesis. Dalton, J. P. (ed.). 6 (3), p.e1558. [Online]. Available at: doi:10.1371/journal.pntd.0001558.g006.

Zipperer, G. R., Arumugam, S., Chirgwin, S. R., Coleman, S. U., Shakya, K. P., KLEI, T. R. (2013). Experimental Parasitology. Experimental parasitology, 135 (2), Elsevier Inc., p.446–455. [Online]. Available at: doi:10.1016/j.exppara.2013.08.007.

Zouré, H. G. M., Wanji, S., Noma, M., Amazigo, U. V., Diggle, P. J., Tekle, A. H., Remme, J. H. F. (2011). The Geographic Distribution of Loa loa in Africa: Results of Large-Scale Implementation of the Rapid Assessment Procedure for Loiasis (RAPLOA). Raso, G. (ed.). PLOS Neglected Tropical Diseases, 5 (6), Public Library of Science, p.e1210–e1211. [Online]. Available at: doi:10.1371/journal.pntd.0001210

Appendix A. Supplementary Tables from Chapter 2.

Table S1. Supplementary list of genes used in qPCR array

Gene Symbol	Gene RefSeq	Gene Symbol	Gene RefSeq
Apcs	NM_011318	Ifng	NM_008337
C3	NM_009778	Ifngr1	NM_010511
C5ar1	NM_007577	Il10	NM_010548
Casp1	NM_009807	Il13	NM_008355
Ccl12	NM_011331	Il17a	NM_010552
Ccl5	NM_013653	Il18	NM_008360
Ccr4	NM_009916	Il1a	NM_010554
Ccr5	NM_009917	Il1b	NM_008361
Ccr6	NM_009835	Il1r1	NM_008362
Ccr8	NM_007720	Il2	NM_008366
Cd14	NM_009841	Il23a	NM_031252
Cd4	NM_013488	Il4	NM_021283
Cd40	NM_011611	Il5	NM_010558
Cd40lg	NM_011616	Il6	NM_031168
Cd80	NM_009855	Irak1	NM_008363
Cd86	NM_019388	Il-33	NM_033439.3
Cd8a	NM_001081110	Irf7	NM_016850
Crp	NM_007768	Itgam	NM_008401
Csf2	NM_009969	Jak2	NM_008413
Cxcl10	NM_021274	Ly96	NM_016923
Cxcr3	NM_009910	Lyz2	NM_017372
Il-3	NM_000588.3	Mapk1	NM_011949
Fasl	NM_010177	Mapk8	NM_016700
Foxp3	NM_054039	Mbl2	NM_010776
Gata3	NM_008091	Mpo	NM_010824
H2-Q10	NM_010391	Eotaxin	NM_011330.3
H2-T23	NM_010398	Myd88	NM_010851
Icam1	NM_010493	Nfkb1	NM_008689
Cxcl1	NM_001511	Nfkbia	NM_010907
Ifnar1	NM_010508	Lta	NM_001159740.2
Cxcl3	NM_002090	Nod1	NM_172729

Gene Symbol	Gene RefSeq	Gene Symbol	Gene RefSeq
Nod2	NM_145857	Tlr2	NM_011905
Rag1	NM_009019	Tlr3	NM_126166
Rorc	NM_011281	Tlr4	NM_021297
Ltb	NM_002341.1	Tlr5	NM_016928
Stat1	NM_009283	Tlr6	NM_011604
Stat3	NM_011486	Tlr7	NM_133211
Stat4	NM_011487	Tlr8	NM_133212
Stat6	NM_009284	Tlr9	NM_031178
Tbx21	NM_019507	Tnf	NM_013693
Ticam1	NM_174989	Traf6	NM_009424
Tlr1	NM_030682	Tyk2	NM_018793

Appendix B. Supplementary results from chapter 3 (functional pathway analysis of machine learning pipeline and WGCNA results)

Table S2. Full list of over-represented Reactome pathways from the WGCNA analysis of Mf immunity. Time points Day -21, Day -14 and Day 67 in the Mf immunity dataset had significantly over-represented pathways. Multiple testing was adjusted for using the Benjamini-Hochberg method to give an adjust P-value, an adjusted P-value < 0.05 and a Q-value < 0.2 was used as the significance cut-off point for significant.

