Transcriptomic Phenotyping of the Macrophage Response to Filarial Nematode Infection

Graham David Thomas

PhD Thesis

The University of Edinburgh

Institute of Immunology and Infection Research,

Ashworth Laboratories

June 2013

DECLARATION

I hereby declare that the contents of this thesis are entirely my own work, unless otherwise stated.

Graham David Thomas

17 June 2013

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

During my PhD studies I have been lucky enough to receive the help and support of many talented people. It goes without saying that I am immensely grateful to my two primary supervisors, Judith Allen and Mark Blaxter, for their continuous support and input throughout my studies.

I arrived in Ashworth with essentially no working knowledge of either immunology or computational biology, and it is only with the tutorage of fellow PhD students and post-docs that I have learned these. In this regard I especially need to thank Sujai Kumar for all things Perl, R and Unix, and Dominik Ruckerl. Dominik in particular has showed an extraordinary capacity for restraint during my wet lab training.

I also must thank my collaborators, Philip Whitfield and Ben Maskrey, at the University of the Highlands and Islands for their contribution. In Chapter 2 we characterize macrophage-derived eicosanoids. Phil and Ben were instrumental in the both design and execution of this part of my PhD, and put together the panel of panel of eicosanoids we went on to measure. When I say 'we', it was actually Ben who ran the mass spectrometer and determined eicosanoid concentrations during this experiment.

Stephen Jenkins, formerly of the Allen Lab, also deserves a word for help during the *L. sigmodontis* microarray project. Whist I was running about flow sorting macrophages, Steve was busy doing the intracellular staining for Ki67 and YM1 (incidentally Steve also analysed the intracellular FACS data for this experiment).

Finally, I really should thank my wife, Vicky, and the rest of my family for putting up with my absenteeism and ill temper over the past few years. Unfortunately, I cannot see that this situation will resolve anytime soon.

ABSTRACT

Helminths are paraphyletic group of parasitic metazoans that collectively consist of the nematodes, trematodes, cestodes and platyhleminthes. Helminth infections are characterized, amongst other things, by IL-4 and IL-13 production from $CD4^+$ Th2 cells. IL-4 and IL-13 induce alternatively activated macrophages (AAMΦ) via their action on the IL-4Rα receptor subunit. AAMΦ are also implicated in numerous other diseases such as fibrosis, allergy, cancer and diabetes.

STAT6 drives IL-4Rα-dependent transcription in AAMΦ, leading to the induction of many genes including Arg-1, YM1 and RELMα. IL-4 also induces macrophage proliferation, a phenotype observed *in vivo*, but not *in vitro*. *In vitro* studies have shown that IL-4 up-regulates macrophage mitochondrial metabolism in a process dependent upon the ligand-dependent transcription factor PPARγ (peroxisome proliferator activator receptor gamma). In spite of the far-reaching effects of IL-4 on macrophage biology and physiology, we are unaware of the true role(s) of AAMΦ during helminth infection.

In this thesis I present two transcriptomic analyses of macrophage responses to filarial nematode infection. In the first case we define IL-4Rα-dependent components of the macrophage response to nematode infection. Using RNA-Seq, WT and IL-4Rα$^{-/-}$ mice implanted with the parasitic nematode *Brugia malayi* are compared against inflammatory-like thioglycollate-elicited macrophages. This has allowed us to infer an anti-inflammatory phenotype for nematode-elicited macrophages (NeMΦ) *in vivo*. We also identify an unanticipated role for NeMΦ in the production of complement components in response to *B. malayi* challenge. During this project I develop a method to map transcription start site (TSS) usage with RNA-Seq data, this algorithm has been developed into the software 'TSS-Predictor' and is presented in chapter 3. Using TSS-Predictor we define TSS usage in our

RNA-Seq dataset, and a subsequent analysis of transcription factor binding sites in AAMΦ-associated promoters confirms a role for PPARs in AAMΦ *in vivo*. Furthermore, we find that PPARδ, but not PPARγ is expressed during helminth infection. A targeted lipid analysis of macrophage-derived eicosanoids identified prostacyclin (PGI2) as a candidate ligand to mediate PPARδ-dependent transcription in NeMΦ during helminth infection *in vivo*.

Finally, we integrate flow cytometry and a microarray gene expression profiling in a time series of macrophages elicited in response to the parasitic nematode *Litomosoides sigmodontis*. We capitalize on the discordant kinetics of alternative activation and proliferation in this model to define gene expression signatures associated with proliferation and AAMΦ. Based upon this analysis we present a working model for IL-4-dependent macrophage proliferation *in vivo* and characterize candidate receptors mediating this process. We also demonstrate *in vivo* that AAMΦ increase mitochondrial mass in response to *L. sigmodontis* infection.

1 INTRODUCTION

1.1 OVERVIEW OF THESIS

Macrophages are workhorses of the innate immune system and serve as both sentinels to detect disease, mediate inflammatory responses via the production of cytokines, and also clear invading pathogens via phagocytosis (Taylor et al., 2005). It is clear that macrophages play a role in the response to helminth infection, a paraphyletic grouping of parasitic nematodes, trematodes and cestodes. In response to helminth infections, activated T-cells produce IL-4 which stimulates macrophages to undergo alternative activation (Kopf et al., 1993; Stein et al., 1992). Alternatively activated macrophages (AAMΦ) are associated with the induction of a distinct gene expression repertoire, and are characterized by up-regulated expression of genes including Arg-1, YM1 and RELMα (Loke et al., 2002). IL-4 has also recently been described to induce macrophage proliferate *in vivo*, and helminth infection is associated with a large expansion of the resident tissue macrophage population (Jenkins et al., 2011).

Although macrophages proliferate and undergo unique changes in gene expression in response to IL-4 *in vivo*, we are unaware of the role(s) of AAMΦ in helminth infection. Furthermore, IL-4 does not induce macrophage proliferation *in vitro*, and the molecular mechanisms driving IL-4-dependent macrophage proliferation remain uncharacterized. In this thesis I apply transcriptomics approaches to address these issues. Firstly, we perform an RNA-Seq analysis of AAMΦ elicited with the parasitic nematode *Brugia malayi* in both WT and IL-4Rα$^{-/-}$ mice. This study provided us with an unprecedented depth and quality of transcriptomic information for *in vivo*-polarized AAMΦ, and has allowed us to define IL-4Rα-dependent gene expression profiles, metabolic pathways, and transcriptional regulatory mechanisms underpinning AAMΦ *in vivo*. Secondly, we integrate flow cytometry and microarray data in a time course analysis of the helminth infection *Litomosoides*

sigmodontis (Allen et al., 2008). The experiment that we present has allowed us to delineate transcriptional repertoires associated with macrophage proliferation and alternative activation by capitalizing on the discordant kinetics of these processes during the response to *L. sigmodontis* infection. I present hypotheses regarding AAMΦ function and regulation during filarial nematode infections, and present a working model for IL-4-dependent macrophage proliferation *in vivo*. My introduction begins with an overview of the innate immune system and macrophage biology, and is followed by an overview of the transcriptomic technologies and data analysis techniques used in this thesis.

1.2 Macrophages and the innate immune system

1.2.1 The Innate Immune System

A challenge faced by all animals is the requirement for protection from an invading parasitic challenge. The innate immune system evolved to identify and eliminate offending foreign materials. The mammalian innate immune system is composed of multiple overlapping cellular and humoral components that interact to oversee the removal or control of a pathogen and the restoration of homeostasis. Any immune response has to be both specific and appropriate, while at the same time minimizing damage to surrounding host tissue. Broadly speaking, the response to all pathogens involves the same set of processes. Non-self-antigen is first recognised by tissue-resident immune cells or stromal cells based upon the presence of pathogen associated molecular patterns (PAMPS). PAMPS are generic chemical moieties that are non-self but are reliably associated with unwanted intruders (Janeway, 1989). These include, for example, lipopolysaccharide (LPS), a major constituent of the surface of gram-negative bacteria. PAMPS are detected by resident tissue immune sentinels and stromal cells by pattern recognition receptors (PRRs). PRRs are ubiquitously

represented around the body and include, but are not restricted to, receptors of the toll like receptor (TLR), nod-like receptor (NOD) and C-type lectin families (Cambi and Figdor, 2003; Kanneganti et al., 2007; Kawai and Akira, 2010).

When accompanied by a 'danger signal' PRR ligation leads to the initiation of an inflammatory response. The type of non-self detected, and both the nature and severity of the accompanying 'danger signal' provides crucial information to prompt an appropriate response (Gallucci and Matzinger, 2001). An appropriate response is mediated via the production of a cocktail of chemokines, cytokines and small molecule inflammatory mediators. These 'immune effectors' shape the subsequent cellular recruitment and activation profile. Chemotactic cytokines, chemokines, operate via endocrine and paracrine mechanisms to bring about the recruitment of leukocytes to sites of inflammation. Effector cell subsets such as neutrophils, eosinophils and monocytes are mobilized into the blood from distant sites such as the bone marrow (BM) and spleen via the interaction between circulating chemokines and cytokines and their cell surface receptors (Moser et al., 2004).

Effector cells are subsequently recruited to sites of inflammation through paracrine functions of immune effector molecules. This inflammatory cascade leads to the modulation of integrin and adhesin expression on vascular epithelial cells, facilitating the recruitment of blood-borne immune cells to an afflicted area (Ley et al., 2007). At the infection site cytokines dictate the polarization and subsequent activity of both recruited and resident immune cells, which ordinarily leads to the coordinated destruction and clearance of invading pathogens and/or recruitment of successor waves of inflammatory components. The activities of the innate immune system occur in conjunction with an adaptive arm of immunity, which is initiated by the innate immune system and serves to direct innate immune function. The adaptive immune system is educated to recognise

specific pathogenic antigens via prior exposure, and accelerates the response to secondary exposure via the activity of helper T-cell, and antibody-mediated B-cell responses.

1.2.2 Complement in innate immunity

The cellular players of innate immunity system operate alongside humoral components, most notably the complement pathway. The complement system is a highly conserved pathway of proteases that circulate in the blood in their inactive forms. The central component of the complement cascade C3 is cleaved and activated by upstream proteases belonging to either the classical, alternative or mannose binding lectin (MBL) complement pathways (Gasque, 2004). C3 is cleaved by a 'C3 convertase' protease to release the C3b sub-unit, C3b binds covalently to the invading foreign body, and initiates a cascade of proteolytic cleavage events that culminate in the formation of the 'membrane attack complex' (MAC), reviewed in (Walport, 2001). This pore-forming protein assemblage is microbicidal, and MAC formation is accompanied by the production of multiple opsonins that promote phagocytosis, the primary function of the complement system. In addition to the opsonisation-promoting intermediaries of the MAC, peptide cleavage products of the complement pathway also act as cytokines and chemokines in their own right (Carroll, 2004).

Over 30 proteins are involved in the activation and regulation of the complement cascade, and a comprehensive overview is beyond the scope of this introduction, for the inclined I would recommend the following two reviews (Carroll, 2004; Walport, 2001). Three pathways lead to the formation of the C3 convertase complexes to activate the complement cascade. The first to be discovered (but last to arise in evolutionary terms) is the classical, C1, pathway. This is induced by the recognition of antigen-bound antibody

complexes by a complex of C1q C1r and C1s proteins. Secondly, the mannose-binding lectin (MBL) pathway is involved in the identification of polysaccharides and is activated by the interaction of carbohydrate-engaged mannose-binding lectin (*Mbl1* or *Mbl2*) or ficolin (*Fcna* or *Fcnb*) proteins (Gasque, 2004). Lastly, the alternative pathway, is constitutively active in blood plasma and serves primarily as a positive feedback loop to amplify C3 conversion initiated by either the C1 or MBL pathways.

1.2.3 MACROPHAGES

Macrophages are the workhorses of the innate immune system. These professional phagocytes are found in all body tissues and play crucial roles in both the maintenance of homeostasis and in the initiation and resolution of inflammatory responses. They abundantly express PRRs, including TLRs and, thus are poised to detect invading pathogens and form a first line of defence against disease (Taylor et al., 2005). They also express complement receptors and Fc-gamma receptors (FcγR) to facilitate the uptake of complement- and antibody-bound particles respectively (Aderem and Underhill, 1999).

Upon stimulation by invading microbial pathogens, macrophages phagocytose and destroy bacteria via the production of nitric oxide (NO). In parallel they produce inflammatory cytokines and chemokines such as TNFα, IL6, IL12, CCL2 and CCL5 and up-regulate expression of major histocompatibility complex (MHC) genes (Ogawa et al., 2005). Thus macrophages detect, engulf and destroy invading pathogens whilst acting to instruct and direct both adaptive and innate immune responses. As professional phagocytes macrophages are much more than innate immune cells. They are essential for the clearance of apoptotic cells and instruct regeneration following injury (Mantovani et al., 2013;

Mukundan et al., 2009) and this endows macrophages with an essential role in the development and in the maintenance of homeostasis.

Macrophages are myeloid-derived cells and until recently most tissue macrophages (with the exception of microglia and Langerhans cells) were believed to originate solely from bone marrow-derived monocytes. A seminal publication from the Geissmann laboratory has recently overturned this dogma, showing that populations of tissue-resident macrophages arise from embryonic yolk sac precursors (Schulz et al., 2012). All macrophages are dependent on either colony stimulating factor 1 (CSF-1, *Csf1*, (Cecchini et al., 1994)) or IL-34 for their development (Wang et al., 2012). Both CSF-1 and IL-34 act through the CSF-1 receptor (CD115, *Csf1r*) to drive the proliferation, maturation and survival of macrophages and their precursors (Hume and MacDonald, 2012). Typically mouse macrophages can be defined and easily distinguished by flow cytometry (FACS) through the expression of the surface markers CD11b, CD115 and F4/80, however, surface marker expression profiles do change depending upon the tissue analysed (Gautier et al., 2012). In many tissues, including the serous cavities, which are the focus of this thesis, macrophages can be identified using FACS on the basis of surface F4/80 expression in conjunction with scatter profiles to determine cell size.

In vivo macrophage populations inhabiting different tissues show a tremendous degree of diversity both transcriptionally and functionally (Gautier et al., 2012; Qian and Pollard, 2010). This likely reflects a combination of differences in the developmental histories and micro-environmental cues received by different macrophage populations. The extent to which macrophage heterogeneity is explained by the irreversible terminal differentiation of 'mature' CSF-1-derived macrophages within tissues, or reflects a reversible phenotypic plasticity remains unresolved. Understanding the diversity of macrophage populations *in*

vivo, and their developmental and functional relationships represents one of the major challenges in myeloid cell biology today (Hashimoto et al., 2011).

1.2.4 Helminth infection and Th2 immunity

Helminths are a polyphyletic group of metazoan parasites that includes members of the phyla Nematoda and Platyhelminthes (Classes Trematoda and Cestoda). Although polyphyletic, 'helminth' provides a convenient nomenclature as these parasites elicit broadly similar immune responses in their mammalian hosts (Allen and Maizels, 2011). It is estimated that one third of the global population are infected by helminths (Hotez et al., 2008) with the most common being the gastro-intestinal infections of hookworm and roundworm (Hall et al., 2008). Helminths are also the causative agents of multiple diseases including elephantiasis (lymphatic filariasis), river blindness (onchocerciasis) and schistosomiasis (Chitsulo et al., 2000; Davies, 1994; Evans et al., 1993). A notable feature of helminth parasites are their lifecycles, which typically involve a juvenile developmental stage occurring within an intermediate non-vertebrate host (Allen and Maizels, 2011). This intermediate host serves as a vector for the transmission of the parasite between mammalian hosts.

Numerous metazoan parasites of rodents have been developed that collectively model the range of helminth diseases afflicting humans, and other hosts of veterinary and agricultural importance. Mouse models for gut-dwelling nematode parasites include *Heligmosomoides polygyrus* and *Nippostrongylus braziliensis* (Maizels et al., 2009). The human parasitic liver fluke *Schistosoma mansoni* readily infects mice, and is used in rodents as a model for itself (Pearce and MacDonald, 2002). Filarial nematodes, the focus of attention in the Allen laboratory and within this thesis, are also modelled by a human parasite, *Brugia malayi*

(McSorley and Maizels, 2012). The infective L3 stage of *B. malayi*, the stage transmitted from the arthropod vector to the mammalian host, does not readily transit from L3 to adult worm within mice. However the Mongolian Jird, *Meriones unguiculatus*, is able to serve as an intermediary for this process. Adult *B. malayi* are extracted from the peritoneal cavity of the Jird, and are subsequently implanted into the peritoneal cavity of mice. This is the *B. malayi* model of lymphatic filariasis and is an effective tool to study immune responses to the adult stage of infection as well as their accompanying microfilarial output (Lawrence et al., 1994).

A more recently developed model for lymphatic filariasis and Onchocerciasis is *Litomosoides sigmodontis*. This natural parasite of the cotton rat, *Sigmodon hispidius*, has the advantage of developing from the infective L3 to adult worm within mice (Allen et al., 2008). Furthermore, depending on the host mouse strain, *L. sigmodontis* may undergo a complete life cycle to produce microfilariae. The transmission stage is reached in the susceptible BALB/c strain, but adult *L. sigmodontis* are killed prior to the onset of sexual maturity (~ day 50 P.I) in the resistant C57BL/6 strain (Le Goff et al., 2002). Thus, *L. sigmodontis* serves as a model to study both the initiation of immune responses to filarial nematode infections, as well as immune mechanisms of resistance and susceptibility.

Helminths have coevolved with their hosts, show remarkable host specificity and in some cases they can live within their hosts for many years (Plotkin et al., 2008). This is accomplished via the evasion and active suppression of the host immune system. Helminths are masters of both these processes (McSorley and Maizels, 2012). Helminth immunosuppression has evolved to counteract protective host responses (which are typically Th2 in nature) and is characterized by the emergence of mast cells, eosinophils, basophils and alternatively activated macrophages (AAMΦ, see below). Th2 type responses

are defined by the production of the cytokines IL-3, IL-4, IL-5, IL-9, IL-10 and IL-13 which drive the above innate-type responses and require the activity of CD4+ Th2 cells (Allen and Maizels, 2011).

Th1 response-inducing microbes can easily be engulfed by phagocytes and destroyed via the production of proinflammatory intermediates. However, this is not an appropriate strategy to deal with the threat of invading metazoan parasites, although Th1 inflammatory responses may effectively kill helminths, these parasites are far too large to be engulfed by phagocytes. It is proposed that collateral tissue damage would bear too greater cost on host fitness for Th1-type anti-helminthic immunity to represent an evolutionarily stable strategy (Allen and Wynn, 2011). Rather Th2 immunity, in addition to protecting against worm infection, is proposed to facilitate tissue repair, indeed, IL-4 and AAMΦ facilitate effective wound healing following colonic punch biopsy (Seno et al., 2009).

The robust induction of Th2 immunity in helminth infection requires the activity of $CD4^+$ T-cells, however the initial 'danger signals' associated with infection, and required to induce the adaptive response, remain elusive. Helminths themselves may be adorned with PAMPS which betray their identity to their hosts, and the identification and characterization of immunogenic helminth products is underway (Hewitson et al., 2008). It is also proposed that damage itself may serve as the danger signal to initiate the innate immune reaction that initiates the Th2 response. Indeed IL-33, a cytokine released by dying epithelial cells, augments Th2 responses, in particular alternative macrophage activation (Kurowska-Stolarska et al., 2009).

1.2.5 AAMΦ

The Th2 cytokines IL-4 and IL-13 polarize macrophages towards a state of 'alternative activation' (AAMΦ). In contrast to classically activated macrophages (CAMΦ) induced by microbial stimulation, AAMΦ are potently induced during helminth infections and allergic inflammation, but may also be found in diverse pathological and physiological settings such as during cancer and in the adipose tissue of lean individuals (Allen and Maizels, 2011; Gordon and Martinez, 2010; Movahedi et al., 2010; Odegaard and Chawla, 2011).

The IL-4Rα chain serves as a common receptor subunit for both IL-4 and IL-13 (Schnyder et al., 1996), and is absolutely required for the induction of AAMΦ. In this thesis we follow Siamon Gordon's definition of alternative activation, which is defined as the IL-4/IL-13-dependent component of macrophage activation (Gordon, 2003). Signalling via either of these cytokines promotes phosphorylation of the IL-4Rα subunit, leading to the activation of the transcription factor STAT6 (Stütz et al., 2003). STAT6 drives alternative activation leading to the up-regulation of numerous genes including ARG1 (arginase 1, *Arg1*), YM1 (*Chi3l3*, chitinase 3 like 3) and RELMα (*Retnla*, resistin like alpha)(Stütz et al., 2003; Welch et al., 2002). YM1 and RELMα were originally described as induced in AAMΦ in response to helminth infection (Loke et al., 2002), and both of these genes serve as excellent markers to define *in vivo*-derived AAMΦ using FACS (Jenkins et al., 2011).

IL-4 has recently been shown to act as a potent macrophage mitogen *in vivo* (Jenkins et al., 2011). In response to IL-4, tissue resident macrophages proliferate and accumulate in the absence of a significant influx of inflammatory macrophages from the bloodstream. This finding has forced a rethink in our understanding of hematopoietic system development, showing that mature tissue-resident macrophages can proliferate in response to a polarizing stimulus. Furthermore, recent unpublished data from the Allen laboratory have

shown that this process is independent of the canonical macrophage mitogen CSF-1 (S Jenkins, unpublished data). Importantly IL-4 does not induce macrophage proliferation *in vitro*, and so the precise molecular mechanisms governing IL-4-dependent macrophage proliferation remain uncharacterized. This gap in our understanding poses a significant unanswered question in macrophage biology and in chapter 7 I develop working model for IL-4α-dependent macrophage proliferation based upon microarray time series data.

At the cellular and organismal level AAMΦ are widely considered to be anti-inflammatory (Gordon and Martinez, 2010). IL-4-stimulated macrophages show an impaired ability to produce pro-inflammatory cytokines (Loke et al., 2002), and the induction of Arg-1 by AAMΦ forms the basis of a functional antagonism between the CAMΦ and AAMΦ phenotypes (Modolell et al., 1995). ARG1 competes with the CAMΦ-induced gene iNOS (inducible nitric oxide synthase) for a common substrate, arginine, which is required by CAMΦ to produce nitrosative intermediaries essential for microbial killing. Cell non-autonomous anti-inflammatory mechanisms of AAMΦ include their ability to suppress effector T-cell proliferation (Liu et al., 2003; Loke et al., 2000a), and promotion of regulatory T-cell profiles (Broadhurst et al., 2012).

AAMΦ also contribute to pathology and are a major constituent of inflammation in multiple models of allergic inflammation (Martinez et al., 2009). AAMΦ-like tumour-associated macrophages (TAM) also correlate with a poor prognosis in multiple cancer models (Ghassabeh et al., 2006; Mantovani et al., 2002). TAM contribute to cancer progression both directly and indirectly by promoting tumour growth, and vascularization as well as aiding metastasis (Qian and Pollard, 2010). TAM however encompass a range of macrophage phenotypes induced by multiple potential stimuli including IL-4/IL-13, IL-10 or glucocorticoids, and, as such, TAM possess a more diverse range of activation states than

the strictly IL-4-dependent AAMΦ (Mantovani et al., 2004). However, in spite of this it is generally accepted that TAMs are anti-inflammatory, and that the range of TAM phenotypes are broadly AAMΦ-like (Van Ginderachter et al., 2006). It will be of interest to determine whether the apparent convergence in function of AAMΦ-like TAM are driven by the processes regulated downstream of the initial polarization stimulus. One such downstream process, and a major proposed physiological role of AAMΦ, is the promotion of wound healing and the resolution of inflammation (Allen and Wynn, 2011). AAMΦ compete with fibroblasts and T-cells during inflammatory responses for extracellular arginine (Pesce et al., 2009). This serves to limit fibrosis by restricting L-proline availability, a catabolite of arginine and prerequisite for collagen synthesis. It is the anti-inflammatory properties of AAMΦ that are postulated to couple with their pro-angiogenic and tissue remodelling abilities to facilitate wound repair and tissue regeneration (Martinez et al., 2009).

Macrophages display a broad range of activation states and have the potential to exert both beneficial and detrimental effects on host physiology. Furthermore, the complexity and range of macrophage activation phenotypes observed *in vivo* is far greater than can be recapitulated *in vitro*. It is thus essential to perform studies of macrophage function *in vivo* to understand their roles during homeostasis and in the response to infection. Although AAMΦ and key marker genes have been characterized in multiple helminth infection settings (Jenkins and Allen, 2010), the role(s) of these cells in helminth infection remains elusive. Defining these roles in helminth infection is the central research theme of the Allen laboratory, and uncovering new genes and molecular pathways associated with AAMΦ during helminth infection is a key aim in this thesis.

1.2.6 MACROPHAGE FUNCTION AND METABOLISM

There is an increasing appreciation that immune processes in general, and macrophages in particular, are intimately integrated with physiological programmes of systemic metabolism (Chawla et al., 2011; Liu et al., 2012). The polar regimes of macrophage activation, CAMΦ and AAMΦ, are associated with obese and lean adipose tissue respectively (Lumeng et al., 2007). Furthermore, these macrophage phenotypes play causative roles in the induction and alleviation of insulin resistance, the hallmark of type-2 diabetes. CAMΦ utilize glycolysis to sustain their activation programme (Tannahill et al., 2013), and CAMΦ derived pro-inflammatory cytokines induce systemic insulin resistance (Xu et al., 2003). This occurs via the activity of pro-inflammatory cytokines on the major metabolic tissues; adipose, muscle and liver (Olefsky and Glass, 2010). Ultimately this serves to increase circulating glucose concentrations, providing a resource for energetically demanding T-helper cell proliferation and innate immune activity. This sets up a positive feedback loop whereby inflammation facilitates the dysregulation of glucose metabolism, driving inflammation. It is chronic inflammation that underpins the progression of diseases associated with the metabolic syndrome including cancer, atherosclerosis and type-2 diabetes (Biswas and Mantovani, 2010; Reilly and Rader, 2003).

AAMΦ exert opposing metabolic effects to CAMΦ have been shown to improve insulin sensitivity by controlling the activities of metabolic tissues (Biswas and Mantovani, 2012). Notably, *in vitro* studies have established a requirement for the oxidative phosphorylation of lipids, rather than glycolysis, in AAMΦ as a means to meet their energetic requirements (Vats et al., 2006). The biochemistry of oxidative metabolism is performed within the mitochondria, and macrophages increase mitochondrial mass in response to IL-4 stimulation *in vitro* (Odegaard et al., 2007). IL-4 stimulates AAMΦ to undergo mitochondrial biogenesis via the activity of the ligand-dependent transcription factors PPARγ and PPARδ,

members of nuclear receptor transcription factor superfamily, and their co-activator protein PCG-1b (*Ppargc1b*, (Vats et al., 2006). Emerging evidence from a variety of *in vitro* studies is suggestive of an intimate cell autonomous relationship between alternative activation and oxidative metabolism. IL-4 leads to the production of PPAR ligands and also induces mRNA expression of PPARγ and PGC-1b in a STAT6-dependent manner (Huang et al., 1999a; Odegaard et al., 2007). In turn, the PPAR response element (PPRE) transcription factor binding site is necessary for IL-4-dependent induction of Arg-1 expression and PPAR ligation suppresses inflammatory gene expression in macrophages (Odegaard et al., 2007; Ogawa et al., 2005). The IL-4-dependent expression of key AAMΦ genes is also disrupted in PPARγ$^{-/-}$ bone-marrow derived macrophages (Odegaard et al., 2007). The relevance of these findings to AAMΦ in the context of helminth infection remains to be determined.

In this thesis I set out to improve our understanding of the role of AAMΦ in helminth infection, and in order to do this I have chosen to apply the transcriptomic approaches discussed below. These approaches have allowed us to interrogate genome-wide gene expression profiles purified macrophage populations, and have been used to address the following outstanding question which form the aims of this thesis: -

What IL-4Rα-dependent molecular pathways and components of macrophage activation are elicited in response to helminth infection?

Do *in vivo* elicited AAMΦ up-regulate mitochondrial metabolism during helminth infection?

What is the relationship between IL-4Rα-dependent alternative activation, and IL-4-dependent macrophage proliferation *in vivo*?

1.3 Transcriptomic analysis of gene expression

1.3.1 Expression profiling as a means of cellular phenotyping

mRNA transcripts within a cell reflect the intermediate state between the genomic information and the final functional protein product. The mouse genome contains 23,158 protein coding genes (Ensembl v70, accessed 24th February 2013), approximately 45% of which are expressed in each cell at any one time (Ramsköld et al., 2009). Roughly 8,000 genes, 75% of all expressed genes, are expressed by every cell (Ramsköld et al., 2009) yet the dynamic range of transcript expression varies by greater than 5 orders of magnitude (Wang et al., 2009). The quantitative assessment of the entire population of mRNA molecules within a cell, or population of cells, thus provides a proxy for functional state. The identification of differences in genome-wide expression profiles between treatments and/or conditions thus provides a means to functionally characterize phenotypic changes arising from a treatment.

Several methods have been developed to facilitate the detection and quantification of mRNA in cells and tissues. The most mature of these technologies is microarray expression profiling, several variants of which exist. However, more recently, advances in second generation sequencing technology have led to the development of direct RNA sequencing (RNA-Seq) which has begun to supersede microarray technology. Both of these technologies are used in this thesis and a specific overview of the methods used to generate workable data from each follows. Although microarrays and RNA-Seq are different technologies that require analysis with different statistical approaches, the fundamental principles of data analysis are identical. Both require a library preparation step to convert RNA into a form that can be measured by the technology. Next, estimates of gene expression are obtained, although this may be a multi-step process. Finally, after

normalization of the primary data, differential expression analysis is used to identify differentially expressed genes.

1.3.2 Illumina BeadArray technology

Microarray technology is the best-established and most mature method for the quantification and analysis of transcriptomic data. Microarrays use a complementary hybridization approach to allow the simultaneous detection and quantification of thousands of genes (Duggan et al., 1999). Many different solutions to this hybridization approach exist and I made use of Illumina BeadArray technology in this thesis (Kuhn et al., 2004).

The BeadArray works by randomly placing beads coated with oligonucleotide sequences complementary to portions of mRNA transcripts in the mouse genome into wells on a silicon array. Each bead contains an individual oligonucleotide sequence (probe) present at approximately 700,000 copies per bead. This provides roughly 3 orders of magnitude dynamic range when probes are hybridized to samples. Attached to each probe is a shorter artificial sequence used to locate the randomly positioned bead by a process of sequential hybridization (Gunderson et al., 2004).

Libraries for analysis on an array are prepared using a PCR based approach. mRNA is converted into double-stranded cDNA that is then used as a template for *in vitro* transcription to produce many copies of biotinylated copy-RNA (cRNA). cRNA is hybridized to the Illumina arrays and signal is detected by developing the arrays with streptavidin-conjugated Cy3 dye (Kuhn et al., 2004). The Mouse WG-6 v2.0 (Ref6) array platform contains comprehensive transcript annotations for the mouse with 45,281 transcripts

represented. Each oligonucleotide probe is present on average 30 times in each array platform. Given the random positioning of each probe within the array, this ensures that at least 5 copies of each probe are present (Kuhn et al., 2004). This provides significant advantages over 'printed' array technologies where probes are located at pre-defined coordinates on a chip because systematic artefacts arising from the influence of neighbouring probes are eliminated with this approach. Furthermore, the redundancy in probe existence on each array improves the precision of expression estimation.

Data analysis for microarray experiments requires the normalization of observed signal intensities across all samples within an experiment and their transformation to remove any relationship between signal intensity and variance (enforcing homoscedasticity). This process allows the application of traditional parametric-based statistical tools and is also an important part of RNA-Seq data analysis, discussed later.

1.3.3 Normalization and Differential Expression analysis for Illumina BeadArrays

While many methods for microarray data analysis exist, standard protocols have emerged to deal with each array technology (Ritchie et al., 2011), these provide the flexibility to ensure an appropriate analysis pipeline for a given experimental design. Data processing for Illumina BeadArrays is a two-step process and takes into account the redundancy in probe numbers on each array chip. The robust spline normalization method (RSN) (Lin et al., 2008) calculates the average expression of each probe in an array after the removal of outliers. Following this, signal intensities are normalized both within and between arrays simultaneously using the variance-stabilizing normalization (VSN) method (Huber et al., 2002a). This normalization procedure ensures that the signal intensity distribution between

arrays in an experiment is roughly equal, and that the variance in expression for all expressed genes is roughly the same across the entire distribution of gene expression levels.

Normalized microarray data of the kind described above can be analysed using parametric methods such as linear regression analysis and t-tests. These are the most frequently used approaches for microarray analysis, and by far the most widely used software package for this is limma (Smyth, 2004) implemented in the Bioconductor environment (Gentleman et al., 2004). The limma package supports generalized linear modelling (GLM) for microarray data in an easy to understand and accessible manner. The flexibility allowed by a GLM approach is that it can incorporate anything from the most trivial to complicated experimental designs. Additionally, limma supports multiple methods for the correction of multiple testing, an important consideration for the elimination of false-positive results in high-throughput data analysis. For both microarray and RNA-Seq analyses presented in this thesis, the Benjamini-Hochberg method to correct for multiple testing is used (Benjamini and Hochberg, 1995). This method controls the type 1 false discovery rate, the likelihood a gene is erroneously classified as differentially expressed, and is typically the default method used by bioinformaticians in this kind of analysis (Dudoit et al., 2003).

1.3.4 Illumina sequencing technology & RNA-Seq

An increasingly preferred method for genome-wide expression profiling is direct highly parallelized sequencing of RNA (RNA-Seq). This technological advancement has been made possible over the past five years with the development of second-generation sequencing technologies, in particular Illumina sequencing (Mardis, 2008). Illumina technology uses a sequencing-by-synthesis approach and allows for the massively parallel sequencing of short

reads of DNA. A brief introduction to the method is given below (for a detailed overview of the technology see (Mardis, 2008)).

Illumina sequencing begins with an adapter-modified DNA library containing size-selected fragments of DNA with specific sequencing adapters ligated to the ends of each fragment. The fragments of this library are annealed to a glass flow cell containing oligonucleotides complementary to the sequencing adapters. Once annealed, the individual molecules of DNA are amplified to form the clusters that are sequenced. Hundreds of thousands of clusters are formed on each slide, each of which is sequenced in parallel. This is achieved by sequentially washing the flow cell with modified oligonucleotides. Each base is conjugated to one of four fluorescent dyes, and contains a 3'-OH group to ensure only one base can be added to a cluster at a time. The base incorporated into each cluster is identified by the fluorescence signal measured after each incorporation cycle, following each incorporation step the dye and blocking 3'-OH group are chemically washed off prior to the next cycle. Sophisticated optics ensures that clusters are reproducibly identified over sequential cycles.

Over the course of this PhD project (2009-2013) the specifications of Illumina sequencing have improved dramatically, both in terms of throughput and sequence read length. We performed Illumina sequencing during project between 2009 and 2010. At this time, a single lane of sequencing data generated on an Illumina GAIIx produced 10-15 million 50 base paired-end reads per sequencing run. Illumina sequencing can be applied to gene profiling. Sequencing libraries are produced from randomly sheared purified mRNA obtained by poly-A selection. A size selected (~200bp) fragment of sequencing adapter-ligated cDNA is amplified and sequenced using Illumina technology as described above.

The short reads during an RNA-Seq experiment represent a random sub sample from the entire population of fragmented mRNA molecules present in the library. Reads are mapped to the reference genome using specialized tools designed to map millions of reads with high fidelity extremely quickly using a Burrows-Wheeler transformation (Langmead et al., 2009). The number of reads that map to a gene reflect the steady state expression level of the gene's constituent transcripts in the parent population. Expression levels are thus quantified by counting reads within exons of known genes based on available annotations.

RNA-Seq has several advantages over microarray technology. Although minor biases in RNA-Seq library generation arise from the differential binding of random hexamer primers during cDNA synthesis (Hansen et al., 2010), RNA-Seq has substantially higher reproducibility and lower and inter-laboratory variability than microarray technology ('t Hoen et al., 2008). Dynamic range is limited only by the depth of sequencing, with a single lane typically providing 5-6 orders of magnitude. This is two orders of magnitude greater than microarray technology and is enough to quantify all biologically relevant transcripts within a homogenous population of cells (Marioni et al., 2008). Because reads map to the genome it is possible to both detect new transcripts and novel splice junctions.

1.3.5 RNA-Seq read mapping and differential expression analysis

Dedicated read mapping software for RNA-Seq, TopHat (Trapnell et al., 2009) has been produced. TopHat maps reads quickly with the aid of the Burrows-Wheeler based algorithms while at the same time supporting mapping at splice junctions and searching for *de novo* splice sites. There are considerations for accurate gene expression quantification in RNA-Seq experiments. Sequences within mammalian genomes are repetitive as many genes

contain paralogues or conserved signalling domains. This means that not all 50-100 base stretches of DNA are unique within genes and that RNA-Seq reads may map equally well to more than a single location on the genome. These so called multi-mapping reads must be discounted prior to gene expression estimation as it is not possible to know which gene these sequences arose from, this ultimately leads to incorrect expression estimation.

RNA-Seq data require normalization prior to the identification of differentially expressed genes. The principal inter-library adjustment required for RNA-Seq data analysis is for sequencing depth. To do this DESeq, the dedicated RNA-Seq analysis program used in this thesis, calculates an effective library size for each sequenced lane based on the median read count per gene (Anders and Huber, 2010). This effective library size is used to adjust for the differences in sequencing depth.

RNA-Seq data are fundamentally different to microarray data and require handling with statistical models that make appropriate distributional assumptions. Owing to the count-based nature of RNA-Seq data, gene expression levels are naively expected to follow a Poisson distribution (Marioni et al., 2008). However it was quickly realised that Poisson statistics capture only the count-based component of RNA-Seq data and fails to account for biological variability (Robinson et al., 2010). DESeq and other RNA-Seq analysis pipelines (Robinson, McCarthy, and Smyth 2010) utilize negative binomial distributions, this model is an over-dispersed Poisson distribution allowing for the incorporation of 'extra variance' and biological variability itself is estimated from the data and incorporated into the statistical procedure. The early implementations of DESeq that are utilized in this thesis relied solely on pairwise comparisons however more recent versions support multi-factorial experimental designs.

1.3.6 Pathway analysis and transcriptional regulation

The usual purpose of a genome-wide expression profiling experiment is to generate testable hypotheses for functional follow-up experiments. Thoughtful experimental design is the most valuable consideration to aid the identification of the contrasts of interest. Once significance values have been assigned to a list of genes multiple approaches may be used to gain biological insight.

The most intuitive approach to analysing transcriptomic data is to look for changes in the expression of genes, or gene families, whose change in expression will yield directly interpretable information. For example in this thesis (discussed below) we explicitly search for changes in the expression profiles of chemokines, cytokines and their receptors. Genes belonging to families, or with a particular classification, can be identified using Gene Ontology (GO) terms (Ashburner et al., 2000). Genes are assigned GO terms belonging to one of three ontologies, defining the 'molecular function', biological process' and 'cellular component' to which a gene belongs. This provides easy access to gene family and biological pathway information. The grouping of genes into pre-defined biological pathways 'gene sets' is the starting point for Gene Set Enrichment Analysis (GSEA) and over-representation analysis. The philosophy underpinning both GSEA and over-representation analysis is that the identification of co-ordinately regulated sets of genes with a common function provides insight into cellular processes occurring in response to treatment.

While both approaches are conceptually similar, they rely on different statistical methods, and provide different qualities of data. GSEA is based on a ranked list of expressed genes and identifies gene sets that are biased towards the extreme ends of the list (i.e. gene sets that are correlated either positively or negatively with a phenotype). An enrichment score is calculated for each gene set and a P-value assigned using a permutation-based method

(Subramanian et al., 2005). Over-representation analysis on the other hand takes a binary approach to identify gene sets whose members are more present than would be expected by chance in a list of differentially expressed genes, given a list of background genes. P-values in this approach are assigned using an exact approach based on the hypergeometric distribution (Cordero et al., 2007). In addition to the Gene Ontology several databases contain curated biological pathway information. Amongst the most widely used are Reactome (Vastrik et al., 2007) and KEGG (Kanehisa and Goto, 2000) both of which contain detail cellular metabolic and signalling pathways.

In a conceptually similar manner to GSEA, transcriptomic data can be combined with genome sequence information to provide insight into mechanisms of transcriptional regulation. These methods correlate differential gene expression with presence or absence patterns of transcription factor binding sites (TFBS) in gene regulatory regions. TFBS libraries such as JASPAR (Bryne et al., 2008) define the relative affinities of transcription factors to cognate DNA sequences, alternatively TF binding occupancy may be determined directly using chromatin immunoprecipitation followed by high throughput sequencing (ChIP-Seq) (Valouev et al., 2008). The presence of TFBS within regulatory regions is then tested using either hypergeometric methods (Heinz et al., 2010) or approaches that take into account the relative affinity of a transcription factor to a TFBS (Frith et al., 2004). Defining a gene's regulatory regions presents a challenge. One naïve approach is to define a promoter as a short sequence (typically ~500 bp) flanking a gene's transcription start site (TSS). Alternatively, regulatory regions defined using ChIP-Seq against epigenetic markers of promoter and enhancer regions may be used provided data is available for a similar cell type to the one under investigation (Heinz et al., 2010).

1.4 Outline of Thesis

In this thesis transcriptomics approaches are taken to define IL-4Rα-dependent components of macrophage activation and proliferation during helminth infection. In chapter 5 RNA-Seq is used to define the transcriptomic profile of FACS-purified AAMΦ induced using the *B. malayi* implant model. I compare *B. malayi*-induced macrophage gene expression profiles against inflammatory-like thioglycollate elicited macrophages to define changes in gene expression resulting from the presence of the nematode and the associated Th2 environment. These data are interpreted alongside profiling of *B. malayi*-elicited macrophages from IL-4Rα$^{-/-}$ mice, which are unable to alternatively activate, in order to identify IL-4Rα-dependent changes in gene expression. Together these data permit identification of IL-4Rα -dependent changes in gene expression whilst controlling for additional effects of *B. malayi*. Using this model I confirm the anti-inflammatory phenotype of AAMΦ and validate the mitochondrial phenotype associated with AAMΦ *in vivo*. Furthermore I discover that PPARγ, required for AAMΦ *in vitro*, is expressed at a low level by *B.* malayi-elicited AAMΦ *in vivo*, although PPARδ is expressed. I validate the role for PPAR transcription factors in AAMΦ *in vivo* by analysing TFBS in macrophage promoters. Gene promoter regions are defined using a novel method for transcription start site identification with RNA-Seq data. The algorithm, TSS-Predictor, is discussed in Chapter 6. I have been able to put forward a putative mechanism for PPARδ-dependent transcriptional regulation in AAMΦ. Using GSEA I identified differential regulation of arachidonic acid metabolism, a source of PPAR ligands, associated with AAMΦ. In collaboration with Philip Whitfield and Ben Maskrey at The University of the Highlands and Islands I went on to characterize AAMΦ-derived eicosanoids and identified the PPARδ ligand PGI2 as abundantly produced in a *B. malayi*-dependent manner.

In chapter 7 of this thesis I present a time course of the macrophage response to *L. sigmodontis* infection. This more natural parasite model than *B. malayi* extends our RNA-Seq analysis from a single time point to capture macrophage gene expression profiles induced early during the response to helminth infection. I use a novel approach to gene expression analysis by integrating gene expression profiling with FACS-based measurements of alternative activation and proliferation. With this approach I define macrophage gene signatures associated with alternative activation and proliferation. An analysis of *cis*-regulatory elements in promoter regions is performed and, surprisingly, no STAT6 signature was found in the promoters of proliferation-associated genes induced in response to IL-4. Using the data obtained from this analysis and our current understanding of IL-4Rα-dependent macrophage proliferation I propose a model to explain these observations. Using this model I identify candidate mediators of IL-4-dependent macrophage proliferation *in vivo*.

2 THE BIOLOGY OF NEMATODE- AND IL-4Rα-DEPENDENT MURINE MACROPHAGE POLARIZATION *IN VIVO* AS DEFINED BY RNA-SEQ AND TARGETED LIPIDOMICS

This chapter contains the unedited version of (Thomas et al., 2012) published in *Blood*. Supplementary information for this chapter can be found in Appendix 1a and supplementary tables can be found in Appendix 1b.

2.1 ABSTRACT

Alternatively activated macrophages (AAMΦ) are a major component of the response to helminth infection however their functions remain poorly defined. To better understand the helminth-induced AAMΦ phenotype we performed a systems-level analysis of *in vivo* derived AAMΦ using an established mouse model. With next-generation RNA sequencing (RNA-Seq) we characterized the transcriptomes of peritoneal macrophages from BALB/c and IL-4Rα$^{-/-}$ mice elicited by the nematode *Brugia malayi*, or via intra-peritoneal thioglycollate injection. We defined expression profiles of AAMΦ-associated cytokines, chemokines and their receptors, providing evidence that AAMΦ contribute towards recruitment and maintenance of eosinophilia. Pathway analysis highlighted complement as a potential AAMΦ-effector function. Up-regulated mitochondrial genes support *in vitro* evidence associating mitochondrial metabolism with alternative activation. We mapped macrophage transcription start sites (TSS), defining over-represented *cis*-regulatory motifs within AAMΦ-associated promoters. These included the binding site for PPAR transcription factors, which maintain mitochondrial metabolism. Surprisingly PPARγ, implicated in the maintenance of AAMΦ, was down-regulated upon infection. PPARδ expression however

was maintained. To explain how PPAR-mediated transcriptional activation could be maintained we used lipidomics to quantify AAMΦ-derived eicosanoids, potential PPAR ligands. We identified the PPARδ ligand PGI2 as the most abundant AAMΦ-derived eicosanoid and propose a PGI2-PPARδ axis maintains AAMΦ during *B. malayi* implantation.

2.2 Introduction

Macrophages display enormous functional diversity determined by signals from their immediate environment. IFNγ stimulation induces classical activation, an essential prerequisite for microbial infection control, whereas IL-4/IL-13 exposure polarizes macrophages towards alternative activation (Gordon and Martinez, 2010). Alternatively activated macrophages (AAMΦ) are now implicated in the promotion of a wide range of diseases including cancer (Mantovani et al., 2002), allergy (Gordon and Martinez, 2010) and fibrosis (Gordon and Martinez, 2010); but also in protection against helminth infection (Allen and Maizels, 2011), diabetes (Odegaard and Chawla, 2011) and obesity (Odegaard and Chawla, 2011).

Despite the flurry of interest in AAMΦ we remain remarkably ignorant of their physiological role(s), partly because classical and alternative activation represent polar-regions in a landscape of activation phenotypes sculpted by multiple factors. Different cellular developmental histories (Jenkins et al., 2011), micro-environmental cues and factor-dependent polarization (Hashimoto et al., 2011) vastly increase the complexity of macrophage phenotypes *in vivo*. IL-4/IL-13 induce canonical alternative activation via IL-4Rα-dependent phosphorylation of STAT6, driving the transcription of a diverse repertoire of genes including *Arg1* (Arginase-1, ARG1), *Chi3l3* (Chitinase 3-like 3, YM1) and *Retnla* (resistin like alpha, RELMα, FIZZ-1). Because exposure to helminths almost universally

induces potent Th2 responses, alternative activation of MΦ is characteristic of these infections (Loke et al., 2002). Indeed, *Chi3l3* and *Retnla* were described as AAMΦ markers associated with challenge by the parasitic nematode *Brugia malayi* (Loke et al., 2002).

An emerging paradigm suggests that cytokine-mediated alterations in cellular metabolism determine cellular life-history (van der Windt et al., 2012) and effector functions (Infantino et al., 2011). For example, a switch from glucose-dependency to mitochondrial metabolism oversees the ability of effector $CD8^+$ T-cells to commit to a memory phenotype (van der Windt et al., 2012). In this context, it is interesting to note that *in vitro* studies suggest that classically activated macrophages (CAMΦ) and AAMΦ are associated with different metabolic profiles. CAMΦ require aerobic glycolysis (Rodríguez-Prados et al., 2010), while AAMΦ couple lipid oxidation with oxidative phosphorylation (Odegaard and Chawla, 2011). Cooperative interactions between STAT6, PPARγ and PGC-1b are considered necessary (Odegaard and Chawla, 2011; Szanto et al., 2010), to induce these metabolic changes in AAMΦ. *In vivo* confirmation of this observation is required, and an improved understanding of the role for mitochondrial metabolism in alternative activation may yield key insights into the effector functions of these cells.

Previous transcriptomic analyses of AAMΦ have used *in vitro* generated cells while *in vivo* studies of Th2 environments have analysed whole tissue (Sandler et al., 2003; Vats et al., 2006). This leaves a gap in our understanding of AAMΦ function during infection. Previously we identified abundantly expressed genes in *in vivo*-derived AAMΦ using an expressed sequence tag approach (Loke et al., 2002). This provided valuable insight into markers expressed by these cells, but lacked the power to critically assess molecular pathways associated with alternative activation *in vivo*. Here, using second-generation Illumina sequencing and mass spectrometry, we combined transcriptomics and lipidomics

to gain a global overview of macrophage IL-4Rα-dependent transcription and transcriptional regulation *in vivo*.

We compared nematode-elicited macrophages (NeMΦ) and inflammatory-like thioglycollate-elicited macrophages (ThioMΦ) from both wild type (WT) and IL-4Rα$^{-/-}$ mice. We sought to define physiological functions of *in vivo*-derived AAMΦ, and have focused on understanding the macrophage response to filarial nematode infection. We defined macrophages as F4/80 positive cells with a negative gating strategy to exclude contaminants. This definition likely includes macrophage sub-populations that contribute differentially towards the overall response, however this comparison allowed us to focus on the most relevant changes in macrophage physiology to filarial nematode challenge. The contrast between WT-NeMΦ and WT-ThioMΦ identified differential gene expression due to the presence of the nematode, or differences in cell origin. Comparing WT-NeMΦ and IL-4Rα$^{-/-}$-NeMΦ revealed IL-4Rα-dependent components of macrophage activation during helminth infection. We followed Siamon Gordon's definition of alternative activation as the IL-4/IL-13-dependent component of macrophage activation (Gordon and Martinez, 2010). IL-4Rα deficiency ablates both IL-4 and IL-13-dependent signalling (Gordon and Martinez, 2010). Thus, co-ordinately differentially expressed (DE) genes in WT-NeMϕ (i.e. *in vivo* generated AAMΦ) relative to both WT-ThioMΦ and IL-4Rα$^{-/-}$-NeMΦ are, by definition, those relevant to alternative activation during helminth infection.

Illumina RNA sequencing (RNA-Seq) provided >5 orders of magnitude dynamic range between the most abundant and lowly expressed genes, delivering the most extensive characterization of *in vivo*-polarized macrophage populations to date. We establish a putative role for AAMΦ in eosinophil recruitment and the complement response during helminth infection. Macrophage transcription start sites (TSSs) were mapped, and, by

characterizing overrepresented *cis*-regulatory elements in AAMΦ promoters, we confirm PPAR-dependent transcription as a major facilitator of alternative activation *in vivo*. Pathway analysis supported these findings by identifying AAMΦ-dependent alterations in lipid and mitochondrial metabolism, key targets of PPAR transcription factors. Finally, liquid chromatography tandem mass spectrometry (LC-MS/MS) was used to define the *B. malayi*-induced repertoire of endogenous eicosanoids, allowing us to propose a mechanism for PPARδ-mediated alternative activation. We thus provide global mechanistic insights into the function and regulation of helminth-elicited AAMΦ and identify putative effector molecules involved in the maintenance and regulation of alternative activation *in vivo*.

2.3 Methods

2.3.1 Mice

BALB/c and IL-4Rα$^{-/-}$ mice bred on the BALB/c background were used in the RNA-Seq experiments. All mice were 8-10 weeks old at the start of the experiments, and were bred in house. All work was conducted in accordance with the Animals (Scientific Procedures) Act of 1986.

2.3.2 Elicitation of macrophage populations for RNA-Seq

Adult *Brugia malayi* were retrieved from the peritoneal cavity of infected Mongolian Jirds obtained from TRS laboratories (Athens, GA). *B. malayi* were washed five times by serial passage into 50 ml RPMI supplemented with 1% Penicillin/Streptomycin (P/S) and 1% L-glutamine.

NeMϕ were obtained by surgically implanting 9 male BALB/c and 9 male IL-4Rα$^{-/-}$ mice each with 4 female and 1 male *B. malayi* adults intraperitoneally. Similarly, ThioMΦ were obtained by intraperitoneal administration of 700 µl 4% Brewer's modified thioglycollate medium in PBS (w/v) (BD Pharmingen) into 9 male BALB/c and 9 male IL-4Rα$^{-/-}$ mice. 21 days post-implantation, or 3 days post-thioglycollate injection, peritoneal exudate cells (PEC) were retrieved by 10 * 1 ml peritoneal lavage with ice-cold RPMI supplemented with 1% P/S, 1% L-glutamine and 5% Fetal Calf Serum (FCS).