Description of Reactome Pathway	Found by ML pipeline
Day -21 (6 hours after 2nd Immunisation)	
Adaptive Immune System	✓
Antigen Presentation: Folding, assembly and peptide loading of class I MHC	✓
Antigen processing-Cross presentation	✓
Class I MHC mediated antigen processing & presentation	✓
Endosomal/Vacuolar pathway	✓
ER-Phagosome pathway	✓
Immune System	✓
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	✓
Interferon gamma signaling	✓
Interferon Signaling	✓
Generation of second messenger molecules	✓
PD-1 signaling	✓
Phosphorylation of CD3 and TCR zeta chains	✓
Translocation of ZAP-70 to Immunological synapse	✓
Activated TLR4 signalling	

ERK/MAPK targets	
ERKs are inactivated	
MAP kinase activation in TLR cascade	
MAPK targets/ Nuclear events mediated by MAP kinases	
MyD88 cascade initiated on plasma membrane	
MyD88 dependent cascade initiated on endosome	
MyD88-independent TLR3/TLR4 cascade	
MyD88:Mal cascade initiated on plasma membrane	
Nuclear Events (kinase and transcription factor activation)	
RHO GTPases Activate Formins	
Toll Like Receptor 10 (TLR10) Cascade	
Toll Like Receptor 2 (TLR2) Cascade	
Toll Like Receptor 3 (TLR3) Cascade	
Toll Like Receptor 4 (TLR4) Cascade	
Toll Like Receptor 5 (TLR5) Cascade	
Toll Like Receptor 7/8 (TLR7/8) Cascade	
Toll Like Receptor 9 (TLR9) Cascade	
Toll Like Receptor TLR1:TLR2 Cascade	
Toll Like Receptor TLR6:TLR2 Cascade	
TRAF6 mediated induction of NFkB and MAP kinases upon TLR7/8 or 9 activation	
TRAF6 Mediated Induction of proinflammatory cytokines	
TRIF-mediated TLR3/TLR4 signaling	
Day -14 (6 hours after 3rd Immunisation)	
Interferon gamma signaling	✓
Interferon Signaling	✓
Day 67	

Endosomal/Vacuolar pathway	
Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	
Interferon gamma signaling	
Interferon Signaling	

Table S3. Full list of over-represented Gene Ontology Terms of biological processes from the WGCNA analysis of Mf immunity. Time point day -21, day 0, day 49 and day 67 in the Mf immunity dataset had significantly over-represented GO terms. Multiple testing was adjusted for using the Benjamini-Hochberg method to give an adjust P-value, an adjusted P-value < 0.05 and a Q-value < 0.2 was used as the significance cut-off.

GO Description	Identified in ML pipeline
Day -21 (6 hours after 2nd Immunisation)	
response to interferon-beta	✓
Day 0 (6 hours post challenge)	
nuclear body organization	
PML body organization	
nucleus organization	
Day 49 (post challenge)	
response to external stimulus	
Day 67 (post challenge)	
antigen processing and presentation	
response to interferon-beta	
cellular response to interferon-beta	
defense response to other organism	
response to external biotic stimulus	

Table S4. Full list of over-represented Gene Ontology Terms of biological processes in the Mf immunity dataset found by the ML Pipeline. Time point day -21, day -14 and day 49 in the Mf immunity dataset had significantly over-represented GO terms. Multiple testing was adjusted for using the Benjamini-Hochberg method to give an adjust P-value, an adjusted P-value < 0.05 and a Q-value < 0.2 was used as the significance cut-off.

BP GO Term ID	BP GO Term	Genes (n)	Adjusted P-value	Genes selected by pipeline
Day -21 (6 hours after 2nd immunisation)				
GO:0035456	response to interferon-beta	7	9.39E-06	Gbp2b, Gbp2, Gbp3, Ifit3, Ifitm3, Igtp, Irf1
GO:0035458	cellular response to interferon-beta	6	4.87E-05	Gbp2b, Gbp2, Gbp3, Ifit3, Igtp, Irf1
GO:0048002	antigen processing and presentation of peptide antigen	6	4.32E-03	B2m, Cd74, H2-D1, H2-Q6, H2-Q8, H2-T23
GO:0071345	cellular response to cytokine stimulus	13	6.38E-03	Cd74, Cdc37, Coro1a, Gbp2b, Gbp2, Gbp3, Ifit3, Ifitm3, Igtp, Irf1, Irf7, Robo1, Tpr
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	5	6.46E-03	B2m, H2-D1, H2-Q6, H2-Q8, H2-T23
GO:0034341	response to interferon-gamma	6	6.86E-03	Cdc37, Gbp2b, Gbp2, Gbp3, Ifitm3, Irf1
GO:0002250	adaptive immune response	11	6.86E-03	B2m, Cd74, Cd79b, Cd8b1, Csk, H2-D1, H2-T23, Irf1, Irf7, Serpina3g, Unc13d
GO:0034112	positive regulation of homotypic cell-cell adhesion	8	6.86E-03	Ank3, Cd74, Coro1a, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0098542	defense response to other organism	13	8.95E-03	B2m, Gbp2b, Gbp2, Gbp3, H2-T23, Ifit3, Ifitm3, Irf1, Nlrp1a, Oasl2, Plac8, Samhd1, Unc13d

GO:0071346	cellular response to interferon-gamma	5	8.95E-03	Cdc37, Gbp2b, Gbp2, Gbp3, Irf1
GO:0042110	T cell activation	12	8.95E-03	B2m, Cd2, Cd3d, Cd74, Coro1a, Csk, Fcgr4, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0070489	T cell aggregation	12	8.95E-03	B2m, Cd2, Cd3d, Cd74, Coro1a, Csk, Fcgr4, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0071593	lymphocyte aggregation	12	8.95E-03	B2m, Cd2, Cd3d, Cd74, Coro1a, Csk, Fcgr4, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0034109	homotypic cell-cell adhesion	13	8.95E-03	Ank3, B2m, Cd2, Cd3d, Cd74, Coro1a, Csk, Fcgr4, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0070486	leukocyte aggregation	12	9.76E-03	B2m, Cd2, Cd3d, Cd74, Coro1a, Csk, Fcgr4, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0045088	regulation of innate immune response	8	1.19E-02	Arf6, Cd74, Cdc37, H2-T23, Irf1, Irf7, Samhd1, Trafd1
GO:0022409	positive regulation of cell-cell adhesion	8	1.20E-02	Ank3, Cd74, Coro1a, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0007159	leukocyte cell-cell adhesion	12	1.46E-02	B2m, Cd2, Cd3d, Cd74, Coro1a, Csk, Fcgr4, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0002483	antigen processing and presentation of endogenous peptide antigen	3	1.46E-02	B2m, H2-D1, H2-T23
GO:0019882	antigen processing and presentation	6	1.46E-02	B2m, Cd74, H2-D1, H2-Q6, H2-Q8, H2-T23