2.3.3 Flow cytometry

All reactions were quenched between all staining steps by the addition of 1 ml FACS buffer (PBS + 0.5% BSA, 2 mM EDTA) followed by centrifugation at 1200 rpm for 5 min, removal of supernatant and brief re-suspension of cells by vortexing. All incubations are carried out on ice and in the dark unless otherwise stated.

2.3.4 Purification of macrophages for RNA-Seq

Prior to RNA sequencing MΦ were purified by flow cytometry. 1*10^7 PEC cells were transferred in 1 ml to FACS tubes and stained with 50 µl Live/Dead-Aqua (Invitrogen, 1/500 dilution in PBS) for 10 min at room temperature. Cells were blocked on ice for 20 min with 300 µl blocking buffer (FACS buffer + 1/300 0.8 mg/ml αCD16/32 + 1/32 mouse serum) before staining with F4/80-biotin (1/100, BioLegend), SiglecF-PE (1/100, eBiosciences), B220-PE (1/100, eBiosciences), CD4-FITC (1/100, eBiosciences) in 300 µl FACS buffer for 20 min. Finally, cells were incubated with 300 µl streptavidin conjugated-APC (1/200, eBiosciences) for 20 min, washed twice with 1 ml FACS buffer and resuspended in 300 µl

FACS buffer before sorting. Martin Waterfall at the Ashworth Laboratories, University of Edinburgh central flow facility performed flow sorting. Immediately after sorting cells were washed in 1ml PBS and stored at -80^{O}C in 700 µl Qiazol (Qiagen) prior to RNA extraction.

2.3.5 Intracellular cytokine staining for RELMα

Intracellular staining for RELMα was performed to validate the activation status of *B. malayi*-implanted BALB/c and IL-4Rα$^{-/-}$ mice. 5*10^{5} PEC were treated with 10 µl each Live/Dead Aqua and blocked with 50 µl blocking buffer as described above. Cell surface staining was performed with 50 µl F4/80-biotin (1/100, BioLegend), SiglecF-PE (1/100, eBiosciences), B220-PE (1/100, eBiosciences), CD3-AF700 (1/300, BioLegend) in FACS buffer for 20 min. Secondary staining was performed with streptavidin conjugated-APC (1/200, eBiosciences), 50 µl incubation for 20 min prior to fixation with 2% PFA at room temperature for 10 min. Intracellular staining was performed by washing cells with PermWash (BD Biosciences) before incubation with rabbit anti-RELMα (1.5 µg/ml, Peprotech) or rabbit-IGG isotype (1.5 µg/ml, Molecular Probes) in PermWash for 45 min. Secondary staining was performed using 50 µl anti-Rabbit-AF488 (1/300, Molecular Probes) and cells were resuspended in 100 µl FACS buffer prior to acquisition on a FACS Canto.

2.3.6 RNA-Seq Library Preparation and sequencing

RNA was extracted using the Qiagen miRNeasy kit according to the manufacturer's instructions and quality assessed using an Agilent Bioanalyzer (RNA Nano 6000). All samples had RIN scores > 8. RNA concentration was quantified using a Q-Bit fluorimeter (Invitrogen).

RNA-Seq libraries were prepared using the Illumina paired-end RNA-Seq library preparation kit in accordance with the manufacturer's guidelines. All reagents are supplied as a part of the Illumina kit unless otherwise stated. We prepared three biological replicates for each condition. 9 mice were used per group with each library being derived from 2 μg total RNA obtained by pooling 0.66 μg from 3 individuals. Poly-Adenylated mRNA was enriched using Sera-Mag Magnetic Oligo(dT) Beads and eluted into 16 μl Tris-HCl. Purified mRNA was then fragmented by incubating at 94 ^{O}C for 5 min using the supplied fragmentation buffer which utilizes divalent cations to catalyse RNA fragmentation. Following mRNA fragmentation first strand cDNA is synthesized using the SuperScript reverse transcriptase and buffer (Invitrogen) and random hexamer primers in a 50 min, 42 ^{O}C incubation. Double stranded cDNA is then synthesized using DNA Polymerase I and GEX second strand buffer. Following clean-up using the QIAquick PCR purification kit (Qiagen) cDNA was eluted into 50 μl Qiagen EB buffer and 'blunt-ended' using Klenow polymerase to remove 3' overhangs and to fill-in 5' overhangs. Following another clean-up step (QIAquick) 3' ends are adenylated using the supplied Klenow exonuclease and Illumina sequencing primers ligated using a T4 DNA ligase, with a further QIAquick clean-up step prior to ligation. cDNA was then eluted into 10 μl Qiagen EB buffer and 200 bp fragments are size selected using a 2% agarose gel is prepared with 1x TAE buffer (BioRad) and Certified Low-Range ultra agarose (BioRad). 10 μl cDNA was run alongside 2 μl 100bp DNA ladder (Invitrogen) at 120 v for 1 hour. A 200 bp +/- 20 bp slice of gel was then excised and cDNA was extracted using the QIAquick gel extraction kit (Qiagen). Finally, the excised cDNA library is amplified by PCR using the high-

fidelity Phusion DNA polymerase, primers specific to the Illumina sequencing adapters, and 18 cycles of a 30 s 65 ^{o}C annealing phase and a 30 s 72 ^{o}C elongation phase. Each RNA-Seq library was then validated on a 2% agarose gel as above and sequenced on a single lane of an Illumina GAIIx at the GenePool genomics facility in Edinburgh, yielding 51 base, paired-end reads.

2.3.7 Gene expression quantification: read mapping and differential expression analysis

Illumina sequencing yielded between 11 and 30 million paired-end reads per lane. Reads were mapped to the mouse reference genome (mm9) using TopHat (Trapnell, Pachter, and Salzberg 2009) (version 1.1.13) and gene expression levels quantified by counting uniquely mapping reads within Ensembl (v58) gene predictions using the intersection-nonempty function of HTSeq-count (version 0.4.7). Raw data and gene expression count tables are available at SRA (accession ERP001255) and ArrayExpress (accession E-MTAB-995). Data quality assessments (QA) and differential expression analyses were performed using the R/Bioconductor package DESeq (version 1.4.1, Anders and Huber 2010). Variance outliers were removed to minimize false-positives, see below for more details.

2.3.8 Identification of technical artefact based upon residual variances

In the course of exploratory data analysis and QA we investigated the expression characteristics of genes with large residual variances (> 20) after DE analysis with DESeq. We identified 190 such protein coding genes (0.83% of the total), > 90% of which were differentially expressed in at least 1 comparison. Hierarchical clustering revealed a

tendency towards higher gene expression amongst this subset of genes between samples BTG3, 4Ne1 and 4Ne2 and the remaining samples (Figure A1-s2). This discrepancy did not affect the overall clustering of samples according to their biological conditions (Figure A1-s3). However, the results do indicate an underlying technical artefact, and were excluded from downstream analysis.

2.3.9 Hierarchical agglomerative clustering and classification of immune effector genes

Hierarchical agglomerative clustering was performed on all genes considered DE between WT-NeMΦ and either IL-4Rα$^{-/-}$-NeMΦ or WT-ThioMΦ. A Euclidean distance metric was calculated between all selected genes using mean-centred variance-stabilized values. Variance stabilization was performed using the transformation provided in DESeq. The resulting matrix was subject to hierarchical agglomerative clustering using Ward's method, and subsequently sub-divided into 20 clusters in the R environment.

Chemokines, chemokine receptors, cytokines and cytokine receptors (GO identifiers GO:0042379, GO:0004950, GO:0005125 and GO:0004896 respectively) were identified using the Bioconductor package 'GO.db', and those present in the differential expression clusters above were identified for further analysis.

2.3.10 Gene-set enrichment analysis

To obtain p-values and t-statistics for gene set enrichment analysis (GSEA), raw data were variance stabilized using the transformation within DESeq. Differential expression was then reanalysed using limma (version 3.6.9, Smyth 2011) after non-specific filtering to enforce

homoscedasticity (variance stabilized value < 2, Figure A1-s4). The effect of this procedure is to remove systematic biases in the data that ordinarily provide greater statistical power to longer and more highly expressed genes (Young et al. 2010). GSEA on biochemical and metabolic pathways as defined by Kyoto Encyclopaedia of Genes and Genomes (KEGG, (Kanehisa and Goto, 2000)) was performed using the method described in Jiang and Gentleman (2007). Briefly, KEGG pathways were assigned a score, defined as the mean directed t-statistic for all expressed genes within the pathway. Disease pathways and those containing < 5 genes were excluded from analysis.

2.3.11 Identification of Transcription Start sites using RNA-Seq data

We have devised a protocol to identify transcription start sites (TSS) accurately using RNA-Seq data, this has now been converted into an application for use by the wider research community. For a detailed overview of the algorithm and how it works see Chapter 7. TSS prediction and promoter region analysis presented in this chapter was performed with a preliminary version of this algorithm, described in brief below.

Method. Accurate identification of transcription start sites is aided by deep sequencing. To maximise the depth of coverage in our experiment we merged reads from all 12 libraries into a single lane and mapped to the mouse reference genome with TopHat (Trapnell et al., 2009) using the default parameters (allowing multi-mapping reads). Aligned reads were then assembled using the reference-guided transcript assembler Cufflinks (Version 0.9.3, (Trapnell et al., 2010)), without the aid of the existing transcript annotations, to aid the discovery of novel transcripts (genes, exons or splice junctions). Cufflinks assembles overlapping sections of mapped reads into transcribed fragments, called 'transfrags'. These

ab initio defined transcripts were subsequently related to known gene models (Ensembl v58) using Cuffcompare from the Cufflinks suite.

TSS prediction operates by revising Cufflinks 'transfrag' predictions to more accurately identify the TSS. We selected the longest Cufflinks transfrag associated with each Ensembl GID (provided it was on the same strand as the reference gene) and calculated the per-base coverage of mapped reads using Samtools (Li et al., 2009). The observed per-base coverage obtained from Samtools and expected coverage for each transfrag provided by Cufflinks were used to trim transfrags to X% of the expected coverage using custom scripts. The trimming strategy and value of X were empirically optimized using chromosome 2 as a test dataset.

Optimization of trimming conditions. Two trimming policies were tested, forward and reverse trimming. Forward trimming begins at the outermost base in an annotation and iterates towards the 3' end of the transcript, removing one base at a time until condition X is met. Reverse trimming begins at the base adjacent to the donor splice site of the first exon and works towards the 5' end of the transcript, one base at a time, until coverage drops below X; all upstream annotation for the gene is then discarded. For both forward and reverse trimming, if threshold X is not met in exon 1 then the process is repeated consecutively for downstream exons until it is. Trimmed annotations are then output in GTF format.

To identify optimal cutoff conditions we identified the longest Cufflinks transfrag for each reference gene (as implemented in *Prep4Trim.pl*) and retained the 50% most abundant transfrags. Using chromosome 2 as a subset, MADs between predicted and nearest annotated TSS were determined at 1%, 5%, 10%, 20%, 30%, 40%, 50% and 60% of the

expected expression for each gene (Figure A1-s5). This optimality criterion assumes that at least one of the Ensembl-predicted TSS is correct, which may not always be the case. The empirically determined optimal trimming condition was 10% of the expected coverage with the reverse-trimming protocol as this provided the minimal MAD value of 32 bp.

Determination of optimal gene expression cut-off. With optimal trimming conditions defined we wanted to rationally define the minimum gene expression level required to accurately predict TSS. A trade-off exists here between maximising the number of promoters for further analysis and introducing false positive TSS predictions resulting from the incomplete assembly of lowly expressed genes.

In order to determine this cutoff we identified TSS for all expressed genes defined by Cufflinks transfrags. Using a similar MAD scoring method to above we allocated genes into one of 20 bins based upon their expression (i.e. 0-5%, 6-10% ... 95-100%) and, for each bin, calculated the MAD from the nearest Ensembl TSS (Figure A1-s6). From this we included the 80% most abundant genes as they show minimal deviation from the expected MAD.

Post-filtering. We tested whether our method allowed us to identify novel promoters in our dataset. We noted that the majority of the TSS a long distance from the nearest Ensembl TSS appeared to result from erroneously selecting Cufflinks transfrag for trimming. Thus, to improve the accuracy of TSS identification we disregarded TSS greater than 500 bp from the nearest Ensembl TSS for the associated gene. Finally, as we required multiple-mapping reads for TSS identification we also removed TSS predictions associated with genes that contained > 30% of multi-mapping reads. Genome-wide this method allowed us to predict 7,817 high-confidence macrophage TSS.

2.3.12 CLOVER ANALYSIS OF OVER-REPRESENTED TRANSCRIPTION FACTOR BINDING SITES

Promoter regions proximal to the identified TSS, 300 bp upstream and 100 bp downstream, were extracted from the mouse genome (mm9) using BEDTools *fastaFromBed*. Overrepresented *cis*-regulatory motifs were identified using Clover (version 2010-02-19 (Frith et al., 2004)) with position weight matrices (PWMs) of transcription factor binding sites (TFBSs) defined by the Jaspar non-redundant vertebrate set (Bryne et al., 2008). Overrepresentation in macrophage promoters was assessed by comparing all macrophage promoters to all mouse promoters (Ensembl v58). Similarly, AAMΦ promoters were assessed by comparing the intersection of genes upregulated in WT-NeMΦ relative to both WT-ThioMΦ and IL-4Rα$^{-/-}$-NeMΦ against all remaining macrophage promoters. P-values were determined by permutation test using 1000 iterations. The number of occurrences of a motif within AAMΦ associated promoters were counted using the online tool RSAT matrix-scan with the default parameter set (Thomas-Chollier et al., 2011).

2.3.13 ELICITATION, CHARACTERIZATION AND ISOLATION OF MACROPHAGE POPULATIONS FOR LIPIDOMICS

Macrophages for the lipidomic study were elicited using exactly the same protocol as described above for RNA-Seq, only 5 mice were used in each group, and we omitted the IL-4Rα$^{-/-}$ thioglycollate group. The first 2 ml of peritoneal lavage culture supernatant was retained and stored at -80 ^{O}C prior to preparation for mass spectrometric analysis. $5*10^5$ PEC from *B. malayi*-implanted mice were retained for flow cytometric cell phenotyping. FACS was performed using the above protocol, our staining panel included, Live/Dead-Aqua (1/500), CD4-FITC (1/200), BioLegend), F4/80-PE (1/200, eBiosciences), Gr1-PerCP (1/200, BioLegend), CD19-APC (1/300, eBiosciences), eosinophils were determined as F4/80

intermediate, SSC high. Samples were fixed for 10 min in 2% PFA prior to acquisition on a FACS Canto. MΦ were adherence purified by plating the remaining PEC into 6 well plates for 4 h at 37 ^{O}C, 5% CO2. Non-adherent cells were removed by washing twice with 1 ml warm culture media (RPMI + 1% L-glut + 1% Pen/Strep + 5% FCS). After 12 hours culture in 1 ml culture media, both adherent macrophages and culture supernatant were collected together and stored at -80 ^{O}C prior to preparation for mass spectrometric analysis.

2.3.14 Isolation and mass spectrometric analysis of eicosanoids

Samples were thawed on ice to which 2 volumes of ice-cold methanol containing the deuterated internal standards 15-HETE-d8, LTB4-d4, PGE2-d4 and 8-iso Prostaglandin F2α-d4 were added to give final concentrations of 2 ng/sample for cell culture samples and 4 ng/sample for peritoneal lavage samples. Following protein precipitation, samples were centrifuged (600 g, 10 min, 4 °C) and the supernatant then diluted to a final concentration of 10% methanol. Samples were acidified to pH 3.5 with 2M HCl and immediately applied to a 6 cc 500 mg C18 SPE Cartridge (Waters) which had been prepared by passing through 2 column volumes of methanol, and 2 volumes of H_2O. Samples were washed with 5 ml H_2O, 2x 5 ml hexane and eluted with 2x 3 ml ethyl acetate. Eluants were dried under a stream of N2 and resuspended in 100 µl of 50:50 HPLC buffer A (H_2O + 0.02% acetic acid):buffer B (methanol + 0.02% acetic acid) and stored at -80°C until analysis. Eicosanoids were separated by injecting 10 µl sample onto a reverse phase HPLC column (Hypersil Gold C18 column; 3 µm x 2.1 x 50 mm; Thermo Scientific) maintained at 40 °C with a gradient from 50% B to 65 at 14 min, increasing to 100% B at 20 min, held at 100% B until 22 mins then recycled to 50% B at 23 mins until 25 mins at a flow rate of 300 µL/min. 6-keto PGF1α was chromatographed on a shorter gradient to obtain optimal peak shape using 60-90% B over

2 min, increased to 100% B at 10 min, then recycled back to 60% B at 11 min and held until 15 min at a flow rate of 400 µl/min. The eluted eicosanoids were detected using an online tandem mass spectrometer (MS/MS) (TSQ Quantum Ultra; Thermo Scientific) operating in negative ion mode with a HESI source using the selected reaction monitoring (SRM) scan mode with argon as the collision gas. Electrospray voltage was -2.5 kV with a vaporizer temperature of 350 °C and capillary temperature of 250 °C. Optimal collision energies and tube lens voltages were optimized individually for each analyte. The analytes and Q1/Q3 transitions were as follows; 15-HETE-d8 (327>226); LTB4-d4 (339>197); PGE2-d4 (355>193); 8-Iso PGF2α-d4 (357>197); 5-OxoETE (317>203); 15-HETE (319>175); 5-HETE (319>115); 12-HETE (319>179); LTB4 (335>195); 6-keto PGF1α (369>163); Thromboxane B2 (369>169); PGF2α (353>193); 20-hydroxy LTB4 (351>195); 13,14-dihydro-15-keto PGE2 (351>235) 13,14-dihydro-15-keto PGD2 (351>207); PGD2 (351>233); PGE2 (351>189) and 8-iso PGF2a (353>193). Data was acquired and peak area integrated using the software (LCquan v2.6; Thermo Scientific). Amounts were quantified by reference to standard curves run in parallel, with extraction efficiency and recovery normalized by reference to the appropriate internal standard. Mass spectrometry and determination of eicosanoid concentrations was performed by Ben Maskrey at UHI (University of the Highlands and Islands, Inverness).

2.4 Results

2.4.1 Generation of AAMΦ and differential expression analysis

We generated NeMΦ and ThioMΦ by implanting BALB/c and IL-4Rα$^{-/-}$ mice with the nematode *B. malayi,* or *via* intraperitoneal administration of thioglycollate to elicit a population of non-polarized, inflammatory-like macrophages. IL-4 stimulation increases F4/80 surface expression on macrophages (Jenkins et al., 2011). In order to enrich for AAMΦ in the implant setting, we collected the brightest F4/80 positive macrophage population in each condition (see Figure A1-s1 and Table A1-ST-1). Macrophage purity was maximized by exclusion of dead cells, doublets, B-cells (B220^{+}), eosinophils (SiglecF^{+}) and CD4^{+} T-cells using negative gating (Figure 2-1a). On average 47% of WT-NeMΦ were RELMα positive, consistent with 30-60% positivity typically seen following *B. malayi* implant (Figure 2-1b,c). RNA-Seq libraries yielded between 11 and 30 million 51 base paired-end reads. Gene expression was quantified by mapping reads to the mouse reference genome using TopHat (Trapnell et al., 2009). 52–70% of reads mapped uniquely to the genome, with 7.7–25 million within exons of known genes (Ensembl v58, Table 2-1). Between 12,853 and 13,520 (56-59%) protein coding genes were expressed in each group, with 12,039 of these common to all four populations (Figure 2-2a). We validated sample purity by assessing the expression of lineage-restricted marker genes for potential contaminants; eosinophils, neutrophils and B-cells (Table A1-ST-2). With this approach we confirmed negligible contamination of neutrophils or eosinophils, however a low level of *Cd19* expression was observed in IL-4Rα$^{-/-}$- NeMΦ.

Library	Total Reads	Number mapped	% mapped	properly paired	% properly paired	singlets	% singlets	unique reads in genes	% unique reads in genes
BALB/c Inf. 1	35646760	24026742	67.4	18883064	78.59	3802882	15.83	13914673	57.9
BALB/c Inf. 2	39663694	26450844	73.3	21349672	80.71	4071578	15.39	15261028	57.7
BALB/c Inf. 3	40960176	26670826	65.1	20637676	77.38	4486602	16.82	15578573	58.4
BALB/c Thio. 1	40398232	26011652	64.3	21817520	83.88	3454426	13.28	14732920	56.6
BALB/c Thio. 2	48366774	33194643	68.6	26782502	80.68	5494173	16.55	19344275	58.2
BALB/c Thio. 3	45552296	28796129	63.2	24350132	84.56	4167411	14.47	16480896	57.2
IL-4Rα-/- Inf. 1	56700552	31428896	55.4	27125044	86.31	4172298	13.28	17799754	56.6
IL-4Rα-/- Inf. 2	23150358	15076797	65.1	12977824	86.08	1911075	12.68	8493659	56.3
IL-4Rα-/- Inf. 3	60376960	42872365	71	32888162	76.71	6826787	15.92	24849426	57.9
IL-4Rα-/- Thio. 1	40492838	24144541	59.6	19772782	81.89	3547653	14.69	13845966	57.3
IL-4Rα-/- Thio. 2	22996482	13033134	56.7	8952742	68.69	2418168	18.55	7725619	59.2
IL-4Rα-/- Thio. 3	36062844	19911073	55.2	14751228	74.09	3874775	19.46	11892859	59.7

Table 2-1. Sequence and mapping statistics for raw Illumina data generated using Samtools flagstat.

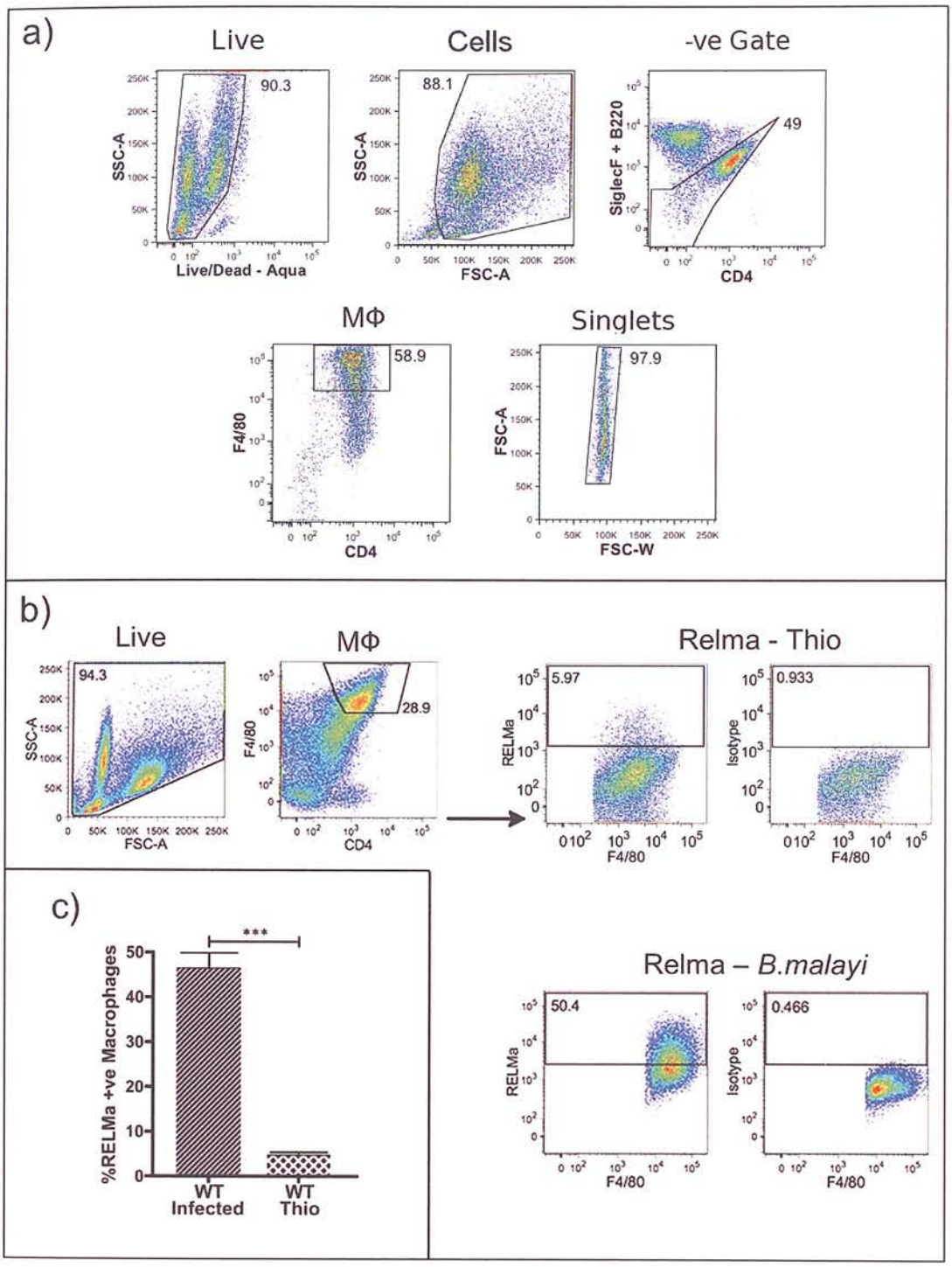

a)
Live
Cells
-ve Gate
90.3
88.1
49
SSC-A
Live/Dead - Aqua
FSC-A
SiglecF + B220
CD4
MΦ
Singlets
58.9
97.9
F4/80
CD4
FSC-A
FSC-W
b)
Live
MΦ
Relma - Thio
94.3
28.9
5.97
0.933
SSC-A
FSC-A
F4/80
CD4
RELMa
Isotype
F4/80
c)
***
%RELMa +ve Macrophages
WT Infected
WT Thio
Relma – B.malayi
50.4
0.466
RELMa
Isotype
F4/80

Figure 2-1. Flow cytometric acquisition of macrophages, and confirmation of alternative activation in WT *B. malayi*-implanted macrophages.

A) Gating strategy used to obtain pure macrophage populations as shown with one representative WT *B. malayi* infected individual. After the removal of dead cells B220, SiglecF and CD4 positive cells were excluded. F4/80 high cells were then selected and doublets removed based upon FSC-W/FSC-A scatter (Forward scatter-width/forward scatter-area).

B) Intracellular cytokine staining for RELMα expression in thioglycollate-elicited macrophages (top) and *B. malayi*-elicited macrophages (bottom). The scatter profile and macrophage gates on the left refer to one representative WT *B. malayi* infected individual.

C) Bar chart showing percentage of RELMα positive macrophages from analysis b) n=9 per group ($p < 0.001$ ***).

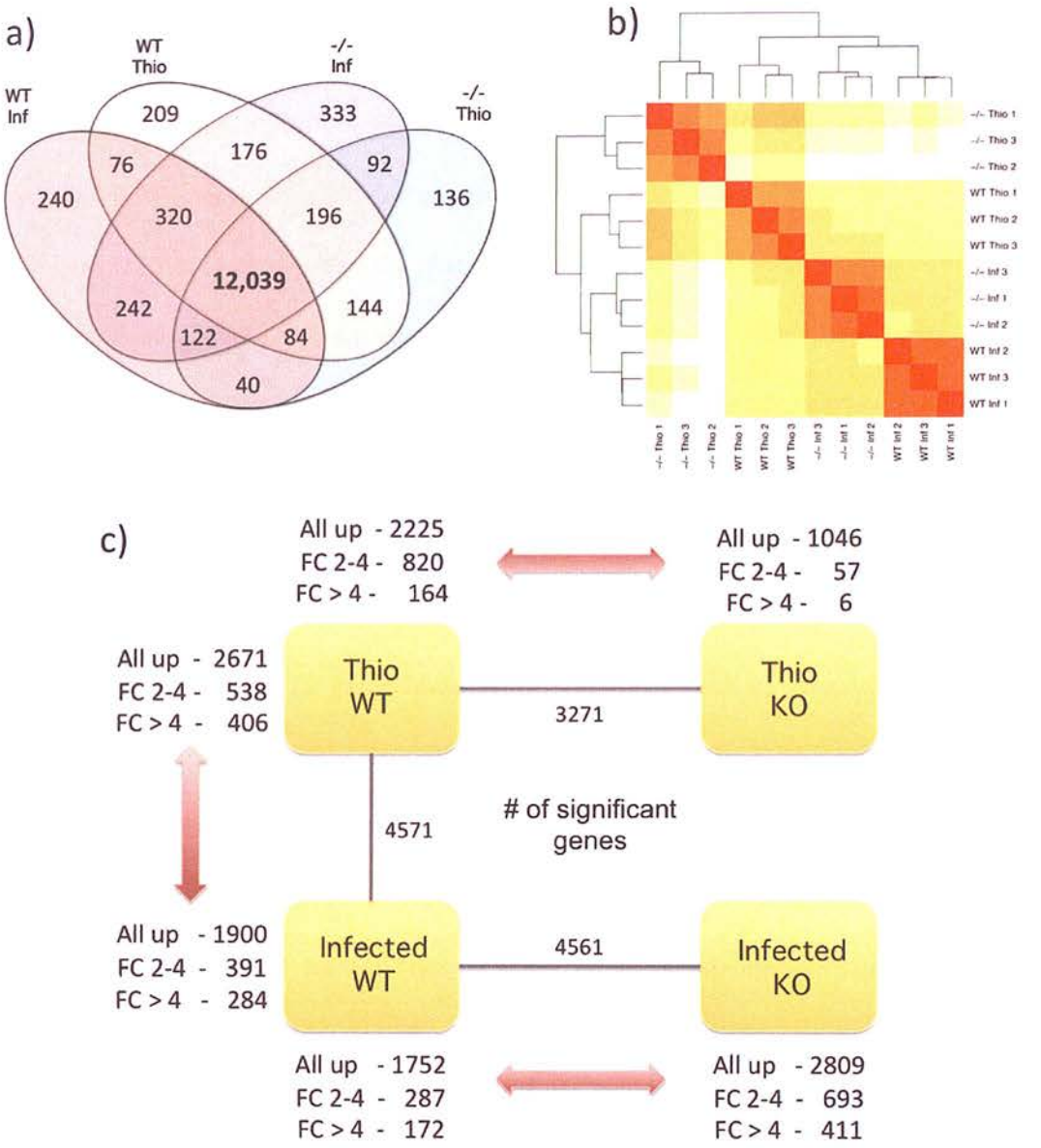


Figure 2-2 Overall gene expression and differential expression analysis.

A) Number of expressed genes in macrophage populations where expressed is considered as at least 1 read mapping to a gene in all 3 replicates of a condition.

B) Unsupervised, hierarchical clustering of individual lanes demonstrating discrete clustering of biological replicates.

C) Summary of DE genes ($p < 0.01$) in each pairwise comparison showing the total number of DE genes (inner), and a breakdown showing the direction of differential expression for both moderately (log 2 fold change ±2-4) and highly (Log 2 fold change > 4) DE genes.

Hierarchical clustering of global gene expression profiles grouped RNA-Seq libraries according to biological condition, reaffirming the quality and reproducibility of our analysis (Figure 2-2b). Following differential expression analysis we identified substantial transcriptional differences between the macrophage populations in the three key comparisons (Figure 2-2c). Using a p-value cutoff of 0.01 after correction for multiple testing (Benjamini-Hochberg method), we identified 4,571 DE genes between WT-NeMΦ and WT-ThioMΦ, 4,561 DE genes between WT-NeMΦ and IL-4-Rα$^{-/-}$-NeMΦ, and 3,271 DE genes between WT-ThioMΦ and IL-4Rα$^{-/-}$-ThioMΦ (Figure 2-2c). A list of all genes discussed in this chapter and associated p-values are provided in supplementary Table A1-ST-3. We observed a higher number of DE genes between WT-NeMΦ and IL-4Rα$^{-/-}$-NeMΦ than between WT- ThioMΦ and IL-4Rα$^{-/-}$-ThioMΦ (Figure 2-2c). Additionally, the magnitudes of these differences were much greater in the infection setting (Figure A1-s7). Furthermore, a similar range of differential expression was observed between WT-NeMΦ and WT-ThioMΦ as for WT-NeMΦ and IL-4Rα$^{-/-}$-NeMΦ. Thus, as expected, IL-4Rα-dependent signalling drove major alterations in the macrophage transcriptional profile in response to Th2-inducing immune stimuli.

We interrogated the expression profiles of DE genes in order to identify those explicitly associated with alternative activation. All DE genes from the WT-NeMΦ vs. WT-ThioMΦ and WT-NeMΦ vs. IL-4Rα$^{-/-}$-NeMΦ comparisons (4,571 and 4,561 respectively) were grouped according to their expression profile using hierarchical agglomerative clustering. The resulting tree was sub-divided into 20 clusters, and the expression profile of genes in each cluster was assessed (Figure A1-s8). We identified five clusters with increased expression levels in WT-NeMΦ relative to WT-ThioMΦ and IL-4Rα$^{-/-}$-NeMΦ. These were

classified as AAMΦ-up (Figure 2-3, Table A1-ST-4). Similarly, five clusters were identified with converse expression profiles, and classified as AAMΦ-down. In total 1,658 genes AAMΦ-up genes and 1,735 AAMΦ-down genes were identified (Figure 2-3). Importantly we can state that expression of genes in AAMΦ-up and AAMΦ-down clusters is IL-4Rα-dependent. However, as the IL-4Rα$^{-/-}$ mice lack IL-4Rα expression on all cells we cannot say whether these effects are cell-autonomous, and in some cases may reflect other changes such as Th2 cell activation.

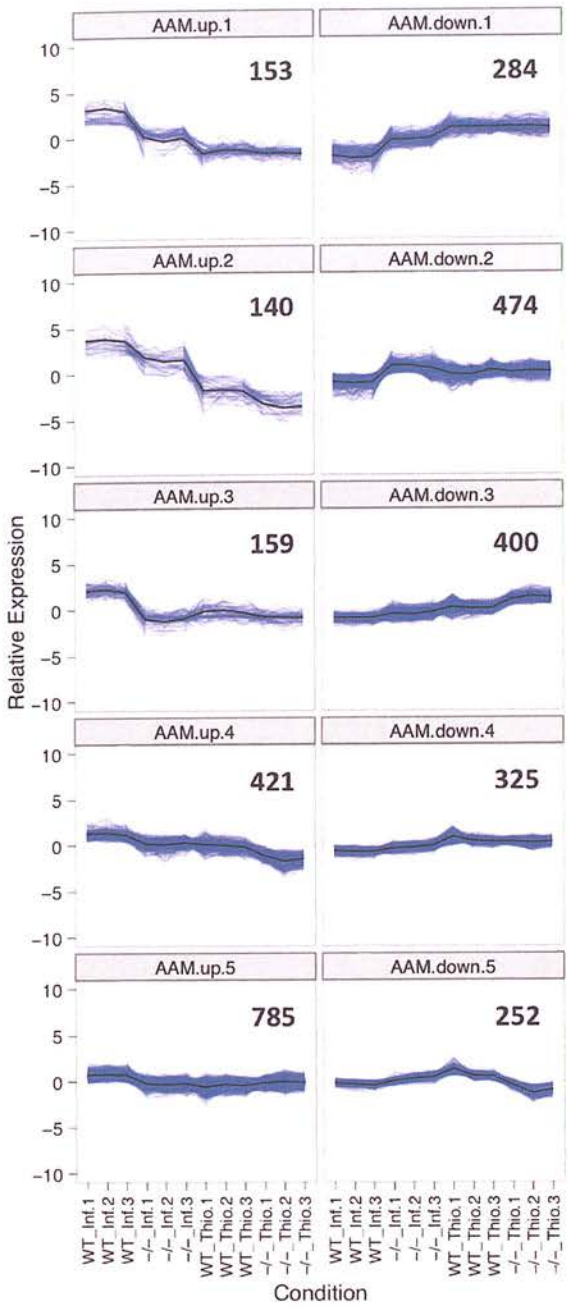


Figure 2-3. Expression profile and GO enrichment of differentially expressed AAMΦ-associated gene sets. The expression profiles of gene sets positively and negatively associated with alternative activation (AAMΦ-up and AAMΦ-down). All statistically significant genes (p<0.01) between WT-NeMΦ and WT-ThioMΦ, and WT-NeMΦ and IL-4rα$^{-/-}$-NeMΦ, were clustered using hierarchical agglomerative clustering. Each gene within an expression cluster is plotted in blue and the mean expression for all genes within each cluster is overlaid in black. The figure in each panel represents the total number of genes in that cluster.

2.4.2 Differential expression of immune effectors

We reasoned that a descriptive, knowledge-based, assessment of immunologically relevant GO terms (encompassing cytokines, chemokines and their receptors; GO terms GO:0005125, GO:0008009, GO:0004896 and GO:0004950 respectively; hereafter called immune effectors) would provide insight into AAMΦ function and regulation. Immune effectors that were present in AAMΦ-up and AAMΦ-down clusters were defined as AAMΦ-associated. Using this criterion we identified AAMΦ-association for expression of 16 cytokines, 11 cytokine receptors, 10 chemokines and 8 chemokine receptors (Figure 2-4). Below we discuss the function of key AAMΦ-associated immune effectors, providing insight into the IL-4Rα-dependent facets of the macrophage response to filarial nematode infection *in vivo*, and defining candidate genes for future investigations.

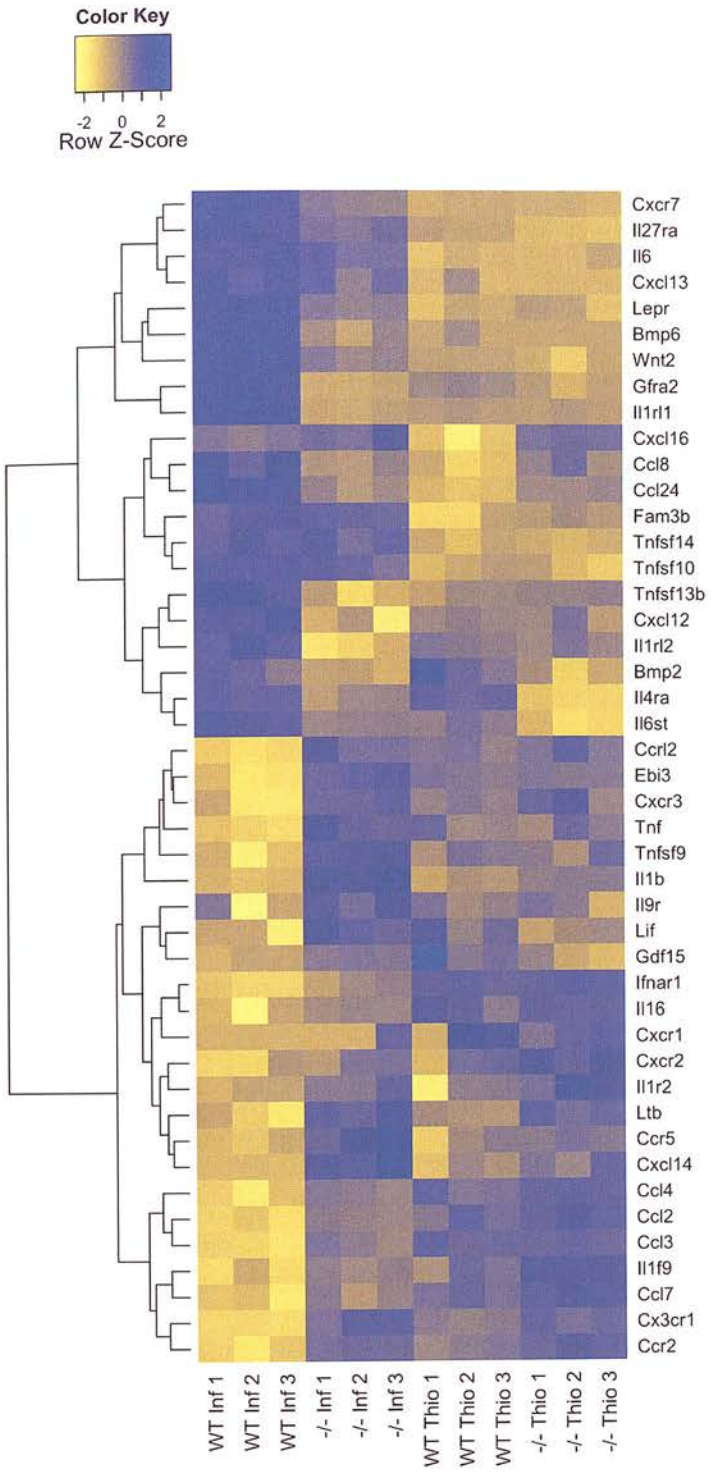


Figure 2-4. Heatmap of alternative activation-modulated immune effectors. Immune effector genes, cytokines, chemokines and their respective receptors (GO:0005125, GO:0008009, GO:0004896 and GO:0004950 respectively), in AAMΦ-associated clusters were identified based upon gene ontology annotations. Hierarchical clustering analysis reveals a unique expression profile of AAMΦ-associated immune effectors.

CHEMOKINE AND CHEMOKINE RECEPTOR EXPRESSION PROFILES. The AAMΦ chemokine expression profile (Figure 2-4) was consistent with macrophage-mediated maintenance of the cellular milieu in the peritoneal cavity of *B. malayi*-implanted mice, which is composed primarily of B-cells, macrophages and eosinophils (Loke et al., 2007). All assayed macrophage populations abundantly and constitutively expressed *Ccl6* and *Ccl9* (Table A1-ST-5), which may serve to maintain macrophage populations within the peritoneal cavity (Romagnani, 2002). In the AAMΦ-up clusters we identified the eosinophil chemoattractants *Ccl24* and *Ccl8* (Romagnani, 2002), as abundantly and moderately expressed respectively. *Cxcl12* and *Cxcl13* were also identified as AAMΦ-associated, however were expressed at much lower levels. Consistent with an anti- or non-inflammatory phenotype of AAMΦ (Gordon and Martinez, 2010), the AAMΦ-down clusters contained *Ccl3* (Mip1α), *Ccl4*, *Ccl2* (MCP-1), *Ccl7* (MCP-3) and *Cxcl14*, all of which are involved in acute phase inflammation, attracting primarily monocytes and neutrophils (Mantovani et al., 2004).

Interestingly, AAMΦ-up clusters were devoid of chemokine receptors with the exception of *Cxcr7* (Figure 2-4). *Cxcr7* responds to the stromal-derived *Cxcl12*, affecting both migration and differentiation in monocytes (Sánchez-Martín et al., 2011). *Ccr1* (MIP-1 α receptor, Table A1-ST-5) was also highly expressed by macrophages, but not classified as AAMΦ-associated. Therefore, *Cxcr7* and/or *Ccr1* may be key determinants of nematode-induced AAMΦ localization(Mantovani et al., 2004; Sánchez-Martín et al., 2011)'. A large number of chemokine receptors were present in the AAMΦ-down clusters. These included *Ccr2*, *Cxcr1*, *Cxcr2*, *Cxcr3*, *Ccr5* and *Ccrl2* implying that AAMΦ are impaired in their capacity to respond to numerous chemokines. This suggests that WT-NeMΦ do not migrate to prime lymphatic T-cell responses, in spite of high MHCII expression and functional antigen presentation (Loke et al., 2000b; Mylonas et al., 2009).

CYTOKINE AND CYTOKINE RECEPTOR EXPRESSION PROFILES. Cytokine receptor expression was typically maintained or enhanced in NeMΦ. AAMΦ-up clusters contained the IL-33 receptor, *Il1rl1* (ST-2) and the IL-27 receptor subunit *Il27ra* (WSX-1), both of which have previously been characterized as AAMΦ-associated (Kurowska-Stolarska et al., 2009; Rückerl et al., 2006). AAMΦ-up clusters also included *Gfra2* (GDNF family receptor alpha-2), involved in neuronal survival and differentiation (Jing et al., 1997), and the leptin receptor (*Lepr*). Leptin is an adipokine involved in energy homeostasis with broad, pleiotropic effects. There is no described role for leptin in Th2 immunity, however, macrophages of *Lepr* deficient mice express higher levels of inflammatory cytokines (Li et al., 2006) suggesting the possibility that the leptin receptor contributes to the anti-inflammatory profile of AAMΦ.

With regard to cytokines WT-NeMΦ produced significantly more *Il6*, *Wnt2* and *Bmp6* than WT-ThioMΦ or IL-4Rα$^{-/-}$-NeMΦ. *Bmp6* is pleiotropic, suppressing B-cell proliferation (Kersten et al., 2005) and promoting macrophage IL-6 production (Lee et al., 2010). *Wnt2* has no defined role in macrophage physiology, although Wnts do influence immune cell fate decisions (Staal et al., 2008). Thus, *Wnt2* and *Bmp6* represent promising novel candidates for future investigations of Th2 immunity.

AAMΦ are described as broadly anti-inflammatory (Gordon and Martinez, 2010), in part due to their production of IL-10. Surprisingly, in the chronic setting of our experiment, WT-NeMΦ did not transcribe any *Il10* (table A1-ST-5). Nonetheless, AAMΦ-down clusters contained genes for many pro-inflammatory cytokines. Amongst these were *Tnf*, *Lif*, *Il1b* and *Il16* (Cruikshank et al., 2000; Rosado and Rodriguez-Sosa, 2011; Tofaris et al., 2002). Additionally, the expression of the macrophage activating factor *Tnfsf9* (Kim et al., 2009)

and the IL-27 and IL-35 subunit *Ebi3* were lowered in AAMΦ. In summary, AAMΦ expressed greater levels of immune effectors associated with eosinophil recruitment and survival, and lower levels of numerous pro-inflammatory agents.

2.4.3 KEGG PATHWAY ANALYSIS

To assess whether metabolic changes in macrophages previously observed during IL-4 treatment *in vitro* also occur *in vivo*, we applied gene set enrichment analysis (GSEA) using the KEGG pathway database alongside modifications for RNA-Seq data (see methods section). By considering the relative expression of all genes, rather than only DE genes, GSEA provides a sensitive metric for identifying differences in biochemical and cell signalling pathways.

THE COMPLEMENT AND COAGULATION CASCADE. Unexpectedly, the complement and coagulation cascade (KEGG pathway mmu04610) was the most differentially regulated pathway between WT-NeMΦ and WT-ThioMΦ (Table A1-ST-6). The most striking feature of complement expression in AAMΦ was transcript abundance (Figure 2-5). For example, C3 and C4 attracted over 100,000 reads each, making them amongst the most abundant NeMΦ-associated transcripts. NeMΦ showed up-regulation of C1q complex genes (*C1qa*, *C1qb* and *C1qc*), and striking over-expression of MBL pathway constituents specifically FicolinA (*FcnA*), *Cfb* and *Cfp* (Figure A1-S9). This suggests a role for FicolinA-dependent complement activity in the response to helminth infection.

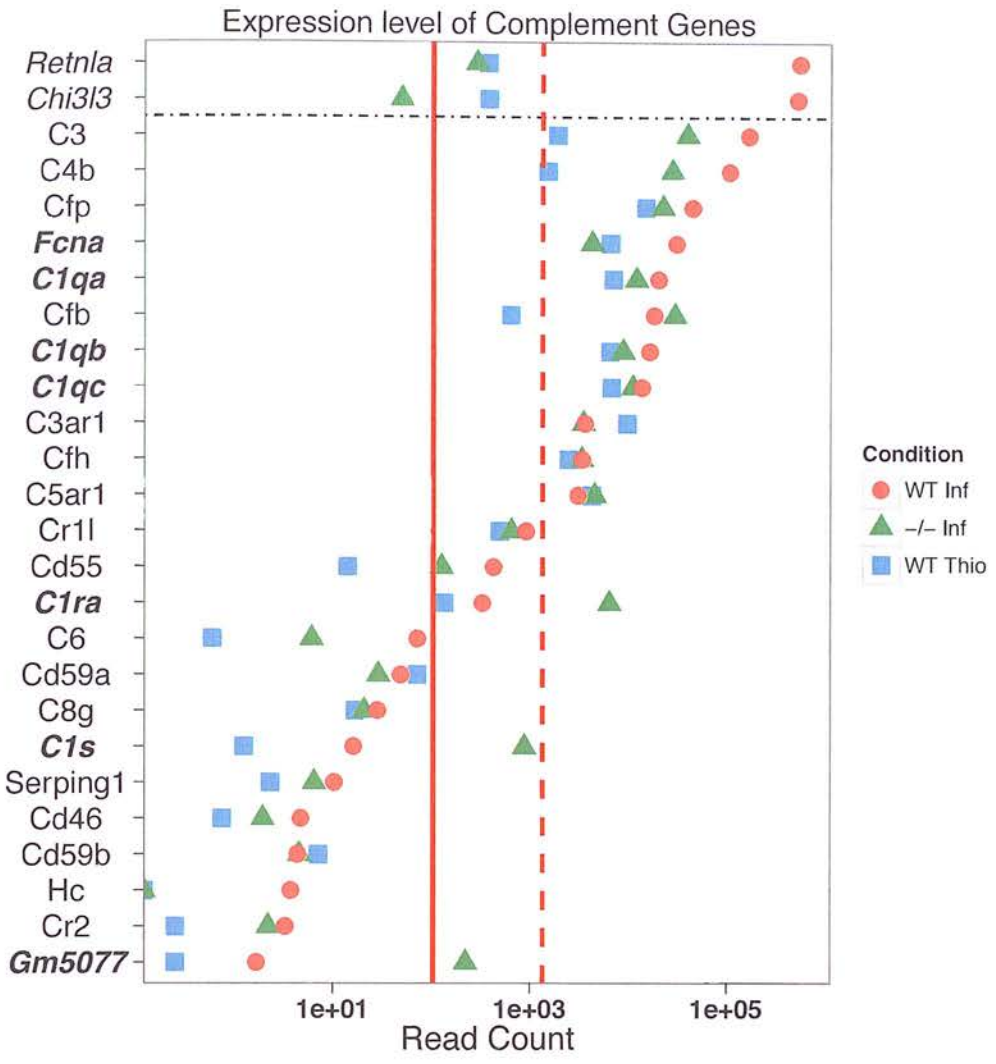


Figure 2-5. Complement components are abundantly expressed in an alternative activation-dependent manner. The expression of genes in the complement and coagulation cascade (KEGG pathway mmu:04610). For reference the median, and 90th percentiles of expression for all expressed genes are included (solid and dashed red lines respectively). Additionally, as highly expressed marker genes *Chi3l3* (YM1) and *Retnla* (RELMα) are included for reference.

MITOCHONDRIAL METABOLISM. Twenty-three of the twenty-seven most differentially regulated pathways between WT-NeMΦ and WT-ThioMΦ were metabolic (Figure A1-s10, Table A1-ST-6). A similar trend was observed between WT-NeMΦ and IL-4Rα$^{-/-}$-NeMΦ, supporting a model wherein a shift in metabolic phenotype is a cardinal feature of alternative activation (Odegaard and Chawla, 2011) (Figure A1-s10). The second most perturbed KEGG pathway between WT-NeMΦ and WT-ThioMΦ was the tricarboxylic acid (TCA) cycle (Table A1-ST-6). Indeed, the majority of expressed TCA cycle genes were expressed at a higher level in WT-NeMΦ relative to both WT-ThioMΦ and IL-4Rα$^{-/-}$-NeMΦ (Figure 2-6a, Figure A1-s11). A similar pattern of gene expression was also observed for genes involved in oxidative phosphorylation (Figure 2-6b). Taken together these findings are consistent with previous studies showing that IL-4 induces expansion of the mitochondrial compartment in macrophages *in vitro* (Odegaard et al., 2007).

PPARγ and PGC-1b, key regulators of mitochondrial metabolism have been described as required for alternative macrophage activation *in vitro* (Odegaard et al., 2007; Vats et al., 2006). Accordingly, the metabolic profile we observe could be explained by transcriptional activity of PPAR family members (PPARα, PPARγ and PPARδ). However, during *B. malayi* infection, macrophage-specific PPARγ$^{-/-}$ mice show no impairment in alternative activation (Dominik Rückerl, unpublished data). Furthermore, although *Ppargc1b* (PGC-1b) expression was augmented in an IL-4Rα- and infection-dependent manner, *Pparg* (PPARγ) levels were substantially lower in NeMΦ than ThioMΦ (Figure 2-6c). *Ppara* (PPARα) was not expressed by any of the assayed macrophage populations (Table A1-ST-5). Only *Ppard* (PPARδ) expression was sustained in NeMΦ, yet it was not up-regulated (Figure 2-6c). This suggests that during helminth infection, PPARδ may compensate, either partially, or completely, for

reduced PPARγ in NeMΦ. Alternatively, PPAR-independent transcription may maintain mitochondrial metabolism in NeMΦ.