GO:0006414	translational elongation	4	1.47E-02	Secisbp2, Eef1a1, Eef1b2, Yrdc
GO:0050870	positive regulation of T cell activation	7	1.47E-02	Cd74, Coro1a, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0034110	regulation of homotypic cell-cell adhesion	9	1.47E-02	Ank3, Cd74, Coro1a, Csk, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0044406	adhesion of symbiont to host	3	1.47E-02	Gbp2b, Gbp2, Gbp3
GO:0050830	defense response to Gram-positive bacterium	5	1.58E-02	B2m, Gbp2b, Gbp2, Gbp3, H2-T23
GO:0050778	positive regulation of immune response	11	1.58E-02	Arf6, B2m, Cd74, Cd79b, Csk, H2-D1, H2-T23, Irf1, Irf7, Lck, Thy1
GO:0019883	antigen processing and presentation of endogenous antigen	3	1.58E-02	B2m, H2-D1, H2-T23
GO:1903039	positive regulation of leukocyte cell-cell adhesion	7	1.58E-02	Cd74, Coro1a, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0035455	response to interferon-alpha	3	2.15E-02	Ifit3, Ifitm3, Tpr
GO:0032845	negative regulation of homeostatic process	7	2.41E-02	Cd74, Coro1a, Csk, Lck, Mcoln1, Rtel1, Thy1
GO:0032844	regulation of homeostatic process	11	2.41E-02	Ank3, B2m, Cd74, Coro1a, Csk, Ets1, Hcar2, Lck, Mcoln1, Rtel1, Thy1
GO:0002819	regulation of adaptive immune response	6	2.41E-02	B2m, Cd74, H2-D1, H2-T23, Irf1, Irf7
GO:0001916	positive regulation of T cell mediated cytotoxicity	3	2.47E-02	B2m, H2-D1, H2-T23
GO:0051497	negative regulation of stress fiber assembly	3	2.47E-02	Clasp2, Dlc1, Pfn1

GO:0060337	type I interferon signaling pathway	3	2.47E-02	Cdc37, Ifitm3, Irf7
GO:0002821	positive regulation of adaptive immune response	5	2.53E-02	B2m, Cd74, H2-D1, H2-T23, Irf1
GO:0071357	cellular response to type I interferon	3	2.66E-02	Cdc37, Ifitm3, Irf7
GO:0001909	leukocyte mediated cytotoxicity	5	2.66E-02	B2m, Coro1a, H2-D1, H2-T23, Unc13d
GO:0050863	regulation of T cell activation	8	2.67E-02	Cd74, Coro1a, Csk, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0051235	maintenance of location	8	2.90E-02	Ank3, Coro1a, Lck, Mcoln1, Pfn1, Sorl1, Thy1, Sun2
GO:1903037	regulation of leukocyte cell-cell adhesion	8	3.38E-02	Cd74, Coro1a, Csk, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0032232	negative regulation of actin filament bundle assembly	3	3.57E-02	Clasp2, Dlc1, Pfn1
GO:0071426	ribonucleoprotein complex export from nucleus	4	3.77E-02	Ddx39b, Rps15, Thoc6, Tpr
GO:0001914	regulation of T cell mediated cytotoxicity	3	3.77E-02	B2m, H2-D1, H2-T23
GO:0035740	CD8-positive, alpha-beta T cell proliferation	2	3.77E-02	H2-T23, Irf1
GO:0002449	lymphocyte mediated immunity	7	3.79E-02	B2m, Cd74, Coro1a, H2-D1, H2-T23, Irf7, Unc13d
GO:0071166	ribonucleoprotein complex localization	4	3.86E-02	Ddx39b, Rps15, Thoc6, Tpr
GO:0045785	positive regulation of cell adhesion	9	3.86E-02	Ank3, Cd74, Coro1a, H2-T23, Irf1, Lck, Tgfbr2, Thy1, Unc13d
GO:0034340	response to type I interferon	3	3.89E-02	Cdc37, Ifitm3, Irf7

GO:0001906	cell killing	5	4.14E-02	B2m, Coro1a, H2-D1, H2-T23, Unc13d
GO:0022407	regulation of cell-cell adhesion	9	4.15E-02	Ank3, Cd74, Coro1a, Csk, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0002460	adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	7	4.15E-02	B2m, Cd74, H2-D1, H2-T23, Irf1, Irf7, Unc13d
GO:0002757	immune response-activating signal transduction	7	4.15E-02	Arf6, Cd79b, Csk, Irf1, Irf7, Lck, Thy1
GO:0042742	defense response to bacterium	7	4.15E-02	B2m, Gbp2b, Gbp2, Gbp3, H2-T23, Nlrp1a, Plac8
GO:0006488	dolichol-linked oligosaccharide biosynthetic process	2	4.15E-02	Alg12, Dpagt1
GO:0043320	natural killer cell degranulation	2	4.15E-02	Coro1a, Unc13d
GO:0046784	viral mRNA export from host cell nucleus	2	4.15E-02	Ddx39b, Thoc6
GO:0051251	positive regulation of lymphocyte activation	7	4.29E-02	Cd74, Coro1a, H2-T23, Irf1, Lck, Tgfbr2, Thy1
GO:0042832	defense response to protozoan	3	4.47E-02	Gbp2b, Gbp2, Gbp3
Day -14 (6 hours after 3rd immunisation)				
GO:0040029	regulation of gene expression, epigenetic	9	3.42E-05	Arid4a, Hist1h2ad, Hist1h2af, Hist1h2ah, Hist1h2ai, Hist1h2an, Hist2h2ac, Spi1, Xist
GO:0045814	negative regulation of gene expression, epigenetic	7	3.42E-05	Hist1h2ad, Hist1h2af, Hist1h2ah, Hist1h2ai, Hist1h2an, Hist2h2ac, Spi1
GO:0035458	cellular response to interferon-beta	5	6.53E-05	Gbp2b, Gbp2, Gbp3, Ube2k, Irf1

GO:0042832	defense response to protozoan	5	6.53E-05	Gbp2b, Gbp2, Gbp3, Irgm2, Slc11a1
GO:0001562	response to protozoan	5	8.76E-05	Gbp2b, Gbp2, Gbp3, Irgm2, Slc11a1
GO:0035456	response to interferon-beta	5	1.00E-04	Gbp2b, Gbp2, Gbp3, Ube2k, Irf1
GO:0006342	chromatin silencing	6	1.67E-04	Hist1h2ad, Hist1h2af, Hist1h2ah, Hist1h2ai, Hist1h2an, Hist2h2ac
GO:0034341	response to interferon-gamma	6	2.19E-04	Gbp2b, Gbp2, Gbp3, Irgm2, Irf1, Slc11a1
GO:0098542	defense response to other organism	12	2.52E-04	B2m, Gbp2b, Gbp2, Gbp3, H2-K1, Irgm2, Irf1, Oasl2, Plac8, Prf1, Slc11a1, Slc25a19
GO:0042742	defense response to bacterium	8	9.22E-04	B2m, Gbp2b, Gbp2, Gbp3, H2-K1, Irgm2, Plac8, Slc11a1
GO:0048002	antigen processing and presentation of peptide antigen	5	1.15E-03	B2m, H2-D1, H2-K1, H2-Q6, Slc11a1
GO:0019885	antigen processing and presentation of endogenous peptide antigen via MHC class I	3	2.20E-03	B2m, H2-D1, H2-K1
GO:0009617	response to bacterium	10	2.32E-03	B2m, Casp1, Cd209b, Gbp2b, Gbp2, Gbp3, H2-K1, Irgm2, Plac8, Slc11a1
GO:0002483	antigen processing and presentation of endogenous peptide antigen	3	2.93E-03	B2m, H2-D1, H2-K1
GO:0002824	positive regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	5	2.93E-03	B2m, H2-D1, H2-K1, Irf1, Slc11a1