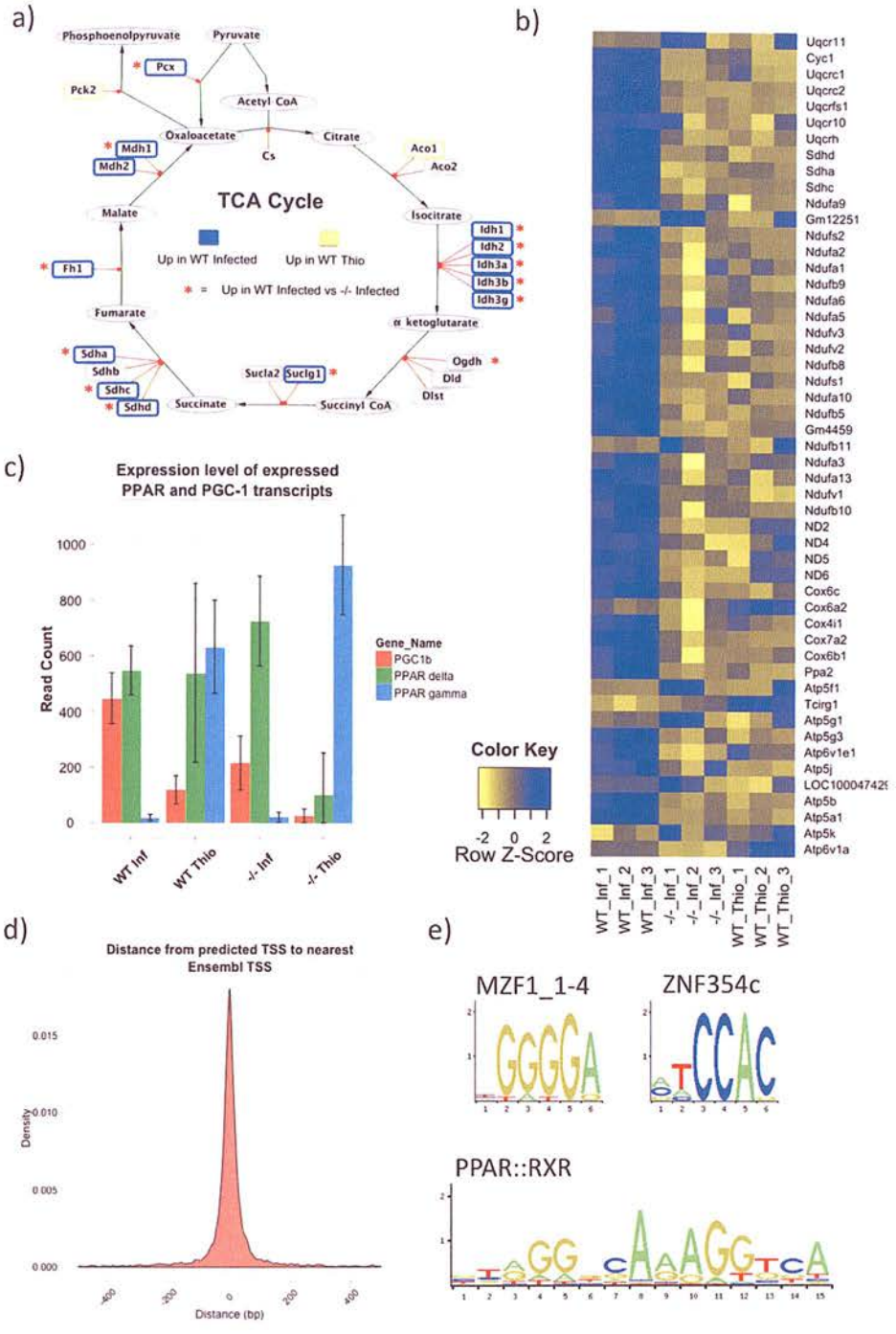

a)
Phosphoenolpyruvate
Pyruvate
Pcx
Pck2
Acetyl CoA
Oxaloacetate
Citrate
Cs
Mdh1
Mdh2
Aco1
Aco2
TCA Cycle
Malate
Isocitrate
Up in WT Infected
Up in WT Thio
* = Up in WT Infected vs -/- Infected
Idh1
Idh2
Idh3a
Idh3b
Idh3g
Fh1
Fumarate
α ketoglutarate
Sdha
Sdhb
Sdhc
Sdhd
Sucla2
Suclg1
Ogdh
Dld
Dlst
Succinate
Succinyl CoA
b)
Uqcr11
Cyc1
Uqcrc1
Uqcrc2
Uqcrfs1
Uqcr10
Uqcrh
Sdhd
Sdha
Sdhc
Ndufa9
Gm12251
Ndufs2
Ndufa2
Ndufa1
Ndufb9
Ndufa6
Ndufa5
Ndufv3
Ndufv2
Ndufb8
Ndufs1
Ndufa10
Ndufb5
Gm4459
Ndufb11
Ndufa3
Ndufa13
Ndufv1
Ndufb10
ND2
ND4
ND5
ND6
Cox6c
Cox6a2
Cox4i1
Cox7a2
Cox6b1
Ppa2
Atp5f1
Tcirg1
Atp5g1
Atp5g3
Atp6v1e1
Atp5j
LOC10004742
Atp5b
Atp5a1
Atp5k
Atp6v1a
WT_Inf_1
WT_Inf_2
WT_Inf_3
-/-_Inf_1
-/-_Inf_2
-/-_Inf_3
WT_Thio_1
WT_Thio_2
WT_Thio_3
Color Key
-2 0 2
Row Z-Score
c)
Expression level of expressed PPAR and PGC-1 transcripts
Read Count
1000
800
600
400
200
0
WT Inf
WT Thio
-/- Inf
-/- Thio
Gene_Name
PGC1b
PPAR delta
PPAR gamma
d)
Distance from predicted TSS to nearest Ensembl TSS
Density
0.015
0.010
0.005
0.000
-400
-200
0
200
400
Distance (bp)
e)
MZF1_1-4
ZNF354c
PPAR::RXR

Figure 2-6. AAMΦ up-regulate mitochondrial tricarboxlyic acid (TCA) cycle genes and show evidence of PPAR-dependent transcription *in vivo*.

A) Schematic of the TCA cycle showing DE genes in WT-NeMΦ relative to WT-ThioMΦ (blue border = up-regulated, yellow border=down-regulated). Additionally, genes up-regulated in WT-NeMΦ relative to IL-4Rα$^{-/-}$-NeMΦ are depicted with a red asterix.

B) Heatmap showing relative expression of mitochondrial electron transport chain (ETC) components between WT-NeMΦ, IL-4Rα$^{-/-}$-NeMΦ and WT-ThioMΦ. The majority of ETC components are most highly expressed in WT-NeMΦ.

C) The expression levels of transcripts encoding for PPARγ, PPARδ and the co-activator protein PGC-1b shows that WT-NeMΦ express high levels of PPARδ and PGC-1b, but not PPARγ.

D) Confirmation of the accuracy of transcription start site (TSS) prediction. The density plot shows the distances from the predicted TSS for each gene to nearest annotated Ensembl TSS (median absolute deviation = 32bp).

E) Consensus motifs for the 3 over-represented transcription factor binding sites in AAMΦ-associated promoter regions identified using clover ($p<0.01$), comparison of AAMΦ-associated promoters (TSS -400, +100 bp) against non-AAMΦ-associated macrophage promoters.

2.4.4 TSS IDENTIFICATION AND PROMOTER REGION ANALYSIS

To address whether PPARs facilitate transcription of IL-4Rα-dependent genes *in vivo* we analysed transcription factor binding sites (TFBSs) in the promoters of AAMΦ-associated genes. We developed and optimized a method to accurately identify TSSs using RNA-Seq data (Figure 2-6d). Briefly, transcripts were defined using Cufflinks (Trapnell et al., 2010), and TSSs identified based upon expected gene expression and the observed sequencing read depth at each position in the gene. Proximal promoter sequences (300 bp upstream and 100 bp downstream of each TSS) were obtained from 7,817 high-confidence TSSs, and analysed for over-represented TFBSs.

We validated our method by identifying overrepresented TFBSs in our promoters relative to all mouse promoters using Clover (Frith et al., 2004), and comparing the found set with the macrophage lineage restricted TFBSs identified by Hume *et al* (2010). In our data we identified 54 motifs that were present significantly more often than expected (p-value < 0.01). Hume *et al* (2010) used a different library of TFBS motifs and identified 27 significant motifs, 23 of which were represented in the library we used. Our analysis identified 18 of these (78%) as overrepresented in our macrophage promoters, reinforcing the findings of the previous study and confirming the accuracy of our predictions.

Promoters were categorized according to the expression characteristics of their parent genes. We identified 682 AAMΦ-associated promoters, derived from genes upregulated in WT-NeMΦ relative to both WT-ThioMΦ and IL-4Rα$^{-/-}$-NeMΦ. The remaining 7,135 promoters were classified as generic macrophage promoters. We compared AAMΦ-associated promoters to the generic set, and identified only three over-represented TFBS motifs: MZF1, ZNF354c and PPARG:RXRA (also known as PPAR response element, PPRE), (Table 2-2 and Figure 2-6e). We used the Regulatory Sequence Analysis Toolkit (Thomas-

Chollier et al., 2011) to count the number of instances of each motif within our set of 682 AAMϕ-associated promoters and found that 113 (17%) contained high-confidence PPRE elements. Because the MZF1_1-4 and ZNF354c consensus motifs are very short we could not quantify these at the default significance threshold ($p < 1e^{-5}$). The observation of an abundant and over-represented PPRE identifies PPAR-mediated transcription as significant component of AAMΦ-associated transcriptional regulation *in vivo*.

Jaspar Motif	Clover P-value
MZF1_1-4 Zinc-coordinating	0
PPARG::RXRA Zinc-coordinating	0.004
ZNF354C Zinc-coordinating	0.006

Table 2-2. The three statistically over-represented position weight matrices in AAMΦ-associated promoters relative to all macrophage promoters. Clover P-value determined by permutation test over 1000 iterations.

2.4.5 Analysis and quantification of NeMΦ-derived eicosanoids

GO and KEGG gene set enrichment analysis identified augmented mitochondrial and metabolic gene expression, specifically the TCA cycle, as infection and IL-4Rα-dependent processes (Figure 2-6a,b). Our analysis of AAMΦ-associated *cis*-regulatory features further supported a role for PPAR involvement (Figure 2-6e). Together these strongly suggest that PPAR-mediated transcription is a driving component of the AAMΦ phenotype *in vivo*. PPARγ and oxidative phosphorylation are described as necessary for alternative activation (Odegaard and Chawla, 2011), yet NeMΦ *Pparg* expression was low (Figure 2-6c). As PPARδ is the major PPAR driving TCA cycle flux in adipocytes (Roberts et al., 2011) we hypothesized that it may compensate for PPARγ in NeMΦ. The PPAR family are ligand-dependent transcription factors. Thus, although *Ppard* was not DE we explored the possibility that increased ligand availability might modulate PPARδ-dependent transcription in NeMΦ.

KEGG pathway analysis identified arachidonic acid (AA) metabolism, a known source of PPAR ligands, as up-regulated in WT-NeMΦ (Table A1-ST-6). Products of the AA enzyme cascade, eicosanoids, comprise a wide variety of pro- and anti-inflammatory mediators, including the prostaglandins, leukotrienes and lipoxins. We characterized a broad range of these compounds produced by NeMΦ and ThioMΦ using LC-MS/MS in an independent, homologous experiment (lacking the IL-4Rα$^{-/-}$ thioglycollate group). Individual eicosanoids (for a full list of species see supplementary methods) were measured in peritoneal lavage fluid and *ex vivo* in 12-hour cultures of adherence purified macrophages. *Ex vivo* cultures showed that NeMΦ produced significantly more eicosanoids than ThioMΦ (Figure 2-7a) irrespective of IL-4Rα expression. This cannot be attributed to enhanced activity of the AA-liberating phospholipase A2 enzymes as Pla2 transcripts were expressed in ThioMΦ at a

comparable level to NeMΦ (Figure A1-s12). This suggests IL-4Rα-independent effects of *B. malayi* drive AA catabolism.

Principal component analysis (PCA) of eicosanoid profiles from cultured macrophages showed that the three treatment groups formed distinct clusters (Figure 2-7b). IL-4Rα$^{-/-}$-NeMΦ produced high levels of pro-inflammatory PGE2 and downstream metabolites that promote acute phase inflammation (Trebino et al., 2003). WT-NeMΦ were characterized by production of the 12/15-lipoxygenase-derived hydroxyeicosatetraenoic acid (HETE) metabolites, predominantly 12-HETE (Figure 2-7c). Interestingly, by far the most abundant cox-derivative in WT-NeMΦ was 6-keto-PGF1α, the auto-oxidation product of the PPARδ agonist prostacyclin (PGI2, Lim and Dey, 2002, Figure 2-7d).

Integrating the lipidomic profiles and gene expression data allowed us to better understand the relationship between transcriptional profiles and metabolic phenotype. We plotted gene expression and metabolite data onto a pathway map of the enzymatic cascade, relating AA catabolism to metabolite production (Dennis et al., 2010). Alterations in synthase gene expression showed high concordance with the relative abundances of daughter metabolites (Figure 2-7e), with one notable exception. Despite robust expression of *Gpx1*, *Alox5* and *Lta4h* we were unable to detect any 5-lipoxygenase derived metabolites, 5-HETE 5-OxoETE, LTB4 or, 20-OH LTB4 (Figure A1-s13). Thus peritoneal macrophages were primed to produce leukotrienes but did not. Substrate competition for AA or higher-level control of leukotriene biosynthesis may explain this observation. In summary, eicosanoid production by macrophages is stimulated in response to *B. malayi*. Rather than affecting total eicosanoid production, IL-4Rα-dependent signalling modulated the architecture of the downstream enzymatic cascade, leading to a more anti-inflammatory eicosanoid profile.

Heterotypic interactions contribute towards the eicosanoid profile *in vivo*. To discern the relative contributions of macrophage and non-macrophage derived eicosanoids, we measured AA products in peritoneal lavage from the mice in the same experiment (Figure A1-s13). PCA plots (Figure 2-7f) showed no distinction between WT thioglycollate-injected and IL-4Rα$^{-/-}$-*B. malayi*-implanted mice. The WT *B. malayi*-implanted individuals however clustered distinctly, showing that the response to *B. malayi* implantation generated a unique, IL-4Rα-dependent, eicosanoid environment. This also implies that *B. malayi* is not a major source of total peritoneal eicosanoids in this model. WT-implanted mice had increased HETE, PGD2 and TxB2 concentrations (Figure 2-7g), the latter two of which are eosinophil chemoattractants (Arima and Fukuda, 2011; Ishizuka et al., 1998)'. It is unlikely that the PGD2 was macrophage-derived as low levels were produced in AAMΦ *ex vivo* cultures. TxB2 concentrations correlated with, and may be explained by, the increase in macrophage numbers associated with infection in a WT environment (Figure A1-s14a, b). We have therefore identified a distinct lipid effector profile in the response to helminth infection. While the activity of 5-lipoxygenase did not contribute towards this profile, the actions of cyclooxygenase and 15-lipoxygenase did.

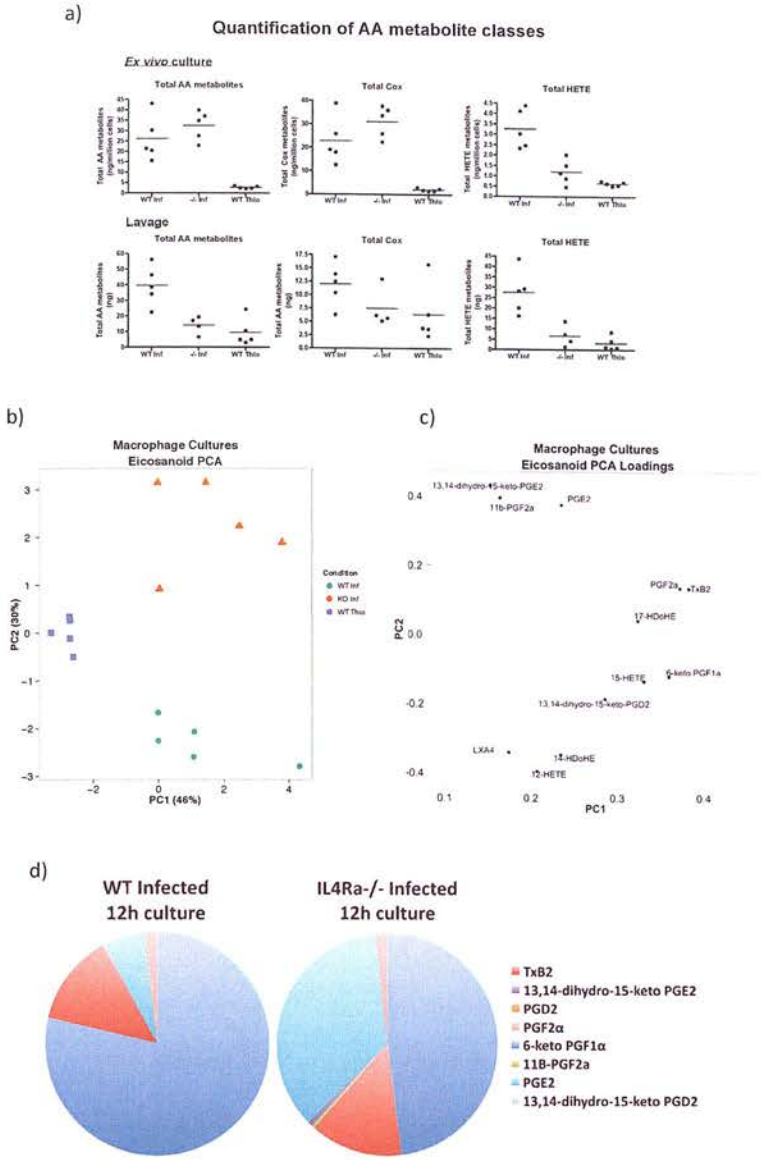


Figure 2-7 Characterization of AAMΦ-derived eicosanoids.

A) Quantification of arachidonic acid (AA) metabolites in 12h cultures of purified macrophages (upper) and in peritoneal lavages (lower).

B) PCA scores plot for 12h cultures of purified macrophages.

C) PCA loading plot for b) showing how individual eicosanoids contribute towards the primary and secondary principal (x and y respectively) axes.

D) Pie chart showing the breakdown in the production of cyclooxygenase metabolites between WT-NeMΦ and IL-4Rα$^{-/-}$-NeMΦ.

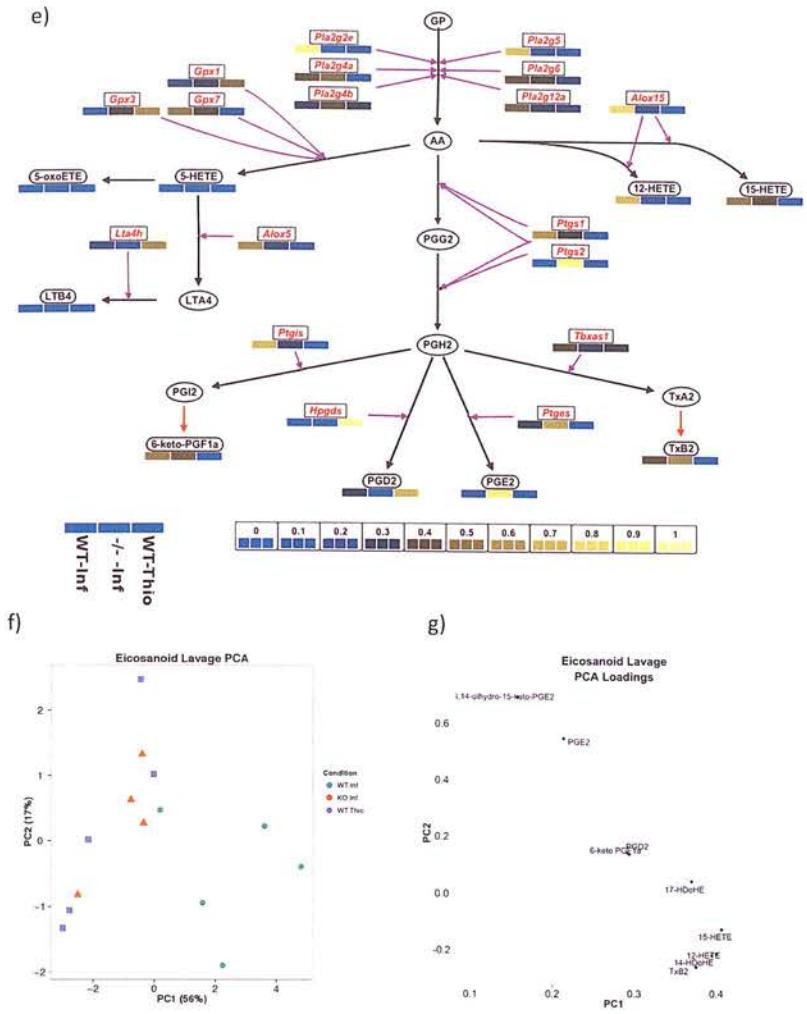


Figure 2-7 Characterization of AAMΦ-derived eicosanoids.

E) Schematic of the AA cascade reproduced using Vanted. Genes are shown in square boxes with red text, measured metabolites are in rounded boxes with black text. Black arrows indicate enzymatic reactions, red arrows auto oxidation. Relative expression values (normalized such that the total over all 3 conditions =1) are beneath the associated gene/metabolite. WT *B. malayi* (left box) IL-4Rα$^{-/-}$ *B. malayi* implanted (centre) and WT-Thioglycollate (right).

F+G) PCA scores and loading plot as in b) and c) for the peritoneal lavages.

2.5 Discussion

We have characterized the phenotype of *B. malayi*-elicited AAMΦ using RNA-Seq and a targeted lipid analysis. Together these reveal the effects of IL-4Rα dependent signalling on macrophage physiology in a chronic Th2 environment. Our findings strongly support the emerging paradigm that alternative activation and inflammation are associated with wholesale alterations in metabolism (Odegaard and Chawla, 2011). Further, GSEA and a detailed analysis of AAMϕ-associated chemokine and cytokine expression provide intriguing clues as to the physiological role(s) of AAMΦ *in vivo*.

One of the more unexpected findings of our analysis was that, collectively, macrophage complement genes were amongst the most abundantly expressed and differentially regulated in the response to helminth challenge, and may represent a previously unrecognised effector axis of AAMΦ. A requirement for complement in anti-helminthic immunity has previously been demonstrated, as C3 mediates clearance during both *Strongyloides stercorallis* (Kerepesi et al., 2006) and *Schistosoma mansoni* infection (La Flamme et al., 2003). In the context of *B. malayi*, complement interacts with microfilariae (Carter et al., 2007) and may contribute towards their clearance. Here we identify AAMΦ-derived FicolinA as a candidate for further study into the regulation of the complement response to filarial nematode infection.

We also identified IL-4Rα-dependent expression of cytokines not currently associated with AAMΦ physiology including *Bmp6* and *Wnt2*. Intriguingly *Bmp6* has been associated with increased MΦ iNOS activity, suggesting that it may curtail alternative activation (Hong et al., 2009). *Wnt2* is a highly abundant AAMΦ-derived cytokine with no identified role in macrophage biology. *Wnt2* does however affect hematopoietic lineage commitment (Wang et al., 2007), and our AAMΦ-up clusters contained *Cdh1* (E-cadherin), an AAMΦ

marker in both humans and mice (Van den Bossche et al., 2009). Given the known convergence between cadherin and Wnt pathways, further work to elucidate the role of *Wnt2* in macrophage biology is warranted. We have observed IL-4Rα and infection dependent production of cytokines, chemokines and eicosanoids suggesting that NeMΦ aid in the orchestration and maintenance of the immune environment. Specifically, NeMΦ potentially regulate eosinophil recruitment via multiple, redundant, mechanisms. Reese *et al* (2007) previously demonstrated chitin-induced eosinophilia as an AAMΦ-dependent process, but whether they fulfil this role in the more complex environment of live parasite infection is unclear. We show here that WT-NeMΦ transcribe eosinophil chemotactic factors *in vivo*. *Ccl8* (Romagnani, 2002) and *Ccl24* (Mantovani et al., 2004) operate as *Ccr3* ligands, while TxA2, the bioactive precursor of TxB2, induces eosinophilia by affecting vascular epithelial integrin expression. Our data are therefore consistent with AAMΦ-mediated recruitment and maintenance of eosinophilia during helminth infection. Interestingly, in contrast to the work by Reese *et al* (2007) our proposed mechanism for eosinophil recruitment is LTB4-independent. Our findings may differ because Reese *et al* used an acute stimulus, whilst we employ a chronic model of helminth infection. Nevertheless, our data suggest cooperative interactions between AAMΦ-derived protein and lipid mediators ensure robust, tightly regulated eosinophil recruitment to sites of infection.

IL-4Rα-dependent signalling led to reduced expression of multiple pro-inflammatory cytokines and chemokines, supporting a model wherein AAMΦ are anti- or non-inflammatory. Macrophage PPAR activity augments alternative activation and reduces the expression of pro-inflammatory cytokines (Odegaard and Chawla, 2011). Indeed thiazolidenediones, anti-diabetic PPARγ agonists, improve insulin sensitivity in obese white

adipose tissue (WAT) partly by curtailing CAMΦ (Odegaard and Chawla, 2011). Global similarities can be drawn between our surveyed NeMΦ and the beneficial AAMΦ in WAT of lean individuals, these include enhanced oxidative metabolism and low pro-inflammatory cytokine expression (Odegaard and Chawla, 2011). By assaying TFBS in NeMΦ promoters we provide compelling evidence that PPAR-mediated transcription is an important facet of the macrophage response to helminth infection *in vivo*. WAT AAMΦ and NeMΦ are however distinct subsets. WAT AAMΦ express PPARγ, *Il10* and *Cd36* (Liang et al., 2004), while *Cd36* and PPARγ were down-regulated, and *Il10* was not expressed, in NeMΦ (Table A1-ST-1). This is consistent with alveolar AAMΦ failing to express IL-10 in response to *Nippostrongylus braziliensis* (Chen et al., 2012). However, the observation that *N. braziliensis* infected mice show improved glucose tolerance (Shapiro et al., 2011) demonstrates that the physiological consequences of modulating alternative activation, either in the response to infectious disease, or during homeostasis, imparts similarities that are greater than the observed differences between WAT AAMΦ and NeMΦ. This is perhaps not surprising as alternative activation in both these disparate environments is STAT6-dependent. An improved understanding of the influence of context-specific (i.e. tissue/infection model) cues in fine-tuning AAMΦ polarization is required to fully appreciate the functional diversity of AAMΦ phenotypes.

We observed low PPARγ expression in NeMΦ, consistent with a recent report of PPARγ expression profiles in different macrophage subsets (Gautier et al., 2012b). Analysis of AAMΦ-associated *cis*-regulatory regions suggested a role for PPARs in WT-NeMΦ, and we therefore sought a mechanism for PPAR-dependent transcription during *B. malayi* infection. By profiling macrophage-derived eicosanoids we found that nematode- but not thioglycollate-elicited macrophages abundantly produced the PPARδ ligand PGI2. Due to

the low expression of PPARγ in NeMΦ, and constitutive PPARδ expression, macrophage PGI2 production provides a feasible and testable explanation for the maintenance of PPAR-dependent transcription in NeMΦ. However it should be noted that because NeMΦ do express PPARγ we cannot rule out a role for both these transcription factors during alternative activation. We have shown that macrophage eicosanoid generation is IL-4Rα independent in this system. This suggests that the liberation of PPAR ligands is *B. malayi*-mediated, and is perhaps achieved via a pathogen associated molecular pattern-dependent mechanism. IL-4Rα-dependent signalling did however alter the expression of AA catabolic enzymes, enhancing the production of anti-inflammatory HETEs (Allen and Maizels, 2011). Hence, extracellular cues (i.e. IL-4/IL-13) manipulated the architecture of the enzymatic cascade, reorganizing the eicosanoid landscape. Ultimately this reduced pro-inflammatory eicosanoid production, and increased the synthesis PGI2 and anti-inflammatory HETEs. We propose a model in which the presence of a 'danger signal' is coupled with cytokine-dependent structural reorganization, providing both an initiator and higher order specificity that contribute towards sustaining the NeMϕ phenotype *in vivo*.

We have defined the phenotype of IL-4Rα-stimulated AAMΦ *in vivo* during the response to a nematode infection. All of the data we present are supportive of anti-inflammatory roles for AAMΦ in this setting. To summarize, WT-NeMΦ broadly down-regulated the expression of numerous pro-inflammatory cytokines and chemokines. The eicosanoid environment generated by NeMΦ is defined by enhanced production of anti-inflammatory HETE metabolites, and also the abundant production of the PPARδ ligand PGI2. We infer that NeMΦ have a decreased migratory capacity, due to the down-regulation of multiple chemokines, and therefore are unlikely to prime lymphatic T-cell responses. Our study has also identified the production of complement as a major component of the macrophage

response to filarial nematode challenge. While complement may aid in the recognition and clearance of either adult *B. malayi* or microfilariae, it is also has extensive immuno-regulatory functions (Schmidt and Gessner, 2005). For example, complement C5a interacts with surface Fc receptors to modulate inflammation (Schmidt and Gessner, 2005). In this context our WT-NeMΦ strongly up-regulated expression of the regulatory *Fcgr2b* (Table A1-ST-1). Thus the complement-Fc receptor axis induced by AAMΦ in the response to nematode challenge may be an additional anti-inflammatory feature of IL-4Rα-dependent macrophage activation.

2.6 Acknowledgements

This work was supported by a Wellcome Trust PhD studentship award to GDT, an MRC programme grant to JEA (MRC-UK G0600818) and MRC core funding to the GenePool facility to MLB (MRC 60900740). BM and PDW are supported by the European Regional Development Fund (ERDF), Highlands and Islands Enterprise (HIE) and the Scottish Funding Council (SFC).

We thank Marian Thompson in the GenePool genomics facility for support in RNA-Seq data generation and Martin Waterfall for expertise with flow sorting. Additionally we would like to thank Tom Freeman, David Hume, Stephen Jenkins and Lucy Jones for critical reading of the manuscript.

2.7 Authorship Contributions

GT, DR, MB and JA devised experiments. GT performed most of the experimental work and analyses. Lipidomics were performed by BM, PW and GT. The manuscript was written by GT, MB and JA, with contributions from all authors.

2.8 Conflicts of Interest

The authors declare no competing financial interests.

3 TSS-PREDICTOR: IDENTIFICATION OF *CIS*-REGULATORY REGIONS WITH RNA-SEQ DATA

3.1 INTRODUCTION

Eukaryotic transcriptional regulation leads to the generation of multiple transcriptional isoforms from individual genes arising from regulated changes in exonic splicing and promoter usage. Differential promoter usage leads to the production of mRNA isoforms with multiple transcription start sites (TSS). The identification of these TSS provides a useful starting point to analyse mechanisms of differential gene expression. By defining *cis*-acting transcription factors laying in promoter sequences proximal to TSS it is possible to gain key insight into the processes governing cellular responses to developmental and environmental cues (Suzuki et al., 2009).

Second-generation sequencing technologies have catalysed a shift in our ability to investigate transcriptomes. Deep sequencing of cDNA (RNA-Seq) is increasingly superseding microarray technology as the tool of choice for transcriptomic profiling, primarily due to the greatly improved sensitivity and dynamic range offered by RNA-Seq (Wang et al., 2009). Methods have been developed to exploit new opportunities offered by RNA-Seq including facilitated transcript discovery (Trapnell et al., 2010), alternative splicing (Shen et al., 2012), RNA editing and SNP calling (Li et al., 2009). RNA-Seq also has the potential to identify major TSS usage for genes. However, TSS identification is hampered by the fact that multiple TSSs exist for individual genes. As a consequence of this dedicated methods for TSS identification from RNA-Seq data do not currently exist.

Over the past decade multiple large consortia, including the FANTOM and ENCODE projects, have contributed significantly towards the functional annotation of high quality genomes such as human and mouse (Carninci et al., 2005; Consortium, 2004). A major

component of the annotation effort has been the high-throughput deployment of CAGE (cap analysis gene expression) technology to map TSS at base-resolution genome wide. CAGE works by capturing and sequencing the first 20 bases of mRNA transcripts from the 5' end (Shiraki et al., 2003). These sequenced 'tags' and then mapped to the reference genome and counted, providing expression profiling and TSS mapping simultaneously. This work has led to the generation of near, if not complete, maps of TSS usage genome wide (Carninci et al., 2005). In this chapter I present TSS-Predictor, a pipeline that leverages high quality gene annotations to predict major TSS usage for genes from RNA-Seq experiments.

TSS-Predictor achieves this by revising gene predictions produced by the reference-guided transcript assembler Cufflinks (Trapnell et al., 2010). Cufflinks assembles overlapping sections of mapped reads into transcribed fragments, called 'transfrags' which are then related to known gene models. In principle this is sufficient to allow TSS identification as, for each gene, the most abundant set of Cuffcompare transcripts sharing a common TSS ought to reflect the primary TSS.

Anecdotal evidence gained from visualizing reads mapped to the reference genome alongside Cufflinks and Ensembl gene predictions indicated that Cufflinks tended to systematically predict TSS upstream of the true TSS, for an example see Fiure 3-5 (p 95). I was therefore not satisfied with Cufflinks TSS predictions and devised a strategy to improve on these. This is now implemented in the publicly available program TSS-Predictor. TSS-Predictor is an embodiment of the method developed in Chapter 2. Significant improvements have been made in the usability of the program, and several steps that were previously performed manually, such as the selection of transfrags for trimming, are now automated. Figures are also automatically generated to help the user assess the success of

TSS trimming, and an optimization step can be performed to identify the optimal parameters for TSS identification.

In this chapter I first describe the scripts in the TSS-Predictor pipeline, explaining how the algorithm works and how the two main programs, *TSS-Optimizer.pl* and *TSS-Predictor.pl*, can be used to generate high-quality TSS predictions from RNA-Seq data. Using the RNA-Seq dataset generated in chapter 2 it is shown that TSS-Predictor improves TSS inferences from Cufflinks. Finally I present the public interface for the TSS-Predictor project (tss-predictor.bio.ed.ac.uk). This website contains links to the project download page and extensive help documentation including a walkthrough example for generating TSS predictions from raw FASTQ data.

3.2 METHODS

3.2.1 DATA INPUT AND PRE-PROCESSING

Prior to using TSS-Predictor transcript models must be predicted from the data. Sequence reads are mapped to a reference genome using TopHat allowing for multiple-mapping reads. Reads that map uniquely to the genome are desirable for expression quantification because biases are introduced by multiple-mapping reads (Mortazavi et al., 2008). However this restricts the 'mappability' of certain repetitive regions of the genome that may include highly conserved signalling motifs, interaction domains, and closely related genes (Koehler et al., 2011). Effective transcript assembly demands coverage across the length of a gene and thus requires the use of multiply mapping reads (Trapnell et al. 2010). Mapped reads are next assembled using Cufflinks and assigned to reference genes using Cuffcompare from the Cufflinks suite. In the working example presented in this chapter mouse macrophage RNA-Seq data generated in Chapter 2 were mapped to the mouse mm9 reference genome and assigned to the Ensembl v67 reference annotation set.

3.2.2 THE TSS-PREDICTOR ALGORITHM.

Below I describe the function of the individual scripts in the TSS-Predictor program. For an overview of how these are linked to produce the final TSS-Predictor program see Figure 3-1.

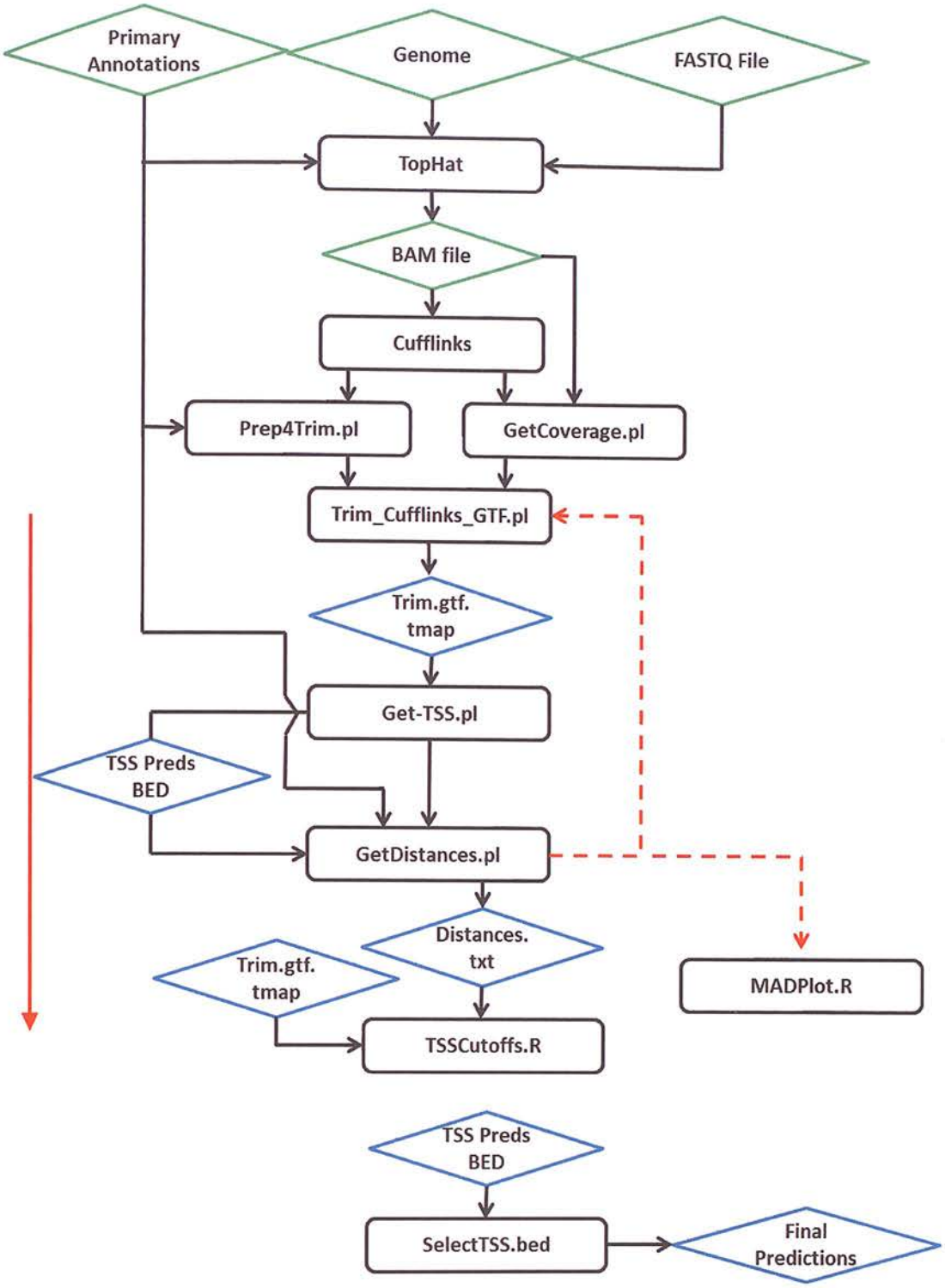


Figure 3-1. Overview of the TSS-Predictor algorithm. Rectangles represent scripts or programs. Diamonds depict data files, green diamonds are input files from external sources whereas blue diamonds are produced by TSS=Predictor. The red solid line shows the portions of the script covered by TSS-Predictor.pl and the red dashed line shows portions of the script covered by TSS-Optimizer.pl

3.2.2.1 *GetCoverage.pl*.

The first step in the TSS-Predictor pipeline (see Figure 3-1) is the calculation of per-base read coverage at each transcribed location on the genome. This is achieved using BEDTools applications (Quinlan and Hall 2010). The user provides the 'combined.gtf' file produced by Cufflinks Cuffcompare, which contains the exon information for all 'transfrags' produced by Cufflinks including their locations, linking them to reference annotations where applicable. *GetCoverage.pl* uses BEDTools *mergeBed* to 'flatten' all predicted gene models into a list of features that are deemed to be transcribed. *GetCoverage.pl* next accesses the BAM file produced by TopHat which contains the positions of all sequenced reads aligning to the reference genome. The BEDTools feature *intersectBed* is used to identify all mapped reads overlapping transcribed regions. This data is then passed to *genomeCoverageBed* which counts the number of observed reads overlapping each transcribed base on the reference genome. The output of *GetCoverage.pl* is a three-column file containing the chromosome, position and observed sequencing depth.

3.2.2.2 *Prep4Trim.pl*.

The second step in TSS-Predictor is the selection of a representative Cufflinks 'transfrag' for each reference gene. The user provides *Prep4Trim.pl* with the reference annotation set used by Cufflinks Cuffcompare (Ensembl v67 annotations) and two Cufflinks output files. The 'transcripts.gtf' annotation file of transfrags produced by Cufflinks, and the Cuffcompare 'gtf.tmap' file containing additional information for to each transfrag including the transcript per-base expression level and the Ensembl transcript ID to which each transfrag is linked. *Prep4Trim.pl* assigns cufflinks transfrags to reference gene using the primary reference annotations, and calculates expected per-base coverage for each

reference gene by summing length-normalized per-base coverage estimates for all transfrags linked to the gene as described in Figure 3-2.

The longest Cufflinks transfrag, on the same strand, is selected for each reference gene. *Prep4Trim.pl* produces two files – 'trim.combined.gtf' and 'trim.gtf.tmap'. 'Trim.combined.gtf' is a reduced version of the Cufflinks 'combined.gtf' file and contains only the set of longest transfrags selected for trimming. 'Trim.gtf.tmap' is an amended version of the 'gtf.tmap' output. This contains a mapping between Cufflinks transfrags that are to be trimmed and the reference gene, as well as the amended coverage estimate for the gene. All other fields in 'trim.gtf.tmap' file are transfrag level information generated by Cuffcompare.

$$C_{(G)} = \sum_{i=1}^{n} \frac{C_{(Ti)} L_{(Ti)}}{L_{(Tmax)}}$$

Figure 3-2. Per-gene coverage calculation. Each expressed reference gene (*G*) is composed of *n* transfrags (*T*), each with coverage (*C*) and length (*L*). Thus, C_{Ti} is the coverage for the i^{th} transfrag of gene *G*, and L_{Ti} is the length of the i^{th} transfrag for gene *G*.

3.2.2.3 *TRIM_CUFFLINKS_GTF.PL*

Trim_Cufflinks_GTF.pl parses the expected gene-level coverage information from the 'trim.gtf.tmap' file and transfrag information from 'trim.combined.gtf' produced by *Prep4Trim.pl*. Using the per-base coverage information calculated by *GetCoverage.pl* each selected isotig is trimmed at the position where the observed expression drops below a pre-defined fraction of the expected coverage for the gene (X). The estimation of a credible value for X is described below.

To do this *Trim_Cufflinks_GTF.pl* iterates, one base at a time, from the last base (3') in the first exon of each selected transfrag 5' towards the first base, provided that the observed coverage at that base is $> X * C_G$. The annotation is trimmed once the per-base coverage of the selected transfrag drops below $X * C_G$. If the observed expression of the final base in the first exon is less than $X * C_G$ then the second exon is examined, so on and so forth. Figure 3-3 provides a graphical representation of the trimming policy implemented by *Trim_Cufflinks_GTF.pl* as well as the tendency for Cufflinks to overestimate transcript length.

Trim_Cufflinks_GTF.pl generates an 'isotigs.trimmed.gtf' file that contains revised transcription start site predictions for each isotig. An additional variable "contain_tss" is added to the 9th field of the GTF file to identify exons predicted to contain transcription start sites.

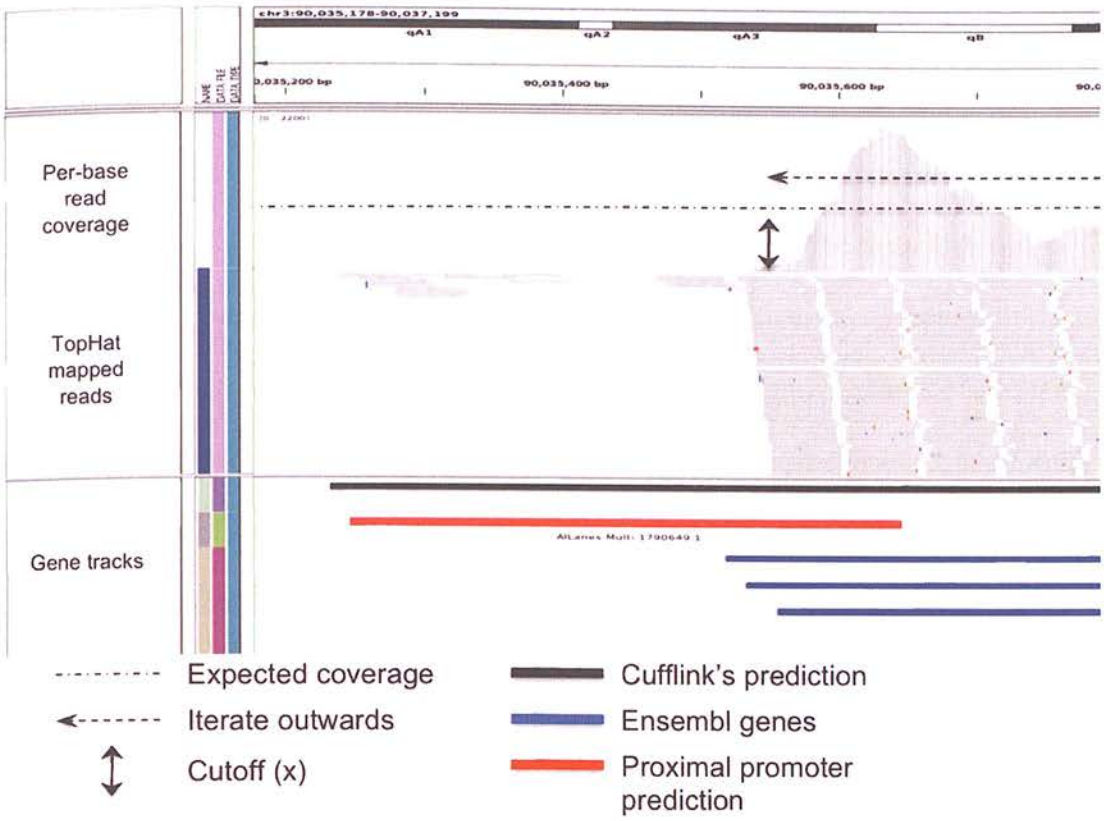


Figure 3-3. Overview of TSS-predictor's reverse-trimming policy as described in section 4.2.2. The top section shows a histogram of per-base read coverage. In the lower panel it can be seen that the Cufflinks gene prediction (black) overestimates the true gene primary TSS by including a few spurious reads in the gene annotation. The true TSS is marked by a precipitous drop in read coverage around the Ensembl-defined TSS (blue). The promoter region defined by TSS-predictor (default value 300bp upstream, 100bp downstream of TSS) predictions provides a substantial improvement on the cufflinks predictions.

3.2.2.4 *GET-TSS.PL AND GETDISTANCES.PL.*

These are the final scripts in the main TSS-Predictor and TSS-Optimizer pipelines. Get-TSS.pl parses the 'isotigs.trimmed.gtf' file and reports transcription start sites based on the presence of the "contain_tss = yes" flag in the 9th fields of the GTF file. By default Get-TSS.pl reports a one base feature identifying the TSS, however to extract promoter sequences of arbitrary length the flags '–u' and '–d' may be set, this produces a BED file containing upstream and downstream regions '–u' and '-d' bases from the TSS. BED format TSS predictions contain chromosome, start position, end position, reference gene and strand information. GetDistances.pl generates a tab-delimited text file containing the reference gene, chromosome, predicted transcription start site, nearest reference transcript, nearest reference transcription start site and the distance between the two predictions.

3.2.2.5 *TSS-OPTIMIZER.PL*

This script is a wrapper for the main TSS prediction algorithm and is used to estimate the cutoff fraction (*X*). *TSS-Optimizer.pl* takes one chromosome, defined by the user and predicts transcription start sites over a range of cutoff fractions specified by the '*-s*' start, '*-e*' end and '*-i*' increment flags. *TSS-Optimizer.pl* reports, for each gene at each cutoff fraction, the distance between the predicted TSS and nearest annotated TSS for the same gene. TSS-Optimizer also R scripts to produce a graph showing the median absolute deviation (MAD) between the predicted and nearest annotated transcription start sites at each cutoff fraction. The cutoff fraction (*X*) that gives the lowest MAD is determined as optimal for use with TSS-Predictor.

A number of files are produced by *TSS-Optimizer.pl*. Firstly, a reduced version of the '.coverage' file is created for the chromosome of interest. This simple step substantially reduces the time taken to iterate over each coverage cutoff. *TSS-Optimizer.pl* produces multiple copies of 'isotigs.trimmed.gtf', 'distances.txt' and 'tss.prediction.bed', one for each cutoff fraction interrogated. Finally a matrix file 'CutoffDistances.txt' containing distance information for each gene at each coverage cutoff is produced.

The most informative output of *TSS-Optimizer.pl* is the graph 'Optimizer-MADPlot.jpg'. This bar chart, produced by the R script '*MADPlot.R*', shows the MAD between TSS predictions and nearest annotated TSS over all cutoff fractions assessed. The cutoff fraction yielding the lowest MAD is shown in red and should be given to '*TSS-Predictor.pl*' to determine TSS usage genome wide. The actual MAD calculated by TSS-Optimizer contains no post-filtering and is therefore typically much greater than the value obtained in the final predictions.

3.2.2.6 *TSS-PREDICTOR.PL*

TSS-Predictr.pl identifies TSS from RNA-Seq data genome-wide. Similar to *TSS-Optimizer.pl* this script acts as a wrapper for *Prep4Trim.pl*, *Trim_Cufflinks_GTF.pl*, *Get-TSS.pl* and *GetDistances.pl*. Accordingly, the user provides *TSS-Predictor.pl* with the reference annotations, the cufflinks directory, Cuffcompare prefix and path to the 'cuffcompare.tmap' file, the cutoff fraction (*X*) determined using *TSS-Optimizer.pl* and the number of bases upstream and downstream of the TSS prediction to report. A BED format 'tss.prediction.bed' file is produced alongside the 'distances.txt' file produced by *GetDistances.pl*. An essential step for obtaining high quality TSS predictions is filtering the data. The two strategies for filtering we implement are depth-based and distance-based

filtering. While filtering is performed using the *'SelectTSS.pl'* script, figures are produced by *'TSS-Predictor.pl'* to aid the selection of appropriate cutoff conditions.

3.2.2.7 *SelectTSS.pl*

This script is used to filter the dataset after running *TSS-Predictor.pl*. The user provides either an expression, or distance based cutoff using the flags *'-e'* or *'-d'*. *Select.tss.pl* reports a BED file containing only those genes with expression greater than the specified cutoff, or predicted TSS less than the specified distance from the nearest annotated TSS for the gene. If both *'-e'* and *'-d'* are specified then both criteria must be met for a gene to be reported. Additionally the user may specify whether to report the predicted TSS, or the TSS of the closest reference transcript using the flag '-r'. When a high quality reference annotation set us used, high quality macrophage TSS are obtained after removing both lowly expressed genes and TSS predictions a 'large' distance from the nearest annotated TSS. As the promoter regions surrounding TSS used for *cis*-regulatory analysis are typically quite small (500-1000 bp), TSS predictions further than the centre of the desired promoter region to the nearest annotated TSS are typically removed (500 bp).

3.2.3 Using TSS-Predictor with macrophage RNA-Seq data

TSS-Predictor was applied to our macrophage RNA-Seq dataset to demonstrate the identification of high-quality TSS predictions. All 12 RNA-Seq libraries were mapped to the reference genome with TopHat using default parameters to permit multi-mapping reads. In order to obtain deep sequencing coverage for transcript discovery and TSS prediction, all mapped read files were merged and sorted by chromosome using SAMTools, aligned reads were then assembled using Cufflinks (Trapnell et al. 2010). Ensembl reference annotations

were downloaded from the Ensembl FTP site and converted from NCBMI37 to UCSC mm9 format using *Ensembl2UCSC.pl*, available at the TSS-Predictor GitHub page. After the creation of a '.coverage' file with *GetCoverage.pl*, *TSS-Optimizer.pl* was run with the default parameters, using chromosome 1 as a subsample. *TSS-Optimizer.pl* predicts TSS for cutoff (*X*) values between and 1% and 20% of the expected coverage for the gene in 1% increments. *TSS-Predictor.pl* was then run at a cutoff of 4% to identify TSS-usage genome-wide as informed by the *TSS-Optimizer.pl* MAD plot. Finally, to generate a list of high-quality TSS predictions, genes with TSS greater than 500bp from the nearest annotated TSS, and with a per-base expression less than 65 were disregarded using *SelectTSS.pl*.

3.3 Results

3.3.1 TSS-Predictor and validation with Macrophage-RNA-Seq data

3.3.1.1 ***TSS-Predictor Application.*** The utility of the TSS prediction method devised for chapter 2 has been extended into an application, TSS-Predictor, which may be used by the wider research community. A detailed overview of the algorithm, script by script, is provided above. Below I present a working example of TSS-Predictor, Appendix 2 contains a list of the software versions used in this example. Extensive help documentation for the program including a walkthrough example of how to generate high-quality TSS-Predictions from raw sequence data using our macrophage RNA-Seq dataset is available on the website and is included in Appendix 3. Here we review the outputs of the various stages of the TSS-Predictor algorithm and describe how they inform user choice to obtain high quality predictions.

3.3.1.2 ***Transcription start site prediction with TSS-Predictor.*** Following read mapping and transcript assembly with Cufflinks, a '.coverage' file was generated and the optimal cutoff fraction determined to be 4% (0.04) using *TSS-Optimizer.pl* (Figure 3-4). The lowest value highlighted in red in Figure 3-4 is chosen as the optimal coverage cutoff and should be provided to *TSS-Predictor.pl* in the next step of the TSS-Prediction process.

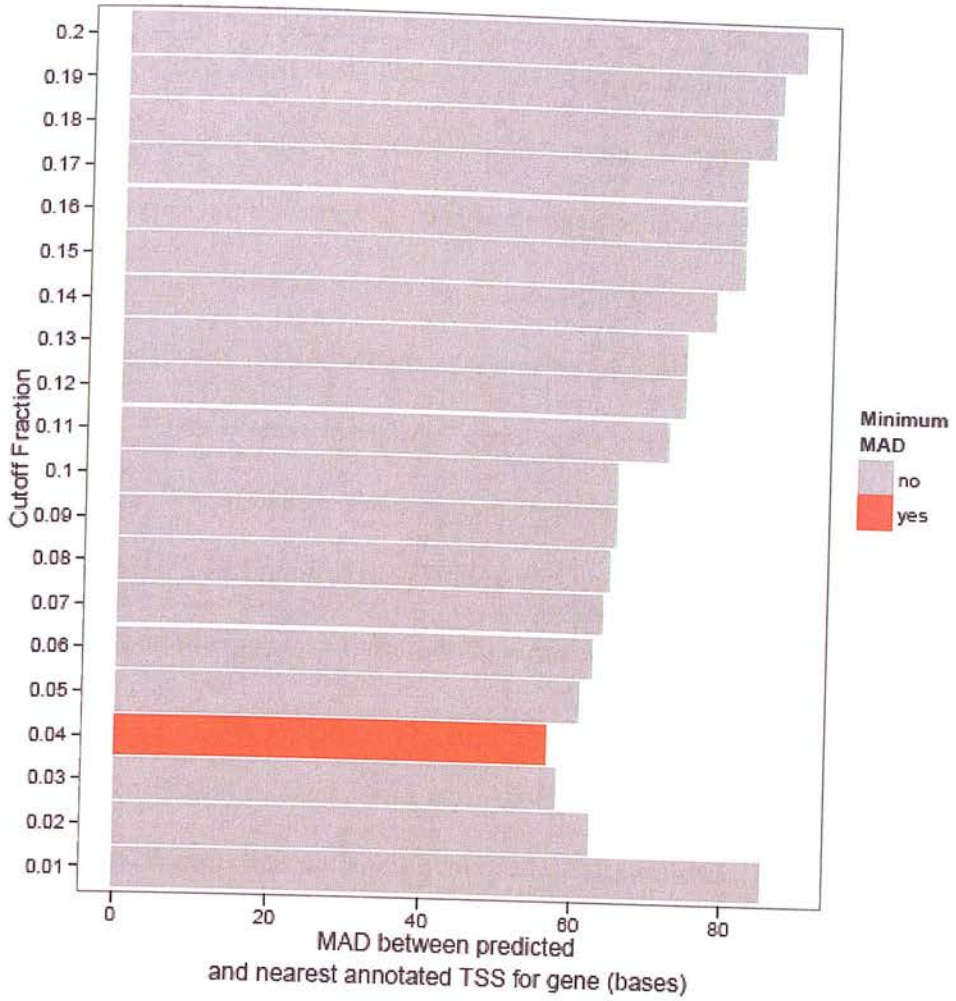


Figure 3-4. Optimizer-MADplot produced by *TSS-Optimizer.pl*. The trimming cutoff condition that generates the minimum MAD between TSS predictions and the nearest annotated TSS is shown in red (0.04, 4%) and should be used during TSS-Predictor.

Following the identification of the optimal cutoff fraction TSS were predicted genome-wide using *TSS-Predictor.pl*. As described in the methods section, *TSS-Predictor.pl* identifies TSS usage genome-wide for all expressed reference gene, i.e. those that have been assigned to at least one cufflinks transfrag. Low abundance transcripts are likely to assemble incorrectly due to incomplete coverage across the length of the gene. *SelectTSS.*pl is designed to filter potentially incorrect TSS predictions by removing lowly expressed genes, but this required user input. *TSS-Predictor.pl* assists user decision making by providing a graphical overview of the MAD for genes binned by expression level (by calling the R script *TSSCutoffs.R*). Genes are binned according to their expression level, and the MAD between TSS predictions and the nearest annotated TSS is calculated for each bin (Figure 3-5). The median expression level of the first bin (lowest expression) with a MAD less than the MAD for the entire dataset is taken as the expression cutoff for depth-based filtering (Figure 3-5). An additional approach to obtaining high-quality TSS predictions is to disregard predictions that lie a large distance away from the nearest annotated TSS. This assumes that the annotation set provided to TSS-Predictor is of high quality. Figure 3-6 shows that the majority of TSS predictions are within 500bp of the nearest annotated TSS. Anecdotal evidence from our RNA-Seq dataset suggested that highly expressed genes with TSS predictions >500 bases from the nearest predicted TSS tended to arise from incorrect transcript assembly, or selection of an incorrect Cufflinks transfrag for trimming. It is possible however that identifying abundant TSS a large distance from the nearest annotated TSS provides a strategy to define novel TSS usage, thus the selection of a distance-based cutoff can be suppressed. To generate our high-quality TSS predictions here I remove TSS belonging to both lowly expressed genes (per-base coverage <65) and whose distance is greater than 500bp from the nearest annotated TSS for the gene.

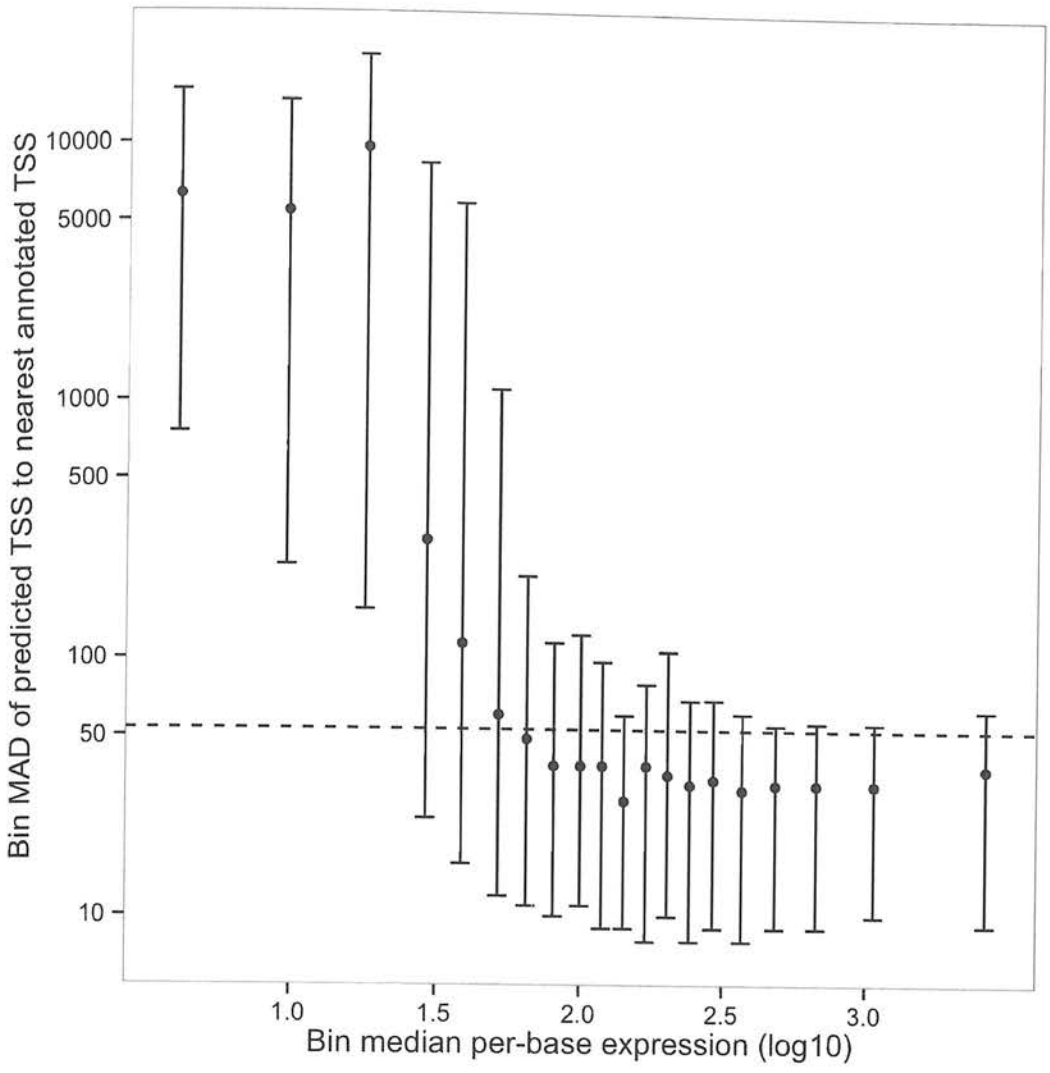


Figure 3-5. MAD-expression plot produced by *TSS-Predictor.pl*, each point shows the median MAD for genes within each bin, grouped by per-base expression level. The dashed line shows the MAD for all TSS predictions over the entire dataset. Error bars depict the interquartile range within each bin.

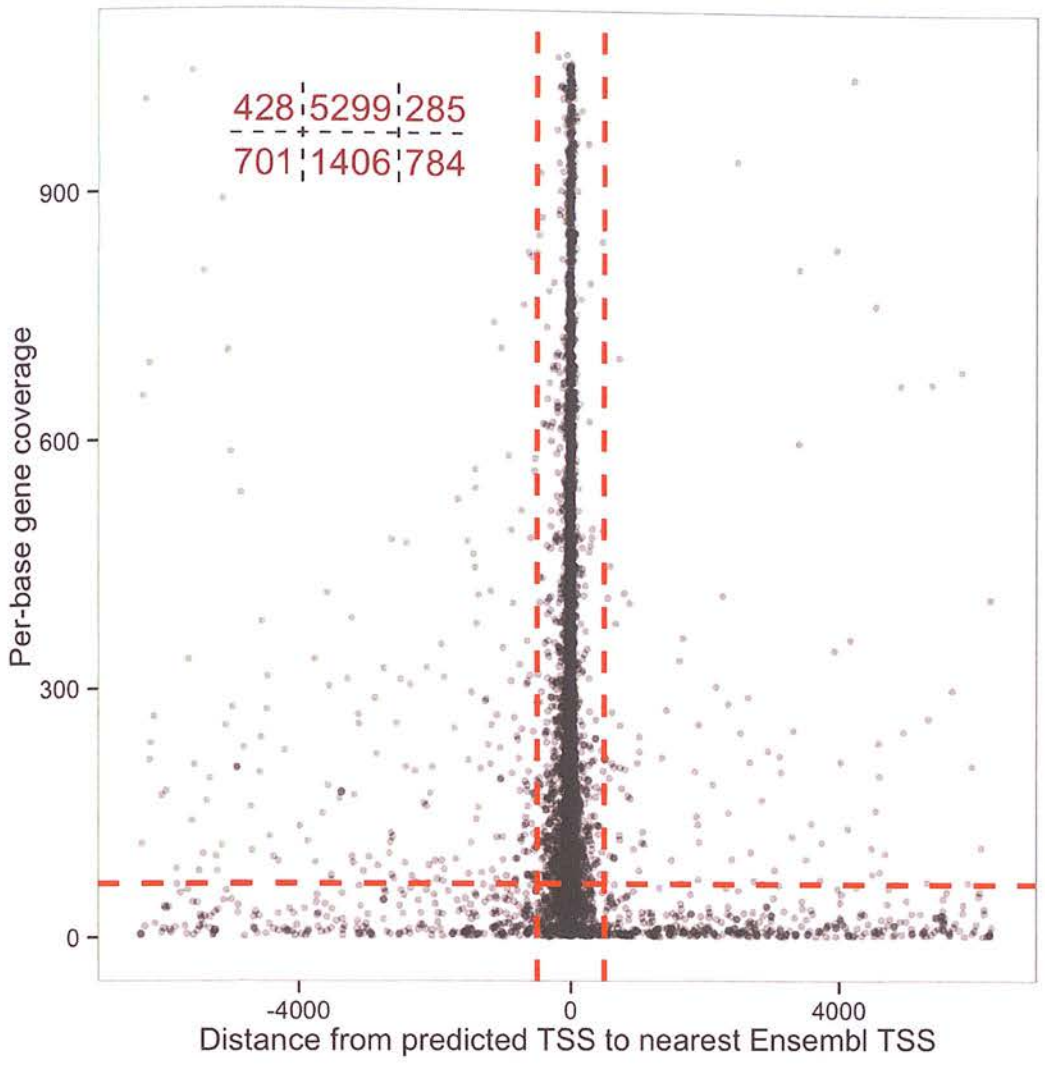


Figure 3-6. The expression cutoff plot produced by TSS-Predictor.pl (x-axis is truncated at +/-6000bp and the y-axis at the median per-base coverage of the bin with the most abundant coverage) the horizontal red line shows the effect of removing lowly-expressed genes with the approach outlined in Section 4.2.2 and depicted in Figure 4-4. The value for the horizontal dashed red line is the median expression level of the first bin with a lower MAD than the population MAD shown in Figure 4-5 and is reported to the screen by *Tss-Predictor.pl*. In this instance the advised per-base coverage cutoff is 65.09. The vertical red line depicts a 500bp window either side of the nearest annotated TSS. The numbers within the embedded table reflect the number of genes in each quadrant of the plot. Note these do not accurately reflect the data shown due to the truncation of the axes.

3.3.1.3 *VALIDATION OF TSS-PREDICTOR ACCURACY.* To validate the improvement provided by TSS-Predictor over Cufflinks we calculated the distances between TSS defined by Cufflinks, and those defined by TSS-Predictor. Assuming that one of the TSS annotations in the reference annotation set is correct, I define TSS-prediction accuracy as the absolute distance between the predicted TSS and the nearest annotated TSS for the associated gene. Distances were calculated between TSS-Predictor-defined TSS and the nearest annotated TSS for the gene. Similarly, these same distances were calculated for the original Cufflinks predictions. Cufflinks predictions were then grouped according to whether more than one Cufflinks gene prediction was generated for each Ensembl gene. Figure 3-7 shows the distances for TSS-Predictor-defined TSS and Cufflinks-defined TSS for all genes with TSS < 500bp from the nearest annotated TSS. We can learn a number of things from this figure:-

A similar number of Cufflinks genes and TSS-Predictor genes have distances <500bp when Cufflinks transfrag-Ensembl gene mapping is 1 to 1.

Very few genes have this relationship when Cufflinks transfrag-Ensembl mapping is X >1.

TSS predictor substantially increases the number and accuracy of predictions less than 100bp from the nearest annotated TSS.

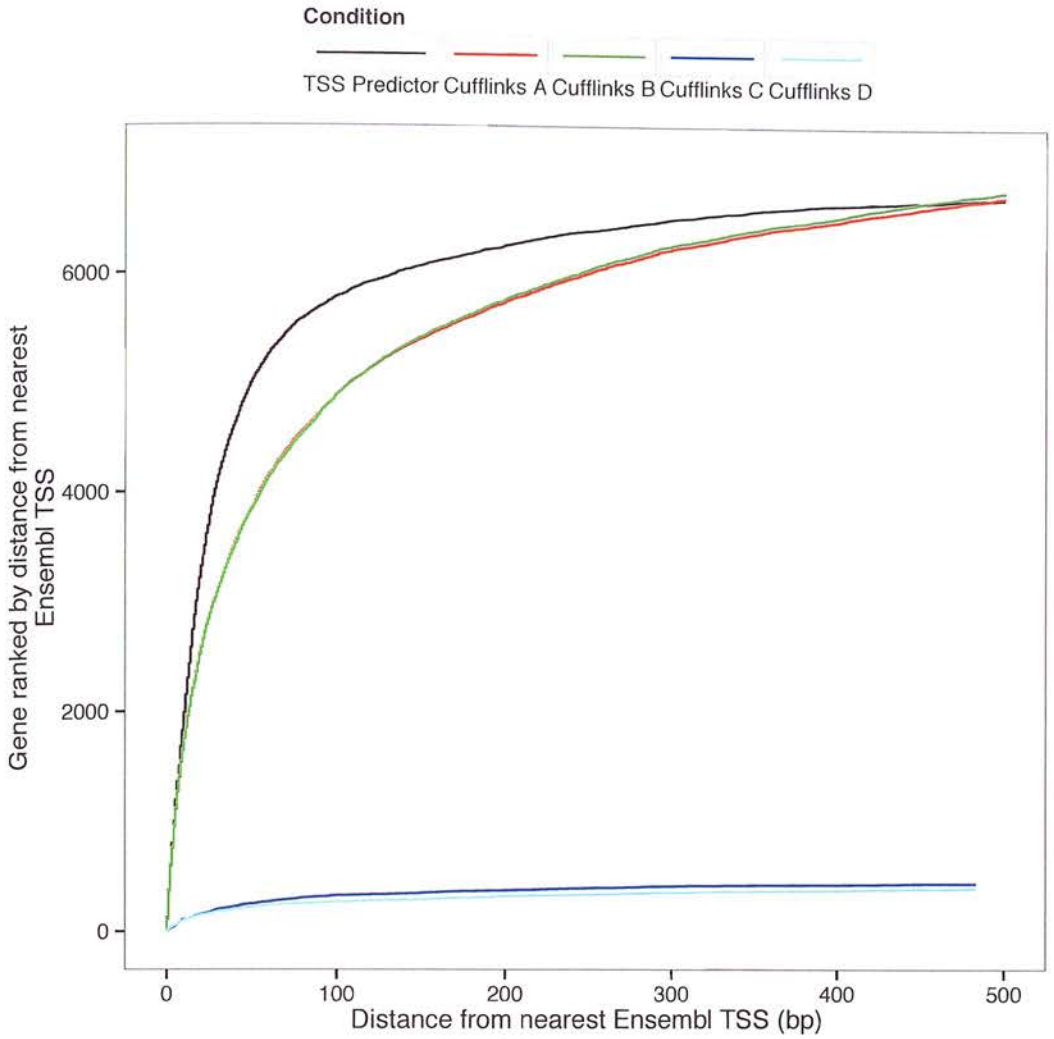


Figure 3-7. TSS-Predictor predictions are closer to the nearest annotated TSS for the gene than Cufflinks predictions. Cufflinks A = Cufflinks gene:EnsGene 1:1 mapping, most abundant transfrag. Cufflinks B = Cufflinks gene:EnsGene 1:1 mapping, longest Cufflinks transfrag. Cufflinks C = Cufflinks gene:EnsGene >1:1 mapping, most abundant transfrag. Cufflinks D = Cufflinks gene:EnsGene >1:1 mapping, longest Cufflinks transfrag.

3.3.1.4 *TSS-Predictor Webpage.*

The TSS-Predictor project is hosted from URL http://tss-predictor.bio.ed.ac.uk (Figure 3-8). The website contains links to the GitHub page https://github.com/GrahamThomas/TSS-Predictor, where the code can be downloaded. Help documentation (Appendix 3) and links to files from the walkthrough example are also provided.

Figure 3-8. Screenshot of the main page for the TSS-predictor project, http://tss-predictor.bio.ed.ac.uk

3.4 Discussion

This chapter presents TSS-Predictor, a novel approach for TSS identification using RNA-Seq data. The method was originally designed for TSS prediction in a macrophage RNA-Seq project (Thomas et al., 2012a) and is here generalized for use with RNA-Seq datasets provided the reference genome and accompanying annotations are of high quality. Implemented in Perl and R, TSS-Predictor also makes use of BEDTools (as such it is designed for use on Unix operating systems and requires a command line for usage). TSS-Predictor is trivial to use for those with basic command line experience. Automated scripts produce figures to guide user decisions necessary for the production of high-quality TSS predictions.

The website http://tss-predictor.bio.ed.ac.uk provides links to the download page for the project and hosts example files as well as extensive help documentation. This includes a step-by-step guide on the production of TSS predictions from raw Illumina RNA-Seq data. The output BED data format is a de facto standard, and is compatible with many other applications for downstream analysis of identified *cis*-regulatory regions.

We have demonstrated the utility of TSS-Predictor with the original RNA-Seq dataset for which it was designed. Prior to publication, future work will require further testing, and validation on a range of mouse and human RNA-Seq datasets downloaded from public repositories. In addition to this TSS-Predictor will also be incorporated into the online bioinformatics workbench GeneProf (Halbritter et al., 2012). GeneProf provides web-based access to command line bioinformatics tools, allowing biologists access to sophisticated tools for systems biology research. By incorporating TSS-Predictor into this service we will widely increase its accessibility within the research community.

4 TIME SERIES ANALYSIS OF MACROPHAGE ACTIVATION IN RESPONSE TO *L. SIGMODONTIS* INFECTION

4.1 INTRODUCTION

In chapter 2, we defined the phenotype of helminth-elicited AAMΦ using *B. malayi*-elicited macrophages purified from BALB/c mice 21 days post implantation. This is by far the most comprehensive profiling of NeMΦ performed to date, uncovering both IL-4Rα-dependent and IL-4Rα-independent components of the macrophage activation programme. Nonetheless, the experimental strategy we devised has drawbacks that limit the scope of conclusions. Firstly, because we assayed the macrophage response to one nematode species we are unable to state whether our findings are generally applicable to NeMΦ phenotypes, or if they reflect a specific component of the response to *B. malayi* in BALB/c mice. Secondly our experiment was performed at a single time-point 21 days post-infection. We decided to use this time point because natural helminth infections are typically chronic, and this would reflect the fully established macrophage response. Furthermore, due to the surgical implant procedure it was necessary to allow sufficient time for complete recovery to ensure analysis of helminth-dependent macrophage phenotypes. We also cannot make any claims regarding temporal profiles of gene expression during NeMΦ induction from our RNA-Seq analysis. Our final concern regarded the use of recruited, thioglycollate-elicited, macrophages as a control group when we know AAMΦ can be derived from the resident population (Jenkins et al., 2011).

The *Litomosoides sigmodontis* infection model provides a solution to the problems outlined above. *L. sigmodontis*, a natural parasite of the cotton rat, is competent to complete its life cycle in the mouse (Hoffmann et al., 2000a). Infective larval L3's are delivered sub-cutaneously, these migrate via the lymphatic's into the pleural cavity at 3-4 days post-infection whereupon they undergo two moults, eventually reaching sexual maturity and producing microfilarae 50-60 days post-infection in the susceptible BALB/c mouse strain (Hoffmann et al., 2000a). However, most mouse strains such as C57BL/6 are resistant to *L. sigmodontis* infection. Resistance correlates with earlier a significant expansion of pleural cavity leukocytes between day 10 and day 30, and adult worms are killed prior to the production of microfilarae (Babayan et al., 2003). Thus, *L. sigmodontis* infection in C57BL/6 mice is a model to understand the immunological traits conferring resistance to filarial nematode infection.

Owing to the physiologically relevant migration into, and establishment of infection in, the pleural cavity, this provides an excellent model for a temporal analysis of the NeMΦ phenotype. The infection site also differs between the *B. malayi* and *L. sigmodontis* models. Thus, with the choice of an appropriate host mouse strain we are able to decouple the strain, site, and species-specific components of the macrophage response to *B. malayi* infection presented in Chapter 2 from those that genuinely reflect the physiological response to nematode infection.

It has recently been reported that IL-4 induces both AAMΦ and proliferation *in vivo* (Jenkins et al., 2011), however only alternative activation but not proliferation can be replicated *in*

vitro. During *L. sigmodontis* infection macrophages accumulate in the pleural cavity between day 8 and day 15 P.I. This accumulation is IL-4-dependent and results from proliferation of the resident macrophage population rather than recruitment from the periphery (Jenkins et al., 2011). An open question of biological significance is how can IL-4 support macrophage proliferation *in vivo* but not *in vitro*? Related to this question is a paradox concerning the mechanisms of IL-4-dependent macrophage proliferation. Our RNA-Seq analysis confirmed *in vivo*, observations previously reported *in vitro*, that IL-4 induces mitochondrial metabolism in AAMΦ (Odegaard et al., 2007). Indeed, A Chawla and colleagues have stated that alternative macrophage activation requires enhanced mitochondrial metabolism (Vats et al., 2006). The prevailing dogma regarding mechanisms of cellular proliferation is the requirement for aerobic metabolism, also known as the Warburg effect (Vander Heiden et al., 2009). It therefore seems counterintuitive that AAMΦ both up-regulate mitochondrial metabolism and proliferate in response to IL-4 *in vivo*, and one immediate outstanding question is whether the two processes can occur concurrently.

In this chapter we present the first transcriptomic time series analysis of highly purified alternatively activated macrophages until the establishment of adult *L. sigmodontis*, day 28 post infection (Hoffmann et al., 2000a). We make use of cDNA microarrays for transcriptomic profiling owing to the lower requirement for input RNA as compared to RNA-Seq libraries. The time series covers both the induction phase of alternative activation as well as the macrophage proliferative burst and has, as a starting point, naïve tissue-resident pleural cavity macrophages. A novel approach to data analysis is presented in which we integrate a FACS-based phenotypic analysis of alternative activation and

proliferation with genome-wide expression profiling. Linear modelling and principal component analysis (PCA) are applied to this dataset allowing us to delineate the transcriptional phenotypes associated with the macrophage alternative activation and proliferation regimes.

4.2 METHODS

4.2.1 LITOMOSOIDES INFECTIONS

35 male C57BL/6 mice, aged 10 weeks, were each infected with 30 *L. sigmodontis* L3s suspended in 100 µl RPMI by sub-cutaneous injection. On days 0, 4, 6, 8, 11, 14, 18 and 28, 5 mice were each sacrificed by terminal anaesthesia (100 µl Domitor/Vetalar 1:1 sub-cut) followed by CO2 after bleeding of the brachial artery. Day 0 mice served as naïve controls. 10 x 1 ml pleural cavity lavages (RPMI + 5% FCS + 2 mM EDTA + 1% penicillin/streptomycin) were taken and retained for flow sorting.

4.2.2 MACROPHAGE PURIFICATION.

Pleural lavage cells were made up to 10 x 10^7/ ml and treated with 1 ml RBC lysis buffer (Sigma). 200k cells were retained for intracellular staining with the proliferation markers Ki67 and YM1 and the remaining cells retained for flow sorting. Surface staining volumes were calculated based on the number of cells available such that block and surface staining was performed in a final volume of 30 µl/million cells.

Cells were sorted on an Aria (BD) by Martin Waterfall at the flow cytometry facility at Ashworth Laboratories. F4/80-Hi resident tissue macrophages were purified on the basis of F4/80-APC (F4/80-biotin, 1:100; Strep-APC, 1:300) and MHCII-PacBl positivity.

We excluded dead cells using Live/Dead-Aqua staining and removed cell doublets with FSC-A/FSC-W doublet gate. Negative gating was also used to exclude B-Cells (CD19-PE, 1:100) and eosinophils (SiglecF-PE, 1:100). Primary and secondary antibody surface staining steps were performed for 25 min each. Sorted macrophages were stored in 700 µl Qiazol (Qiagen) prior to RNA extraction.

4.2.3 Intracellular staining

Pleural cavity intracellular staining for YM1 and Ki67 expression was performed by Stephen Jenkins. 200k pleural cavity lavage cells were retained prior to FACS sorting (see section 7.2.1). Cells were surface stained with F4/80-PeCy7 (1/200), CD11c-PE (1/200), CD19-APC 1/200) and MHCII-PacBl (1/400) at the indicated concentrations in a final volume of 50 µl for 25 min. Following this cells were washed twice in FACS buffer (5% FCS, 2 mM EDTA in PBS) and fixed overnight at 4°C using Fixation/Permiabilization solution (eBiosciences). Intracellular staining was performed in BD Perm/Wash. Ki67-FITC (7.5 µl/sample), YM1-biot (1/50), Rabbit-anti-mouse-RELMα (1/75) were added to cells at the indicated concentrations in a final volume of 25 µl and incubated for 45 min on ice. Cells were washed twice with PermWash before incubation with anti-rabbit-PacBl (1/300) and strep-PerCP (1/300) in 25 µl for 30 min. Cells were washed twice in PermWash and resuspended in 80 µl FACS buffer before acquisition on a FACS Canto. Data were analysed by Stephen Jenkins using the FlowJo software.

4.2.4 Microarray library prep and hybridizations

RNA was extracted from purified macrophages by phenol-chloroform extraction followed by clean-up over RNeasy MinElute columns (Qiagen) according to the manufacturer's instructions and eluted into a final volume of 15 µl. RNA extractions were performed in five batches of eight in order to maximise the speed of RNA extractions. Batches were created by extracting one biological replicate from each treatment group together, making the identification of batch effects tractable. RNA concentrations and integrity were assessed using an RNA-pico chip on the Agilent Bioanalyzer (Agilent Technologies). Samples had an

RNA integrity >8 and were suitable for microarray library preparation. Total RNA was prepared for hybridization onto Illumina mouse Ref6-WG2 arrays by conversion into Biotinylated amplified RNA (cRNA) using the Illumina TotalPrep RNA amplification kit (Ambion) in accordance with the manufacturer's guidelines. cRNA libraries were deemed to be of sufficient quality if the fragment size distribution was 250-5500 nucleotides, and if the final amount of cRNA was > 1.5 µg. Concentration and fragment length distribution were assayed on a Bioanalyzer RNA nano chip (Agilent Technologies) according to the manufacturer's instructions. We did not obtain sufficient cRNA for some libraries with a low starting cell number, however successful hybridizations were obtained for all 24 successful libraries are shown in Table 4-1 (results section 7.3.1). Microarray hybridizations were performed onto MouseWG-6 v2 arrays (Illumina) at the Edinburgh Wellcome Trust Clinical research Facility.

4.2.5 Microarray normalization, ArrayQualityMetrics validation and differential expression analysis

Microarray pre-processing and normalization.

Microarray data processing and analysis were performed in R. Raw data intensities were first normalized between arrays using the robust spline normalization method and homoscedasticity was enforced using the variance stabilizing transformation (Huber et al., 2002b; Lin et al., 2008), both using the Bioconductor package Beadarray (Dunning et al., 2007). Beadarray was used to convert data into a MIAME compliant ExpressionSetIllumina object for future submission to online gene expression repositories. Non-expressed probes were filtered prior to statistical analysis using Illumina's method implemented in the Beadarray package (Dunning et al., 2007), a detection score < 0.05 defined probes as non-expressed. These data were used for all subsequent analyses. The quality of normalized

data were assessed using arrayQualityMetrics (Kauffmann et al., 2009), all hybridizations were confirmed to be of high quality. For the arrayQualityMetrics report see Appendix 4.

TRANSFORMATION OF FACS DATA FOR STATISTICAL ANALYSES.

Percentage values of YM1 and $Ki67^{hi}$ positivity for each sample, as determined by flow cytometry, were imported into R and tested for normality using the Shapiro-Wilk Normality Test (R function shapiro.test). %YM1 positive data were found to satisfy the assumption of normality ($P<0.01$), however $\%Ki67^{hi}$ frequencies did not. Square root transformation of $\%Ki67^{hi}$ frequencies yielded normally distributed values suitable for statistical modelling.

PRINCIPAL COMPONENT ANALYSIS AND LINEAR MODELLING.

Principal component analysis of normalized gene expression data were performed using the base R function 'prcomp'. Differential expression analysis of microarray data was performed using the limma package (Smyth). A simple linear model was fit to simultaneously incorporate %YM1 positive and transformed $\%Ki67^{hi}$ positive values. P-values were calculated based on the significance of the correlation of probe expression relative to each independent variable (YM1 or root-transformed $Ki67^{hi}$), and corrected for multiple testing using the Benjamini-Hochberg method. We selected a significance cutoff based on the log-odds ratio of differential expression (B). This Bayesian estimator is more conservative than the parametric approach (Smyth). A B value cutoff of 4.6 was chosen corresponding to a <1% likelihood that a correlation between dependent and independent variables has occurred by chance.

4.2.6 KEGG GSEA (BROAD METHOD)

Gene Set Enrichment Analysis (GSEA) on microarray data was performed using the Broad method (Subramanian et al., 2005) using the downloadable Java applet (version 2.0.10). Mouse WG Ref 6 array annotations and KEGG pathway annotations (v3.0) were obtained from the Broad Institute MsigDB (Liberzon et al., 2011). Genes were ranked for GSEA according to the Pearson's correlation coefficient between gene expression profile and %YM1 positive, or root transformed %Ki67^{hi} levels. The most abundantly expressed probe for each gene was considered in this analysis. P-values for each gene set were calculated by 1000 permutations of sample labels. A nominal P-value of 0.05 and FDR of 25% were used for significance cut-offs as suggested by the authors of the program (Subramanian et al., 2005).

4.2.7 ORDEREDLIST ANALYSIS

The Bioconductor package OrderedList (Lottaz et al., 2006) was used to identify similarities in gene expression patterns between microarray and RNA-Seq experiments by comparing ranked lists of gene expression using the default parameters. Microarray data were ordered according to the significance of the correlation of gene expression with respect to YM1 level, taking into account the direction of the correlation (most significant upregulated < Not DE > most significant down-regulated). This list of genes was compared against two similar rankings of genes from the RNA-Seq experiment (see chapter2), firstly WT-Infected vs. WT-Thioglycollate and secondly, WT-Infected vs. IL-4Rα$^{-/-}$ infected. P-values were taken from the differential expression test performed on the RNA-Seq data in limma for use in GSEA in chapter 2.

4.2.8 Heatmaps and Figures

Heatmaps were drawn using the heatmap.2 function in the gplots package (Warnes et al., 2009) and all other graphics in R were produced using ggplot2, Prism (GraphPad) or automatically using the GSEA software.

4.2.9 Cis-regulatory analysis with Homer

To identify over-represented TFBS we first assigned macrophage TSS defined in Chapter 2 to expressed genes identified on Illumina microarrays using NCBI gene IDs. TSS were next assigned as AAMΦ-associated or Ki67^{hi}-associated if the parent gene was deemed significantly correlated with respect to the appropriate covariate in the limma analysis (above). Promoter regions were defined for these TSS by downloading a BED file of H3K4Me3 peaks from the lab website (http://biowhat.ucsd.edu/homer/data/ucsc/sheinz-10-05-28/ThioMac-H3K4me3.alignment.mm8.bed.gz, Heinz et al. 2010). ChIP-Seq peaks were merged using BEDTools slop and merge respectively. All ChIP-Seq peaks within ±1 kb of a gene TSS were identified and retained using BEDTools intersect. The resulting file was used for motif enrichment analysis using Homer (Heinz et al., 2010). Default parameters were run with Homer's findMotifsGenome.pl script and the mouse mm9 genome was provided at the reference. Background sequences for overrepresentation analysis were selected by Homer to be of equal length and similar GC content to each test sequence.

4.3 RESULTS

4.3.1 ACTIVATION AND PROLIFERATION KINETICS OF MACROPHAGES IN RESPONSE TO *L. SIGMODONTIS* INFECTION

In order to study the kinetics of macrophage proliferation and alternative activation in the context of parasite infection we infected C57BL/6 with 30 *L. sigmodontis* L3 larvae each. We have previously defined YM1 as an IL-4Rα-dependent marker of alternative activation in macrophages (Loke et al., 2002) and Ki67^{hi} expression as a surrogate for actively proliferating cells by FACS (Jenkins et al., 2011). At each time point, days 0 (naïve), 4, 6, 8, 11, 14, 18 and 28, we measured alternative activation and proliferation by FACS using YM1 and Ki67 respectively.

L. sigmodontis enter the pleural cavity at day 4 P.I. after migrating through the lymphatic vasculature (Hoffmann et al., 2000a). Consistent with previous observations from our lab (S Jenkins – unpublished) there was a notable difference in the kinetics of proliferation vs. alternative activation in the macrophage compartment. An increase in proliferation was observed by day 6, yet we did not begin to see enhanced YM1 expression until day 8 (Figure 4-1). We have taken advantage of this discordant response to discriminate between genes involved in macrophage proliferation and alternative activation in response to *L. sigmodontis* infection *in vivo*. In this experiment a subsample of cells (200k) from each mouse were used for flow cytometric phenotyping while the remainder were retained for microarray expression profiling. The microarray expression profiles reflect the average expression of each gene over the entire population of F4/80^{hi} macrophages analysed, whereas the flow cytometry data provides information on the frequency of proliferating and alternatively activated cells within the sample as a whole. Because we have paired (i.e. flow and microarray) data for each mouse this experimental design allows us to correlate

population-level FACS data against gene expression values to explicitly identify expression profiles associated with each phenotype.

Our macrophage sorting strategy was similar to that performed during our RNA-Seq analysis of *B. malayi*-elicited macrophages. Cell doublets were excluded based on scatter profiles, dead cells were disregarded by Live/Dead-Aqua staining and eosinophils (SiglecF$^+$), B-cells (CD19$^+$), and T-cells (CD3$^+$) removed based on negative gating for PE. F4/80 and MHCII were used to sort the F4/80 brightest macrophage population over our time course. Figure 4-2 shows representative sorts at each time point. RNA was extracted from the sorted macrophage populations and microarray libraries were prepared (See methods section). We ensured the highest possible standard in our data acquisition, and with the exception of 2 libraries collected at day 8 all assayed samples were sorted to >85% purity (average 94.5% excluding day 8) and had RIN (RNA integrity) values >8 (Table 4-1).

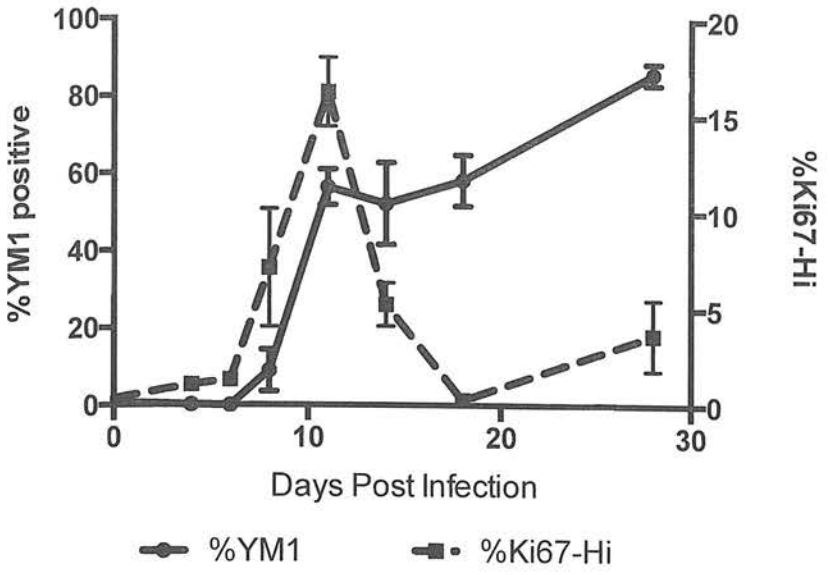


Figure 4-1. The frequency of %YM1 positive (solid line) and %Ki67^{hi} (dashed line) pleural cavity F4/80^{Hi} resident tissue macrophages during time series of *L. sigmodontis* infection in this chapter. YM1 serves as a marker of AAMΦ and Ki67^{hi} macropahges are deemed to be those in cell cycle as previuously publeihed by Jenkins *et al* (2011). Error bars represent SEM with 5 mice per time point.

Microarray libraries were prepared as describe in the methods section and hybridized to Illumina Whole Genome Mouse Ref6 arrays to quantitate mRNA expression levels. Due to the low numbers of macrophages present at early time points (Table 4-1 and Figure 4-3) we had limited success obtaining microarray libraries of sufficient concentration and quality for hybridization. Consequently data were only available for one naïve individual, two mice at day 4, and we obtained no successful libraries from day 6 post-infection. In total 24 successful libraries and hybridizations were performed (see Table 4-1). The viability of a large number of microarray experiments is undermined by the inadvertent 'aliasing' of technical and biological variability. For example, if one performs independent RNA extractions on biological replicates then there is no way of discriminating between changes in gene expression due to treatment and those due to RNA extraction. Throughout this experiment it was ensured that any potential technical artefacts could be identified during data QC. RNA was extracted using a balanced-block design where a single replicate from each time point were extracted together, small RNA extraction groups also expedited each isolation, maximizing RNA quality. Microarray libraries were prepared in a single batch to minimize technical variability at this critical stage. Finally samples were hybridized to arrays (each array takes 6 samples) in a manner that minimized the overlap between RNA extraction groups, day post infection and array number.

Prior to statistical analysis of differential expression in microarray experiments, data are normalized both within- and between-arrays. Using the Bioconductor package beadarray (Dunning et al., 2007) we implemented an optimal pipeline for the normalization of Illumina microarray data (Lin et al., 2008). The robust spline normalization (RSN) method (Lin et al., 2008) was used to accurately define signal intensities for microarray probes, this is specifically designed to take advantage of the replicated and random probe distribution

on Illumina arrays (Lin et al., 2008). Raw signal intensities were then scaled using the VSN method (Huber et al., 2002b) which stabilizes the mean-variance relationship between signal intensities across the range of expression values both within and between arrays. This approach to data normalization allows the application of traditional parametric statistical techniques, such as linear modelling (Huber et al., 2002b; Smyth). After normalization, data QC was performed using the battery of analyses provided in the Bioconductor package arrayQualityMetrics (Kauffmann et al., 2009). This package performs three kinds of test for array data quality and highlights hybridizations that fail any one of these tests. All of the samples we analysed passed every test. Between array comparisons are made by assessing pairwise distances between arrays and using PCA, the distributions of normalized array intensities are assessed using box- and density-plots. Finally individual array quality is assessed by calculating Hoeffding's statistic (D_a) for each array, this assesses systematic biases in relative expression (M) as a function of absolute expression (A) over each array, an indicator or microarray hybridization quality. A printout of the arrayQualityMetrics report for this project is included in Appendix 4.

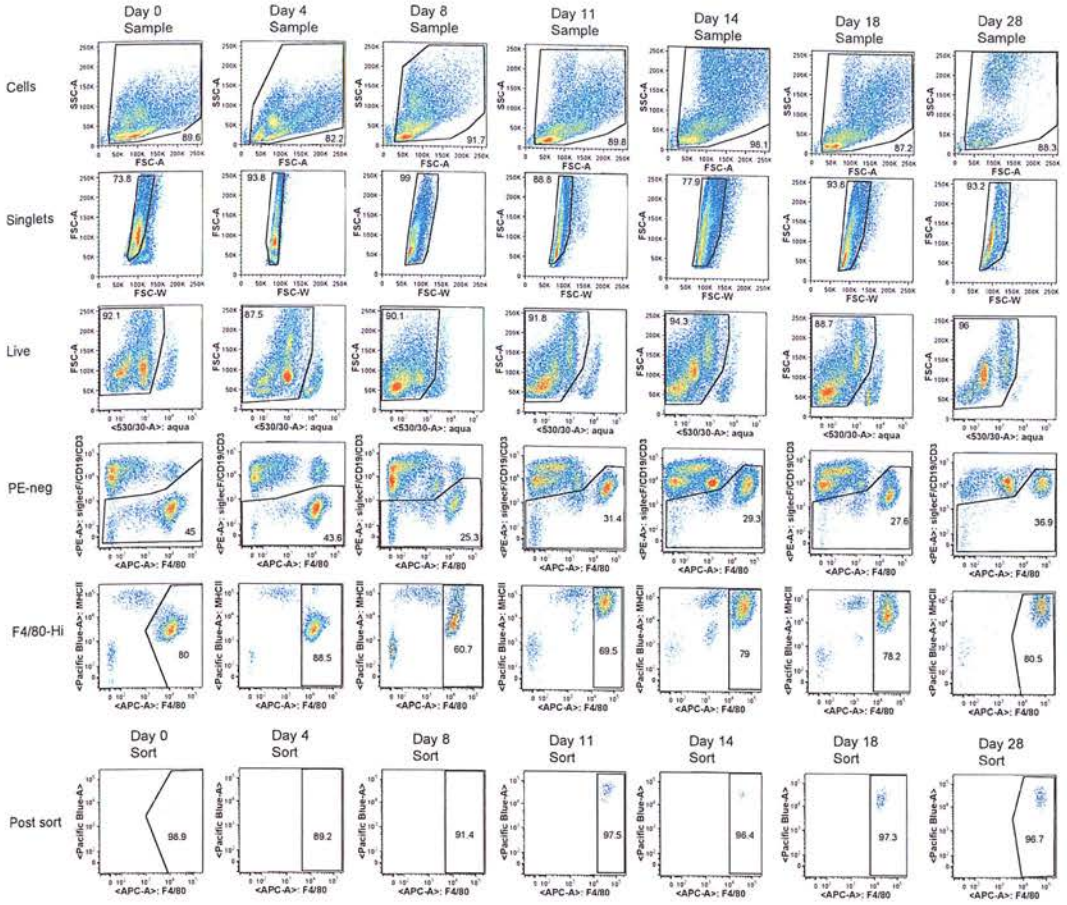


Figure 4-2. Sorting strategy for macrophage isolation during *L. sigmodontis* time-course. Cells were first identified based on FSC-A, SSC-A scatter profiles and cell-cell doublets excluded by FSC-A and FSC-W scatter profiles. Following this dead cells were excluded using Live/Dead-aqua staining and T-cells, B-cells and eosinophils excluded by negative gating for CD3-PE, CD19-PE and SiglecF-PE respectively. Finally macrophages were idneitified based upon F4/80-APC and MHCII-PacBl positivity. One representative mouse is shown for each time point.

Class	Day	Total Cells (*10^6)	Total F4/80^{hi} (*10^6)	% F4/80^{hi}	Sorted F4/80^{hi}	F4/80^{hi} purity (%)	Total RNA (µg)	RIN	Array No.	Total cRNA
Naïve	0	0.82	0.223	27.2	56,105	97.8	2.28	9.6	2	2.28
Infected	4	0.93	0.208	22.4	46,704	92.8	1.74	9.7	1	1.74
Infected	4	1.50	0.585	39.0	86,593	89.2	1.60	9.1	3	1.60
Infected	8	5.84	0.439	7.5	45,722	67.5	2.40	9.9	1	2.40
Infected	8	1.78	0.285	16.0	52,745	65.7	4.16	9.5	2	4.16
Infected	8	3.92	0.459	11.7	122,674	88.2	1.90	8.8	4	1.90
Infected	8	2.58	0.490	19.0	78,722	89.2	2.78	9.1	3	2.78
Infected	11	20.4	2.917	14.3	367,132	98.8	4.48	9.6	2	4.48
Infected	11	13.8	2.636	19.1	656,723	93.6	5.26	9.6	3	5.26
Infected	11	11.8	3.269	27.7	1,081,281	97.5	2.98	9.7	3	2.98
Infected	11	17.8	2.830	15.9	276,928	91.7	2.98	9.6	1	2.98
Infected	11	14.6	2.073	14.2	211,033	96.2	2.32	9.5	4	2.32
Infected	14	16.8	4.939	29.4	327,051	86.1	1.92	9.5	1	1.92
Infected	14	17.8	4.948	27.8	347,283	93.1	2.32	9.8	4	2.32
Infected	14	12.6	1.940	15.4	147,448	98.8	1.58	10	2	1.58
Infected	18	11.8	2.631	22.3	265,261	92.6	3.00	9.2	4	3.00
Infected	18	8.40	2.402	28.6	189,438	89.6	3.34	8.9	2	3.34
Infected	18	16.8	5.796	34.5	589,505	96.3	2.74	9.9	3	2.74
Infected	18	15.4	4.050	26.3	517,888	97.3	2.56	9.2	4	2.56
Infected	18	24.6	11.193	45.5	1,433,706	96.5	5.02	9.6	1	5.02
Infected	28	69.5	20.781	29.9	610,803	98.0	2.38	9.4	4	2.38
Infected	28	41.5	7.429	17.9	425,410	98.8	2.64	9.4	2	2.64
Infected	28	61.5	9.963	16.2	315,393	97.2	4.12	9.5	3	4.12
Infected	28	65.0	14.950	23.0	327,241	98.6	3.68	9.5	1	3.68

Table 4-1. Summary statistics for each successful cDNA sample profiled by microarray in this chapter. The tabl detailis the day post infection with *L. sigmodontis*, both the total number of pleural lavage calls and F4/80^{hi} macrophages and the number of sorted F4/80^{hi} macrophages, the final amount of RNA extracted and bioanlayzer RNA integrity (RIN). Additionally the array number the sample was profiled on and the final cRNA yield prior to array analysis.

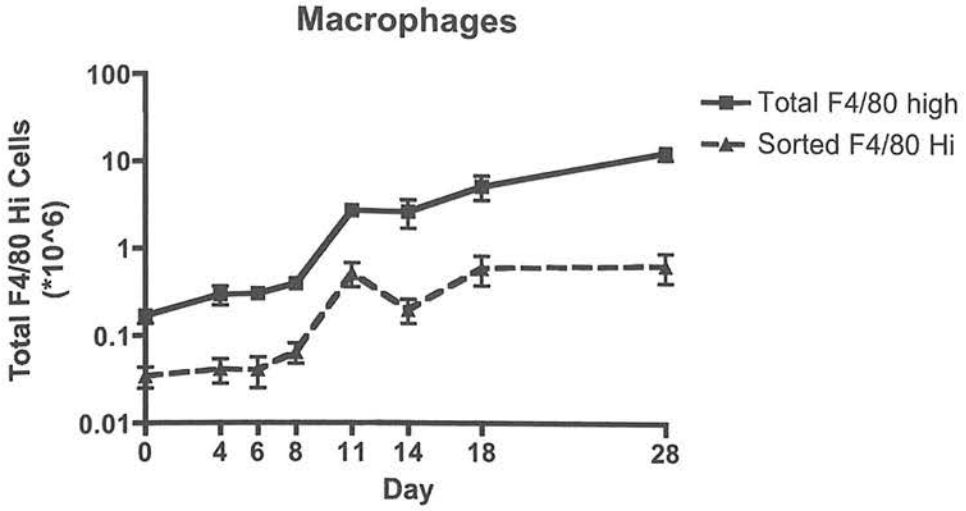


Figure 4-3. Increase in macrophage cell numbers in the response to *L. sigmodontis* infection. The total number of F4/80^{hi} tissue resident macrophages (solid line), and the number of FACS sorted F4/80^{hi} macrophages (dashed line) obtained from the pleural cavities of *L. sigmodontis*-infected mice for microarray profiling. Day 0 mice were naïve. Error bars represent SEM with 5 mice per time point.

4.3.2 MICROARRAY PCA ANALYSIS - MACROPHAGE PROLIFERATION AND ALTERNATIVE ACTIVATION ARE CHARACTERIZED BY INDEPENDENT TRANSCRIPTIONAL PROFILES

We performed principal component analysis (PCA) to understand the processes controlling the variability in macrophage gene expression during the expansion and activation response to *L. sigmodontis* infection. PCA is an unsupervised dimension reduction technique that groups genes into bins that explain the variability in a dataset. As an unsupervised technique, PCA is naïve to experimental design and/or potential confounding artefacts. With an appropriate experimental design (such as discussed in section c.i in this chapter) PCA can be used to confirm data quality or identify potential confounding covariates. Each principal component (PC) is assigned an 'eigenvalue', which serves as a measure of the proportion of variability in the overall dataset that is explained by that component. The more variability explained by a principal component the higher the eigenvalue. Furthermore, a gene's 'loading' within each principal component defines the relative contribution of that gene to the principal component's eigenvalue. Thus, the PC explaining the greatest variability in a dataset is identified as having the largest eigenvalue, and the genes that contribute most within each PC are identified as having the largest loading.

Our initial principal component analysis shows that the majority of variability, 61.9%, in our dataset is explained by the first two principal components, accounting for 36.4% and 25.5% respectively (Figure 4-4a). We wanted to understand if the variance observed in PC1 and PC2 can be explained by biology, rather than technical artefact. A plot of PC1 against PC2 with samples coloured by day (Figure 4-4b) shows a general tendency for gene expression profiles to cluster according to day post infection. Interestingly the naïve (Day 0) and two Day 4 samples cluster tightly together indicating no differences exist in the transcriptional profiles of these populations between at this early time points. Linear regression analysis of

gene expression profiles between day 0 and day 4 samples confirmed that day 4 samples were essentially naïve. There was a greater degree of similarity between day 0 and mean day 4 gene expression (R=0.989) than between the two independent biological replicates sampled at day 4 (R =0.983, Figure7-5a,b).