GO:0044406	adhesion of symbiont to host	3	3.37E-03	Gbp2b, Gbp2, Gbp3
GO:0002474	antigen processing and presentation of peptide antigen via MHC class I	4	3.47E-03	B2m, H2-D1, H2-K1, H2-Q6
GO:0002821	positive regulation of adaptive immune response	5	3.47E-03	B2m, H2-D1, H2-K1, Irf1, Slc11a1
GO:0019883	antigen processing and presentation of endogenous antigen	3	3.48E-03	B2m, H2-D1, H2-K1
GO:0016458	gene silencing	6	4.38E-03	Hist1h2ad, Hist1h2af, Hist1h2ah, Hist1h2ai, Hist1h2an, Hist2h2ac
GO:0001916	positive regulation of T cell mediated cytotoxicity	3	6.33E-03	B2m, H2-D1, H2-K1
GO:0071346	cellular response to interferon-gamma	4	7.62E-03	Gbp2b, Gbp2, Gbp3, Irf1
GO:0019882	antigen processing and presentation	5	7.68E-03	B2m, H2-D1, H2-K1, H2-Q6, Slc11a1
GO:0001914	regulation of T cell mediated cytotoxicity	3	1.07E-02	B2m, H2-D1, H2-K1
GO:0002822	regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	5	1.07E-02	B2m, H2-D1, H2-K1, Irf1, Slc11a1
GO:0002819	regulation of adaptive immune response	5	1.57E-02	B2m, H2-D1, H2-K1, Irf1, Slc11a1
GO:0050830	defense response to Gram-positive bacterium	4	1.64E-02	B2m, Gbp2b, Gbp2, Gbp3
GO:0002456	T cell mediated immunity	4	2.26E-02	B2m, H2-D1, H2-K1, Slc11a1
GO:0002711	positive regulation of T cell mediated immunity	3	2.44E-02	B2m, H2-D1, H2-K1
GO:0002484	antigen processing and presentation of endogenous	2	2.44E-02	H2-D1, H2-K1

	peptide antigen via MHC class I via ER pathway			
GO:0002485	antigen processing and presentation of endogenous peptide antigen via MHC class I via ER pathway, TAP-dependent	2	2.44E-02	H2-D1, H2-K1
GO:0001913	T cell mediated cytotoxicity	3	2.47E-02	B2m, H2-D1, H2-K1
GO:0002250	adaptive immune response	7	2.70E-02	B2m, H2-D1, H2-K1, Irf1, Nedd4, Serpina3g, Slc11a1
GO:0001912	positive regulation of leukocyte mediated cytotoxicity	3	4.20E-02	B2m, H2-D1, H2-K1
GO:0001771	immunological synapse formation	2	4.21E-02	Lgals3, Prf1
Day 49 (49 days after challenge)				
GO:0006342	chromatin silencing	5	4.99E-02	Hist1h2ad, Hist1h2af, Hist1h2ai, Hist1h2an, Hist2h2ac
GO:0045814	negative regulation of gene expression, epigenetic	5	4.99E-02	Hist1h2ad, Hist1h2af, Hist1h2ai, Hist1h2an, Hist2h2ac

Table S5. Full list of over-represented Gene Ontology Terms of biological processes in the *O. volvulus* endemic area dataset, identified by the ML Pipeline. 58 BP GO terms were found over-represented in the *O. volvulus* dataset. Multiple testing was accounted for using the Benjamini-Hochberg method to give an adjust P-value, an adjusted P-value < 0.05 and a Q-value < 0.2 was used as the significance cut-off.

BP GO Term ID	BP GO Term	Genes (n)	Adjusted P-value	Genes selected by pipeline
GO:0001906	cell killing	8	1.14E-03	CAMP, CTSG, DEFA1, DEFA1B, DEFA3, DEFA4, ELANE, GNLY
GO:0006261	DNA-dependent DNA replication	10	3.18E-04	CCNE2, CDC45, CDT1, DACH1, GINS2, LONP1, MCM2, MCM4, SLBP, TOP2A
GO:0032508	↳ DNA duplex unwinding	5	3.81E-02	CDC45, GINS2, MCM2, MCM4, TOP2A
GO:0006270	↳ DNA replication initiation	6	5.38E-04	CCNE2, CDC45, CDT1, GINS2, MCM2, MCM4
GO:0006268	↳ DNA unwinding involved in DNA replication	3	8.23E-03	MCM2, MCM4, TOP2A
GO:0033260	↳ nuclear DNA replication	5	6.28E-04	CDC45, CDT1, DACH1, GINS2, SLBP
GO:0044786	↳ cell cycle DNA replication	5	2.05E-03	CDC45, CDT1, DACH1, GINS2, SLBP
GO:0006260	↳ DNA replication	12	2.61E-03	CCNE2, CDC45, CDT1, DACH1, GINS2, KIAA0101, LONP1, MCM2, MCM4, RBBP4, SLBP, TOP2A
GO:0006959	humoral immune response	8	1.70E-02	C4BPA, CAMP, CD46, DEFA1, DEFA1B, DEFA3, DEFA4, PHB
GO:0031640	killing of cells of other organism	8	3.56E-07	CAMP, CTSG, DEFA1, DEFA1B, DEFA3, DEFA4, ELANE, GNLY
GO:0051873	↳ killing by host of symbiont cells	3	5.28E-03	CAMP, CTSG, ELANE