The overall tendency appears to be for a correlation between day P.I. and the first principal component. The frequency of YM1 positive macrophages as measured by FACS was used as a surrogate marker for AAMΦ in this experiment. This is because we know that macrophage YM1 expression is IL-4Rα-dependent in this system. The percentage of YM1 positive macrophages from our flow analysis were correlated against PC1 to explore whether PC1 captures a gene expression profile consistent with alternative activation (Figure 4-4c). This is indeed the case and a linear regression of PC1 against %YM1 positive reveals a strong, highly significant, correlation between the two variables (R=0.78, P<0.0001). From this we can conclude that roughly a third (36.4%) of the total variation in macrophage transcriptional activity during *L. sigmodontis* infection is associated with the alternative activation regime. Because the $Ki67^{hi}$ macrophage population represents actively proliferating cells (Jenkins et al., 2011), we performed a similar correlation analysis between $\%Ki67^{hi}$ macrophages measured by FACS and the second PC (Figure 4-4e and f). The results of this analysis revealed a similar, highly significant, correlation between $Ki67^{hi}$ levels and PC2 gene expression profiles (R=0.46, P<0.001). Similarly, roughly a quarter (25.5%) of the total variation in macrophage gene expression in this dataset correlates with macrophage proliferation. Thus, approximately 60% of the total variability in gene expression can be attributed to either alternative activation or proliferation, which further validates the quality of the dataset in this analysis. Importantly, as by definition, there is no correlation between PC1 and PC2, these data show that the expression profiles associated

with proliferation and alternative activation are independent. Importantly, neither the correlation between YM1 and PC2, or $Ki67^{hi}$ and PC1 were significant (P=0.1238 and P=0.1289 respectively). This is surprising considering that both alternative activation and proliferation are predominantly IL-4-dependent in this model.

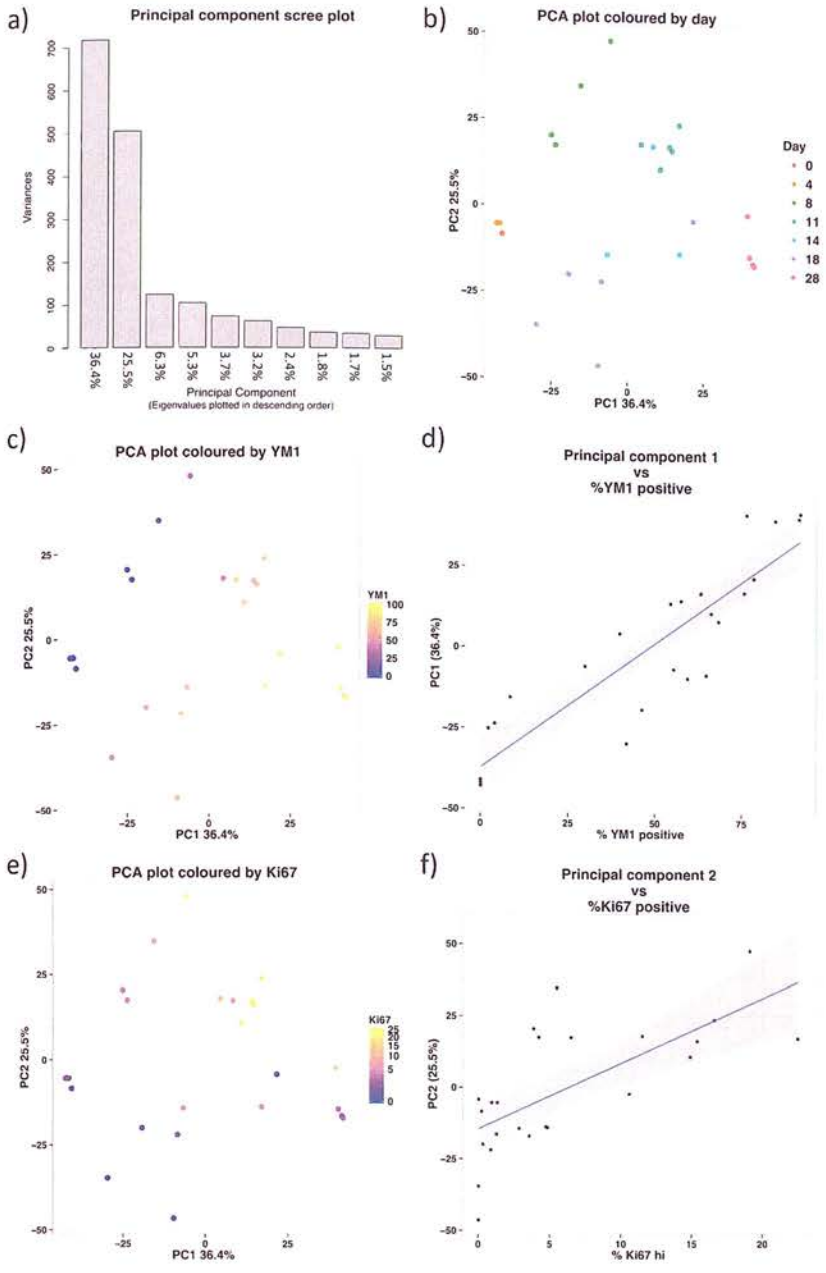


Figure 4-4. Principal component analysis (PCA) of microarray data showing that the first principal component (PC) correlates with AAMΦ and that the second principal component correlates with macrophage proliferation. PCA was performed on all expressed genes in the microarray dataset. A) Scree plot showing the relative proportion of variance explained by each of the 10 largest principal components. Scatter plots of PC1 vs PC2 loadings coloured by b) day, c) %YM1 positivity and e) %Ki67^{hi} positivity as measured by FACS. D) Linear regression analysis of PC1 values against %YM1 positivity. F) Linear regression analysis of PC2 values against %Ki67^{hi} positivity.

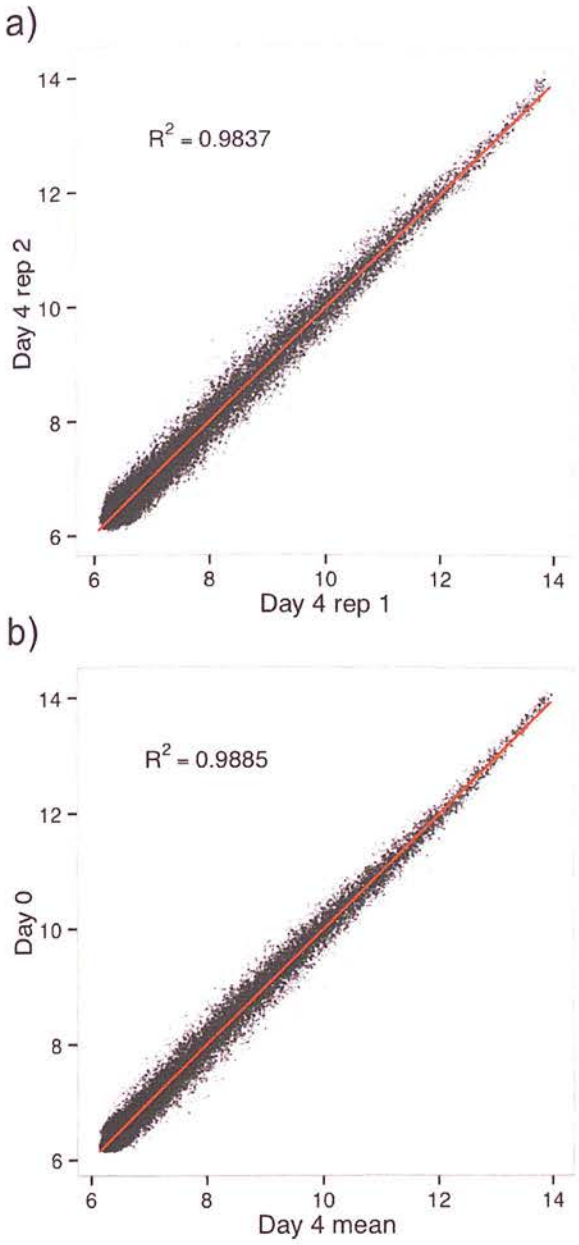


Figure 4-5. F4/80^{hi} macropahges at day 4 post infection are transcriptionally indistinguishable from naïve pleural cavity F4/80^{hi} macrophages. A) Linear regression of gene expression profiles of the two independent biological replicates sampled at day 4 post infection (P.I.) showing a high degree of similarity between the two biological replicates. B) Linear regression analysis of mean gene expression of samples acquired at day 4 P.I. against the single successful day 0, naïve, microarray hybridization.

4.3.3 Linear modelling to identify gene expression association with YM1 and Ki67

A generalized linear modelling (GLM) approach was taken to identify statistically significant changes in gene associated with %YM1 and %Ki67^{hi} macrophages as defined by FACS. Firstly we determined that the expression profiles of our explanatory (FACS) variables adhered to the requirements of linear modelling and were normally distributed. A Shapiro-Wilk normality test confirmed that YM1 was normally distributed (i.e. failed to reject the null hypothesis at our chosen cutoff of $P<0.01$, $P=0.02$). The distribution of Ki67^{hi} levels on the other hand were not Gaussian ($P<0.001$), this was corrected using a square root transform of %Ki67^{hi} levels ($P=0.17$).

We confirmed the fit of our linear model by analysing residual errors after fitting the linear model. For successful linear modelling the explanatory data must provide a good fit to the observed data. This assumption requires that residual errors must be normally distributed after fitting the linear model. The analysis of residual errors from microarray experiments can identify systematic biases to help identify outlying patterns of gene expression. Firstly, to compare residuals between the many genes in a microarray experiment they must be normalized, and studentization is the method used for this (Oron et al., 2008). This approach involves dividing each residual for a gene by the standard deviation of the residuals for the gene. Hierarchical clustering of studentized residuals resulted in the production of a heatmap showing little, if any, structure (Figure 4-6). This tells us that a linear model describing gene expression as a function of YM1 and Ki67^{hi} captures the variation in the dataset well. If additional temporal features of alternative activation were detectable over the time course and interval range we have sampled we would expect to have observed systematic biases in the residual errors of the linear model.

A log-odds ratio of 4.6 was used as our significance cutoff, this corresponds to a < 1% chance of a correlation between our explanatory and response variables (FACS data and gene expression data) occurring by chance. The log-odds ratio calculated by limma provides a more accurate measure of significance when few samples are present, and is derived by borrowing information on the variance in gene expression of similarly expressed probes across all genes (Smyth, 2004). The B-value of 4.6 corresponds to adjusted P-values of $1.1x10^{-5}$ for YM1 and $1.2x10^{-4}$ for YM1 associated genes and is therefore more conservative than the parametric method. We identified 685 significant probes associated with YM1, and 569 probes associated with Ki67. Multiple probes may map to a single gene, these results corresponded to a significant association of 500 genes with YM1, and 399 genes with Ki67 (Figure 4-7a). Supporting the findings of our principal component analysis, that there is little overlap in significance profiles associated with alternative activation and proliferation. Thirteen significant genes were shared between the two lists (Figure 4-7a, b) including two genes, *Ccbl1* and *Cd209b*, whose expression is positively associated with alternative activation, but negatively associated with proliferation. Notably the expression profile of these genes is down-regulated (relative to naïve Day 0 and Day 4) at day 11, the peak of IL-4-dependent proliferation and a time point with significant alternative activation.

Nine of the 13 genes significantly associated with both YM1 and Ki67 were up-regulated during proliferation, and significantly down-regulated later during alternative activation (Figure 4-7b). This list of genes includes the proliferation-associated gene *Cdca4* (Hayashi et al., 2006), the histone cluster protein *Hist2h2ab*, and the nucleosome assembly protein *Nap1l1* (Keck and Pemberton, 2012) and is suggestive of chromatin remodelling, perhaps reflecting a maturation phase post peak proliferation. Intriguingly, genes associated with cellular migration and adhesion *Coro1a* and *Smtn* and *Sell* (L-selectin, Cd62l) are down-

regulated over the time series of *L. sigmodontis* infection (Gatfield et al., 2005; Grailer et al., 2009). Together, these data support our previous conclusion (Thomas et al., 2012a) that AAMΦ are associated with a reduced migratory capacity. We also observe the reverse expression profile, where genes negatively correlate with proliferation and positively correlate with alternative activation. These include the cell-surface marker *Cd209b* (DC-SIGN, (Cheong et al., 2010) which is expressed at a lower level in Day 11 proliferating macrophages (the peak of IL-4-dependent proliferation) relative to both naïve and later time points. If we assume that the F4/80^{hi} macrophage population we assayed during this time course is derived from a resident population, these data imply that additional changes occur in the macrophage phenotype post IL-4-dependent proliferation. Furthermore, *Cd290b* expression may reflect this maturation or terminal differentiation step in AAMΦ as it is described to do so in dendritic cells (Cheong et al., 2010).

The distribution of statistically significant positive and negative correlations with YM1 is roughly equal with 269 positive and 231 negative genes; similarly, the distribution of fold changes is roughly equal (Figure 4-8) showing that alternative activation is associated as much with up-regulation as down-regulation (Table 4-2). Although the microarray analysis has identified a lower number of DE genes than the RNA-Seq analysis presented in chapter 2, the same trend of roughly equal up- and down-regulated genes was observed between both WT-Infected and WT-Thioglycollate and WT- and IL-4Rα$^{-/-}$-infected groups.

The DE profiles of Ki67-associated genes differ from YM1. Of the 399 statistically significant correlations 231 genes positively correlated, while only 80 genes negatively correlated with Ki67 (Table 4-2). Furthermore there was a tendency for up-regulated genes to be more highly correlated with Ki67 than down-regulated genes were negatively correlated (if this was categorical we would say the fold changes are greater, Figure 4-9). This implies that

commitment to proliferation is associated with the induction of genes associated with proliferation, rather than the repression of inhibitors of proliferation.

We have identified substantial changes in dynamic gene expression associated with macrophage proliferation and alternative activation in the response to *L. sigmodontis* infection. Both PCA and linear modelling approaches support the concept that proliferation and alternative activation, while both IL-4-dependent are transcriptionally independent processes.

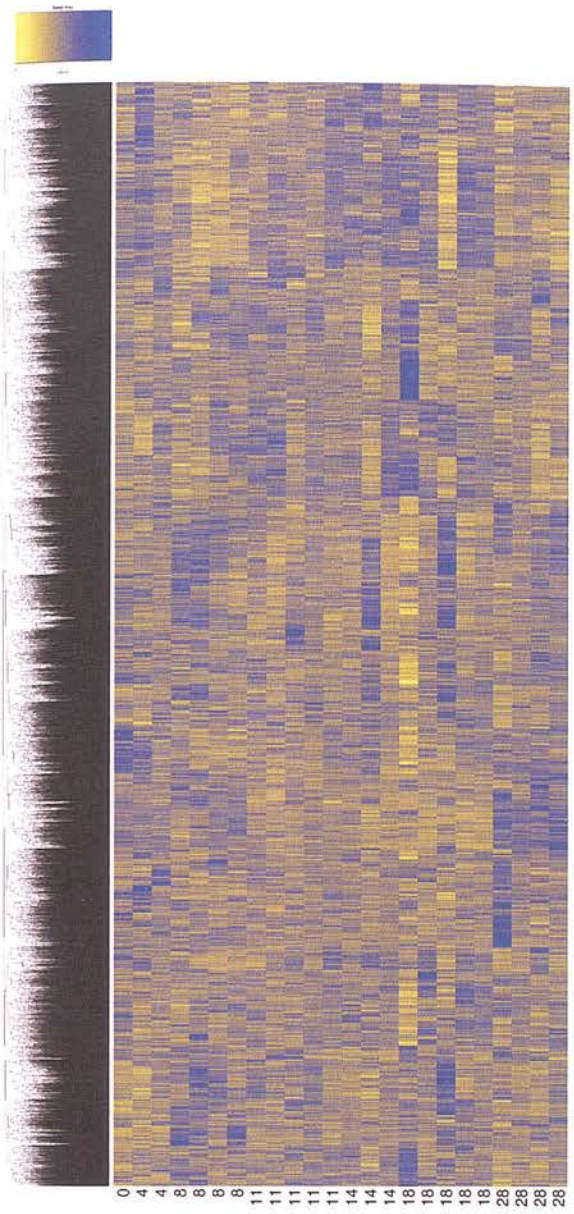


Figure 4-6.No obvious transcriptional signature remains after accounting for gene expression associated with alternative activation (YM1) and proliferation (Ki67^{hi}). Heatmap of internally studentized residual variances after fitting the linear model to describe gene expression as a function of %YM1 positive and %Ki67^{hi} positive macrophages. Hierarchical clustering was performed using a eucledian distance metric.

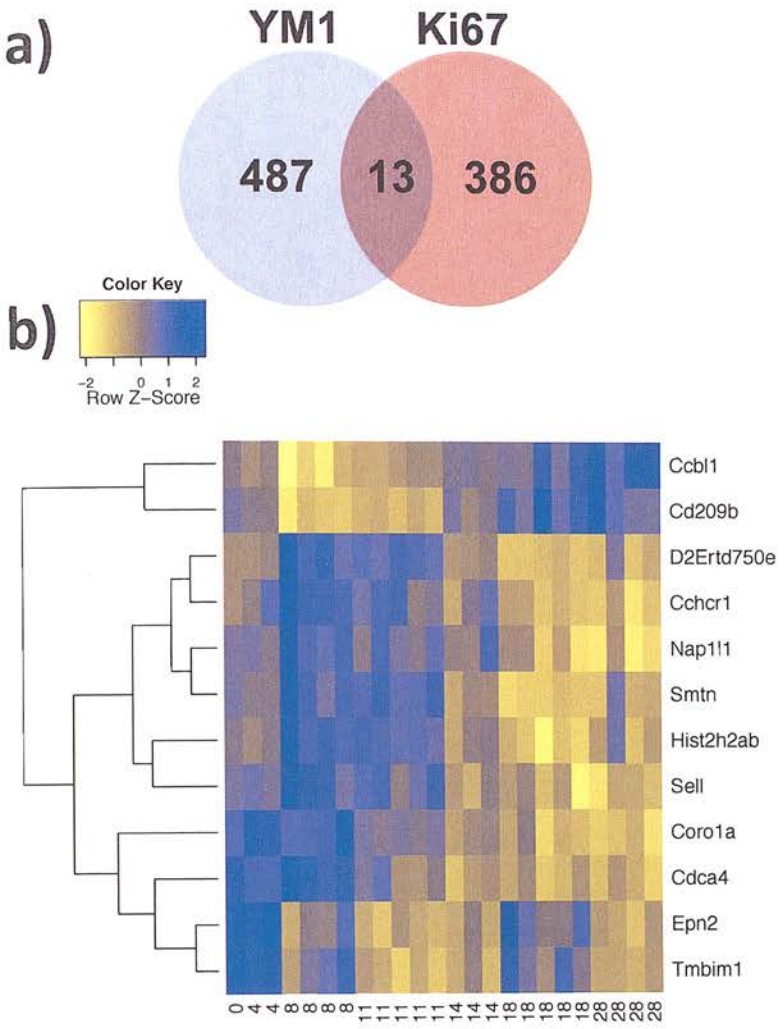


Figure 4-7.Few genes significantly associate with both alternative activation and proliferation. A) Venn diagram showing the number of genes significantly correlated with %YM1 positive and %Ki67^{hi} levels, and the overlap between the two. B) Gene expression profiles of those genes significantly correlated with both %YM1 positive and %Ki67^{hi} levels.

	Positive correlation	Negative correlation
YM1	270	231
$Ki67^{hi}$	320	80

Table 4-2. Genes that correlate significantly with either %YM1 positivity of $Ki67^{hi}$ levels as determined by FACS analysis showing the split between positive and negative correlation.

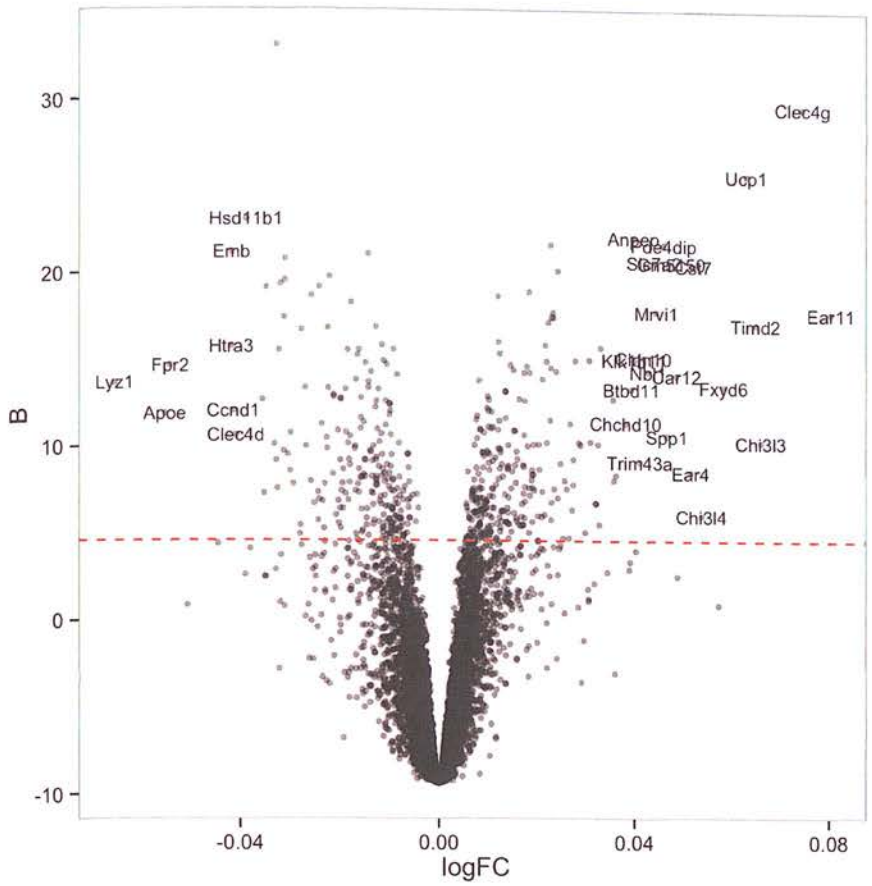


Figure 4-8. Volcano plot showing the degree of correlation of gene expression with %YM1 positivity (x axis) against statistical significance (y axis). Statistically significant genes are above the red dashed line (B=4.6). The top 30 most highly correlated, positively or negatively, sorted by logFC on all significant genes, are annotated on the graph.

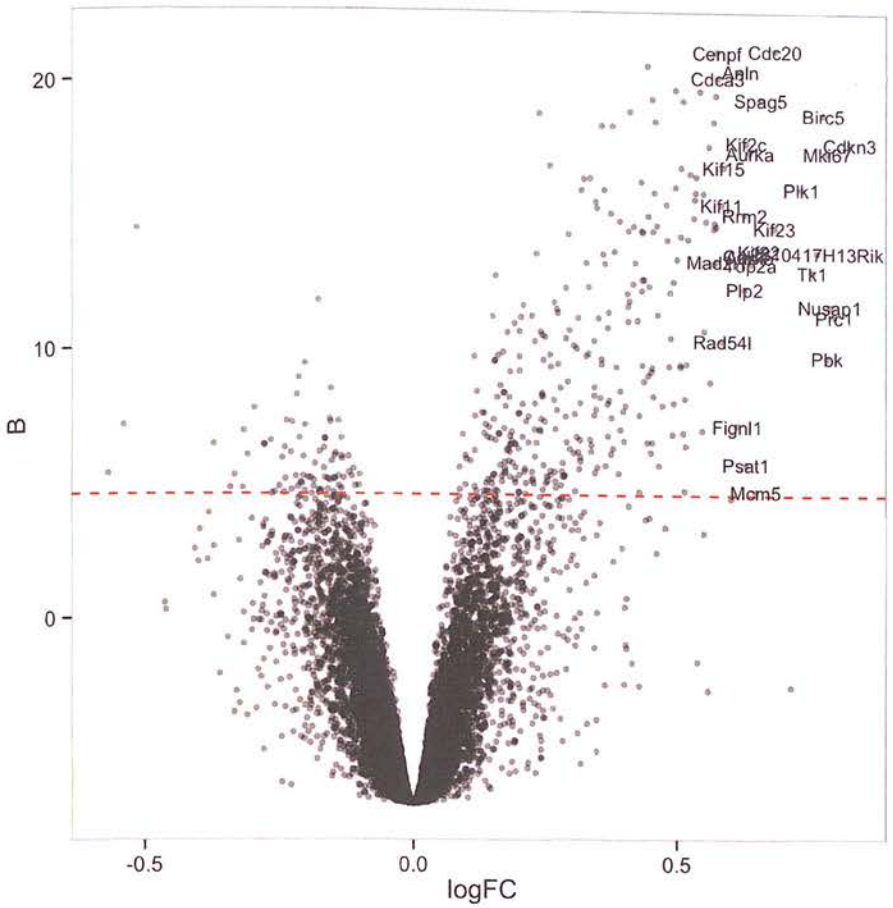


Figure 4-9. Volcano plot showing the degree of correlation of gene expression with %Ki67^{hi} positivity (x axis) against statistical significance (y axis). Statistically significant genes are above the red dashed line (B=4.6). The top 30 most highly correlated, positively or negatively, sorted by logFC on all significant genes, are annotated on the graph.

4.3.4 GSEA AND FOLLOW-UP OF KI67-ASSOCIATED GENES HELPS DEFINE THE AAMΦ PROLIFERATION TRANSCRIPTIONAL PROFILE

GSEA.

We have inferred the expression of genes involved in macrophage proliferation using the frequency of $Ki67^{hi}$ macrophages measured by FACS as a surrogate for proliferation. As a starting point to better understand the regulatory processes underpinning macrophage proliferation *in vivo* we performed GSEA. GSEA was applied to $Ki67^{hi}$ associated genes as described in the methods section in order to identify differentially regulated KEGG pathways. Using significance cutoffs suggested by the GSEA authors (nominal P-value 0.05, FDR 25%) fifteen KEGG pathways were positively correlated with $Ki67^{hi}$ levels, and none were significantly negatively correlated (Table 4-3).

Significant pathways broadly fall into two classes. The most numerous set of significant pathways relate to the structural aspects of the cell-cycle and include 'meiosis', 'mitosis' and 'DNA replication', as well as additional pathways associated with DNA processing, for example, 'DNA excision and repair' and 'homologous recombination'. A second set of pathways constitute metabolic processes supporting proliferation, in particular the production and processing of purine and pyrimidine nucleosides. This set includes 'purine metabolism', 'pyrimidine metabolism' and 'drug metabolism other enzymes'. This analysis confirms that our Ki67-associated gene expression signature is indeed a cell-cycle associated gene expression profile.

KEGG Pathway	Gene Set Size	ES	NES	Nominal p-val	FDR q-val	FWER p-val
KEGG OOCYTE MEIOSIS	89	0.5	1.844	0.004	0.075	0.071
KEGG CELL CYCLE	105	0.6	1.840	0.002	0.040	0.075
KEGG DRUG METABOLISM OTHER ENZYMES	22	0.5	1.760	0.004	0.083	0.196
KEGG PROGESTERONE MEDIATED OOCYTE MATURATION	68	0.4	1.751	0.002	0.070	0.218
KEGG HOMOLOGOUS RECOMBINATION	22	0.7	1.730	0.025	0.074	0.264
KEGG DNA REPLICATION	33	0.9	1.701	0.006	0.090	0.345
KEGG NUCLEOTIDE EXCISION REPAIR	43	0.6	1.670	0.039	0.110	0.436
KEGG SPLICEOSOME	89	0.6	1.670	0.029	0.097	0.437
KEGG CYSTEINE AND METHIONINE METABOLISM	22	0.6	1.658	0.010	0.096	0.467
KEGG PROTEASOME	38	0.8	1.653	0.014	0.091	0.482
KEGG PYRIMIDINE METABOLISM	78	0.5	1.619	0.049	0.116	0.571
KEGG PURINE METABOLISM	126	0.4	1.618	0.010	0.107	0.574
KEGG MISMATCH REPAIR	22	0.7	1.618	0.039	0.099	0.575
KEGG AMINOACYL TRNA BIOSYNTHESIS	40	0.6	1.582	0.039	0.117	0.675
KEGG P53 SIGNALING PATHWAY	55	0.4	1.487	0.047	0.193	0.871

Table 4-3. Gene set enrichment analysis results of KEGG pathways positively correlated with %Ki67^{hi} expression showing that a cell-cycle signature is associated with Ki67 protein expression levels. ES = enrichment score, NES = normalized enrichment score, FDR q-val= Benjamini-Hochberg corrected p-value, FWER p-val = family wise error rate corrected p-value.

CIS-REGULATORY ANALYSIS.

We have confirmed that our linear modelling approach correctly integrates our microarray dataset with the FACS analysis to identify genes associated with the macrophage proliferation program. Next we wanted to identify candidate factors that may regulate this proliferation program. In order to do this we analysed the proximal promoters of Ki67-associated genes for over-represented TFBS using Homer (Heinz et al., 2010).

We took a slightly different, more biologically informed, approach to *cis*-regulatory analysis to the previous analysis presented in chapter 2. Firstly, 7,319 of the peritoneal macrophage TSS predictions generated in Chapter 2 were assigned to expressed genes from this microarray experiment. A set of TSS belonging to proliferation-associated genes was created by assigning 225 TSS belonging to genes that significantly positively correlated with %Ki67^{hi} levels. Similarly, an AAMΦ-associated TSS set was made by identifying the 182 genes with TSS predictions that positively correlated with %YM1 positivity. H3K4 trimethylation (H3K4me3) is an epigenetic marker for promoter and enhancer regions (Ghisletti et al., 2010). H3K4me3 peaks defined by ChIP-Seq in peritoneal thioglycollate elicited macrophages (Heinz et al., 2010) were downloaded, and promoter regions flanking TSS were identified based on these peaks. Following the protocols in Heinz et al. (2010) we defined a gene's promoter as any H3K4Me3 peak within ±1kb of a predicted TSS. In total 1,098 H3K4me3 peaks were found to mark the promoters of YM1-associated genes, and 1,325 H3K4me3 peaks marked the promoters of Ki67-associated genes. To identify over-represented TFBS associated with Ki67 and YM-1, hypergeometric analysis for over-represented TFBS was performed on these TSS regions using Homer (Heinz et al., 2010) with the mouse genome serving as a background control.

31 known TFBS were over-represented in YM1-associated promoter regions (Table 4-4), and 40 TFBS were found to be over-represented in Ki67^{hi}-associated promoter regions (Table 4-5). 13 TFBS motifs were common to both Ki67^{hi}-associated and YM1-associated promoters (Table 4-6), including the downstream mediator of M-CSF signalling Elk1 (Yao et al., 2005), the AAMΦ marker Klf4 (Liao et al., 2011) and multiple binding sites for ETS family transcription factors implicated in macrophage development (Wu et al., 1994). STAT6 controls AAMΦ by mediating IL-4-dependent signalling (Takeda et al., 1996). Interestingly, the STAT6 motif was found to be statistically overrepresented in the promoters of YM1-associated genes, but was not found in Ki67^{hi}-associated promoters. This observation supports our understanding that STAT6 is required for AAMΦ. Importantly, the absence of a STAT6 motif in the promoters of proliferation-associated genes further suggests that STAT6 does not directly engage with the machinery of cellular proliferation. This is consistent with our PCA and linear modelling analyses, and supports a model wherein STAT6-dependent transcription and IL-4-dependent proliferation are mechanistically independent processes.

Motif name	YM1 P value	YM1 q value	#YM1 targets with motif	% YM1 targets with motif	# BG seqs with motif	% BG seqs with motif
SPDEF(ETS)	1.00E-14	0	577	52.55%	17772	40.68%
Tbx5(T-box)	1.00E-10	0	971	88.43%	35493	81.25%
IRF4(IRF)	1.00E-08	0	206	18.76%	5488	12.56%
ISRE(IRF)	1.00E-08	0	49	4.46%	741	1.70%
BORIS(Zf)	1.00E-07	0	262	23.86%	7545	17.27%
Mef2c(MADS)	1.00E-07	0	150	13.66%	3830	8.77%
Egr	1.00E-06	0	267	24.32%	7873	18.02%
EKLF(Zf)	1.00E-06	0	178	16.21%	4918	11.26%
PRDM1	1.00E-05	1.00E-04	249	22.68%	7640	17.49%
NFAT(RHD)	1.00E-05	1.00E-04	324	29.51%	10401	23.81%
SCL/HPC7	1.00E-04	3.00E-04	1055	96.08%	40722	93.22%
Znf263(Zf)	1.00E-04	5.00E-04	878	79.96%	32742	74.95%
Srebp1a(HLH)	0.001	0.0018	118	10.75%	3371	7.72%
FOXA1:AR	0.001	0.0024	43	3.92%	962	2.20%
CEBP(bZIP)	0.001	0.0044	204	18.58%	6532	14.95%
USF1(HLH)	0.001	0.0048	270	24.59%	8980	20.56%
STAT6	0.001	0.0058	202	18.40%	6503	14.89%
STAT5(Stat)	0.001	0.0066	146	13.30%	4505	10.31%

Table 4-4. Homer results showing over-represented TFBS unique to AAMΦ-associated promoters. AAMΦ-associated promoters were defined as H3K4me3 peaks present within ±1kb of TSS that were significantly positively correlated with %YM1 positive cells. In total 1,098 AAMΦ-associated H3K4me3 peaks were defined. BG seqs = background sequences.

Motif name	Ki67 P value	Ki67 q value	# Ki67 targets with motif	% Ki67 targets with motif	# BG seqs with motif	% BG seqs with motif
E2F7(E2F)	1.00E-108	0	459	34.64%	4713	11.39%
E2F4(E2F	1.00E-98	0	845	63.77%	14542	35.13%
E2F(E2F)	1.00E-75	0	249	18.79%	1960	4.74%
NFY(CCAAT)	1.00E-71	0	762	57.51%	13814	33.37%
E2F6(E2F)	1.00E-56	0	808	60.98%	16278	39.32%
GFY-Staf	1.00E-40	0	168	12.68%	1565	3.78%
CHR	1.00E-25	0	377	28.45%	6919	16.72%
GFY	1.00E-17	0	122	9.21%	1615	3.90%
BMYB(HTH)	1.00E-14	0	804	60.68%	20630	49.84%
MYB(HTH)	1.00E-13	0	913	68.91%	24325	58.77%
JunD(bZIP)	1.00E-10	0	107	8.08%	1659	4.01%
ETS:RUNX	1.00E-07	0	108	8.15%	1981	4.79%
TEAD4(TEA)	1.00E-05	0	432	32.60%	11186	27.02%
CRE(bZIP)	1.00E-05	0	218	16.45%	5029	12.15%
Nanog (Hox)	1.00E-05	0	1180	89.06%	34997	84.55%
GFX	1.00E-04	4.00E-04	44	3.32%	721	1.74%
FOXA1	1.00E-04	4.00E-04	479	36.15%	12878	31.11%
ETS1(ETS)	0.001	8.00E-04	658	49.66%	18454	44.58%
Lhx3 (Hox)	0.001	0.0011	401	30.26%	10698	25.85%
Atf1(bZIP)	0.001	0.0022	345	26.04%	9142	22.09%
ZBTB33	0.001	0.0012	98	7.40%	2108	5.09%
Foxh1 (Forkhead)	0.001	0.0048	273	20.60%	7137	17.24%
Pax7(Hox)	0.001	0.001	21	1.58%	261	0.63%
Tcf4(HMG)	0.01	0.0067	223	16.83%	5733	13.85%
bHLHE40(HLH)	0.01	0.0073	144	10.87%	3506	8.47%
EWS:FLI1-fusion(ETS)	0.01	0.0072	391	29.51%	10689	25.82%
NF1:FOXA1	0.01	0.0059	32	2.42%	546	1.32%

Table 4-5. Homer results showing over-represented TFBS in promoters of %Ki67^{hi} associated transcripts. %Ki67^{hi}-associated promoters were defined as H3K4me3 peaks within ±1kb of TSS that were significantly positively correlated with %Ki67^{hi}. In total 1,325 AAMΦ-associated H3K4me3 peaks were defined. BG seqs = background sequences.

Motif name	YM1 P value	YM1 q value	#YM1 targets with motif	% YM1 targets with motif	# BG seqs with motif	% BG seqs with motif
Sp1(Zf)	1.00E-37	0	598	54.46%	15425	35.31%
Elk1(ETS)	1.00E-35	0	596	54.28%	15569	35.64%
Klf4(Zf)	1.00E-26	0	472	42.99%	12159	27.83%
ELF1(ETS)	1.00E-21	0	514	46.81%	14271	32.67%
Fli1(ETS)	1.00E-21	0	719	65.48%	22318	51.09%
Elk4(ETS)	1.00E-21	0	552	50.27%	15767	36.09%
GABPA(ETS)	1.00E-08	0	551	50.18%	18137	41.52%
ETS(ETS)	1.00E-07	0	302	27.50%	9109	20.85%
Maz(Zf)	1.00E-07	0	892	81.24%	32567	74.55%
NRF1	1.00E-05	1.00E-04	244	22.22%	7442	17.04%
HIF2a(HLH)	1.00E-04	5.00E-04	249	22.68%	7865	18.00%
NRF1(NRF)	0.001	0.0031	231	21.04%	7460	17.08%
E2F1(E2F)	0.001	0.0033	302	27.50%	10108	23.14%

Motif name	Ki67 P value	Ki67 q value	# Ki67 targets with motif	% Ki67 targets with motif	# BG seqs with motif	% BG seqs with motif
Sp1(Zf)	1.00E-53	0	714	53.89%	15425	35.31%
Elk1(ETS)	1.00E-28	0	672	50.72%	15569	35.64%
Klf4(Zf)	1.00E-23	0	516	38.94%	12159	27.83%
ELF1(ETS)	1.00E-21	0	602	45.43%	14271	32.67%
Fli1(ETS)	1.00E-11	0	803	60.60%	22318	51.09%
Elk4(ETS)	1.00E-37	0	713	53.81%	15767	36.09%
GABPA(ETS)	1.00E-09	0	665	50.19%	18137	41.52%
ETS(ETS)	1.00E-16	0	411	31.02%	9109	20.85%
Maz(Zf)	0.001	0.001	997	75.25%	32567	74.55%
NRF1	1.00E-14	0	340	25.66%	7442	17.04%
HIF2a(HLH)	1.00E-08	0	321	24.23%	7865	18.00%
NRF1(NRF)	1.00E-18	0	358	27.02%	7460	17.08%
E2F1(E2F)	1.00E-76	0	599	45.21%	10108	23.14%

Table 4-6. Homer results detailing over-represented TFBS common in both 1,098 AAMΦ and 1,325 %Ki67^{hi} associated promoters (H3K4me3 peaks within ±1kb of gene TSS). BG seqs = background sequences.

A ROLE FOR E2F TRANSCRIPTION FACTORS IN MACROPHAGE PROLIFERATION?

The most over-represented TFBS's associated with Ki67 associated genes are for E2F family transcription factors which modulate DNA synthesis and cell cycle progression (Table 4-5, (Dyson, 1998). E2F family members are highly conserved, in mouse a total of 8 E2F family members (*E2f1, E2f2, E2f3, E2f4, E2f5, E2f6, E2f7* and *E2f8*) play both positive and negative roles in cell cycle regulation (Dyson, 1998). The transcription factors *Tfdp1* (Transcription factor DP1) or *Tfdp2* (Transcription factor DP2) and *Eapp* (E2F-associated phosphoprotein) interact with E2F family members to mediate E2F transcriptional activity (Dyson, 1998).

We explored the dynamic expression profiles E2F complex members to ascertain whether these factors may govern the commitment of IL-4-dependent cell cycle in macrophages (Figure 4-10). Only *E2f2* and *E2f8* showed a statistically significant correlation with Ki67, and none correlated significantly with YM1. We did however observe a general tendency for the expression of the DNA binding E2F transcription factors (*E2f1, E2f2* and *E2f3*) to correlate with Ki67, consistent with their role in cell cycle control and DNA synthesis (Dyson, 1998). Interestingly the peak of *E2f1* and *E2f3* expression is day 8, and this correlates with our expectation for the peak of IL-4-independent proliferation during *L. sigmodontis* infection (S Jenkins unpublished). *E2f2* levels were induced by day 8 but did not peak until day 11 (Figure 4-10). This profile that is consistent with the IL-4-dependent component of proliferation. Thus there is an apparent divergence in correlation between E2F family members and the IL-4-independent and IL-4-dependent components macrophage proliferation. This raises the possibility that differing proliferative programmes require different E2F family complexes. To test this hypothesis we induced macrophages to proliferate *in vivo* in response to IL-4 or recombinant M-CSF, the mechanism underlying the IL-4 independent phase of macrophage proliferation (S Jenkins, unpublished). RT-PCR

analysis of FACS purified macrophages confirmed that E2F transcription factor gene expression was associated with proliferation. There was however no clear distinction in the E2F transcriptional profiles between M-CSF and IL-4 induced proliferation (data not shown). While this supports the idea that E2F factors drive macrophage proliferation, from this experiment we are unable to state whether distinct E2F transcriptional profiles are driven in response to IL-4 and M-CSF.

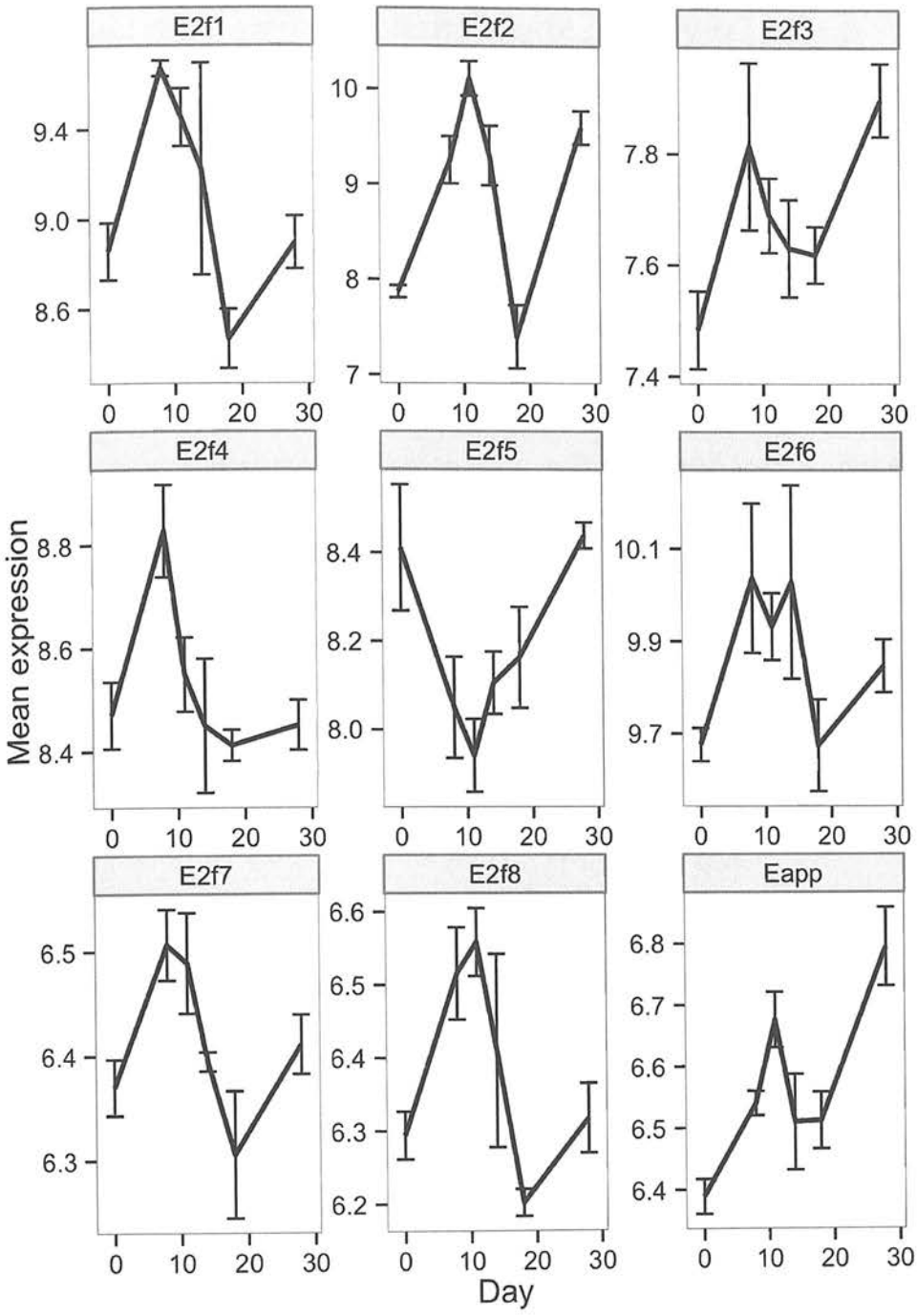


Figure 4-10. Expression profiles of E2F gene family members in F4/80^{hi} macropahges over the time course of *L. sigmodontis* infection. E2F1 and E2F2 expression profiles appear to peak with the M-CSF and IL-4-dependent components of proliferation at days 8 and 11 respectively. Error bars represent SEM calculated for each day independently. Days 0 and 4 were pooled to form naïve group (day 0).

4.3.5 COMPARISONS BETWEEN *L. SIGMODONTIS* AND *B. MALAYI*-ELICITED AAMΦ - IDENTIFICATION OF A 'CORE' NeMΦ EXPRESSION PROFILE

We have inferred nematode-induced IL-4Rα-dependent macrophage gene expression profiles both directly and indirectly in three separate experiments in this thesis. Using the *B. malayi* implant model in the BALB/c strain we have directly identified IL-4Rα-dependent gene expression by comparing gene expression profiles between WT and IL-4Rα$^{-/-}$ infected mice. Using the same model we have identified differential gene expression by comparing WT thioglycollate-elicited macrophages to WT *B. malayi*-elicited AAMΦ, a substantial component of which is IL-4Rα-dependent (Thomas et al., 2012a). Finally, using the *L. sigmodontis* infection model we have identified AAMΦ-associated genes by correlating gene expression profiles with YM1 levels defined by FACS analysis.

We set out to definitively compile a list of NeMΦ-associated genes by integrating the results of our microarray and RNA-Seq analyses. By identifying common NeMΦ-associated genes using experiments performed in different host strains (BALB/c and C57BL/6), and with different parasite species (*B. malayi* and *L. sigmodontis*) that localize to different body cavities we are able to define a core list of such genes by removing species- and platform-specific biases.

No analytically meaningful way exists to directly compare gene expression profiles measured by RNA-Seq and microarray profiling as the data types are fundamentally different. We can however compare results from these two data types using a non-parametric rank-based approach. The OrderedList meta-analysis technique (Lottaz et al., 2006) determines the degree of similarity between two ranked gene lists by computing a 'similarity score'. The similarity score is based on the overlap between two ranked lists, weighting genes towards the top of the list, importantly the score depends upon rank

rather than expression level. A p-value is assigned to the similarity score by means of a permutation test. This is achieved by randomising one of the lists and re-calculating the similarity score multiple times. The final results are a p-value and a list of the genes that contribute most to the similarity score (Lottaz et al., 2006).

To identify genes positively associated with alternative activation we performed two list comparisons. Genes from our microarray analysis were ranked according to the significance of their correlation with YM1, this list was compared against a list of gene rankings derived from the WT-NeMΦ vs. WT-ThioMΦ comparison from our RNA-Seq analysis. We observed a highly significant similarity between the two lists (P value < 0.0001, based on 100,000 permutations) with 390 genes contributing towards this similarity (Figure 4-11a). The same analysis was performed, only this time we compared YM1-associated genes against a list of rankings drawn from the WT-NeMΦ vs. IL-4Rα$^{-/-}$-NeMΦ comparison from our RNA-Seq analysis. Again, we found a high degree of similarity between the two lists (P value < 0.0001, based on 100,000 permutations), with 343 genes contributing to this similarity score (Figure 4-11a). In total 241 genes were found to overlap between the two lists, and these comprise our 'core' NeMΦ-associated genes (Table 4-7). Within this list are all of the major genes we would expect to see, including *Chi3l3*, *Retnla*, *Arg1* and *Ppargc1b* (Loke et al., 2002; Pesce et al., 2009; Vats et al., 2006). Additionally multiple genes previously characterized as AAMΦ-associated 'immune effectors', including the chemokine receptor *Cxcr7*, cytokine receptors *Il1rl1* and *Il27ra*, as well as the cytokines *Bmp6* and *Wnt2* (Kersten et al., 2005; Kurowska-Stolarska et al., 2009; Rückerl et al., 2006; Sánchez-Martín et al., 2011; Staal et al., 2008). We also identified the complement component *C3* and inhibitory Fc-gamma receptor *Fcgr2b* as core AAMΦ-associated genes. Surface markers to facilitate the isolation and characterization of *in* vivo-derived AAMΦ are relatively scarce. The C-type

lectins mMGL1 and mMGL2 have previously been described as candidate AAMΦ-associated surface markers (Raes et al., 2005). mMGL2 (*mgl2*) but not mMGL1 (*Clec10a*) is identified amongst the list of 'core' AAMΦ-associated genes. This was due to the fact that *Clec10a* was not significantly correlated with YM1 expression in our microarray time course. Interestingly, other C-type lectins including *Clec4g* and *Clec4a3* were amongst the core list (Table 4-7) and may warrant further investigation as AAMΦ-associate marker proteins. Finally we have also identified a number of core NeMΦ genes that are associated with alternative activation in human monocyte-derived macrophages including *Alox5*, *Cd209* and *Irf4* (Martinez et al., 2006; Satoh et al., 2010). *Irf4* has previously been shown to regulate the transcription of AAMΦ marker genes (Satoh et al., 2010), our *cis*-regulatory analysis also identified an *Irf4* signature, this may represent a key conserved component of AAMΦ transcriptional regulation.

The analysis was performed for down-regulated genes by reversing the agene lists. We found 383 common down-regulated genes between WT *B. malayi*-elicited macrophages and YM1 correlated genes in the response to *L. sigmodontis* infection (Figure 4-11b). Interestingly, there was relatively less overlap (68 genes) between the genes negatively associated with YM1 from the microarray analysis and down-regulated genes between WT and IL-4Rα$^{-/-}$ NeMΦ, although the majority of these, 55, were shared with the 383 down-regulated genes from the set above. This comprises our list of 'core' down-regulated NeMΦ genes and can be found in Table 4-8.

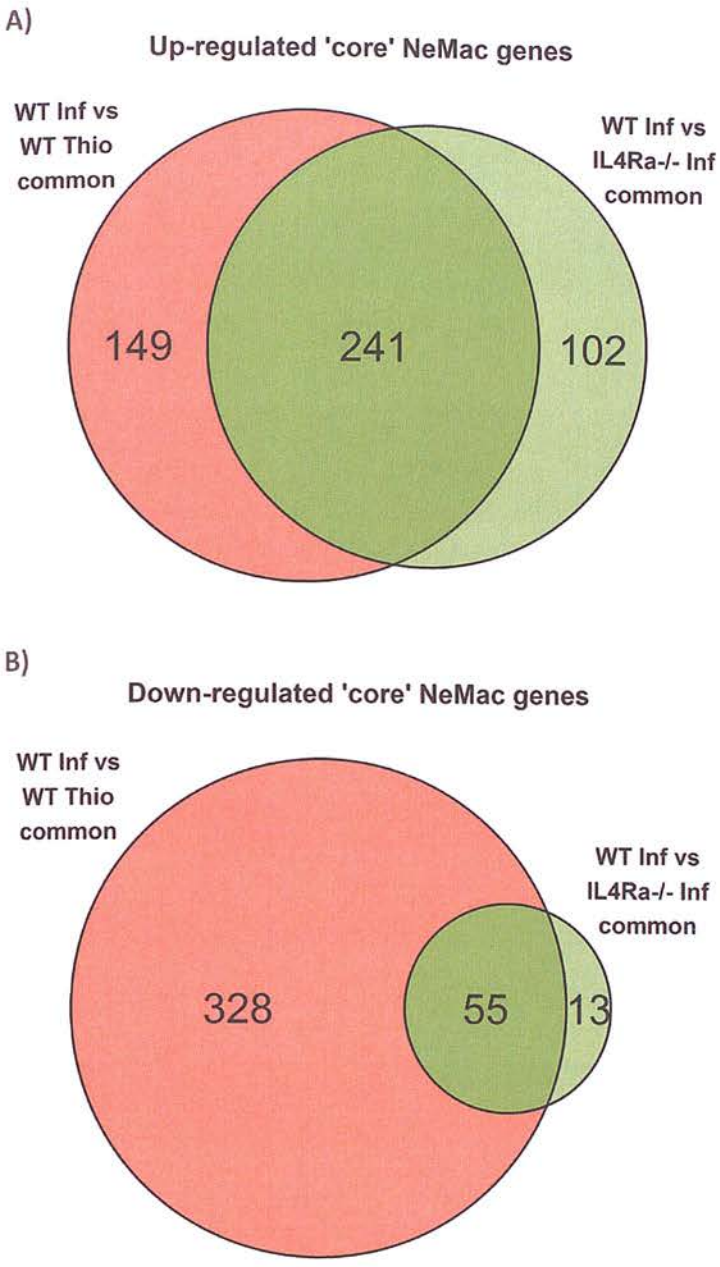


Figure 4-11. Summary of OrderedList analyses defining 'core'-NeMΦ genes. Venn diagrams showing the number of a) up-regulated, and b) down-regulated, genes determined by OrderedList analysis. Genes ranked by correlation with %YM1 positivity in the microarray analysis were compared against gene lists derived by ranking both the WT-NeMΦ vs. WT-ThioMΦ (left circles) and WT-NeMΦ vs. IL-4Rα$^{-/-}$-NeMΦ (right circles) comparisons from the RNA-Seq analyses.

Symbol	Entrez ID	Lito vs. YM1	WT Brugia vs. WT Thio	WT Brugia vs. IL-4Rα$^{-/-}$ Brugia	Description
Clec4g	75863	1.19E-13	5.67E-130	1.53E-225	C-type lectin domain family 4, member g
Ucp1	22227	3.47E-12	3.55E-75	5.75E-73	uncoupling protein 1 (mitochondrial, proton carrier)
Pde4dip	83679	7.21E-11	1.73E-26	2.40E-28	phosphodiesterase 4D interacting protein (myomegalin)
Slc7a2	11988	1.04E-10	1.69E-52	4.78E-32	solute carrier family 7 (cationic amino acid transporter, y+ system), member 2
Gm5150	381484	1.04E-10	2.93E-32	7.53E-51	predicted gene 5150
Cst7	13011	1.11E-10	1.47E-83	1.23E-77	cystatin F (leukocystatin)
Hfe	15216	2.82E-10	4.99E-47	4.85E-43	hemochromatosis
Mrvi1	17540	7.09E-10	5.83E-51	8.46E-79	MRV integration site 1
Tln2	70549	7.09E-10	2.55E-70	7.59E-41	talin 2
Ear11	93726	7.09E-10	1.19E-125	8.68E-223	eosinophil-associated, ribonuclease A family, member 11
Clec4a3	73149	8.86E-10	1.05E-24	1.00E-10	C-type lectin domain family 4, member a3
Arl4d	80981	3.00E-09	1.61E-34	1.05E-42	ADP-ribosylation factor-like 4D
Mgl2	216864	3.26E-09	1.30E-90	2.59E-173	macrophage galactose N-acetyl-galactosamine specific lectin 2
Prkaa2	108079	5.15E-09	8.92E-50	5.73E-76	protein kinase, AMP-activated, alpha 2 catalytic subunit
Nrg4	83961	5.35E-09	1.16E-08	9.56E-10	neuregulin 4
Dzip1	66573	6.38E-09	8.01E-24	4.48E-28	DAZ interacting protein 1
Prkcq	18761	6.48E-09	1.69E-34	7.71E-18	protein kinase C, theta
Nbl1	17965	8.00E-09	2.15E-43	1.48E-73	neuroblastoma, suppression of

					tumorigenicity 1
A530064D06Rik	328830	8.00E-09	4.54E-37	1.49E-36	RIKEN cDNA A530064D06 gene
Gsto1	14873	8.25E-09	3.29E-13	5.59E-15	glutathione S-transferase omega 1
Car12	76459	9.04E-09	8.37E-51	1.45E-66	carbonic anyhydrase 12
Fxyd6	59095	1.48E-08	1.47E-83	1.76E-77	FXYD domain-containing ion transport regulator 6
Btbd11	74007	1.65E-08	1.61E-58	1.11E-77	BTB (POZ) domain containing 11
Fut8	53618	1.80E-08	1.02E-07	8.98E-16	fucosyltransferase 8
Rab38	72433	1.82E-08	6.36E-05	8.84E-05	RAB38, member of RAS oncogene family
Bcat2	12036	2.47E-08	6.02E-07	2.57E-09	branched chain aminotransferase 2, mitochondrial
Tcf23	69852	2.47E-08	3.25E-76	3.97E-77	transcription factor 23
Aldh1l1	107747	2.61E-08	4.00E-35	2.37E-14	aldehyde dehydrogenase 1 family, member L1
Pdp2	382051	4.17E-08	2.88E-02	3.80E-03	pyruvate dehyrogenase phosphatase catalytic subunit 2
Lrrc1	214345	5.72E-08	3.60E-05	1.22E-04	leucine rich repeat containing 1
Baiap2	108100	6.39E-08	3.60E-04	8.91E-05	brain-specific angiogenesis inhibitor 1-associated protein 2
Chchd10	103172	7.02E-08	7.82E-25	1.09E-33	coiled-coil-helix-coiled-coil-helix domain containing 10
Slc46a1	52466	9.45E-08	1.78E-07	3.16E-11	solute carrier family 46, member 1
Agl	77559	9.45E-08	1.85E-03	1.47E-04	amylo-1,6-glucosidase, 4-alpha-glucanotransferase
Aldh5a1	214579	9.87E-08	2.15E-05	9.67E-09	aldhehyde dehydrogenase family 5, subfamily A1
Hibadh	58875	1.05E-07	3.46E-13	3.82E-10	3-hydroxyisobutyrate dehydrogenase
Sirt5	68346	1.14E-07	1.42E-05	9.85E-07	sirtuin 5 (silent mating type information regulation 2 homolog) 5

					(S. cerevisiae)
Chi3l3	12655	1.58E-07	5.97E-34	3.35E-67	chitinase 3-like 3
Cd209a	170786	1.58E-07	2.49E-90	1.82E-153	CD209a antigen
Prkcdbp	109042	1.76E-07	2.63E-14	8.56E-25	protein kinase C, delta binding protein
Sla	20491	1.76E-07	5.28E-31	1.30E-11	src-like adaptor
Isyna1	71780	1.86E-07	5.03E-40	1.42E-29	myo-inositol 1-phosphate synthase A1
Prodh	19125	2.35E-07	9.58E-15	1.27E-07	proline dehydrogenase
Lpar1	14745	3.09E-07	3.47E-34	4.80E-09	lysophosphatidic acid receptor 1
Tanc1	66860	3.30E-07	4.42E-23	1.95E-15	tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 1
Ppargc1b	170826	3.34E-07	6.76E-14	8.48E-07	peroxisome proliferative activated receptor, gamma, coactivator 1 beta
Brwd1	93871	3.91E-07	3.30E-17	9.77E-17	bromodomain and WD repeat domain containing 1
C77370	245555	4.00E-07	9.17E-45	1.77E-12	expressed sequence C77370
Cpt2	12896	4.88E-07	1.63E-02	2.22E-05	carnitine palmitoyltransferase 2
Bmp6	12161	5.13E-07	2.19E-44	1.04E-60	bone morphogenetic protein 6
St7	64213	5.56E-07	3.77E-12	6.62E-08	suppression of tumorigenicity 7
Mical2	320878	6.48E-07	6.19E-12	2.89E-25	microtubule associated monoxygenase, calponin and LIM domain containing 2
Cadps2	320405	7.70E-07	3.91E-46	3.81E-21	Ca2+-dependent activator protein for secretion 2
Cd209d	170779	8.25E-07	2.38E-75	3.10E-113	CD209d antigen
Plxna2	18845	8.25E-07	5.19E-12	2.07E-47	plexin A2
Fbxo18	50755	8.33E-07	3.41E-03	3.13E-03	F-box protein 18
Cdc14b	218294	1.12E-06	3.77E-12	9.10E-06	CDC14 cell division cycle 14B
C3	12266	1.15E-06	1.74E-32	1.66E-07	complement component 3

Pla2g7	27226	1.17E-06	5.71E-24	1.21E-12	phospholipase A2, group VII (platelet-activating factor acetylhydrolase, plasma)
Pigx	72084	1.20E-06	2.05E-08	1.79E-06	phosphatidylinositol glycan anchor biosynthesis, class X
Tmprss13	214531	1.54E-06	8.81E-23	4.72E-21	transmembrane protease, serine 13
Cpne8	66871	1.63E-06	1.36E-31	2.73E-24	copine VIII
Cd55	13136	1.72E-06	1.19E-36	7.33E-14	CD55 antigen
Lpcat3	14792	1.74E-06	3.53E-04	2.09E-03	lysophosphatidylcholine acyltransferase 3
Nphp1	53885	1.82E-06	6.73E-44	2.03E-26	nephronophthisis 1 (juvenile) homolog (human)
Pon3	269823	1.94E-06	1.08E-10	3.26E-15	paraoxonase 3
Trim72	434246	2.03E-06	1.76E-27	3.75E-15	tripartite motif-containing 72
Pfkfb4	270198	2.19E-06	3.93E-08	3.06E-13	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4
Rhou	69581	2.24E-06	2.79E-14	1.73E-06	ras homolog gene family, member U
Trmt2b	215201	2.46E-06	1.25E-05	1.83E-07	TRM2 tRNA methyltransferase 2 homolog B (S. cerevisiae)
Sft2d2	108735	2.70E-06	2.20E-09	2.71E-24	SFT2 domain containing 2
Pvrl2	19294	2.89E-06	7.30E-22	6.30E-09	poliovirus receptor-related 2
Dym	69190	2.91E-06	1.69E-05	4.53E-05	dymeclin
Bfsp1	12075	3.09E-06	1.91E-07	9.53E-14	beaded filament structural protein 1, in lens-CP94
Oxnad1	218885	3.09E-06	3.95E-06	6.30E-06	oxidoreductase NAD-binding domain containing 1
Lrg1	76905	3.53E-06	4.11E-35	5.64E-13	leucine-rich alpha-2-glycoprotein 1
Chi3l4	104183	3.69E-06	7.87E-36	1.69E-57	chitinase 3-like 4
Sdsl	257635	3.92E-06	8.67E-16	3.05E-14	serine dehydratase-like

Emr1	13733	4.19E-06	6.69E-14	9.10E-10	EGF-like module containing, mucin-like, hormone receptor-like sequence 1
Erlec1	66753	4.19E-06	7.45E-04	1.64E-03	endoplasmic reticulum lectin 1
Nln	75805	4.97E-06	9.17E-08	5.63E-10	neurolysin (metallopeptidase M3 family)
Ramp3	56089	5.00E-06	8.29E-90	2.31E-83	receptor (calcitonin) activity modifying protein 3
Nudt19	110959	5.26E-06	1.44E-15	7.53E-08	nudix (nucleoside diphosphate linked moiety X)-type motif 19
BC026585	226527	5.26E-06	8.83E-04	1.20E-08	cDNA sequence BC026585
Slc41a2	338365	5.29E-06	1.64E-06	8.31E-14	solute carrier family 41, member 2
Stim2	116873	5.36E-06	1.62E-03	2.05E-03	stromal interaction molecule 2
Ust	338362	5.41E-06	2.14E-58	1.20E-44	uronyl-2-sulfotransferase
Acss1	68738	5.46E-06	1.64E-08	1.71E-41	acyl-CoA synthetase short-chain family member 1
Cd209b	69165	5.52E-06	4.26E-93	6.95E-157	CD209b antigen
Pi16	74116	5.75E-06	1.07E-61	3.27E-39	peptidase inhibitor 16
Pcca	110821	6.04E-06	1.01E-05	6.89E-06	propionyl-Coenzyme A carboxylase, alpha polypeptide
Atp5a1	11946	6.24E-06	9.51E-04	8.36E-04	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1
Nt5e	23959	7.55E-06	1.21E-81	1.11E-64	5' nucleotidase, ecto
Gca	227960	7.63E-06	1.70E-11	4.47E-05	grancalcin
Setd7	73251	8.48E-06	7.76E-13	3.44E-05	SET domain containing (lysine methyltransferase) 7
Fech	14151	9.60E-06	1.29E-04	7.24E-08	ferrochelatase
Ndufaf1	69702	9.60E-06	2.40E-03	1.23E-03	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, assembly

					factor 1
Aif1l	108897	9.94E-06	2.07E-15	9.50E-12	allograft inflammatory factor 1-like
Adamts2	216725	1.02E-05	3.75E-69	1.11E-59	a disintegrin-like and metallopeptidase (reprolysin type) with thrombospondin type 1 motif, 2
Srxn1	76650	1.02E-05	3.37E-04	4.75E-08	sulfiredoxin 1 homolog (S. cerevisiae)
Fgfr1	14182	1.07E-05	5.44E-35	1.55E-18	fibroblast growth factor receptor 1
Prr13	66151	1.09E-05	3.55E-10	3.61E-09	proline rich 13
Pdxdc1	94184	1.11E-05	2.23E-02	1.03E-03	pyridoxal-dependent decarboxylase domain containing 1
Foxred1	235169	1.25E-05	1.68E-04	2.14E-02	FAD-dependent oxidoreductase domain containing 1
Prdx3	11757	1.27E-05	1.30E-16	7.47E-12	peroxiredoxin 3
Etf1	225363	1.32E-05	1.78E-09	2.05E-04	eukaryotic translation termination factor 1
Sdhc	66052	1.40E-05	1.33E-06	3.53E-10	succinate dehydrogenase complex, subunit C, integral membrane protein
Coq9	67914	1.40E-05	6.70E-04	3.48E-07	coenzyme Q9 homolog (yeast)
9130008F23Rik	71583	1.47E-05	8.89E-55	2.22E-56	RIKEN cDNA 9130008F23 gene
Gm826	329554	1.63E-05	4.44E-09	8.53E-15	predicted gene 826
Cd244	18106	1.69E-05	8.34E-42	1.01E-11	CD244 natural killer cell receptor 2B4
Fars2	69955	1.69E-05	2.82E-03	3.17E-03	phenylalanine-tRNA synthetase 2 (mitochondrial)
Meis1	17268	1.88E-05	2.74E-07	1.61E-08	Meis homeobox 1
Dixdc1	330938	1.97E-05	2.44E-73	6.13E-108	DIX domain containing 1
Gnmt	14711	2.10E-05	1.67E-06	1.21E-03	glycine N-methyltransferase
Insr	16337	2.14E-05	3.53E-33	4.83E-41	insulin receptor
Igf1	16000	2.15E-05	3.47E-17	4.20E-115	insulin-like growth

					factor 1
Chn2	69993	2.26E-05	3.31E-22	1.31E-17	chimerin (chimaerin) 2
Fcgr2b	14130	2.28E-05	7.04E-20	1.25E-20	Fc receptor, IgG, low affinity IIb
Slc22a23	73102	2.29E-05	3.35E-20	3.61E-28	solute carrier family 22, member 23
Ube2e2	218793	2.34E-05	1.33E-07	9.02E-04	ubiquitin-conjugating enzyme E2E 2
Pigz	239827	2.53E-05	9.66E-08	1.92E-08	phosphatidylinositol glycan anchor biosynthesis, class Z
Lyplal1	226791	2.63E-05	1.32E-09	1.33E-13	lysophospholipase-like 1
Arg1	11846	2.86E-05	9.96E-11	3.31E-09	arginase, liver
Uqcrc2	67003	2.88E-05	1.03E-06	6.43E-07	ubiquinol cytochrome c reductase core protein 2
Pfkm	18642	2.95E-05	1.97E-07	1.92E-03	phosphofructokinase, muscle
Casp6	12368	3.13E-05	2.27E-04	1.62E-03	caspase 6
Tm7sf3	67623	3.17E-05	1.86E-05	1.89E-06	transmembrane 7 superfamily member 3
Acad10	71985	3.64E-05	1.34E-05	2.77E-06	acyl-Coenzyme A dehydrogenase family, member 10
Trpm3	226025	3.66E-05	4.88E-03	1.45E-04	transient receptor potential cation channel, subfamily M, member 3
Pglyrp1	21946	3.99E-05	9.56E-78	6.62E-38	peptidoglycan recognition protein 1
Tns4	217169	4.06E-05	2.75E-10	4.13E-06	tensin 4
Ndufa10	67273	4.06E-05	2.58E-04	6.54E-07	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex 10
Cela1	109901	4.09E-05	5.02E-03	9.65E-11	chymotrypsin-like elastase family, member 1
Aldh6a1	104776	4.80E-05	5.38E-03	1.38E-08	aldehyde dehydrogenase family 6, subfamily A1
Atp5g3	228033	4.81E-05	4.25E-07	9.91E-06	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit C3

					(subunit 9)
Abcd2	26874	4.97E-05	4.51E-28	2.02E-24	ATP-binding cassette, sub-family D (ALD), member 2
Sdhd	66925	5.06E-05	1.02E-08	1.35E-06	succinate dehydrogenase complex, subunit D, integral membrane protein
Kcnk6	52150	5.23E-05	7.46E-06	1.60E-13	potassium inwardly-rectifying channel, subfamily K, member 6
Aqp3	11828	5.33E-05	4.72E-55	1.30E-81	aquaporin 3
Tlr4	21898	5.51E-05	2.19E-08	3.70E-13	toll-like receptor 4
Acad11	102632	5.83E-05	2.84E-04	1.59E-03	acyl-Coenzyme A dehydrogenase family, member 11
Fam118b	109229	5.88E-05	4.23E-03	4.25E-06	family with sequence similarity 118, member B
Kcnn3	140493	6.98E-05	5.81E-88	5.17E-54	potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3
Slc34a2	20531	7.37E-05	2.06E-46	1.62E-69	solute carrier family 34 (sodium phosphate), member 2
A430107O13Rik	214642	8.87E-05	2.68E-14	3.03E-18	RIKEN cDNA A430107O13 gene
Tmem143	70209	9.94E-05	4.71E-04	4.04E-07	transmembrane protein 143
Trap1	68015	1.02E-04	3.15E-03	3.80E-03	TNF receptor-associated protein 1
Ppl	19041	1.06E-04	5.27E-44	1.44E-43	periplakin
Lysmd4	75099	1.06E-04	2.52E-02	2.18E-04	LysM, putative peptidoglycan-binding, domain containing 4
Thsd4	207596	1.10E-04	2.73E-13	4.84E-21	thrombospondin, type I, domain containing 4
Limk1	16885	1.14E-04	8.42E-19	4.36E-12	LIM-domain containing, protein kinase
Mapk11	19094	1.16E-04	1.59E-15	2.91E-05	mitogen-activated protein kinase 11
Cdkn2a	12578	1.16E-04	2.98E-39	1.39E-22	cyclin-dependent kinase

					inhibitor 2A
Idh3g	15929	1.17E-04	7.37E-05	1.76E-03	isocitrate dehydrogenase 3 (NAD+), gamma
Ipmk	69718	1.17E-04	8.14E-03	1.33E-03	inositol polyphosphate multikinase
Catsper2	212670	1.17E-04	8.26E-03	4.91E-02	cation channel, sperm associated 2
Oat	18242	1.37E-04	3.60E-08	3.75E-10	ornithine aminotransferase
Pecam1	18613	1.46E-04	2.23E-60	4.99E-36	platelet/endothelial cell adhesion molecule 1
Bzw2	66912	1.62E-04	4.95E-34	6.14E-18	basic leucine zipper and W2 domains 2
Timp1	21857	1.70E-04	2.72E-10	8.02E-15	tissue inhibitor of metalloproteinase 1
Adora3	11542	1.77E-04	5.05E-01	6.59E-01	adenosine A3 receptor
Adora3	11542	1.77E-04	7.98E-06	3.64E-03	adenosine A3 receptor
1700026L06Rik	69987	1.80E-04	2.61E-04	2.43E-03	RIKEN cDNA 1700026L06 gene
Lyve1	114332	1.83E-04	8.76E-49	1.97E-164	lymphatic vessel endothelial hyaluronan receptor 1
Mrps25	64658	1.89E-04	4.44E-05	1.89E-08	mitochondrial ribosomal protein S25
Acy3	71670	1.89E-04	1.73E-16	7.49E-08	aspartoacylase (aminoacylase) 3
Unc5b	107449	1.90E-04	3.24E-10	1.30E-13	unc-5 homolog B (C. elegans)
Dnajc3	100037258	1.96E-04	1.22E-12	4.82E-11	DnaJ (Hsp40) homolog, subfamily C, member 3
Dfna5	54722	2.01E-04	2.23E-05	8.95E-13	deafness, autosomal dominant 5 (human)
Idh2	269951	2.15E-04	1.09E-04	8.86E-11	isocitrate dehydrogenase 2 (NADP+), mitochondrial
Dusp23	68440	2.27E-04	9.52E-04	1.35E-03	dual specificity phosphatase 23
Fh1	14194	2.28E-04	7.63E-06	9.56E-06	fumarate hydratase 1
Ift52	245866	2.28E-04	1.14E-02	5.21E-03	intraflagellar transport 52
Ttc28	209683	2.33E-04	3.66E-25	1.01E-31	tetratricopeptide repeat domain 28

Il1rl1	17082	2.41E-04	1.97E-69	2.87E-130	interleukin 1 receptor-like 1
Kcnk13	217826	2.49E-04	1.09E-09	1.52E-21	potassium channel, subfamily K, member 13
Cux1	13047	2.72E-04	4.67E-05	1.08E-05	cut-like homeobox 1
Vmn2r26	56552	2.85E-04	5.85E-43	2.00E-20	vomeronasal 2, receptor 26
Zadh2	225791	2.88E-04	2.42E-16	1.70E-13	zinc binding alcohol dehydrogenase, domain containing 2
Idh1	15926	3.06E-04	6.19E-06	8.83E-06	isocitrate dehydrogenase 1 (NADP+), soluble
Gclm	14630	3.13E-04	8.74E-03	1.49E-02	glutamate-cysteine ligase, modifier subunit
Irf4	16364	3.55E-04	1.03E-20	4.30E-26	interferon regulatory factor 4
Ndufb5	66046	3.64E-04	2.56E-07	2.26E-08	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5
Cox7a2	12866	3.93E-04	1.90E-12	1.28E-11	cytochrome c oxidase, subunit VIIa 2
Hsp90b1	22027	4.33E-04	1.68E-06	3.56E-05	heat shock protein 90, beta (Grp94), member 1
Angptl2	26360	4.41E-04	7.56E-74	5.31E-54	angiopoietin-like 2
2410016O06Rik	71952	4.52E-04	3.98E-12	2.02E-07	RIKEN cDNA 2410016O06 gene
Pcdh7	54216	4.71E-04	2.35E-10	1.20E-07	protocadherin 7
Retnla	57262	5.05E-04	1.50E-33	2.57E-50	resistin like alpha
Pim3	223775	5.74E-04	7.36E-14	9.23E-10	proviral integration site 3
Srms	20811	5.97E-04	4.33E-04	2.79E-05	src-related kinase lacking C-terminal regulatory tyrosine and N-terminal myristylation sites
Golm1	105348	6.05E-04	1.94E-10	1.74E-12	golgi membrane protein 1
Ctse	13034	6.20E-04	1.37E-27	8.83E-72	cathepsin E
Ggcx	56316	6.73E-04	1.51E-02	3.70E-04	gamma-glutamyl carboxylase
Ikbkg	16151	6.81E-04	1.54E-02	5.15E-03	inhibitor of kappaB kinase gamma
Gstt3	103140	6.90E-04	2.01E-13	2.31E-17	glutathione S-

					transferase, theta 3
Cxcr7	12778	7.08E-04	3.45E-53	5.36E-39	chemokine (C-X-C motif) receptor 7
Slc13a3	114644	7.19E-04	4.20E-10	1.09E-04	solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 3
Prnp	19122	7.61E-04	1.47E-04	2.01E-32	prion protein
Ndufv2	72900	7.66E-04	8.97E-06	7.28E-06	NADH dehydrogenase (ubiquinone) flavoprotein 2
Gatm	67092	7.77E-04	4.49E-30	3.40E-05	glycine amidinotransferase (L-arginine:glycine amidinotransferase)
1300001I01Rik	74148	8.33E-04	5.38E-08	1.87E-10	RIKEN cDNA 1300001I01 gene
Prickle1	106042	9.06E-04	4.31E-31	3.45E-46	prickle homolog 1 (Drosophila)
Atp5b	11947	9.29E-04	8.17E-03	8.25E-03	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit
Cds2	110911	9.46E-04	7.57E-07	5.38E-12	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 2
A630007B06Rik	213993	9.48E-04	1.43E-06	1.34E-05	RIKEN cDNA A630007B06 gene
Ftsjd1	234728	9.56E-04	3.70E-10	5.96E-07	FtsJ methyltransferase domain containing 1
Gstk1	76263	9.60E-04	3.53E-06	1.40E-06	glutathione S-transferase kappa 1
Tjp1	21872	9.75E-04	1.39E-23	9.60E-18	tight junction protein 1
Bcam	57278	1.00E-03	1.18E-57	3.78E-39	basal cell adhesion molecule
Txnrd3	232223	1.01E-03	1.12E-03	2.17E-03	thioredoxin reductase 3
Alox5	11689	1.07E-03	1.22E-15	1.58E-08	arachidonate 5-lipoxygenase
Ttc25	74407	1.11E-03	6.29E-40	2.66E-38	tetratricopeptide repeat domain 25
Spint2	20733	1.13E-03	3.43E-79	6.07E-61	serine protease inhibitor, Kunitz type 2
Maml2	270118	1.15E-03	9.13E-05	5.53E-15	mastermind like 2 (Drosophila)

Ppat	231327	1.21E-03	4.48E-11	1.18E-06	phosphoribosyl pyrophosphate amidotransferase
Dcaf6	74106	1.27E-03	7.93E-05	4.67E-05	DDB1 and CUL4 associated factor 6
Ak3	56248	1.30E-03	9.20E-08	1.10E-12	adenylate kinase 3
Uaca	72565	1.38E-03	5.37E-22	1.27E-16	uveal autoantigen with coiled-coil domains and ankyrin repeats
Mdh1	17449	1.46E-03	1.61E-14	1.36E-12	malate dehydrogenase 1, NAD (soluble)
Tmem119	231633	1.50E-03	4.15E-17	3.58E-18	transmembrane protein 119
Gnai1	14677	1.56E-03	3.24E-23	6.66E-33	guanine nucleotide binding protein (G protein), alpha inhibiting 1
Hebp2	56016	1.58E-03	6.38E-12	2.88E-20	heme binding protein 2
0610009O20Rik	66839	1.61E-03	9.75E-06	1.05E-03	RIKEN cDNA 0610009O20 gene
Clec4a1	269799	1.63E-03	1.49E-14	5.13E-04	C-type lectin domain family 4, member a1
Il27ra	50931	1.64E-03	2.77E-35	6.08E-12	interleukin 27 receptor, alpha
Aldh1a2	19378	1.81E-03	2.88E-18	1.14E-18	aldehyde dehydrogenase family 1, subfamily A2
Btk	12229	2.11E-03	3.60E-05	2.46E-03	Bruton agammaglobulinemia tyrosine kinase
St3gal4	20443	2.12E-03	1.14E-27	5.20E-16	ST3 beta-galactoside alpha-2,3-sialyltransferase 4
Acpp	56318	2.15E-03	9.59E-44	1.93E-26	acid phosphatase, prostate
Qdpr	110391	2.17E-03	1.30E-03	5.62E-09	quinoid dihydropteridine reductase
Cisd1	52637	2.18E-03	3.35E-09	1.68E-08	CDGSH iron sulfur domain 1
Gng4	14706	2.30E-03	6.90E-03	2.56E-05	guanine nucleotide binding protein (G protein), gamma 4
Mttp	17777	2.38E-03	2.93E-04	6.00E-07	microsomal triglyceride transfer protein

1110002N22Rik	68550	2.38E-03	5.90E-02	2.68E-02	RIKEN cDNA 1110002N22 gene
Add3	27360	2.49E-03	3.98E-05	4.58E-03	adducin 3 (gamma)
Ptcd2	68927	2.75E-03	5.02E-04	1.13E-02	pentatricopeptide repeat domain 2
Pdk2	18604	2.89E-03	1.28E-05	1.00E-06	pyruvate dehydrogenase kinase, isoenzyme 2
Adcyap1r1	11517	2.96E-03	4.16E-16	1.51E-19	adenylate cyclase activating polypeptide 1 receptor 1
Wnt2	22413	3.07E-03	1.77E-90	9.57E-79	wingless-related MMTV integration site 2

Table 4-7. Core up-regulated NeMΦ genes defined by OrderedList analysis. Genes ranked by correlation with %YM1 positivity in the microarray analysis were compared against gene lists derived by ranking both the WT-NeMΦ vs. WT-ThioMΦ and WT-NeMΦ vs. IL-4Rα$^{-/-}$-NeMΦ comparisons from the RNA-Seq analyses. The intersection of genes defined as common between both comparisons are the 'core' NeMΦ genes included in this table. P-values shown are corrected for multiple testing (Benjamini-Hochberg method).

Symbol	Entrez ID	Lito vs YM1	WT Brugia vs WT Thio	WT Brugia vs -/- Brugia	Description
Lcp1	18826	8.32E-15	6.55E-182	6.44E-157	lymphocyte cytosolic protein 1
Emb	13723	8.95E-11	1.72E-64	2.30E-109	embigin
Coro1a	12721	1.04E-10	1.18E-187	1.92E-189	coronin, actin binding protein 1A
Ebi3	50498	1.70E-10	2.12E-11	5.39E-22	Epstein-Barr virus induced gene 3
Ngfrap1	12070	2.42E-10	1.92E-66	2.23E-08	nerve growth factor receptor (TNFRSF16) associated protein 1
Fkbp1a	14225	2.42E-10	1.62E-53	8.50E-21	FK506 binding protein 1a
P2ry6	233571	3.40E-10	1.27E-132	2.71E-42	pyrimidinergic receptor P2Y, G-protein coupled, 6
Lst1	16988	1.47E-09	4.00E-43	4.94E-71	leukocyte specific transcript 1
Rassf2	215653	3.55E-09	5.19E-42	1.05E-16	Ras association (RalGDS/AF-6) domain family member 2
Lsr	54135	3.55E-09	3.80E-68	8.68E-34	lipolysis stimulated lipoprotein receptor
Trim3	55992	5.99E-09	5.04E-21	2.61E-09	tripartite motif-containing 3
Trpv2	22368	1.70E-08	4.43E-97	2.85E-28	transient receptor potential cation channel, subfamily V, member 2
Glipr1	73690	1.70E-08	4.64E-55	1.03E-05	GLI pathogenesis-related 1 (glioma)
Ptms	69202	2.68E-08	1.03E-177	6.60E-65	parathymosin
Ogfrl1	70155	1.35E-07	5.03E-19	8.42E-14	opioid growth factor receptor-like 1
Irf8	15900	1.92E-07	2.19E-172	9.30E-166	interferon regulatory factor 8
Adrb2	11555	2.17E-07	6.63E-37	2.54E-21	adrenergic receptor, beta 2
Anxa5	11747	2.40E-07	2.90E-59	2.17E-04	annexin A5
Smox	228608	2.75E-07	3.31E-24	1.48E-41	spermine oxidase
Arid3a	13496	5.18E-07	3.23E-70	1.99E-48	AT rich interactive domain 3A (BRIGHT-

					like)
Tle1	21885	7.58E-07	4.49E-18	5.63E-10	transducin-like enhancer of split 1, homolog of Drosophila E(spl)
Fcgr1	14129	1.17E-06	3.28E-229	0.00E+00	Fc receptor, IgG, high affinity I
Gng2	14702	1.40E-06	1.81E-83	1.32E-11	guanine nucleotide binding protein (G protein), gamma 2
Endod1	71946	1.46E-06	9.53E-24	3.67E-04	endonuclease domain containing 1
Slc11a1	18173	2.77E-06	1.32E-15	1.80E-12	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 1
Slc16a10	72472	4.22E-06	1.83E-36	1.69E-18	solute carrier family 16 (monocarboxylic acid transporters), member 10
Ccrl2	54199	4.47E-06	1.00E-38	5.36E-66	chemokine (C-C motif) receptor-like 2
Mpeg1	17476	5.35E-06	2.38E-56	6.01E-26	macrophage expressed gene 1
Map4k4	26921	6.32E-06	1.93E-33	1.89E-19	mitogen-activated protein kinase kinase kinase kinase 4
Cd300a	217303	6.56E-06	1.20E-22	4.16E-04	CD300A antigen
Cyp27a1	104086	6.60E-06	3.98E-32	1.10E-03	cytochrome P450, family 27, subfamily a, polypeptide 1
Znrf1	170737	8.39E-06	3.68E-41	5.85E-08	zinc and ring finger 1
Lgals1	16852	8.48E-06	9.87E-111	4.32E-35	lectin, galactose binding, soluble 1
Sgms2	74442	1.02E-05	2.17E-66	8.32E-35	sphingomyelin synthase 2
Hopx	74318	1.23E-05	8.89E-32	5.13E-16	HOP homeobox
Lpl	16956	1.33E-05	5.29E-135	1.15E-37	lipoprotein lipase
Agpat3	28169	1.34E-05	3.19E-13	1.01E-04	1-acylglycerol-3-phosphate O-acyltransferase 3
Nuak1	77976	1.39E-05	1.15E-10	5.31E-27	NUAK family, SNF1-like kinase, 1
Tsc22d3	14605	1.68E-05	1.00E-16	2.78E-08	TSC22 domain family,

					member 3
S100a4	20198	2.07E-05	4.13E-43	3.91E-19	S100 calcium binding protein A4
Pid1	98496	2.21E-05	8.21E-31	2.83E-28	phosphotyrosine interaction domain containing 1
Cryba4	12959	2.42E-05	1.19E-05	2.70E-03	crystallin, beta A4
Tnfrsf13b	57916	2.83E-05	3.83E-24	5.71E-14	tumor necrosis factor receptor superfamily, member 13b
Lamc1	226519	2.99E-05	1.31E-33	5.67E-10	laminin, gamma 1
Nrp1	18186	3.66E-05	4.59E-44	1.77E-05	neuropilin 1
Dgkg	110197	4.05E-05	1.43E-09	8.41E-47	diacylglycerol kinase, gamma
Dcakd	68087	4.05E-05	5.96E-11	7.08E-06	dephospho-CoA kinase domain containing
Pilra	231805	4.06E-05	1.43E-35	1.10E-187	paired immunoglobin-like type 2 receptor alpha
Apobec1	11810	4.21E-05	1.02E-42	2.04E-09	apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1
Rap2a	76108	4.55E-05	3.15E-33	2.22E-18	RAS related protein 2a
Ccl3	20302	5.91E-05	7.67E-133	1.48E-68	chemokine (C-C motif) ligand 3
Evl	14026	6.38E-05	1.37E-80	9.74E-54	Ena-vasodilator stimulated phosphoprotein
Dtx3	80904	6.92E-05	3.63E-49	1.60E-25	deltex 3 homolog (Drosophila)
Mef2c	17260	6.97E-05	1.29E-33	5.89E-09	myocyte enhancer factor 2C
Nt5dc2	70021	7.43E-05	1.67E-49	1.54E-09	5'-nucleotidase domain containing 2

Table 4-8. Core down-regulated NeMΦ genes defined by OrderedList analysis. Genes ranked by reverse correlation with %YM1 positivity in the microarray analysis were compared against gene lists derived by reverse ranking both the WT-NeMΦ vs. WT-ThioMΦ and WT-NeMΦ vs. IL-4Rα$^{-/-}$-NeMΦ comparisons from the RNA-Seq analyses. The intersection of genes defined as common between both comparisons are the 'core' NeMΦ genes included in this table. P-values shown are corrected for multiple testing (Benjamini-Hochberg method).

4.3.6 KINETIC EXPRESSION PROFILES OF PPAR-PROSTACYCLIN AXIS GENES

A drawback of our RNA-Seq analysis is that we cannot state whether certain genes are required in a transient manner during alternative activation. We were particularly interested in *Pparg*, which as been described as required for AAMΦ mitochondrial metabolism (Odegaard et al., 2007). However, our RNA-Seq analysis did not support a role for *Pparg* in the AAMΦ response to *B. malayi* implant 21 days P.I. However, it is still possible that *Pparg* is required early during alternative activation and down-regulated later on. Our time course data allowed us to assess gene expression changes over time, and argued against an early up-regulation of *Pparg* (Figure 4-12). In chapter 2 we propose that a PGI2-PPARδ axis maintains PPAR-dependent gene expression in AAMΦ in the absence of PPARγ expression. In the microarray time course the expression profiles of genes involved in PGI2 synthesis and PPAR activity were entirely consistent with this hypothesis. *Pparg* was expressed at a much lower level than *Ppard throughout the time course*, and both constituents for PGI2 synthesis *Ptgs1* and *Ptgis* are highly and constitutively expressed (Figure 4-12). As expected, the expression profile of the PPAR co-activator protein *Ppargc1b* significantly correlated with YM1 levels. Thus these data support our previous interpretation that *Ppard* compensates for low *Pparg* expression in resident tissue AAMΦ by enhanced production of PGI2 (Figure 4-12).

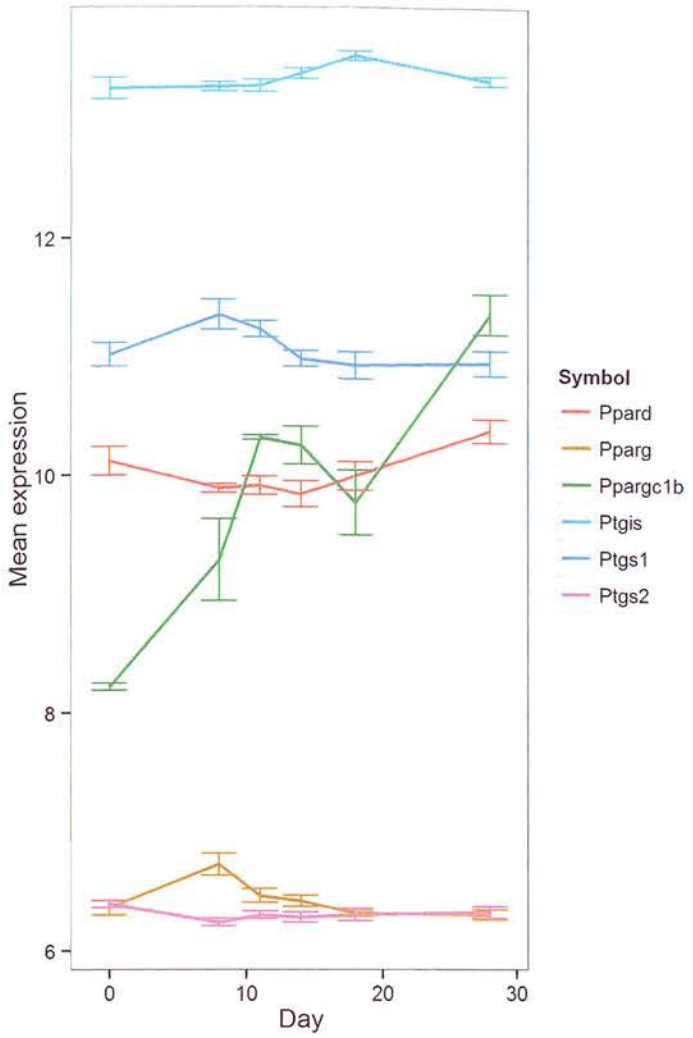


Figure 4-12. Expression profiles of PPAR transcription factors, the PPAR co-activator protein *Ppargc1b*, and genes associated with PGI2 synthesis. The gene expression profile in *L. sigmodontis*-induced pleural cavity AAMΦ is consistent with the peritoneal *B. malayi*-induced AAMΦ profile presented in Chapter 2 for these genes, and is consistent with a prostacyclin – PPARδ transcriptional axis maintaining PPAR-dependent gene expression in AAMΦ derived from both infection models. Day 0 and day 4 expression values were pooled, and considered as day 0 to generate error bars. Error bars represent standard error (SEM) values.

4.3.7 *L. SIGMODONTIS* INFECTION LEADS TO AN UP-REGULATION OF MITOCHONDRIAL METABOLIC PATHWAYS

GSEA was performed on YM1-asosociated genes to identify molecular pathways that correlate with YM1 expression. This analysis recapitulated the findings of our RNA-Seq study, showing marked up-regulation of molecular pathways associated with cellular metabolism. In particular we observed a gene expression profile consistent with increased mitochondrial metabolism, including the TCA cycle (Table 4-9, Figure 4-13a).

Curiously our *cis*-enrichment analysis did not identify over-represented PPAR motifs in the promoters of YM1-associated genes. This is in contradiction to both the RNA-Seq analysis presented in Chapter 2 and the GSEA results here. The expression profile of TCA cycle-associated genes shows a marked positive correlation with %YM1 levels (Figure 4-13b). Furthermore, an analysis of PPAR motif presence in promoters showed that 70% (19/27) of TCA cycle genes contained PPAR response elements, relative to a background frequency of 34.7% as determined by Homer. This supports the notion that PPAR transcription factors are required for AAMΦ mitochondrial metabolism. To assess whether the observed gene expression profiles correlate with changes in mitochondrial mass we used the mitochondrial dye MitoTracker Green. This fluorescent dye accumulates within the mitochondria of live cells, and provides a direct measure of mitochondrial mass (Pendergrass et al., 2004). Confirming our GSEA results we observed a marked increase in MitoTracker green intensities in pleural cavity macrophages at 28 days post *L. sigmodontis* infection relative to naïve tissue resident macrophages (Figure 4-13c).

NAME	SIZE	Enrichment Score	NES	Nominal P-val	FDR q-val
KEGG CITRATE CYCLE TCA CYCLE	25	0.801	1.658	0.010	0.089
KEGG VALINE LEUCINE AND ISOLEUCINE DEGRADATION	37	0.748	1.927	0.000	0.047
KEGG PYRUVATE METABOLISM	28	0.719	1.855	0.000	0.056
KEGG PROPANOATE METABOLISM	23	0.711	1.743	0.012	0.068
KEGG OXIDATIVE PHOSPHORYLATION	93	0.701	1.698	0.012	0.078
KEGG PENTOSE PHOSPHATE PATHWAY	19	0.668	1.687	0.000	0.078
KEGG PARKINSONS DISEASE	88	0.667	1.718	0.023	0.079
KEGG BUTANOATE METABOLISM	23	0.659	1.755	0.000	0.085
KEGG FATTY ACID METABOLISM	31	0.656	1.786	0.004	0.096
KEGG GALACTOSE METABOLISM	20	0.654	1.646	0.016	0.094
KEGG GLYCOLYSIS GLUCONEOGENESIS	39	0.620	1.716	0.010	0.072
KEGG FRUCTOSE AND MANNOSE METABOLISM	26	0.611	1.643	0.014	0.090
KEGG HUNTINGTONS DISEASE	132	0.599	1.785	0.002	0.072
KEGG ALZHEIMERS DISEASE	128	0.551	1.754	0.008	0.073
KEGG VIBRIO CHOLERAE INFECTION	42	0.503	1.565	0.022	0.154
KEGG TRYPTOPHAN METABOLISM	23	0.501	1.612	0.016	0.112
KEGG AMINO SUGAR AND NUCLEOTIDE SUGAR METABOLISM	36	0.491	1.492	0.047	0.225
KEGG SNARE INTERACTIONS IN VESICULAR TRANSPORT	29	0.489	1.556	0.024	0.156
KEGG PEROXISOME	67	0.487	1.660	0.015	0.096

Table 4-9. Gene set enrichment analysis results of KEGG pathways significantly positively correlated with %YM1 expression as measured by FACS showing metabolic pathways associated with alternative activation (nominal p-value < 0.05, FDR < 0.25). ES = enrichment score, NES = normalized enrichment score, FDR q-val= Benjamini-Hochberg corrected p-value.

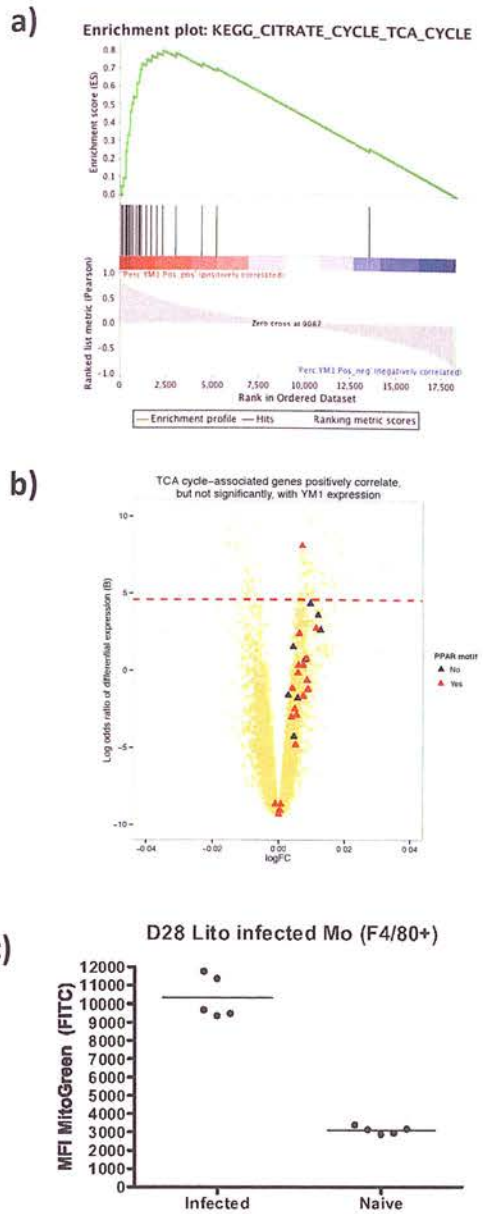


Figure 4-13. A) GSEA plot showing the positions of KEGG TCA-cycle associated genes in a list of genes ranked according to Pearson's correlation between gene expression and %YM1 positivity as measured by FACS.

B) Volcano plot showing correlation of gene expression against %YM1 positivity (y-axis truncated at 10 for clarity). TCA cycle genes are shown as triangles. TCA cycle genes with predicted PPAR motifs in the proximal promoter sequence are coloured in red, those without are shown in black. The dotted red line depicts the significance threshold B=4.6.

C) MitoTracker green staining showing mitochondrial dye uptake in F4/80 positive pleural cavity macrophages from *L. sigmodontis* infected mice 28 days post infection.

4.3.8 Both proliferating and non-proliferating AAMΦ are associated with increased mitochondrial metabolism.

IL-4 induces both alternative activation and macrophage proliferation *in vivo*. Proliferating cells typically utilize aerobic glycolysis, as evidenced by both proliferating cancer cells (Vander Heiden et al., 2009) and T-cells undergoing clonal expansion (Fox et al., 2005). Thus it is curious that IL-4 enhances both mitochondrial metabolism associated with AAMΦ as well as proliferation. We set out to address whether macrophages induced to proliferate in response to IL-4 also demonstrate enhanced mitochondrial respiration. The MitoTracker orange CMH2xROS dye is redox-sensitive and provides a measure of mitochondrial membrane potential and, hence, mitochondrial respiration (Arita et al., 2006). It is selectively taken up by mitochondria and oxidised to generate the fluorescent molecule that is measured. In contrast to the MitoTracker Green dye this small molecule dye can be fixed, and is compatible with intracellular staining and flow cytometry. We therefore optimized a method to allow concurrent visualization of mitochondrial abundance, cellular proliferation and AAMΦ.

We applied this staining protocol to pleural cavity macrophages elicited during a time series of IL-4c treatment. Consistent with previous studies within the lab YM1 was induced by 12h (Figure 4-14a), however we did not observe robust proliferation until 24h (Figure 4-14b). Furthermore IL-4c treated macrophages were competent to express AAMΦ markers whilst up-regulating the marker of proliferation Ki67 (Figure 4-14c). IL-4c treatment also increased mitochondrial abundance associated with alternative activation (Figure 4-14d,e). $Ki67^{hi}$ proliferating macrophages possessed an increased mitochondrial mass relative to naive and PBS treated controls (Figure 4-14f). These data do not preclude the possibility that IL-4-dependent macrophage proliferation is facilitated by aerobic glycolysis, they do however

tell us that the mitochondrial phenotype associated with AAMΦ is not mutually exclusive with IL-4-dependent proliferation.

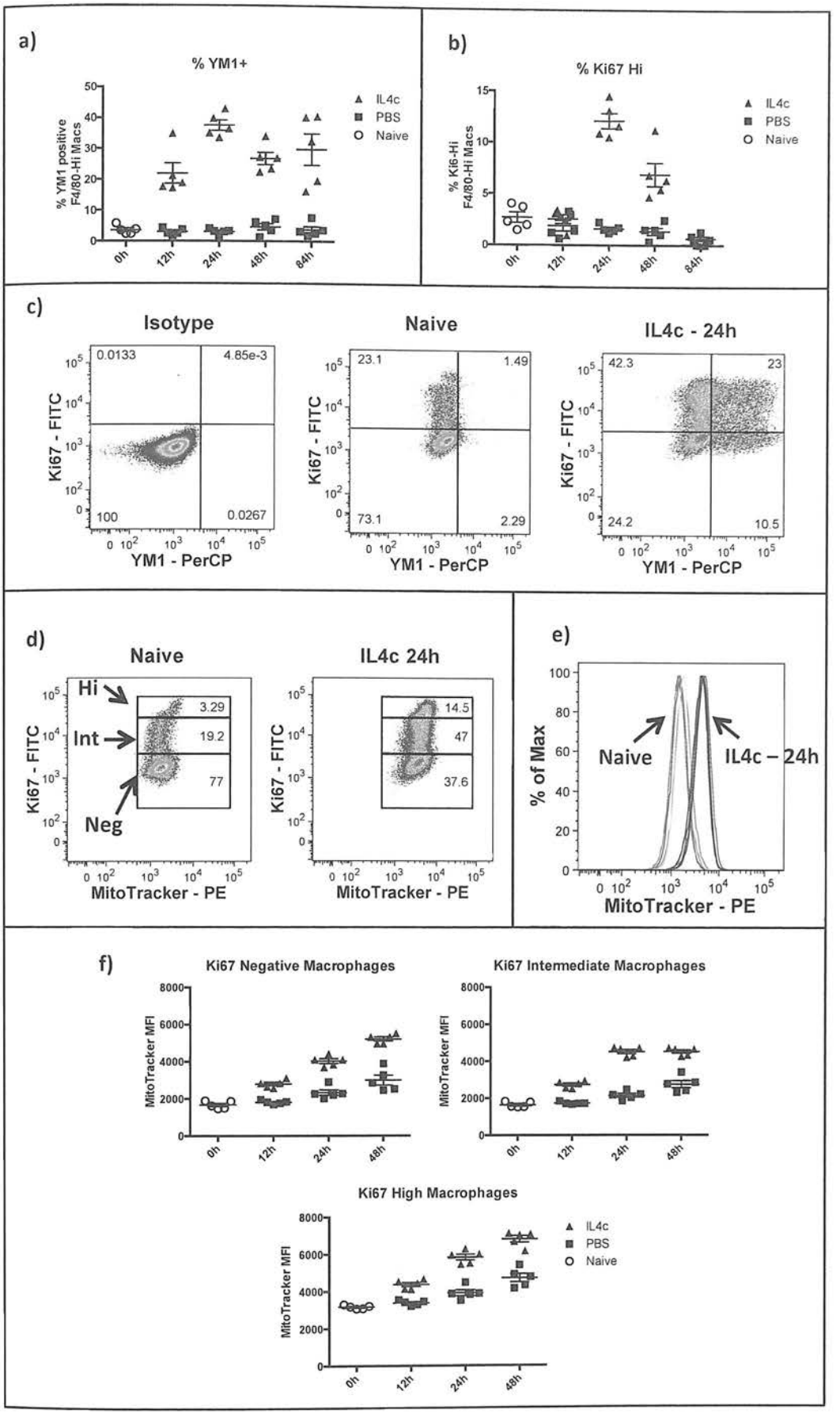

a)
% YM1+
% YM1 positive F4/80-Hi Macs
IL4c
PBS
Naive
0h
12h
24h
48h
84h
b)
% Ki67 Hi
% Ki6-Hi F4/80-Hi Macs
c)
Isotype
Naive
IL4c - 24h
Ki67 - FITC
YM1 - PerCP
0.0133
4.85e-3
100
0.0267
23.1
1.49
73.1
2.29
42.3
23
24.2
10.5
d)
Naive
IL4c 24h
Hi
Int
Neg
3.29
19.2
77
14.5
47
37.6
MitoTracker - PE
e)
% of Max
Naive
IL4c – 24h
f)
Ki67 Negative Macrophages
Ki67 Intermediate Macrophages
Ki67 High Macrophages
MitoTracker MFI

Figure 4-14. Pleural cavity macrophage alternative activation, proliferation and mitochondrial mass in a time-course of IL-4 complex (IL-4c) treatment.
A) Frequency of $YM1^{+}$ $F4/80^{hi}$ pleural cavity MΦ, b) frequency of proliferating, $Ki67^{hi}$ $F4/80^{hi}$ macrophages, c) YM1 and Ki67 staining in naïve and IL-4c-treated macrophages 24h after stimulation showing concurrent proliferation and alterative activation. D) Representative gating for $Ki67^{hi}$, $Ki67^{int}$ and $Ki67^{low}$ staining against MitoTracker CM-H2Xros intensity in naïve and IL-4c-treated macrophages, e) Histogram showing MitoTracker CM-H2Xros intensities in naïve and IL-4c-treated macrophages 24 hours after IL-4c administration. F) Breakdown of MitoTracker CM-H2Xros intensities in proliferating and non-proliferating macrophages in response to IL-4c treatment.