GO:0051883	↳	killing of cells in other organism involved in symbiotic interaction	3	1.28E-02	CAMP, CTSG, ELANE
GO:0051852	↳	disruption by host of symbiont cells	3	5.28E-03	CAMP, CTSG, ELANE
GO:0051818	↳	disruption of cells of other organism involved in symbiotic interaction	3	1.28E-02	CAMP, CTSG, ELANE
GO:0044364	↳	disruption of cells of other organism	8	3.56E-07	CAMP, CTSG, DEFA1, DEFA1B, DEFA3, DEFA4, ELANE, GNLY
GO:0000082		G1/S transition of mitotic cell cycle	12	5.38E-04	AURKA, BRD7, CCNE2, CDC45, CDCA5, CDKN3, CDT1, FBXO31, ID2, MCM2, MCM4, TYMS
GO:0098813	↳	nuclear chromosome segregation	12	1.63E-03	AURKB, CDC20, CDCA5, CENPI, DLGAP5, KIFC1, NCAPG, NUSAP1, PRC1, PTTG3P, TOP2A, UBE2C
GO:0045840	↳	positive regulation of mitotic nuclear division	5	5.29E-03	AURKA, CDCA5, DLGAP5, NUSAP1, UBE2C
GO:0044772	↳	mitotic cell cycle phase transition	16	1.14E-03	AURKA, BRD7, CCNB2, CCNE2, CDC45, CDCA5, CDKN3, CDT1, DLGAP5, FBXO31, HMMR, ID2, MCM2, MCM4, TYMS, UBE2C
GO:0044770	↳	cell cycle phase transition	16	2.04E-03	AURKA, BRD7, CCNB2, CCNE2, CDC45, CDCA5, CDKN3, CDT1, DLGAP5, FBXO31, HMMR, ID2, MCM2, MCM4, TYMS, UBE2C
GO:0000070	↳	mitotic sister chromatid segregation	8	5.14E-03	AURKB, CDCA5, DLGAP5, KIFC1, NCAPG, NUSAP1, PRC1, UBE2C
GO:0000083	↳	regulation of transcription involved in G1/S transition of mitotic cell cycle	4	8.23E-03	CDC45, CDT1, ID2, TYMS

GO:0045931	↳ positive regulation of mitotic cell cycle	6	4.42E-02	AURKA, CDC45, CDCA5, DLGAP5, NUSAP1, UBE2C
GO:0007067	↳ mitotic nuclear division	14	5.92E-03	AURKA, AURKB, CCNB2, CDC20, CDCA2, CDCA5, DLGAP5, KIFC1, NCAPG, NUSAP1, OIP5, PRC1, UBE2C, ZC3HC1
GO:0007076	↳ mitotic chromosome condensation	3	1.82E-02	CDCA5, NCAPG, NUSAP1
GO:0044843	↳ cell cycle G1/S phase transition	12	6.28E-04	AURKA, BRD7, CCNE2, CDC45, CDCA5, CDKN3, CDT1, FBXO31, ID2, MCM2, MCM4, TYMS
GO:0000819	↳ sister chromatid segregation	11	1.14E-03	AURKB, CDC20, CDCA5, CENPI, DLGAP5, KIFC1, NCAPG, NUSAP1, PRC1, TOP2A, UBE2C
GO:0051785	↳ positive regulation of nuclear division	5	1.06E-02	AURKA, CDCA5, DLGAP5, NUSAP1, UBE2C
GO:0071103	DNA conformation change	11	5.85E-03	CDC45, CDCA5, CENPI, GINS2, MCM2, MCM4, NCAPG, NUSAP1, OIP5, RBBP4, TOP2A
GO:0032392	↳ DNA geometric change	5	4.71E-02	CDC45, GINS2, MCM2, MCM4, TOP2A
GO:0006323	↳ DNA packaging	8	2.73E-02	CDCA5, CENPI, MCM2, NCAPG, NUSAP1, OIP5, RBBP4, TOP2A
GO:0030261	↳ chromosome condensation	4	1.84E-02	CDCA5, NCAPG, NUSAP1, TOP2A
GO:0032091	negative regulation of protein binding	5	4.54E-02	ACE, AURKA, AURKB, CAMK1, CCL23
GO:0007059	chromosome segregation	14	5.38E-04	AURKB, CDC20, CDCA2, CDCA5, CENPI, DLGAP5, KIFC1, NCAPG, NUSAP1, OIP5, PRC1, PTTG3P, TOP2A, UBE2C
GO:0002251	organ or tissue specific immune response	4	1.75E-02	CAMP, DEFA1, DEFA1B, DEFA3