4.4 Discussion

In this chapter we present the first transcriptomic time-course analysis of the macrophage response to a nematode infection using the filarial parasite *L. sigmodontis*. It has previously been shown by our lab that IL-4 stimulates macrophages both to alternatively activate and proliferate *in vivo* (Jenkins et al., 2011), however treatment of macrophages with IL-4 *in vitro* does not drive this proliferative response. We sought to investigate this phenomenon to generate testable hypotheses explaining this process. Further, this experiment has allowed us to validate findings inferred from our RNA-Seq analysis in chapter 2, and assess whether differences in AAMΦ responses can be observed between early and late time points.

We had noted previously that the kinetics of proliferation and alternative activation are discordant in the response to filarial nematode infection, and this observation was confirmed in this experiment (Figure 4-1). There is initially a proliferative phase, which we had found to be independent of IL-4, followed by an IL-4-dependent proliferative burst (Jenkins et al., 2011). The peak of IL-4-dependent proliferation coincides with robust induction of AAMΦ at day 11 (Figure 4-1, S Jenkins unpublished). In our time course AAMΦ persisted from day 11 however proliferation dropped after this point (Figure 4-1). We capitalized on these differential kinetics by array profiling FACS-purified F4/80^{hi} macrophages, and combining this with a single cell phenotypic analysis. By correlating global gene expression profiles with YM1 and Ki67^{hi} levels determined by FACS we have defined transcriptional repertoires associated with alternative activation and proliferation respectively. To our knowledge this is the first time microarray and FACS has been used to deconvolute transcriptional signatures from heterogeneous populations.

4.4.1 ARE ALL F4/80^{HI} MACROPHAGES EQUAL?

Thirteen genes were significantly correlated with both YM1 and Ki67^{hi} levels. The majority, nine, of these genes were more highly expressed in proliferating than in naïve cells, and were down-regulated later in the time course. This list contained genes that were involved in chromosomal remodelling and proliferation. Because these genes positively correlated with Ki67^{hi} and negatively correlated with YM1, this may suggest macrophages do not have an indefinite capacity to proliferate in response to IL-4. Indeed at day 18 P.I. we observed a crash in proliferation, with levels dropping well below baseline (Figure 4-1). This may reflect exhaustion driven by the depletion of factors required for proliferation. Yet because alternative activation continues to increase at day 18 P.I. this factor may not be IL-4. Alternatively, any limited capacity for macrophages to proliferate in response to IL-4 may be cell intrinsic. Either of these possibilities could explain how AAMΦ are sustained without continued proliferation during the later phases of *L. sigmodontis* infection.

It is important to consider that our analysis has profiled a population of cells, and that due to the stochastic nature of gene expression, functional heterogeneity always exists within a 'pure' populations of cells (Raj and van Oudenaarden, 2008). This property, by definition, restricts the potential of a population of cells to perform a given task (in response to a stimulus) to a sub-population at any given moment in time. This will include the ability of a cell with mitogenic properties to retain self-renewal capacity or differentiate (Yamanaka, 2009). We have observed the differential regulation of chromatin-remodelling genes associated with YM1 and Ki67^{hi} levels. It is possible that this differential expression is driven by a cell-intrinsic limitation in the ability of tissue-resident macrophages to proliferate *in situ*. If so, then the frequency of macrophages competent to proliferate will change during the *L. sigmodontis* infection time course and may explain the crash in proliferation that we observe at day 18 P.I. It follows that distinct 'early' and 'late' populations of F4/80^{hi}

macrophages that differ in their mitogenic capacity may exist within the pleural cavity. Cell-surface markers that correlate, positively or negatively, with the expression profile of these chromatin remodelling genes may serve as effective markers for these populations and would provide a starting point to explore this hypothesis. It is worth noting that this hypothesis refers to the 'life history' traits of macrophages, and can be independent of alternative activation *per se*.

Included in the list of genes that positively correlated with Ki67^{hi}, and negatively correlated with YM1, was the cell surface marker *Sell* (L-selectin, Rosen, 2004) and migration-associated marker *Coroa1* (Gatfield et al., 2005). Following on from the hypothesis presented above, L-selectin may provide a surface marker for 'early' tissue MΦ. To define *Sell* positive macrophages as 'early' MΦ, which are competent to proliferate, it would be necessary to see if *Sell* positive macrophages incorporated BrdU in advance of other F4/80^{hi} tissue macrophages in response to IL-4c treatment. BrdU is a tractable DNA analogue that is used to ascertain whether cells have proliferated during administration (Yu et al., 1992). A BrdU pulse delivered over the initial phase of macrophage proliferation in response to *L. sigmodontis* infection would label all proliferating macrophages, and I predict this would include a high proportion of all *Sell* positive macrophages. The down-regulation of *Sell* and *Coro1a* later in infection supports our previous interpretation that AAMΦ have decreased migratory capacity (Thomas et al., 2012). Two genes, *Cd209b* and *Ccbl1*, negatively correlated with Ki67^{hi} and positively correlated with YM1 expression as determined by FACS. The cell surface marker *Cd209b* (DC-SIGN) has previously been defined as a marker for mature dendritic cells (Cheong et al., 2010) and may also identify 'late' tissue MΦ, i.e. a final tissue resident MΦ population to develop that is refractory to *in situ* proliferation. *Cd209b* could thus be a useful marker to determine whether IL-4-dependent proliferation is

associated with terminal differentiation. Using the same BrdU experiment suggested above, an early pulse of BrdU during infection would label the first round of proliferating macrophages induced during *L. sigmodontis* infection. My prediction is that this initial labelling would not include many *Cd209b* positive macrophages. Cessation of the BrdU pulse, followed by assessment of BrdU positive macrophages later during infection, would enable the characterization of 'late' tissue MΦ. As stated above, my hypothesis would be that these 'late' tissue MΦ are *Cd209b* positive and are no longer responsive to IL-4-induced proliferation.

4.4.2 No evidence for an early induction of PPARγ or dymnamic modulation of IL-4Rα-dependent genes

We were specifically interested in the PPARγ gene expression profile as this has been shown to be required for AAMΦ and the associated metabolic phenotype *in vitro*. We therefore performed a direct assessment of the expression profiles of genes associated with our previous hypothesis, that a PPARδ-PGI2 axis maintains alternative activation in the presence of low PPARγ expression (Chapter 2). The results of this time course supported this hypothesis showing that there was no early spike in PPARγ expression. A more global analysis of temporal gene expression profiles after accounting for YM1 and Ki67^{hi} expression determined by FACS was also performed. Hierarchical clustering of studentized residual errors after fitting the linear model did not reveal any additional temporal gene expression features in the dataset. This argues against the presence of 'early' and 'late' waves of alternative activation with unique IL-4Rα-dependent gene expression profiles. Specifically, there was no obvious cluster of genes with outlying residual errors during the time course that would identify an additional gene signature that was not already captured by the linear model. However, a more focussed analysis of temporal differential expression

profiles has not yet been performed. An alternative, perhaps more sensitive, approach to identifying temporal changes in gene expression would be to apply correlation network analysis, such as implemented in BioLayout Express 3D (Theocharidis et al., 2009). This approach is capable of identifying small numbers of genes with shared expression profiles, and could unveil more subtle temporal gene expression profiles. Such analysis could be applied to either normalized expression data or to residual errors, and is an outstanding objective for this microarray data analysis project.

4.4.3 A PUTATIVE MECHANISM FOR IL-4-DEPENDENT MACROPHAGE PROLIFERATION

Several lines of evidence support the idea that IL-4-dependent macrophage proliferation and alternative activation are transcriptionally independent.

- The kinetics of the two processes are discreet.
- The two major axes of variation in the dataset, as defined by PCA, are explained by alternative activation (YM1) and proliferation ($Ki67^{hi}$) respectively. Furthermore, these do not correlate. GSEA confirmed that $Ki67^{hi}$-associated gene expression profiles represent a cell cycle signature.
- Computational analysis of over-represented *cis*-regulatory motifs in the promoters of genes correlated with YM1 and $Ki67^{hi}$ identified STAT6 motifs in the promoters of YM1-associated, but not $Ki67^{hi}$-associated promoters.

Cis-regulatory analysis of cell cycle-associated genes showed that macrophage proliferation was associated with E2F, NFY and MYB motifs (Table 4-5). This is consistent with a previous analysis of cell cycle regulatory motif activity in the promoters of monocyte-like THP1 cells (Suzuki et al., 2009). We have unveiled a possible link between IL-4-dependent proliferation

and TFBS for E2F family transcription factors. The dynamics of E2F transcription factor gene expression indicate that IL-4-dependent and IL-4-independent phases of proliferation may be governed by different E2F family members. We attempted to confirm this observation using FACS-purified macrophages induced to proliferate in response to M-CSF or IL-4 *in vivo*. Our objective was to correlate M-CSF and IL-4-dependent proliferative profiles with differing E2F gene expression profiles, however RT-PCR analysis of these samples did not support this initial finding. It is still possible that different E2F family members drive M-CSF and IL-4-dependent proliferation and if this were the case it would show the two processes are mechanistically independent. At this stage all we can state is that E2F gene expression profiles correlate with a proliferative signature that is also associated with over-represented E2F TFBS. Therefore these factors probably mediate IL-4-dependent macrophage proliferation.

How can it be that IL-4 drives proliferation *in vivo* but not *in vitro*? In addition to the evidence presented above, data from our lab using whole-body lethally irradiated C57BL/6 mice engrafted with *Stat6*$^{-/-}$ bone marrow has demonstrated a requirement for STAT6 in IL-4-dependent macrophage proliferation (S Jenkins, unpublished). Furthermore, IL-4c treated mice up-regulate the expression of AAMΦ markers within a few hours but macrophages are not induced to proliferate after 1 day (Figure 4-14). These data support a model wherein IL-4-dependent proliferation is downstream of STAT6. The most parsimonious explanation for this is one wherein IL-4/STAT6 sensitizes macrophages to proliferate *in vivo* via the induction of a second factor (factor *X*, Figure 4-15). The *in vivo*, but not *in vitro*, environment would provide constitutive access to a co-factor for factor *X* that governs the commitment to proliferate. This would explain how in the presence of IL-4 macrophages proliferate in all bodily sites at which we have so far looked (Jenkins et al., 2011).

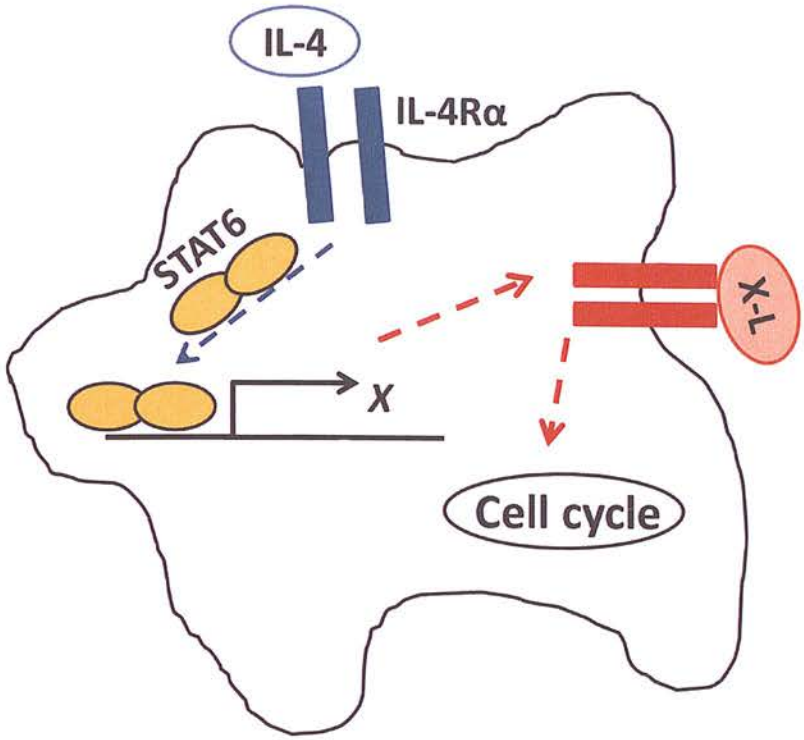


Figure 4-15. Proposed model for IL-4-dependent macrophage proliferation. STAT6 is activated by IL-4Rα-dependent signalling and drives the expression of a second gene *X*, postulated to be a receptor. In turn factor *X* permits the macrophage to divide in response to a second signal provided by the external environment.

4.4.4 Cellular mitochondrial metabolism in alternative activation

Finally, we have shown that *in vivo* AAMΦ induced by IL-4c injection, or *L. sigmodontis* infection, increase mitochondrial mass as measured by MitoTracker fluorescence intensity. This was expected based upon the GSEA in our time course associating the TCA cycle with YM1 expression. This is also consistent with the pathway enrichment and analysis of *cis*-regulatory elements presented in our RNA-Seq analysis. Surprisingly however, the *cis*-enrichment analysis presented alongside this time course failed to define a PPAR signature associated with YM1 expression. Szanto *et al* have reported that macrophage PPARγ-dependent gene expression is augmented by STAT6 in response to IL-4 (Szanto et al., 2010). However, in the same paper, the transcription of AAMΦ STAT6-dependent target genes was relatively unaffected by the absence of PPARγ (Szanto et al., 2010). This latter finding directly contradicts the findings of the Chawla laboratory who show that the same AAMΦ target genes are impaired in response to IL-4 in macrophage-specific *Pparg*$^{-/-}$ mice (Odegaard et al., 2007). Regardless, one of the main differences between the RNA-Seq and microarray analyses we have presented is the number of DE genes we were able to detect, presumably due to the enhanced sensitivity and dynamic range provided by RNA-Seq. We have shown that TCA-cycle associated genes are induced in AAMΦ, but many were not statistically significantly correlated with YM1 expression in our microarray (Figure 4-13). These genes were also shown to have over-represented PPAR motif presence within promoter regions. Therefore, one possible explanation for these findings is that STAT6 does augment PPAR-dependent gene expression (Szanto et al., 2010), and these genes include key AAMΦ-associated metabolic pathways. However, not all STAT6-dependent changes in gene expression are PPARγ-dependent (indeed it is hard to imagine how this could be the case). At the transcriptional level STAT6 may alter PPAR-dependent gene expression subtly, and these changes may not readily be detected by microarray. However, RNA-Seq may

provide enough sensitivity to identify these genes as significantly DE. This interpretation reinforces our understanding that the most inducible changes in IL-4-dependent transcription are driven by STAT6. It is clear that alterations in macrophage cellular metabolism induced by PPARs and STAT6 can have profound effects on systemic metabolism, and perhaps wider aspects of AAMΦ physiology. The extent to which transcriptional changes in AAMΦ are mediated directly by STAT6, PPARs, or are downstream of either of these processes remains to be elucidated. ChIP-Seq methods would be needed to directly correlate gene expression profiles with TF-DNA occupancy to address this. However, the induction of the TCA cycle in $CD8^+$ T-cells has profound effects on cellular life history characteristics (Pearce et al., 2009). It is thus possible that alterations in mitochondrial metabolism may exert their effects on AAMΦ less so by modulating macrophage gene expression, but rather population structure *in vivo*.

We have shown that proliferating AAMΦ also have increased mitochondrial mass. This does not mean that IL-4-dependent proliferation does not utilize glycolysis and, in this context, it may be of interest to note that the insulin receptor, responsible for the control of glucose uptake, was amongst the list of core-NeMΦ genes we identified. The data we have presented do however seem to contradict the generally accepted notion that cellular proliferation 'prefers' aerobic glycolysis to oxidative metabolism as a means to maximise carbon availability, and to minimise DNA damage caused by mitochondrial reactive oxygen species (Vander Heiden et al., 2009). An alternative mechanism for glycolysis-independent, glutaminolysis-dependent proliferation has recently been reported in B-cell like P493 cells (Le et al., 2012). Importantly, this process occurred in glucose-deprived conditions and depended upon the mitochondrial TCA cycle (Le et al., 2012). There is one report of low level proliferation of primary macrophages induced by IL-4 *in vitro*, which was reported to

have occurred under low glucose conditions (Odegaard et al., 2008). It is tempting to speculate that IL-4-dependent macrophage proliferation relies upon a similar glutaminolysis-dependent mechanism. However because all proliferating cells require glutamine for the production of intermediate metabolites the hypothesis would be difficult to test (Matés et al., 2002, 2009). An *in vitro* system will be required to directly study the metabolic requirements of IL-4-dependent macrophage proliferation. This is also necessary to further our understanding of the cellular role of mitochondrial metabolism in alternative activation. Thus, the advancement of this line of research will be greatly aided by the definition of additional co-factors required to induce IL-4-dependent macrophage proliferation *in vitro*.

5 Discussion

In this thesis I have presented two independent systems-level analyses detailing the physiological responses of macrophages to filarial nematode infections. The results that I present in Chapter 2 combine RNA-Seq gene expression profiling, TSS mapping and *cis*-regulatory analysis with targeted lipidomics to characterize IL-4Rα-dependent molecular pathways associated with AAMΦ during *B. malayi* infection. This particular analysis necessitated the development of a novel method for TSS mapping with RNA-Seq data. This leveraged the high-quality genomic annotation information available for mouse (and other models for basic mouse biology such as humans), and was developed into the software tool TSS-Predictor. The algorithmic logic and performance of TSS-Predictor is presented in Chapter 3. In Chapter 4 I present a novel approach to microarray time series expression profiling. Linear modelling was used to integrate FACS-based cellular phenotyping with genome-wide gene expression profiling during a time course of the response to *L. sigmodontis* infection.

The analyses presented in this thesis support the currently accepted view that AAMΦ are anti-inflammatory, or at least are not pro-inflammatory (Gordon and Martinez, 2010). The gene expression profiles of IL-4Rα-dependent AAMΦ-derived cytokines, chemokines and their receptors have been characterized, some of which have unknown function, in the response to nematode infection (Thomas et al., 2012). I hope that these data and analyses serve as a resource to catalyse the study of these molecules. I have also combined NeMΦ gene expression signatures associated with filarial nematode infection from two separate experiments to define a 'core' list of IL-4Rα-dependent NeMΦ-associated genes. This list was derived from macrophage gene expression measurements taken in response to different filarial parasites (*B. malayi* and *L. sigmodontis*) that occupy different physiological

sites (peritoneal and pleural cavity) in different host strains (BALB/c and C57BL/6), and measured using different technologies (RNA-Seq and microarray). This list may not be all encompassing, as technical artefacts may have led to the omission of additional 'true' core-NeMΦ genes. My hope that this stringent approach has led in the generation of a high confidence list of IL-4Rα-dependent NeMΦ-associated genes and will thus have wider utility. For example highly expressed cell surface genes may serve as useful markers to identify polarized NeMΦ by FACS.

In addition to the gene-level characterizations above I have confirmed that nematode infection, and specifically IL-4, induces metabolic pathways associated with mitochondrial metabolism in AAMΦ *in vivo*. This has been shown indirectly with GSEA and directly using MitoTracker staining, and is further supported by *cis*-regulatory element analysis. *In vitro* studies have previously demonstrated that AAMΦ-associated mitochondrial metabolism is mediated by PPARγ (Odegaard et al., 2007). Furthermore, *in vitro,* macrophage *Pparg* itself is induced in response to IL-4 (Huang et al., 1999). It therefore came as a surprise that NeMΦ in the highly IL-4 dependent context of *B. malayi* infection did not express appreciable levels of *Pparg*. However, *B. malayi*-induced NeMΦ abundantly expressed *Ppard*, showed a PPAR transcriptional signature in the genes of IL-4Rα-dependent promoters, and up-regulated mitochondrial metabolic pathways. Based on these data I set out to define putative PPARδ ligands. Because NeMΦ showed differential regulation of arachidonic acid metabolism, a key pathway that serves as a source of PPAR ligands, we performed targeted lipidomics to screen AAMΦ-derived eicosanoids in collaboration with colleagues in Inverness (Philip Whitfield and Ben Maskrey, University of the Highlands and Islands). With this technology we identified the PPARδ ligand PGI2 (Forman et al., 1997) as the most abundant AAMΦ-derived eicosanoid and propose that a PPARδ-PGI2 axis

maintains PPAR-dependent transcription in the NeMΦ response to *B. malayi* infection. The patterns of gene expression in our *L. sigmodontis* time course experiment supported this initial hypothesis.

The *L. sigmodontis* time course experiment proved useful in characterizing the dynamic expression profiles of NeMΦ genes, and provided strong support for the conclusions drawn from my RNA-Seq experiment. However, the primary purpose of this experiment was to gain insight into the differing roles of IL-4 in macrophage proliferation and alternative activation. Linear modelling was used to deconvolute macrophage transcriptional responses associated with alternative activation and proliferation and was achieved combining FACS measurements of YM1 and Ki67 expression with microarray profiling. From this I was able to show that it is unlikely that STAT6 drives macrophage proliferation directly by binding the promoter regions of cell-cycle-associated genes. Based on the findings of this analysis and our understanding that IL-4-dependent macrophage proliferation is STAT6-dependent I have proposed a working model for IL-4-dependent macrophage proliferation.

A pictorial overview of the main findings arising from the work presented in this thesis can be found in Figure 5-1 below. In the remainder of this discussion I will identify the main outstanding biological themes arising from this thesis, and discuss approaches I would take to develop each of these. Finally I consider the utility of systems-wide computational approaches in advancing our understanding of immune processes.

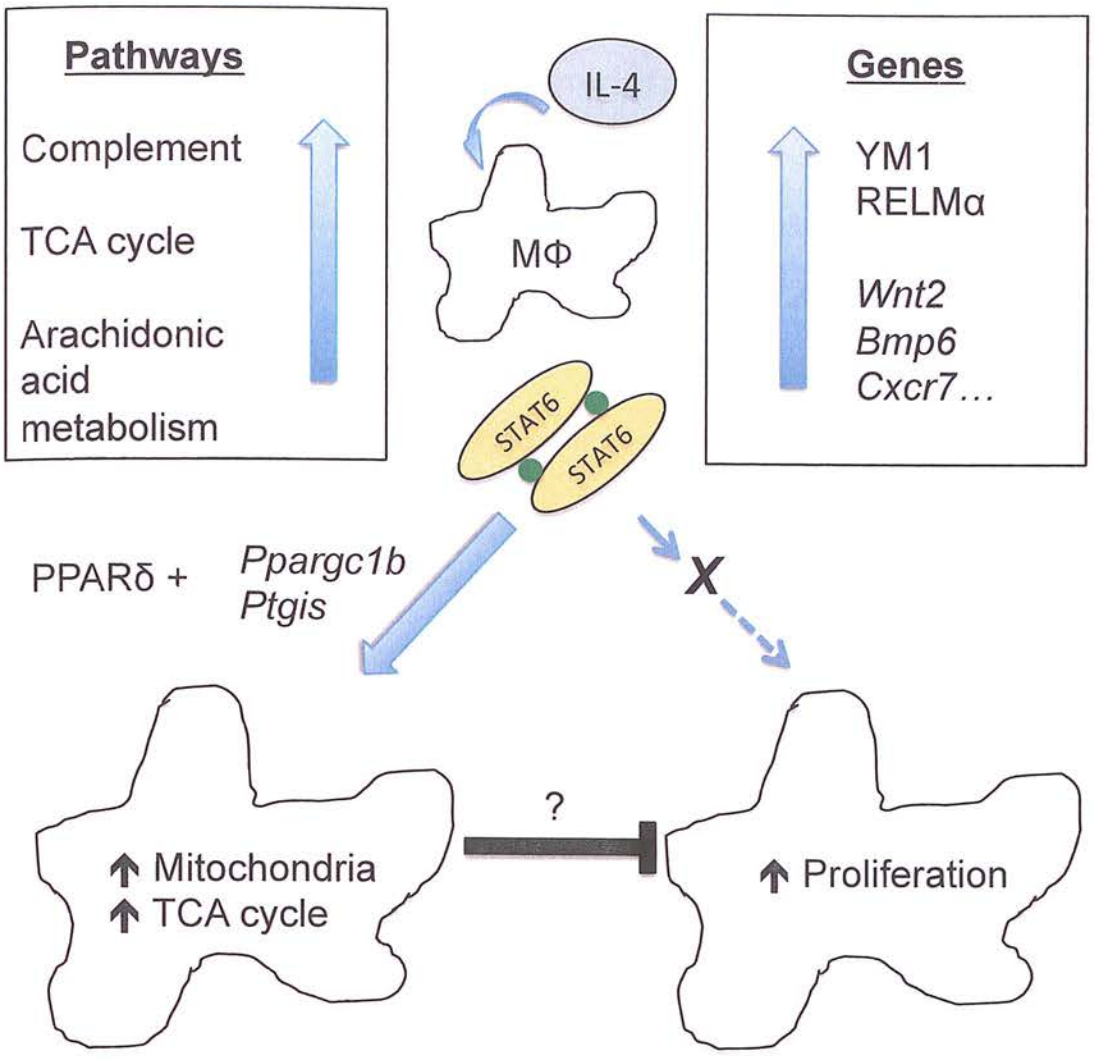


Figure 5-1. Schematic overview of the findings in this thesis. Using the nematode infection models *B. malayi* and *L. sigmodontis* I have characterized IL-4Rα-dependent gene expression profiles and metabolic pathways. Additionally I have investigated mechanisms of IL-4Rα-dependent macrophage proliferation. NeMΦ up-regulate pathways associated with complement activity and mitochondrial metabolism. I have shown that NeMΦ increase mitochondrial mass and metabolism, however do not express high levels of the PPAR transcription factor *Pparg*. Instead, NeMΦ express the related *Ppard* gene and abundantly produce the PPARδ ligand prostacyclin via the up-regulation of genes in the biosynthetic arachidonic acid metabolism pathway. My analysis of IL-4Rα-dependent macrophage proliferation supports does not suggest that STAT6 drives macrophage proliferation directly, and may act via an intermediate protein, perhaps a receptor. The activity of this intermediate factor is predicted to mediate IL-4Rα-dependent proliferation via E2F family transcription factors.

5.1.1 Preparing TSS-Predictor for publication

TSS-Predictor (Chapter 3) was developed to predict the primary TSS usage for a gene from RNA-Seq data. It was necessary to develop this tool because Cufflinks, the only other application that has this capability (Trapnell et al., 2010), generated predictions that were considered inaccurate, largely because of over extension of the 5' end of putative transcriptional units. The initial decision to make TSS-Predictor was based upon a qualitative judgement informed by observing read mapping to the reference genome alongside annotations and Cufflinks gene predictions. As expected, I have since shown that TSS-Predictor does outperform Cufflinks using our dataset. Prior to publication however it will be desirable to test TSS-Predictor on a wider range of RNA-Seq datasets derived from model organisms. Many such datasets are now deposited online, and are available at the European Nucleotide Archive (ENA). A simple search for 'RNA-Seq' at the ENA returned 819 results (http://www.ebi.ac.uk/ena/, 22 March 2013).

TSS will be predicted for RNA-Seq datasets from a range or RNA-Seq experiments across species with 'high quality' TSS predictions, such as those in the ENCODE and modENCODE projects (Consortium, 2004; Gerstein et al., 2010; Roy et al., 2010). TSS predictions will be compared against TSS defined by Cufflinks in relation to reference annotations, as done in Chapter 6. TSS-Predictor relies on the transcript annotations to determine high quality transcription start sites. The distance between the nearest annotated TSS for a gene and the predicted TSS also serves as the metric for algorithm performance. I decided that this was appropriate because it was clear that one of the Ensembl-predicted TSS was correct in the vast majority of cases. Furthermore, Cufflinks had a tendency to systematically predict transcript TSS that were upstream of the 'true' TSS. This justification however will need to be supported by empirical evidence prior to the publication of TSS-Predictor. In order to obtain this I plan to identify an experimental condition (i.e. particular tissue or cell line) for

which both CAGE and RNA-Seq are available. This will allow me to compare TSS predictions derived from Cufflinks and TSS-Predictor against an independent gold standard.

5.1.2 The Complement Cascade in AAMΦ

Pathway analysis from our RNA-Seq dataset revealed a previously uncharacterized role for the complement cascade in the AAMΦ response to *B. malayi* infection. Macrophages are known to be a source of complement (Nathan, 1987), and complement components have been shown to interact with, and contribute towards the killing of microfilariae (Carter et al., 2007). I was struck by the high levels of expression level of complement components, in particular C3, by AAMΦ. Additionally my analyses identified components of the mannose binding lectin pathway, specifically *Fcna*, as highly up-regulated in the response to *B. malayi* implant. My subsequent analysis of macrophages from *L. sigmodontis* infected mice did not reveal a complement signature to the same extent, although C3 was placed within the list of 'core' NeMΦ genes, suggesting that this is induced in response to IL-4.

One possible explanation for this discrepancy is that complement up-regulation is restricted to macrophages in the peritoneal cavity. A key background control for the RNA-Seq experiment is the thioglycollate-elicited macrophage group. Macrophages are acutely recruited from the bloodstream in response to thioglycollate (Hopper, 1986) and may not yet have been induced to express resident peritoneal levels of complement. The peritoneal cavity occupies a particularly sensitive niche in body due to its proximity with the gut. This may necessitate the provision of innate immune components such as complement by a resident population of macrophages, in order to minimize the flow of material between the peritoneal cavity and systemic circulation. This mechanism may not be required within the pleural cavity.

An alternative explanation is that complement components are induced in response to microfilaria, the offspring produced by female filarial nematodes that circulate in the bloodstream. Microfilariae are known to induce a more inflammatory, Th1-biased, immune response than adult filarial nematodes (Lawrence et al., 1994; MacDonald et al., 2003) and presence/absence patterns of microfilariae are a major difference between the *B. malayi* and *L. sigmodontis* models (Le Goff et al., 2002). The adult *B. malayi* implanted during the RNA-Seq experiment constitutively produce microfilariae throughout the course of the experiment. However this will not have been the case during the *L. sigmodontis* time course. *L. sigmodontis* do not produce microfilariae until 50 days post infection (Hoffmann et al., 2000b), and indeed never do in the resistant C57BL/6 strain that we profiled (Hoffmann et al., 2001). A final possibility of course is that these differences in complement expression are reflective of mouse strain differences (C57BL/6 vs. BALB/c). This may be of interest regarding the susceptibility of BALB/c mice. Although complement is well established as a mediator of host defence, it is increasingly being appreciated as a regulator of immunity also (Heeger and Kemper, 2012). Thus it is also plausible that complement induction by AAMΦ in BALB/c, but not C57BL/6, mice may actually confer susceptibility to *L. Sigmodontis* infection in the BALB/c strain.

To drive this line of research forwards it would first be necessary to determine the conditions under which macrophages produce complement in response to filarial nematode infection. The role of *B. malayi* relative to *L. sigmodontis*, and the relative contribution of microfilariae could be tested using an implant model. Young (< day 50) and old (> day 50) adult *L. sigmodontis*, microfilariae producing and non-producing respectively, could be surgically implanted into the peritoneal cavities of BALB/c mice. A control group for this experiment would be *B. malayi* implantation and would serve as a direct reference

to the RNA-Seq experiment that we have performed. This would enable the contrast between parasite species and the presence of microfilariae in the context of the peritoneal cavity.

RT-PCR analysis of various complement components (C3, C1 complex and MBL constituents) produced by purified macrophages would need to be performed. Depending on the outcome of this experiment it would be of interest to determine whether macrophages are induced to produce complement in the pleural cavity response to *L. sigmodontis* microfilariae. After determining the conditions under which complement is produced I would then identify a role for complement in the response to filarial nematode infection. As discussed above, potential roles for complement in helminth infection range from conferring host susceptibility to mediating anti-microfilarial responses. Complement inhibitors and/or genetic approaches would be needed to determine the functions and roles of the relative components of the complement system in the response to nematode infection. The liver and macrophages are amongst the most abundant producers of complement. An attractive strategy to define roles of macrophage-derived complement in the response to filarial nematode infection would be to make use of bone marrow chimeras in which complement deficient bone marrow was grafted into lethally irradiated wild-type hosts. This would allow specific depletion of complement within the hematopoietic system. This approach is feasible as a range of complement-deficient mice are available on both C57BL/6 and BALB/c genetic backgrounds (Trendelenburg et al., 2005).

5.1.3 Understanding the mechanisms behind IL-4-dependent macrophage proliferation is a central question in macrophage biology

Tissue resident macrophages occupy almost all bodily tissues and display a tremendous degree of transcriptional heterogeneity (Gautier et al., 2012). While all macrophages require *Csf1r* for their development (Hume and MacDonald, 2012), the relative importance of additional polarizing factors downstream of *Csf1r* remains an underexplored area. Indeed, the major challenges in myeloid cell biology today are underlain by our ignorance of the ontological relationships between, and the functions of, the various tissue resident macrophage subsets (Hashimoto et al., 2011). It is proposed that macrophages acquire their tissue-specific properties as a result of the microenvironment they experience. For example, migratory Langerhans cells, tissue resident macrophages of the epidermis, display a unique phenotype. Upon migration to the lymph nodes, these same Langerhans cells acquire a cell surface expression profile indistinguishable to migratory dendritic cells when measured by FACS (Merad et al., 2008). This observation proves either phenotypic plasticity, or the continuation of a macrophage developmental programme that is influenced in a tissue specific context.

Many resident macrophage populations require *Csf1r*-dependent signalling for their maintenance in the steady state (MacDonald et al., 2010). However, in addition to *Csf1r*, tissue macrophage cellular identity can require specific factors. For example *Spic* is necessary only for the development of splenic red pulp macrophages (Kohyama et al., 2009). We know from our own experiments that *Csf1r*-dependent signalling is not required for IL-4-dependent macrophage proliferation (S Jenkins unpublished), although it may or may not be necessary for the long term maintenance of steady state peritoneal and pleural cavity populations. Therefore the identification of a second factor driving IL-4-dependent macrophage proliferation, and thus the establishment and maintenance of a Th2 context-

dependent macrophage population, has implications for our understanding of basic macrophage biology that extend beyond the fields of alternative activation and Th2 immunity.

5.1.4 AN EXPERIMENTAL PLAN TO IDENTIFY FACTOR *X*

Based upon the evidence outlined in Chapter 4 I have presented a model for IL-4-induced macrophage proliferation. My hypothesis predicts that IL-4 drives the expression of a second factor that is required to allow macrophages to divide in response to a signal provided by the *in vivo* environment. This relatively simple mechanism would explain why macrophages proliferate in response to IL-4 *in vivo*, but not *in vitro*.

The programmes of cellular proliferation and growth in mammalian cells are governed by growth factors and cytokines (Lemmon and Schlessinger, 2010), and these factors operate through broad family of evolutionarily conserved receptor tyrosine kinases (RTK) (Schlessinger, 2000). Thus, RTKs are responsible for relaying cell extrinsic growth and differentiation signals to cellular metabolic and signalling cascades. Amongst the RTK family is *Csf1r*, the canonical receptor driving the macrophage proliferation profile (Hume and MacDonald, 2012). Studies from within the Allen lab have already discounted the possibility that *Csf1r* is responsible for IL-4-dependent macrophage proliferation (S Jenkins unpublished). I hypothesise that factor X, induced by STAT6, belongs to the large family of RTK receptors. For this hypothesis to hold true, factor X must be part of the list of 'core' NeMΦ genes that I have identified. However, it is also plausible that the candidate RTK is constitutively expressed, and that an adapter protein or co-factor is present in the 'core' list of NeMΦ genes. I have identified candidate RTKs by examining the expression profile of all RTK family members (GO term GO:0004714) in both the RNA-Seq and microarray

experiments I have performed (Appendix 5). This has resulted in the identification of 7 candidate receptors for downstream analysis (Table 5-1). All of these receptors were moderately to highly expressed in both experiments, but not necessarily up-regulated in AAMΦ. Only two of these factors *Insr* (insulin receptor), and *Fgfr1* (fibroblast growth factor receptor 1) were present in the 'core' NeMΦ-associated gene list.

In order to test these candidates my approach would be to deliver IL-4c to mice alongside selective inhibitors of the candidate receptors to identify RTK pathways that are required for IL-4-dependent macrophage proliferation *in vivo*. An important prediction regarding the phenotype of the knockdown/inhibition of candidate *X* is that proliferation, but not alternative activation, would be impaired. Because off-target inhibition is a common problem with pharmacological approaches, this will serve as a starting point to narrow down candidates for *in vitro* proliferation assays. *In vitro* assays would identify ligands that induce macrophage proliferation in the presence of IL-4, identifying these would represent a significant advance in our understanding of macrophage biology. With the development of an *in vitro* system for IL-4-dependent macrophage proliferation siRNA or similar knockdown approaches could then be used to confirm the functional relevance of individual receptor genes to IL-4-dependent macrophage proliferation.

Gene Symbol	RTK Family
Axl	MER/AXL/TYRO3
Csf1r	PDGF
Epha2	EPH
Fgfr1	FGF
Flt1	VEGF
Insr	INSR
Mertk	MER/AXL/TYRO3
Nrp1	NRP

Table 5-1. Candidate receptors mediating IL-4Rα-dependent macrophage proliferation. Transmembrane receptors with tyrosine kinase activity (GO term GO:0004714) expressed by both *B. malayi*-elicited and *L. sigmodontis*-elicited AAMΦ.

5.1.5 Mitochondrial metabolism and the physiology of Th2 immune responses

AAMΦ orchestrate major changes in whole body metabolism by the preferential uptake of lipids, and their catabolism via oxidative phosphorylation in a PPAR-dependent fashion (Vats et al., 2006). This is accompanied by the systemic redistribution of glucose and lipids, mediated via paracrine activities of AAMΦ on the major metabolic tissues (Odegaard and Chawla, 2011) with the consequence of enhancing insulin action. In the context of type-2 diabetes and the metabolic syndrome, IL-4 and AAMΦ help to counteract the adverse metabolic consequences of insulin resistance (Odegaard et al., 2007). We know that helminth infection leads to a large expansion of a resident macrophage population. Furthermore, the helminth parasite *N. braziliensis* ameliorates HFD-induced insulin resistance (Wu et al., 2011). Our transcriptomic data show that helminth-elicited AAMΦ increase *Insr* expression. Thus contrary to the dogma that AAMΦ preferentially oxidise lipids (Vats et al., 2006), the *Insr* expression profile implies that AAMΦ may also act as a significant glucose sink. It would be of considerable interest to determine if, and to characterize how, AAMΦ behaviour is differentially affected by the utilization of different carbon sources.

A great deal of work still remains to be done to understand the molecular relationship(s) between oxidative metabolism and AAMΦ immunogenic properties. I have established that AAMΦ are characterized by an increase in mitochondrial mass, and that this is induced by IL-4 *in vivo*. This is important as previous studies have only shown this *in vitro*. From *in vivo* studies we know obese white adipose tissue (WAT) is associated with inflammatory macrophages (Weisberg et al., 2003), and that lean WAT has an AAMΦ signature and enhanced expression of TCA cycle-associated gene expression (Lumeng et al., 2007; Vats et

al., 2006). However, a direct comparison between obese and lean WAT macrophage-associated gene expression is not valid as obese WAT is characterized by macrophage hyperplasia (Weisberg et al., 2003).

I have already considered that mitochondrial metabolism in AAMΦ may contribute to AAMΦ immune activity by affecting systemic metabolism. Furthermore, oxidative metabolism could extend AAMΦ lifespan, affecting macrophage 'population structure', with numerous possible consequences for immunity. There is precedence for the assertion that metabolic pathways of resource utilization at the cellular level can profoundly affect cellular inflammatory properties. The mitochondrial citrate carrier *Slc25a1* catalyses the export of citrate, a key TCA cycle intermediary, from the mitochondria and is required for robust LPS-induced inflammation in macrophage-like U937 cells (Infantino et al., 2011). Furthermore, in a manner dependent upon glycolysis, the TCA cycle intermediary succinate drives inflammatory gene expression by stabilizing the transcription factor HIF1α (Tannahill et al., 2013). Chemical inhibitors of mitochondrial activity ablate AAMΦ function, at least as measured by arginase activity (Vats et al., 2006), and PPAR activation suppresses inflammatory gene expression (Ogawa et al., 2005; Straus and Glass, 2007).

My findings are consistent with the above models, namely that mitochondrial metabolism and PPAR activity impart anti-inflammatory gene expression profiles in macrophages. It will be important however to determine which functions of AAMΦ are directly STAT6-dependent, which are PPAR-dependent, and which are downstream of other regulatory mechanisms induced during alternative activation. Furthermore, an unresolved debate exists as to whether PPARs require STAT6 or if STAT6 requires PPARs to drive AAMΦ-associated gene expression (Odegaard et al., 2007; Szanto et al., 2010). In reality the situation is probably not quite so black and white that we might expect one transcription

factor to require the other for its activity. Nor is it necessary to assume that Th2 responses and AAMΦ need a particular metabolic phenotype, although the evident coupling between AAMΦ anti-inflammatory gene expression and oxidative metabolism suggests that this may be the case. To date, studies of AAMΦ polarization and metabolism tend to have focussed on obesity and Th2 diabetes as models. It is indisputable that, under these circumstances, alternative activation of adipose tissue macrophages is intimately associated with a 'lean' metabolic phenotype. I believe however that the questions that ought to be asked in the context of helminth infection are: How are immune and metabolic processes are intertwined?; What components of AAMΦ are affected by alterations in metabolism, both systemic and cellular?; And, how do the metabolic components of macrophage activation relate to wider aspects of Th2 immunity and defence against disease?

5.1.6 Informatics approaches to immune cell biology

A combination of systems biology and immune cell biology approaches have been used to characterize and define molecular pathways associated with AAMΦ *in vivo*. Such *in vivo* approaches are absolutely essential to understand the true roles of immune cell subsets as the complexity of live parasite infection, and the associated heterotypic cellular responses, cannot be recapitulated *in vitro*. The success of the transcriptomic experiments in this thesis has been contingent upon the isolation of highly purified macrophage populations using FACS. FACS is an essential tool for the phenotypic screening and functional analysis of cellular subsets derived *in vivo*. However, FACS is restricted by limits to the number of variables that can be simultaneously observed, and indeed the availability of appropriate, specific antibodies for sorting.

By virtue of the ability to measure gene expression profiles genome wide, the combination of FACS and transcriptomic approaches used in this thesis have provided fundamental new insights into IL-4Rα-dependent molecular pathways associated with AAMΦ during nematode infection. Furthermore, the analysis of regulatory elements in the promoters of AAMΦ and proliferation-associated genes has allowed us to infer the transcriptional regulatory mechanisms that may underlie these pathways. Thus, systems biology approaches are highly complementary to widely used immune cell biology methods. Together these provide a powerful engine for hypothesis generation and testing, facilitating functional studies to elucidate the roles of immune cell subsets and their products.

Several large consortia are involved in the systems-level classification of mammalian regulatory elements and immune cell types. The ENCODE and FANTOM consortia are engaged in the large-scale characterization of transcriptional activity and epigenetic modifications across the genomes of target species using RNA-Seq, ChIP-Seq and variants thereof. The objectives FANTOM and ENCODE are to catalogue all genomic regulatory elements (Consortium, 2004; Kawaji et al., 2010). The aim of these programmes is to elucidate fundamental mechanisms of transcriptional regulation, and cell subset specification using both cell lines and mammalian tissues. The Immgen consortium is undertaking the transcriptomic characterization of many immune cell types to describe their gene expression repertoires, and to determine developmental and functional relationships between immune cell subsets. These are large and labour-intensive projects that provide the wider research community with valuable resources, and access to methods that would not otherwise be available. However, these large-scale projects focus on the description and classification of steady state or development in tissue or cell populations, or established cell lines.

A niche remains available for researchers with informatics skills to address biological questions that involve the study of specific cells, or molecules, in the context of a disease or environmental perturbation. Indeed the vast majority of basic biology takes this form. Thus, an unprecedented window of opportunity is provided by genomic technologies to interrogate fundamental cell biological processes at a genome wide level. Several key unanswered questions regarding the regulation of immune cell types can now be addressed. For example, transcription factors such as STATs, PPARS, and their related family members play overarching roles in immune regulation. Such factors operate in multiple immune cell types, and 'direct' or polarize immune responses in a coordinated manner. Yet, the transcriptional responses induced by individual transcription factors are cell-specific. Such cell-specific action is regulated through coordinated, often layered, activities of different transcription factor combinations. We are now able to begin to decipher how these interactions regulate the function of their host cell types. In doing this we may begin to generate a predictive understanding of effects of molecular or pharmacological perturbations on immune responses, and apply such predictive science to medical interventions in health and disease.

6 BIBLIOGRAPHY

't Hoen, P.A.C., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H.A.M., de Menezes, R.X., Boer, J.M., van Ommen, G.-J.B., and den Dunnen, J.T. (2008). Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res. *36*, e141.

Aderem, A., and Underhill, D.M. (1999). Mechanisms of phagocytosis in macrophages. Annu. Rev. Immunol. *17*, 593–623.

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., et al. (2008). Of mice, cattle, and humans: the immunology and treatment of river blindness. PLoS Negl Trop Dis *2*, e217.

Allen, J.E., and Maizels, R.M. (2011). Diversity and dialogue in immunity to helminths. Nat. Rev. Immunol. *11*, 375–388.

Allen, J.E., and Wynn, T.A. (2011). Evolution of Th2 immunity: a rapid repair response to tissue destructive pathogens. PLoS Pathog. *7*, e1002003.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Arima, M., and Fukuda, T. (2011). Prostaglandin D_2 and T(H)2 inflammation in the pathogenesis of bronchial asthma. Korean J. Intern. Med. *26*, 8–18.

Arita, Y., Harkness, S.H., Kazzaz, J.A., Koo, H., Joseph, A., Melendez, J.A., Davis, J.M., Chander, A., and Li, Y. (2006). Mitochondrial localization of catalase provides optimal protection from H2O2-induced cell death in lung epithelial cells. Am J Physiol Lung Cell Mol Physiol *290*, L978–L986.

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., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. *25*, 25–29.

Babayan, S., Ungeheuer, M.-N., Martin, C., Attout, T., Belnoue, E., Snounou, G., Rénia, L., Korenaga, M., and 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. Infect. Immun. *71*, 6820–6829.

Benjamini, Y., and 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*, 289–300.

Biswas, S.K., and Mantovani, A. (2010). Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. Nat. Immunol. *11*, 889–896.

Biswas, S.K., and Mantovani, A. (2012). Orchestration of Metabolism by Macrophages. Cell Metabolism *15*, 432–437.

Van den Bossche, J., Bogaert, P., van Hengel, J., Guérin, C.J., Berx, G., Movahedi, K., Van den Bergh, R., Pereira-Fernandes, A., Geuns, J.M.C., Pircher, H., et al. (2009). Alternatively activated macrophages engage in homotypic and heterotypic interactions through IL-4 and polyamine-induced E-cadherin/catenin complexes. Blood *114*, 4664–4674.

Broadhurst, M.J., Leung, J.M., Lim, K.C., Girgis, N.M., Gundra, U.M., Fallon, P.G., Premenko-Lanier, M., McKerrow, J.H., McCune, J.M., and Loke, P. (2012). Upregulation of retinal dehydrogenase 2 in alternatively activated macrophages during retinoid-dependent type-2 immunity to helminth infection in mice. PLoS Pathog. *8*, e1002883.

Bryne, J.C., Valen, E., Tang, M.-H.E., Marstrand, T., Winther, O., da Piedade, I., Krogh, A., Lenhard, B., and Sandelin, A. (2008). JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. Nucleic Acids Res *36*, D102–106.

Cambi, A., and Figdor, C.G. (2003). Dual function of C-type lectin-like receptors in the immune system. Current Opinion in Cell Biology *15*, 539–546.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563.

Carroll, M.C. (2004). The complement system in regulation of adaptive immunity. Nat. Immunol. *5*, 981–986.

Carter, T., Sumiya, M., Reilly, K., Ahmed, R., Sobieszczuk, P., Summerfield, J.A., and Lawrence, R.A. (2007). Mannose-binding lectin A-deficient mice have abrogated antigen-specific IgM responses and increased susceptibility to a nematode infection. J. Immunol. *178*, 5116–5123.

Cecchini, M.G., Dominguez, M.G., Mocci, S., Wetterwald, A., Felix, R., Fleisch, H., Chisholm, O., Hofstetter, W., Pollard, J.W., and Stanley, E.R. (1994). Role of colony stimulating factor-1 in the establishment and regulation of tissue macrophages during postnatal development of the mouse. Development *120*, 1357–1372.

Chawla, A., Nguyen, K.D., and Goh, Y.P.S. (2011). Macrophage-mediated inflammation in metabolic disease. Nat. Rev. Immunol. *11*, 738–749.

Chen, F., Liu, Z., Wu, W., Rozo, C., Bowdridge, S., Millman, A., Van Rooijen, N., Urban, J.F., Jr, Wynn, T.A., and Gause, W.C. (2012). An essential role for T(H)2-type responses in limiting acute tissue damage during experimental helminth infection. Nat. Med. *18*, 260–266.

Cheong, C., Matos, I., Choi, J.-H., Dandamudi, D.B., Shrestha, E., Longhi, M.P., Jeffrey, K.L., Anthony, R.M., Kluger, C., Nchinda, G., et al. (2010). Microbial stimulation fully differentiates monocytes to DC-SIGN/CD209(+) dendritic cells for immune T cell areas. Cell *143*, 416–429.

Chitsulo, L., Engels, D., Montresor, A., and Savioli, L. (2000). The global status of schistosomiasis and its control. Acta Tropica *77*, 41–51.

Consortium, T.E.P. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636–640.

Cordero, F., Botta, M., and Calogero, R.A. (2007). Microarray data analysis and mining approaches. Brief Funct Genomic Proteomic *6*, 265–281.

Cruikshank, W.W., Kornfeld, H., and Center, D.M. (2000). Interleukin-16. J. Leukoc. Biol. *67*, 757–766.

Davies, J.B. (1994). Sixty Years of Onchocerciasis Vector Control: A Chronological Summary with Comments on Eradication, Reinvasion, and Insecticide Resistance. Annual Review of Entomology *39*, 23–45.

Dennis, E.A., Deems, R.A., Harkewicz, R., Quehenberger, O., Brown, H.A., Milne, S.B., Myers, D.S., Glass, C.K., Hardiman, G., Reichart, D., et al. (2010). A mouse macrophage lipidome. J. Biol. Chem. *285*, 39976–39985.

Dudoit, S., Shaffer, J.P., and Boldrick, J.C. (2003). Multiple Hypothesis Testing in Microarray Experiments. Statistical Science *18*, 71–103.

Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P., and Trent, J.M. (1999). Expression profiling using cDNA microarrays. Nat Genet *21*, 10–14.

Dunning, M.J., Smith, M.L., Ritchie, M.E., and Tavaré, S. (2007). beadarray: R classes and methods for Illumina bead-based data. Bioinformatics *23*, 2183–2184.

Dyson, N. (1998). The regulation of E2F by pRB-family proteins. Genes Dev. *12*, 2245–2262.

Evans, D.B., Gelband, H., and Vlassoff, C. (1993). Social and economic factors and the control of lymphatic filariasis: A review. Acta Tropica *53*, 1–26.

La Flamme, A.C., MacDonald, A.S., Huxtable, C.R., Carroll, M., and Pearce, E.J. (2003). Lack of C3 Affects Th2 Response Development and the Sequelae of Chemotherapy in Schistosomiasis. J Immunol *170*, 470–476.

Forman, B.M., Chen, J., and Evans, R.M. (1997). Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta. Proc. Natl. Acad. Sci. U.S.A. *94*, 4312–4317.

Fox, C.J., Hammerman, P.S., and Thompson, C.B. (2005). Fuel feeds function: energy metabolism and the T-cell response. Nat. Rev. Immunol *5*, 844–852.

Frith, M.C., Fu, Y., Yu, L., Chen, J.-F., Hansen, U., and Weng, Z. (2004). Detection of functional DNA motifs via statistical over-representation. Nucleic Acids Res *32*, 1372–1381.

Gallucci, S., and Matzinger, P. (2001). Danger signals: SOS to the immune system. Curr. Opin. Immunol. *13*, 114–119.

Gasque, P. (2004). Complement: a unique innate immune sensor for danger signals. Mol. Immunol. *41*, 1089–1098.