GO:0031145	anaphase-promoting complex-dependent proteasomal ubiquitin-dependent protein catabolic process	5	4.42E-02	AURKA, AURKB, CDC20, FBXO31, UBE2C
GO:0002443	leukocyte mediated immunity	10	1.92E-02	ACE, C4BPA, CD46, CLC, CTSG, ELANE, FES, GAPT, KDM5D, UNG
GO:0050832	defense response to fungus	7	1.82E-05	CTSG, DEFA1, DEFA1B, DEFA3, DEFA4, ELANE, GNLY
GO:0009620	↳ response to fungus	7	1.54E-04	CTSG, DEFA1, DEFA1B, DEFA3, DEFA4, ELANE, GNLY
GO:0009617	↳ response to bacterium	13	4.85E-02	ARG1, CAMP, CD24, CEBPE, CSF3, CTSG, DEFA1, DEFA1B, DEFA3, DEFA4, ELANE, GNLY, IDO1
GO:0042742	↳ defense response to bacterium	9	2.27E-02	CAMP, CEBPE, CTSG, DEFA1, DEFA1B, DEFA3, DEFA4, ELANE, GNLY
GO:0050830	↳ defense response to Gram-positive bacterium	6	8.23E-03	CAMP, CTSG, DEFA1, DEFA1B, DEFA3, DEFA4
GO:0019730	↳ antimicrobial humoral response	5	8.23E-03	CAMP, DEFA1, DEFA1B, DEFA3, DEFA4
GO:0019731	↳ antibacterial humoral response	5	5.85E-03	CAMP, DEFA1, DEFA1B, DEFA3, DEFA4
GO:0035821	modification of morphology or physiology of other organism	9	3.56E-04	CAMP, CTSG, DEFA1, DEFA1B, DEFA3, DEFA4, ELANE, GNLY, TYMS
GO:0044146	negative regulation of growth of symbiont involved in interaction with host	3	2.09E-02	CAMP, CTSG, ELANE
GO:0044116	↳ growth of symbiont involved in interaction with host	3	3.78E-02	CAMP, CTSG, ELANE
GO:0044110	↳ growth involved in symbiotic interaction	3	3.78E-02	CAMP, CTSG, ELANE

GO:0044144	↳	modulation of growth of symbiont involved in interaction with host	3	2.44E-02	CAMP, CTSG, ELANE
GO:0002227		innate immune response in mucosa	4	5.39E-03	CAMP, DEFA1, DEFA1B, DEFA3
GO:0002385	↳	mucosal immune response	4	1.47E-02	CAMP, DEFA1, DEFA1B, DEFA3
GO:0002526		acute inflammatory response	7	1.75E-02	ADAM8, C4BPA, CD46, ELANE, MRGPRX1, PHB, PTGER3
GO:0002673	↳	regulation of acute inflammatory response	5	2.77E-02	ADAM8, C4BPA, CD46, PHB, PTGER3