Gatfield, J., Albrecht, I., Zanolari, B., Steinmetz, M.O., and Pieters, J. (2005). Association of the leukocyte plasma membrane with the actin cytoskeleton through coiled coil-mediated trimeric coronin 1 molecules. Mol. Biol. Cell *16*, 2786–2798.

Gautier, E.L., Shay, T., Miller, J., Greter, M., Jakubzick, C., Ivanov, S., Helft, J., Chow, A., Elpek, K.G., Gordonov, S., et al. (2012a). Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. Nat. Immunol. *13*, 1118–1128.

Gautier, E.L., Chow, A., Spanbroek, R., Marcelin, G., Greter, M., Jakubzick, C., Bogunovic, M., Leboeuf, M., van Rooijen, N., Habenicht, A.J., et al. (2012b). Systemic Analysis of PPARγ in Mouse Macrophage Populations Reveals Marked Diversity in Expression with Critical Roles in Resolution of Inflammation and Airway Immunity. J. Immunol. *189*, 2614–2624.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Gerstein, M.B., Lu, Z.J., Van Nostrand, E.L., Cheng, C., Arshinoff, B.I., Liu, T., Yip, K.Y., Robilotto, R., Rechtsteiner, A., Ikegami, K., et al. (2010). Integrative analysis of the Caenorhabditis elegans genome by the modENCODE project. Science *330*, 1775–1787.

Ghassabeh, G.H., De Baetselier, P., Brys, L., Noël, W., Van Ginderachter, J.A., Meerschaut, S., Beschin, A., Brombacher, F., and Raes, G. (2006). Identification of a common gene signature for type II cytokine-associated myeloid cells elicited in vivo in different pathologic conditions. Blood *108*, 575–583.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.-L., et al. (2010). Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity *32*, 317–328.

Van Ginderachter, J.A., Movahedi, K., Hassanzadeh Ghassabeh, G., Meerschaut, S., Beschin, A., Raes, G., and De Baetselier, P. (2006). Classical and alternative activation of mononuclear phagocytes: Picking the best of both worlds for tumor promotion. Immunobiology *211*, 487–501.

Le Goff, L., Lamb, T.J., Graham, A.L., Harcus, Y., and Allen, J.E. (2002). IL-4 is required to prevent filarial nematode development in resistant but not susceptible strains of mice. Int. J. Parasitol. *32*, 1277–1284.

Gordon, S. (2003). Alternative activation of macrophages. Nat. Rev. Immunol *3*, 23–35.

Gordon, S., and Martinez, F.O. (2010). Alternative Activation of Macrophages: Mechanism and Functions. Immunity *32*, 593–604.

Grailer, J.J., Kodera, M., and Steeber, D.A. (2009). L-selectin: role in regulating homeostasis and cutaneous inflammation. J. Dermatol. Sci. *56*, 141–147.

Gunderson, K.L., Kruglyak, S., Graige, M.S., Garcia, F., Kermani, B.G., Zhao, C., Che, D., Dickinson, T., Wickham, E., Bierle, J., et al. (2004). Decoding randomly ordered DNA arrays. Genome Res. *14*, 870–877.

Halbritter, F., Vaidya, H.J., and Tomlinson, S.R. (2012). GeneProf: analysis of high-throughput sequencing experiments. Nat. Methods *9*, 7–8.

Hall, A., Hewitt, G., Tuffrey, V., and de Silva, N. (2008). A review and meta-analysis of the impact of intestinal worms on child growth and nutrition. Matern Child Nutr *4 Suppl 1*, 118–236.

Hansen, K.D., Brenner, S.E., and Dudoit, S. (2010). Biases in Illumina transcriptome sequencing caused by random hexamer priming. Nucleic Acids Res.

Hashimoto, D., Miller, J., and Merad, M. (2011). Dendritic cell and macrophage heterogeneity in vivo. Immunity *35*, 323–335.

Hayashi, R., Goto, Y., Ikeda, R., Yokoyama, K.K., and Yoshida, K. (2006). CDCA4 is an E2F transcription factor family-induced nuclear factor that regulates E2F-dependent transcriptional activation and cell proliferation. J. Biol. Chem. *281*, 35633–35648.

Heeger, P.S., and Kemper, C. (2012). Novel roles of complement in T effector cell regulation. Immunobiology *217*, 216–224.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hewitson, J.P., Harcus, Y.M., Curwen, R.S., Dowle, A.A., Atmadja, A.K., Ashton, P.D., Wilson, A., and Maizels, R.M. (2008). The secretome of the filarial parasite, Brugia malayi: proteomic profile of adult excretory-secretory products. Mol. Biochem. Parasitol. *160*, 8–21.

Hoffmann, W., Petit, G., Schultz-Key, H., Tayor, D., Bain, O., and La Goff, L. (2000). Litomosoides sigmodontis in Mice: Reappraisal of an Old Model for Filarial Research. Parasitology Today *16*, 387–389.

Hoffmann, W.H., Pfaff, A.W., Schulz-Key, H., Soboslay, P.T., and Soboslav, P.T. (2001). Determinants for resistance and susceptibility to microfilaraemia in Litomosoides sigmodontis filariasis. Parasitology *122*, 641–649.

Hong, J.H., Lee, G.T., Lee, J.H., Kwon, S.J., Park, S.H., Kim, S.J., and Kim, I.Y. (2009). Effect of bone morphogenetic protein-6 on macrophages. Immunology *128*, e442–450.

Hopper, K.E. (1986). Kinetics of macrophage recruitment and turnover in peritoneal inflammatory exudates induced by Salmonella or thioglycollate broth. J. Leukoc. Biol. *39*, 435–446.

Hotez, P.J., Brindley, P.J., Bethony, J.M., King, C.H., Pearce, E.J., and Jacobson, J. (2008). Helminth infections: the great neglected tropical diseases. Journal of Clinical Investigation *118*, 1311–1321.

Huang, J.T., Welch, J.S., Ricote, M., Binder, C.J., Willson, T.M., Kelly, C., Witztum, J.L., Funk, C.D., Conrad, D., and Glass, C.K. (1999). Interleukin-4-dependent production of PPAR-gamma ligands in macrophages by 12/15-lipoxygenase. Nature *400*, 378–382.

Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics *18 Suppl 1*, S96–104.

Hume, D.A., Summers, K.M., Raza, S., Baillie, J.K., and Freeman, T.C. (2010). Functional clustering and lineage markers: Insights into cellular differentiation and gene function from large-scale microarray studies of purified primary cell populations. Genomics *95*, 328–338.

Hume, D.A., and MacDonald, K.P.A. (2012). Therapeutic applications of macrophage colony-stimulating factor-1 (CSF-1) and antagonists of CSF-1 receptor (CSF-1R) signaling. Blood *119*, 1810–1820.

Infantino, V., Convertini, P., Cucci, L., Panaro, M.A., Di Noia, M.A., Calvello, R., Palmieri, F., and Iacobazzi, V. (2011). The mitochondrial citrate carrier: a new player in inflammation. Biochem. J. *438*, 433–436.

Ishizuka, T., Kawakami, M., Hidaka, T., Matsuki, Y., Takamizawa, M., Suzuki, K., Kurita, A., and Nakamura, H. (1998). Stimulation with thromboxane A2 (TXA2) receptor agonist enhances ICAM-1, VCAM-1 or ELAM-1 expression by human vascular endothelial cells. Clin. Exp. Immunol. *112*, 464–470.

Janeway, C.A. (1989). Approaching the Asymptote? Evolution and Revolution in Immunology. Cold Spring Harb Symp Quant Biol *54*, 1–13.

Jenkins, S.J., and Allen, J.E. (2010). Similarity and diversity in macrophage activation by nematodes, trematodes, and cestodes. J. Biomed. Biotechnol. *2010*, 262609.

Jenkins, S.J., Ruckerl, D., Cook, P.C., Jones, L.H., Finkelman, F.D., van Rooijen, N., MacDonald, A.S., and Allen, J.E. (2011). Local macrophage proliferation, rather than recruitment from the blood, is a signature of TH2 inflammation. Science *332*, 1284–1288.

Jing, S., Yu, Y., Fang, M., Hu, Z., Holst, P.L., Boone, T., Delaney, J., Schultz, H., Zhou, R., and Fox, G.M. (1997). GFRalpha-2 and GFRalpha-3 are two new receptors for ligands of the GDNF family. J. Biol. Chem. *272*, 33111–33117.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Kanneganti, T.-D., Lamkanfi, M., and Núñez, G. (2007). Intracellular NOD-like Receptors in Host Defense and Disease. Immunity *27*, 549–559.

Kauffmann, A., Gentleman, R., and Huber, W. (2009). arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. Bioinformatics *25*, 415–416.

Kawai, T., and Akira, S. (2010). The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol *11*, 373–384.

Kawaji, H., Severin, J., Lizio, M., Forrest, A.R.R., van Nimwegen, E., Rehli, M., Schroder, K., Irvine, K., Suzuki, H., Carninci, P., et al. (2010). Update of the FANTOM web resource: from mammalian transcriptional landscape to its dynamic regulation. Nucleic Acids Research *39*, D856–D860.

Keck, K.M., and Pemberton, L.F. (2012). Histone chaperones link histone nuclear import and chromatin assembly. Biochim. Biophys. Acta *1819*, 277–289.

Kerepesi, L.A., Hess, J.A., Nolan, T.J., Schad, G.A., and Abraham, D. (2006). Complement component C3 is required for protective innate and adaptive immunity to larval strongyloides stercoralis in mice. J. Immunol. *176*, 4315–4322.

Kersten, C., Sivertsen, E.A., Hystad, M.E., Forfang, L., Smeland, E.B., and Myklebust, J.H. (2005). BMP-6 inhibits growth of mature human B cells; induction of Smad phosphorylation and upregulation of Id1. BMC Immunol. *6*, 9.

Kim, D.-K., Lee, S.C., and Lee, H.-W. (2009). CD137 ligand-mediated reverse signals increase cell viability and cytokine expression in murine myeloid cells: involvement of mTOR/p70S6 kinase and Akt. Eur. J. Immunol. *39*, 2617–2628.

Koehler, R., Issac, H., Cloonan, N., and Grimmond, S.M. (2011). The uniqueome: a mappability resource for short-tag sequencing. Bioinformatics *27*, 272–274.

Kohyama, M., Ise, W., Edelson, B.T., Wilker, P.R., Hildner, K., Mejia, C., Frazier, W.A., Murphy, T.L., and Murphy, K.M. (2009). Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. Nature *457*, 318–321.

Kopf, M., Le Gros, G., Bachmann, M., Lamers, M.C., Bluethmann, H., and Köhler, G. (1993). Disruption of the murine IL-4 gene blocks Th2 cytokine responses. Nature *362*, 245–248.

Kuhn, K., Baker, S.C., Chudin, E., Lieu, M.-H., Oeser, S., Bennett, H., Rigault, P., Barker, D., McDaniel, T.K., and Chee, M.S. (2004). A novel, high-performance random array platform for quantitative gene expression profiling. Genome Res. *14*, 2347–2356.

Kurowska-Stolarska, M., Stolarski, B., Kewin, P., Murphy, G., Corrigan, C.J., Ying, S., Pitman, N., Mirchandani, A., Rana, B., van Rooijen, N., et al. (2009). IL-33 amplifies the polarization of alternatively activated macrophages that contribute to airway inflammation. J. Immunol. *183*, 6469–6477.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lawrence, R.A., Allen, J.E., Osborne, J., and Maizels, R.M. (1994). Adult and microfilarial stages of the filarial parasite Brugia malayi stimulate contrasting cytokine and Ig isotype responses in BALB/c mice. J Immunol *153*, 1216–1224.

Le, A., Lane, A.N., Hamaker, M., Bose, S., Gouw, A., Barbi, J., Tsukamoto, T., Rojas, C.J., Slusher, B.S., Zhang, H., et al. (2012). Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. Cell Metab. *15*, 110–121.

Lee, G.T., Kwon, S.J., Lee, J.-H., Jeon, S.S., Jang, K.T., Choi, H.Y., Lee, H.M., Kim, W.-J., Kim, S.J., and Kim, I.Y. (2010). Induction of interleukin-6 expression by bone morphogenetic protein-6 in macrophages requires both SMAD and p38 signaling pathways. J. Biol. Chem. *285*, 39401–39408.

Lemmon, M.A., and Schlessinger, J. (2010). Cell signaling by receptor-tyrosine kinases. Cell *141*, 1117–1134.

Ley, K., Laudanna, C., Cybulsky, M.I., and Nourshargh, S. (2007). Getting to the site of inflammation: the leukocyte adhesion cascade updated. Nat Rev Immunol *7*, 678–689.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, S., Reddy, M.A., Cai, Q., Meng, L., Yuan, H., Lanting, L., and Natarajan, R. (2006). Enhanced proatherogenic responses in macrophages and vascular smooth muscle cells derived from diabetic db/db mice. Diabetes *55*, 2611–2619.

Liang, C.-P., Han, S., Okamoto, H., Carnemolla, R., Tabas, I., Accili, D., and Tall, A.R. (2004). Increased CD36 protein as a response to defective insulin signaling in macrophages. J. Clin. Invest. *113*, 764–773.

Liao, X., Sharma, N., Kapadia, F., Zhou, G., Lu, Y., Hong, H., Paruchuri, K., Mahabeleshwar, G.H., Dalmas, E., Venteclef, N., et al. (2011). Krüppel-like factor 4 regulates macrophage polarization. J. Clin. Invest. *121*, 2736–2749.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Lim, H., and Dey, S.K. (2002). Minireview: A Novel Pathway of Prostacyclin Signaling—Hanging Out with Nuclear Receptors. Endocrinology *143*, 3207 –3210.

Lin, S.M., Du, P., Huber, W., and Kibbe, W.A. (2008). Model-based variance-stabilizing transformation for Illumina microarray data. Nucleic Acids Res. *36*, e11.

Liu, T.F., Brown, C.M., El Gazzar, M., McPhail, L., Millet, P., Rao, A., Vachharajani, V.T., Yoza, B.K., and McCall, C.E. (2012). Fueling the flame: bioenergy couples metabolism and inflammation. J. Leukoc. Biol. *92*, 499–507.

Liu, Y., Van Ginderachter, J.A., Brys, L., De Baetselier, P., Raes, G., and Geldhof, A.B. (2003). Nitric oxide-independent CTL suppression during tumor progression: association with arginase-producing (M2) myeloid cells. J. Immunol. *170*, 5064–5074.

Loke, P., MacDonald, A.S., Robb, A., Maizels, R.M., and Allen, J.E. (2000a). Alternatively activated macrophages induced by nematode infection inhibit proliferation via cell-to-cell contact. Eur. J. Immunol. *30*, 2669–2678.

Loke, P., MacDonald, A.S., and Allen, J.E. (2000b). Antigen-presenting cells recruited by Brugia malayi induce Th2 differentiation of naïve CD4(+) T cells. Eur. J. Immunol. *30*, 1127–1135.

Loke, P., Nair, M., Parkinson, J., Guiliano, D., Blaxter, M., and Allen, J. (2002). IL-4 dependent alternatively-activated macrophages have a distinctive in vivo gene expression phenotype. BMC Immunology *3*, 7.

Loke, P., Gallagher, I., Nair, M.G., Zang, X., Brombacher, F., Mohrs, M., Allison, J.P., and Allen, J.E. (2007). Alternative activation is an innate response to injury that requires CD4+ T cells to be sustained during chronic infection. J. Immunol *179*, 3926–3936.

Lottaz, C., Yang, X., Scheid, S., and Spang, R. (2006). OrderedList—a bioconductor package for detecting similarity in ordered gene lists. Bioinformatics *22*, 2315–2316.

Lumeng, C.N., Bodzin, J.L., and Saltiel, A.R. (2007). Obesity induces a phenotypic switch in adipose tissue macrophage polarization. J. Clin. Invest *117*, 175–184.

MacDonald, A.S., Loke, P., Martynoga, R., Dransfield, I., and Allen, J.E. (2003). Cytokine-dependent inflammatory cell recruitment patterns in the peritoneal cavity of mice exposed to the parasitic nematode Brugia malayi. Med. Microbiol. Immunol *192*, 33–40.

MacDonald, K.P.A., Palmer, J.S., Cronau, S., Seppanen, E., Olver, S., Raffelt, N.C., Kuns, R., Pettit, A.R., Clouston, A., Wainwright, B., et al. (2010). An antibody against the colony-stimulating factor 1 receptor depletes the resident subset of monocytes and tissue- and tumor-associated macrophages but does not inhibit inflammation. Blood *116*, 3955–3963.

Maizels, R.M., Pearce, E.J., Artis, D., Yazdanbakhsh, M., and Wynn, T.A. (2009). Regulation of pathogenesis and immunity in helminth infections. J Exp Med *206*, 2059–2066.

Mantovani, A., Sozzani, S., Locati, M., Allavena, P., and Sica, A. (2002). Macrophage polarization: tumor-associated macrophages as a paradigm for polarized M2 mononuclear phagocytes. Trends in Immunology *23*, 549–555.

Mantovani, A., Sica, A., Sozzani, S., Allavena, P., Vecchi, A., and Locati, M. (2004). The chemokine system in diverse forms of macrophage activation and polarization. Trends Immunol *25*, 677–686.

Mantovani, A., Biswas, S.K., Galdiero, M.R., Sica, A., and Locati, M. (2013). Macrophage plasticity and polarization in tissue repair and remodelling. J. Pathol. *229*, 176–185.

Mardis, E.R. (2008). Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet *9*, 387–402.

Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. *18*, 1509–1517.

Martinez, F.O., Gordon, S., Locati, M., and Mantovani, A. (2006). Transcriptional profiling of the human monocyte-to-macrophage differentiation and polarization: new molecules and patterns of gene expression. J. Immunol *177*, 7303–7311.

Martinez, F.O., Helming, L., and Gordon, S. (2009). Alternative activation of macrophages: an immunologic functional perspective. Annu. Rev. Immunol. *27*, 451–483.

Matés, J.M., Pérez-Gómez, C., Núñez de Castro, I., Asenjo, M., and Márquez, J. (2002). Glutamine and its relationship with intracellular redox status, oxidative stress and cell proliferation/death. Int. J. Biochem. Cell Biol. *34*, 439–458.

Matés, J.M., Segura, J.A., Campos-Sandoval, J.A., Lobo, C., Alonso, L., Alonso, F.J., and Márquez, J. (2009). Glutamine homeostasis and mitochondrial dynamics. Int. J. Biochem. Cell Biol. *41*, 2051–2061.

McSorley, H.J., and Maizels, R.M. (2012). Helminth infections and host immune regulation. Clin. Microbiol. Rev. *25*, 585–608.

Merad, M., Ginhoux, F., and Collin, M. (2008). Origin, homeostasis and function of Langerhans cells and other langerin-expressing dendritic cells. Nat. Rev. Immunol. *8*, 935–947.

Modolell, M., Corraliza, I.M., Link, F., Soler, G., and Eichmann, K. (1995). Reciprocal regulation of the nitric oxide synthase/arginase balance in mouse bone marrow-derived macrophages by TH 1 and TH 2 cytokines. European Journal of Immunology *25*, 1101–1104.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods *5*, 621–628.

Moser, B., Wolf, M., Walz, A., and Loetscher, P. (2004). Chemokines: multiple levels of leukocyte migration control. Trends Immunol. *25*, 75–84.

Movahedi, K., Laoui, D., Gysemans, C., Baeten, M., Stangé, G., Van den Bossche, J., Mack, M., Pipeleers, D., In't Veld, P., De Baetselier, P., et al. (2010). Different tumor microenvironments contain functionally distinct subsets of macrophages derived from Ly6C(high) monocytes. Cancer Res. *70*, 5728–5739.

Mukundan, L., Odegaard, J.I., Morel, C.R., Heredia, J.E., Mwangi, J.W., Ricardo-Gonzalez, R.R., Goh, Y.P.S., Eagle, A.R., Dunn, S.E., Awakuni, J.U.H., et al. (2009). PPAR-delta senses and orchestrates clearance of apoptotic cells to promote tolerance. Nat. Med. *15*, 1266–1272.

Mylonas, K.J., Nair, M.G., Prieto-Lafuente, L., Paape, D., and Allen, J.E. (2009). Alternatively activated macrophages elicited by helminth infection can be reprogrammed to enable microbial killing. J. Immunol. *182*, 3084–3094.

Nathan, C.F. (1987). Secretory products of macrophages. J. Clin. Invest. *79*, 319–326.

Odegaard, J.I., and Chawla, A. (2011). Alternative macrophage activation and metabolism. Annu Rev Pathol *6*, 275–297.

Odegaard, J.I., Ricardo-Gonzalez, R.R., Goforth, M.H., Morel, C.R., Subramanian, V., Mukundan, L., Red Eagle, A., Vats, D., Brombacher, F., Ferrante, A.W., et al. (2007). Macrophage-specific PPARgamma controls alternative activation and improves insulin resistance. Nature *447*, 1116–1120.

Odegaard, J.I., Ricardo-Gonzalez, R.R., Red Eagle, A., Vats, D., Morel, C.R., Goforth, M.H., Subramanian, V., Mukundan, L., Ferrante, A.W., and Chawla, A. (2008). Alternative M2 activation of Kupffer cells by PPARdelta ameliorates obesity-induced insulin resistance. Cell Metab. *7*, 496–507.

Ogawa, S., Lozach, J., Benner, C., Pascual, G., Tangirala, R.K., Westin, S., Hoffmann, A., Subramaniam, S., David, M., Rosenfeld, M.G., et al. (2005). Molecular determinants of crosstalk between nuclear receptors and toll-like receptors. Cell *122*, 707–721.

Olefsky, J.M., and Glass, C.K. (2010). Macrophages, inflammation, and insulin resistance. Annu. Rev. Physiol. *72*, 219–246.

Oron, A.P., Jiang, Z., and Gentleman, R. (2008). Gene set enrichment analysis using linear models and diagnostics. Bioinformatics *24*, 2586–2591.

Pearce, E.J., and MacDonald, A.S. (2002). The immunobiology of schistosomiasis. Nat Rev Immunol *2*, 499–511.

Pearce, E.L., Walsh, M.C., Cejas, P.J., Harms, G.M., Shen, H., Wang, L.-S., Jones, R.G., and Choi, Y. (2009). Enhancing CD8 T-cell memory by modulating fatty acid metabolism. Nature *460*, 103–107.

Pendergrass, W., Wolf, N., and Poot, M. (2004). Efficacy of MitoTracker Green and CMXrosamine to measure changes in mitochondrial membrane potentials in living cells and tissues. Cytometry A *61*, 162–169.

Pesce, J.T., Ramalingam, T.R., Mentink-Kane, M.M., Wilson, M.S., El Kasmi, K.C., Smith, A.M., Thompson, R.W., Cheever, A.W., Murray, P.J., and Wynn, T.A. (2009). Arginase-1-expressing macrophages suppress Th2 cytokine-driven inflammation and fibrosis. PLoS Pathog. *5*, e1000371.

Plotkin, S., Diemert, D.J., Bethony, J.M., and Hotez, P.J. (2008). Hookworm Vaccines. Clin Infect Dis. *46*, 282–288.

Qian, B.-Z., and Pollard, J.W. (2010). Macrophage diversity enhances tumor progression and metastasis. Cell *141*, 39–51.

Raes, G., Brys, L., Dahal, B.K., Brandt, J., Grooten, J., Brombacher, F., Vanham, G., Noël, W., Bogaert, P., Boonefaes, T., et al. (2005). Macrophage galactose-type C-type lectins as novel markers for alternatively activated macrophages elicited by parasitic infections and allergic airway inflammation. J. Leukoc. Biol *77*, 321–327.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216–226.

Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. PLoS Comput. Biol. *5*, e1000598.

Reese, T.A., Liang, H.-E., Tager, A.M., Luster, A.D., Van Rooijen, N., Voehringer, D., and Locksley, R.M. (2007). Chitin induces accumulation in tissue of innate immune cells associated with allergy. Nature *447*, 92–96.

Reilly, M.P., and Rader, D.J. (2003). The metabolic syndrome: more than the sum of its parts? Circulation *108*, 1546–1551.

Ritchie, M.E., Dunning, M.J., Smith, M.L., Shi, W., and Lynch, A.G. (2011). BeadArray Expression Analysis Using Bioconductor. PLoS Comput Biol *7*, e1002276.

Roberts, L.D., Murray, A.J., Menassa, D., Ashmore, T., Nicholls, A.W., and Griffin, J.L. (2011). The contrasting roles of PPARdelta and PPARgamma in regulating the metabolic switch between oxidation and storage of fats in white adipose tissue. Genome Biol *12*, R75.

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

Rodríguez-Prados, J.-C., Través, P.G., Cuenca, J., Rico, D., Aragonés, J., Martín-Sanz, P., Cascante, M., and Boscá, L. (2010). Substrate fate in activated macrophages: a comparison between innate, classic, and alternative activation. J. Immunol. *185*, 605–614.

Romagnani, S. (2002). Cytokines and chemoattractants in allergic inflammation. Mol. Immunol. *38*, 881–885.

Rosado, J. de D., and Rodriguez-Sosa, M. (2011). Macrophage migration inhibitory factor (MIF): a key player in protozoan infections. Int. J. Biol. Sci. *7*, 1239–1256.

Rosen, S.D. (2004). Ligands for L-selectin: homing, inflammation, and beyond. Annu. Rev. Immunol. *22*, 129–156.

Roy, S., Ernst, J., Kharchenko, P.V., Kheradpour, P., Negre, N., Eaton, M.L., Landolin, J.M., Bristow, C.A., Ma, L., Lin, M.F., et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. Science *330*, 1787–1797.

Rückerl, D., Hessmann, M., Yoshimoto, T., Ehlers, S., and Hölscher, C. (2006). Alternatively activated macrophages express the IL-27 receptor alpha chain WSX-1. Immunobiology *211*, 427–436.

Sánchez-Martín, L., Estecha, A., Samaniego, R., Sánchez-Ramón, S., Vega, M.Á., and Sánchez-Mateos, P. (2011). The chemokine CXCL12 regulates monocyte-macrophage differentiation and RUNX3 expression. Blood *117*, 88–97.

Sandler, N.G., Mentink-Kane, M.M., Cheever, A.W., and Wynn, T.A. (2003). Global gene expression profiles during acute pathogen-induced pulmonary inflammation reveal divergent roles for Th1 and Th2 responses in tissue repair. J. Immunol. *171*, 3655–3667.

Satoh, T., Takeuchi, O., Vandenbon, A., Yasuda, K., Tanaka, Y., Kumagai, Y., Miyake, T., Matsushita, K., Okazaki, T., Saitoh, T., et al. (2010). The Jmjd3-Irf4 axis regulates M2 macrophage polarization and host responses against helminth infection. Nat Immunol *11*, 936–944.

Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. Cell *103*, 211–225.

Schmidt, R.E., and Gessner, J.E. (2005). Fc receptors and their interaction with complement in autoimmunity. Immunol. Lett. *100*, 56–67.

Schnyder, B., Lugli, S., Feng, N., Etter, H., Lutz, R.A., Ryffel, B., Sugamura, K., Wunderli-Allenspach, H., and Moser, R. (1996). Interleukin-4 (IL-4) and IL-13 bind to a shared heterodimeric complex on endothelial cells mediating vascular cell adhesion molecule-1 induction in the absence of the common gamma chain. Blood *87*, 4286–4295.

Schulz, C., Gomez Perdiguero, E., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobsen, S.E.W., Pollard, J.W., et al. (2012). A lineage of myeloid cells independent of Myb and hematopoietic stem cells. Science *336*, 86–90.

Seno, H., Miyoshi, H., Brown, S.L., Geske, M.J., Colonna, M., and Stappenbeck, T.S. (2009). Efficient colonic mucosal wound repair requires Trem2 signaling. Proc. Natl. Acad. Sci. U.S.A. *106*, 256–261.

Shapiro, H., Lutaty, A., and Ariel, A. (2011). Macrophages, meta-inflammation, and immuno-metabolism. ScientificWorldJournal *11*, 2509–2529.

Shen, S., Park, J.W., Huang, J., Dittmar, K.A., Lu, Z., Zhou, Q., Carstens, R.P., and Xing, Y. (2012). MATS: a Bayesian framework for flexible detection of differential alternative splicing from RNA-Seq data. Nucl. Acids Res. *40*, e61–e61.

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., et al. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. Proc. Natl. Acad. Sci. U.S.A. *100*, 15776–15781.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.

Smyth, G.K. limma: Linear Models for Microarray Data. In Bioinformatics and Computational Biology Solutions Using R and Bioconductor, R. Gentleman, V.J. Carey, W. Huber, R.A. Irizarry, and S. Dudoit, eds. (New York: Springer-Verlag), pp. 397–420.

Staal, F.J.T., Luis, T.C., and Tiemessen, M.M. (2008). WNT signalling in the immune system: WNT is spreading its wings. Nat. Rev. Immunol. *8*, 581–593.

Stein, M., Keshav, S., Harris, N., and Gordon, S. (1992). Interleukin 4 potently enhances murine macrophage mannose receptor activity: a marker of alternative immunologic macrophage activation. J. Exp. Med. *176*, 287–292.

Straus, D.S., and Glass, C.K. (2007). Anti-inflammatory actions of PPAR ligands: new insights on cellular and molecular mechanisms. Trends Immunol. *28*, 551–558.

Stütz, A.M., Pickart, L.A., Trifilieff, A., Baumruker, T., Prieschl-Strassmayr, E., and Woisetschläger, M. (2003). The Th2 cell cytokines IL-4 and IL-13 regulate found in inflammatory zone 1/resistin-like molecule alpha gene expression by a STAT6 and CCAAT/enhancer-binding protein-dependent mechanism. J. Immunol. *170*, 1789–1796.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U.S.A. *102*, 15545–15550.

Suzuki, H., Forrest, A.R.R., van Nimwegen, E., Daub, C.O., Balwierz, P.J., Irvine, K.M., Lassmann, T., Ravasi, T., Hasegawa, Y., de Hoon, M.J.L., et al. (2009). The transcriptional network that controls growth arrest and differentiation in a human myeloid leukemia cell line. Nat. Genet. *41*, 553–562.

Szanto, A., Balint, B.L., Nagy, Z.S., Barta, E., Dezso, B., Pap, A., Szeles, L., Poliska, S., Oros, M., Evans, R.M., et al. (2010). STAT6 transcription factor is a facilitator of the nuclear receptor PPARγ-regulated gene expression in macrophages and dendritic cells. Immunity *33*, 699–712.

Takeda, K., Tanaka, T., Shi, W., Matsumoto, M., Minami, M., Kashiwamura, S., Nakanishi, K., Yoshida, N., Kishimoto, T., and Akira, S. (1996). Essential role of Stat6 in IL-4 signalling. Nature *380*, 627–630.

Tannahill, G.M., Curtis, A.M., Adamik, J., Palsson-McDermott, E.M., McGettrick, A.F., Goel, G., Frezza, C., Bernard, N.J., Kelly, B., Foley, N.H., et al. (2013). Succinate is an inflammatory signal that induces IL-1β through HIF-1α. Nature *496*, 238–242.

Taylor, P.R., Martinez-Pomares, L., Stacey, M., Lin, H.-H., Brown, G.D., and Gordon, S. (2005). Macrophage receptors and immune recognition. Annu. Rev. Immunol. *23*, 901–944.

Theocharidis, A., van Dongen, S., Enright, A.J., and Freeman, T.C. (2009). Network visualization and analysis of gene expression data using BioLayout Express(3D). Nat Protoc *4*, 1535–1550.

Thomas, G.D., Rückerl, D., Maskrey, B.H., Whitfield, P.D., Blaxter, M.L., and Allen, J.E. (2012). The biology of nematode- and IL-4Rα-dependent murine macrophage polarization in vivo as defined by RNA-Seq and targeted lipidomics. Blood.

Thomas-Chollier, M., Defrance, M., Medina-Rivera, A., Sand, O., Herrmann, C., Thieffry, D., and van Helden, J. (2011). RSAT 2011: regulatory sequence analysis tools. Nucleic Acids Res. *39*, W86–91.

Tofaris, G.K., Patterson, P.H., Jessen, K.R., and Mirsky, R. (2002). Denervated Schwann cells attract macrophages by secretion of leukemia inhibitory factor (LIF) and monocyte chemoattractant protein-1 in a process regulated by interleukin-6 and LIF. J. Neurosci. *22*, 6696–6703.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol *28*, 511–515.

Trebino, C.E., Stock, J.L., Gibbons, C.P., Naiman, B.M., Wachtmann, T.S., Umland, J.P., Pandher, K., Lapointe, J.-M., Saha, S., Roach, M.L., et al. (2003). Impaired inflammatory and pain responses in mice lacking an inducible prostaglandin E synthase. Proc. Natl. Acad. Sci. U.S.A. *100*, 9044–9049.

Trendelenburg, M., Fossati-Jimack, L., Cortes-Hernandez, J., Turnberg, D., Lewis, M., Izui, S., Cook, H.T., and Botto, M. (2005). The Role of Complement in Cryoglobulin-Induced Immune Complex Glomerulonephritis. J Immunol *175*, 6909–6914.

Valouev, A., Johnson, D.S., Sundquist, A., Medina, C., Anton, E., Batzoglou, S., Myers, R.M., and Sidow, A. (2008). Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat. Methods *5*, 829–834.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. Science *324*, 1029–1033.

Vastrik, I., D'Eustachio, P., Schmidt, E., Joshi-Tope, G., Gopinath, G., Croft, D., de Bono, B., Gillespie, M., Jassal, B., Lewis, S., et al. (2007). Reactome: a knowledge base of biologic pathways and processes. Genome Biol. *8*, R39.

Vats, D., Mukundan, L., Odegaard, J.I., Zhang, L., Smith, K.L., Morel, C.R., Wagner, R.A., Greaves, D.R., Murray, P.J., and Chawla, A. (2006). Oxidative metabolism and PGC-1beta attenuate macrophage-mediated inflammation. Cell Metab *4*, 13–24.

Walport, M.J. (2001). Complement. First of two parts. N. Engl. J. Med. *344*, 1058–1066.

Wang, H., Gilner, J.B., Bautch, V.L., Wang, D.-Z., Wainwright, B.J., Kirby, S.L., and Patterson, C. (2007). Wnt2 coordinates the commitment of mesoderm to hematopoietic, endothelial, and cardiac lineages in embryoid bodies. J. Biol. Chem. *282*, 782–791.

Wang, Y., Szretter, K.J., Vermi, W., Gilfillan, S., Rossini, C., Cella, M., Barrow, A.D., Diamond, M.S., and Colonna, M. (2012). IL-34 is a tissue-restricted ligand of CSF1R required for the development of Langerhans cells and microglia. Nat. Immunol. *13*, 753–760.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet *10*, 57–63.

Warnes, G.R., Bolker, B., and Lumley, T. (2009). gplots: Various R programming tools for plotting data. R Package Version *2*.

Weisberg, S.P., McCann, D., Desai, M., Rosenbaum, M., Leibel, R.L., and Ferrante, A.W. (2003). Obesity is associated with macrophage accumulation in adipose tissue. J. Clin. Invest *112*, 1796–1808.

Welch, J.S., Escoubet-Lozach, L., Sykes, D.B., Liddiard, K., Greaves, D.R., and Glass, C.K. (2002). TH2 cytokines and allergic challenge induce Ym1 expression in macrophages by a STAT6-dependent mechanism. J. Biol. Chem. *277*, 42821–42829.

Van der Windt, G.J.W., Everts, B., Chang, C.-H., Curtis, J.D., Freitas, T.C., Amiel, E., Pearce, E.J., and Pearce, E.L. (2012). Mitochondrial respiratory capacity is a critical regulator of CD8+ T cell memory development. Immunity *36*, 68–78.

Wu, D., Molofsky, A.B., Liang, H.-E., Ricardo-Gonzalez, R.R., Jouihan, H.A., Bando, J.K., Chawla, A., and Locksley, R.M. (2011). Eosinophils sustain adipose alternatively activated macrophages associated with glucose homeostasis. Science *332*, 243–247.

Wu, H., Moulton, K., Horvai, A., Parik, S., and Glass, C.K. (1994). Combinatorial interactions between AP-1 and ets domain proteins contribute to the developmental regulation of the macrophage scavenger receptor gene. Mol. Cell. Biol. *14*, 2129–2139.

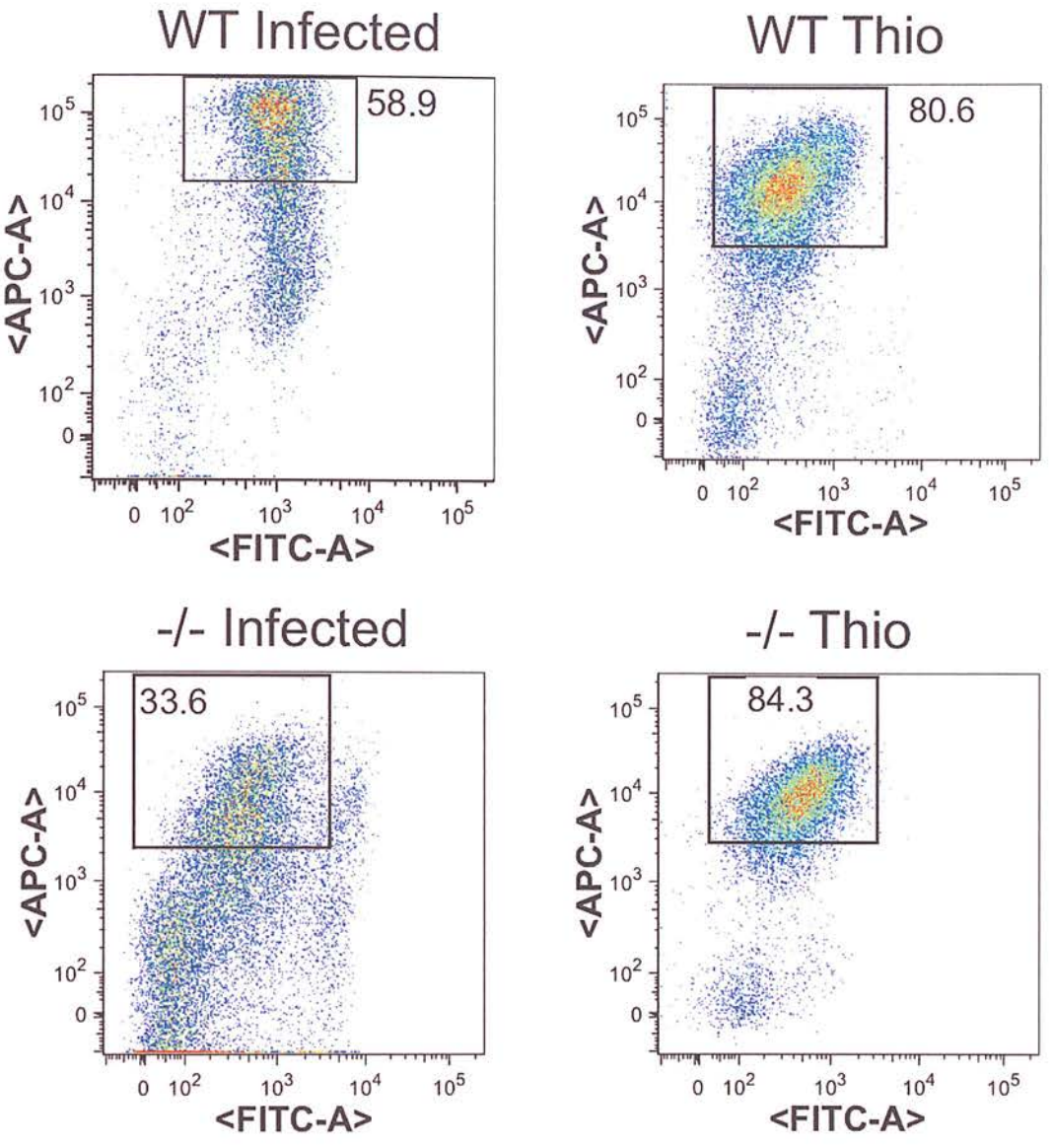
Xu, H., Barnes, G.T., Yang, Q., Tan, G., Yang, D., Chou, C.J., Sole, J., Nichols, A., Ross, J.S., Tartaglia, L.A., et al. (2003). Chronic inflammation in fat plays a crucial role in the development of obesity-related insulin resistance. J. Clin. Invest. *112*, 1821–1830.

Yamanaka, S. (2009). Elite and stochastic models for induced pluripotent stem cell generation. Nature *460*, 49–52.

Yao, G.-Q., Itokawa, T., Paliwal, I., and Insogna, K. (2005). CSF-1 induces fos gene transcription and activates the transcription factor Elk-1 in mature osteoclasts. Calcif. Tissue Int. *76*, 371–378.

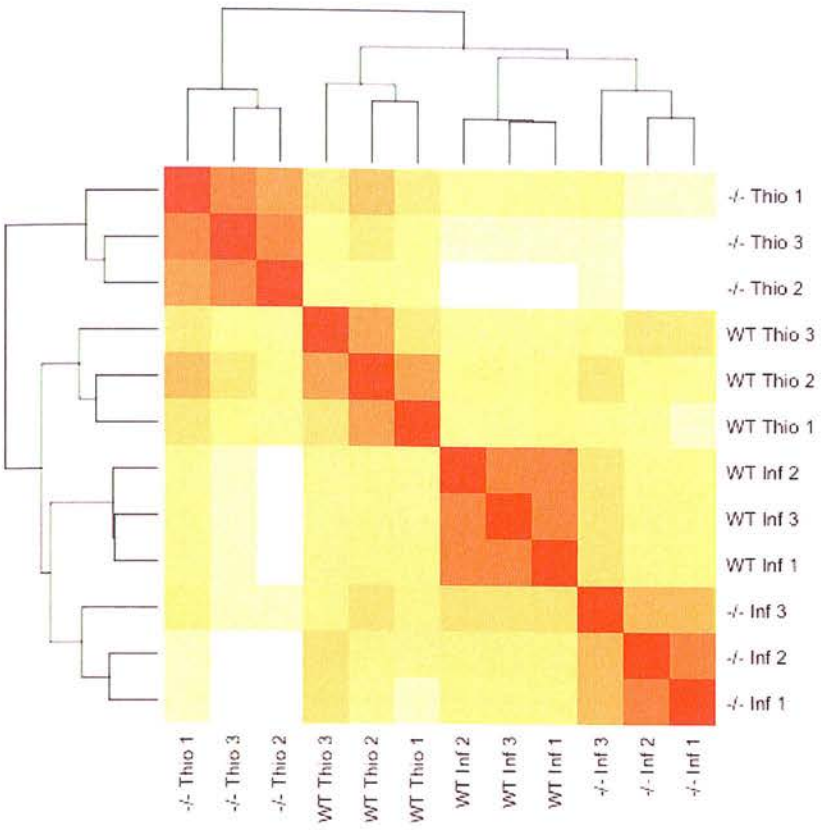
Yu, C.C., Woods, A.L., and Levison, D.A. (1992). The assessment of cellular proliferation by immunohistochemistry: a review of currently available methods and their applications. Histochem. J. *24*, 121–131.

Appendix 1A - Supplementary Figures for Chapter 2

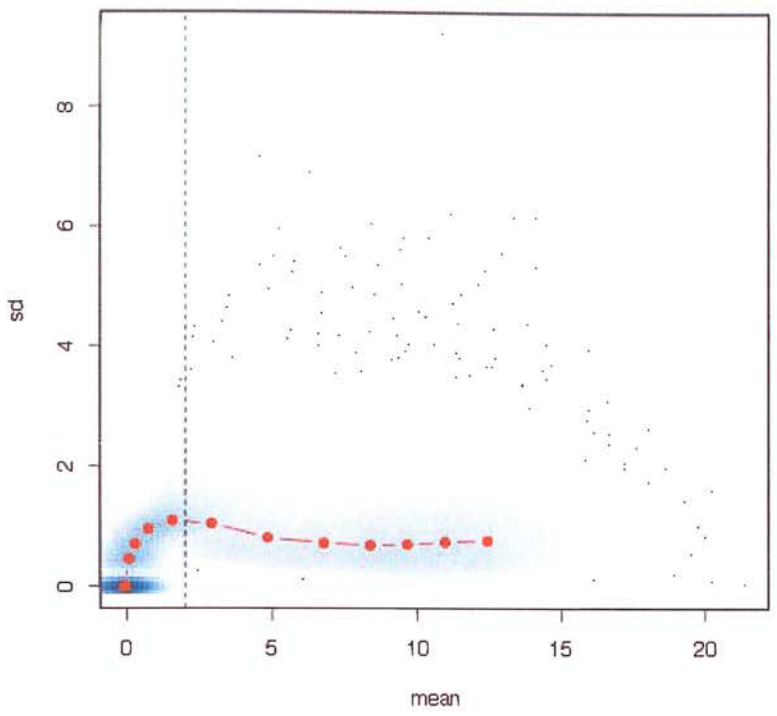


A1-s1. Representative gates showing F4/80 high macrophage populations obtained during flow sorting for WT and IL-4rα$^{-/-}$, *B. malayi*-elicited and thioglycollate-elicited macrophages. APC = F4/80, FITCc = CD4.

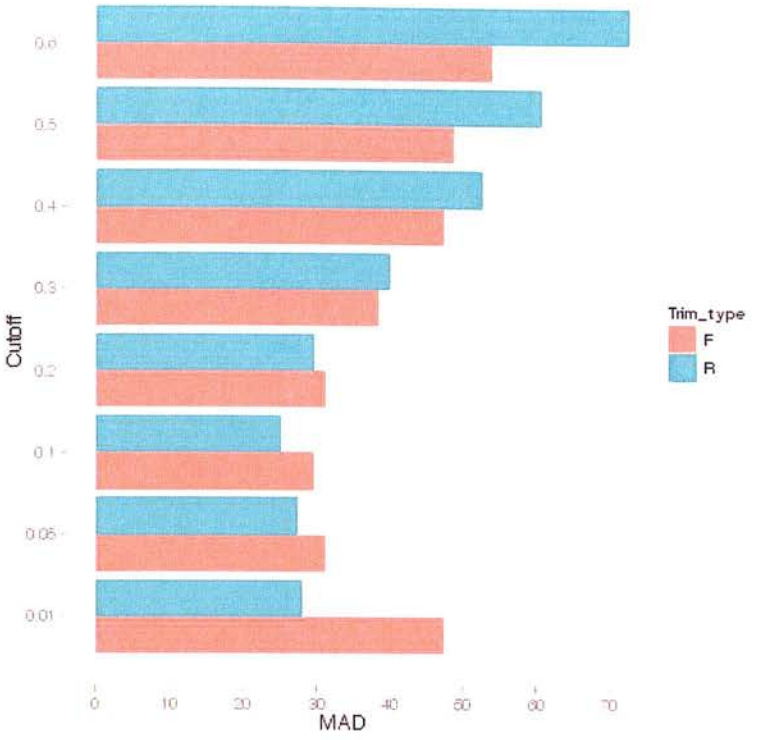
A1-s2. Hierarchical cluster analysis of genes with residual variance >20. Replicates WT-thioglycollate 3, IL-4rα$^{-/-}$ - infected 1 and IL-4rα$^{-/-}$ -infected 2 cluster distinctly, and separately, from the remaining samples, all of which show similar expression profiles.



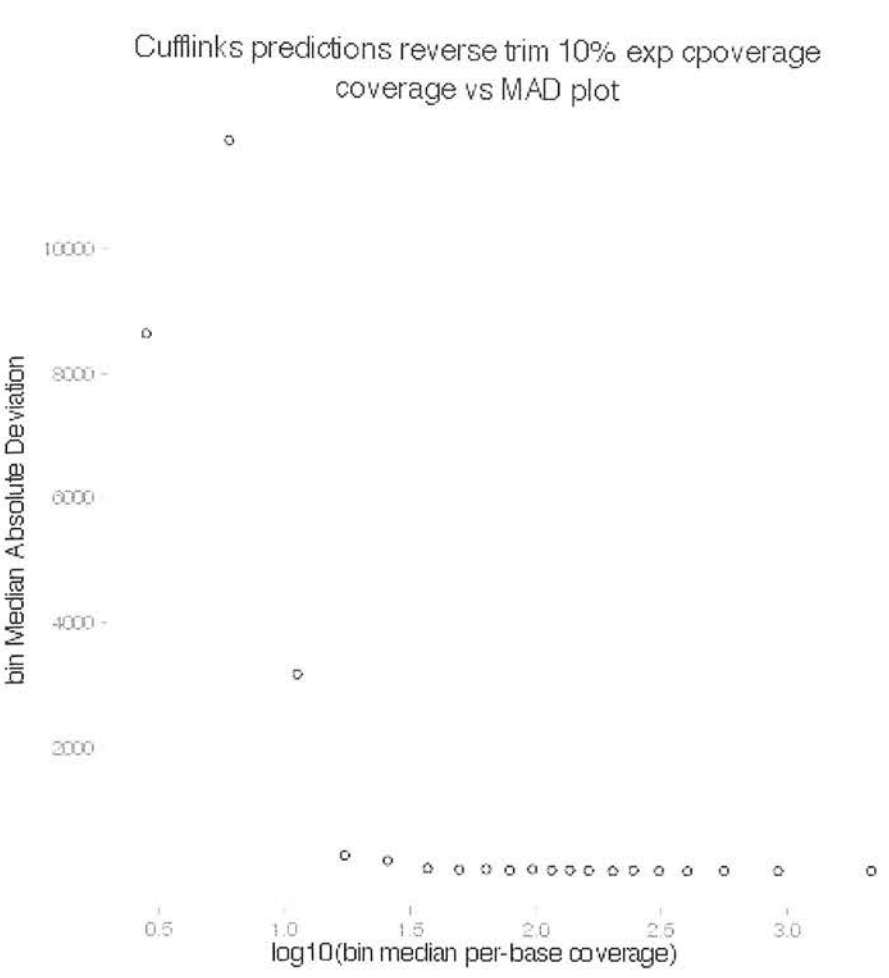
A1-s3. Hierarchical sample clustering with all Ensembl genes prior to outlier removal. This shows that the measured gene expression profiles are dominated by biologically, rather than technically, induced variation.



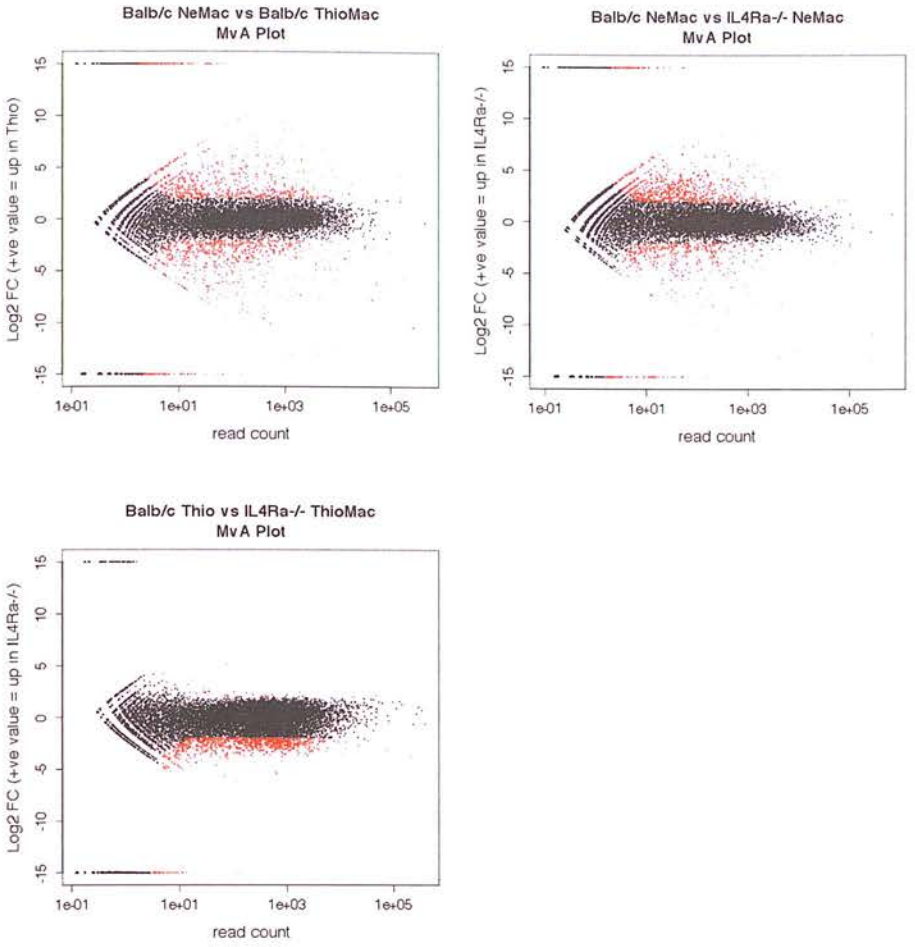
A1-s4. Mean vs standard deviation plot for genes after application of variance-stabilization. The black dotted line represents the non-specific cut-off filter used to enforce homoscedasticity prior to limma analysis.



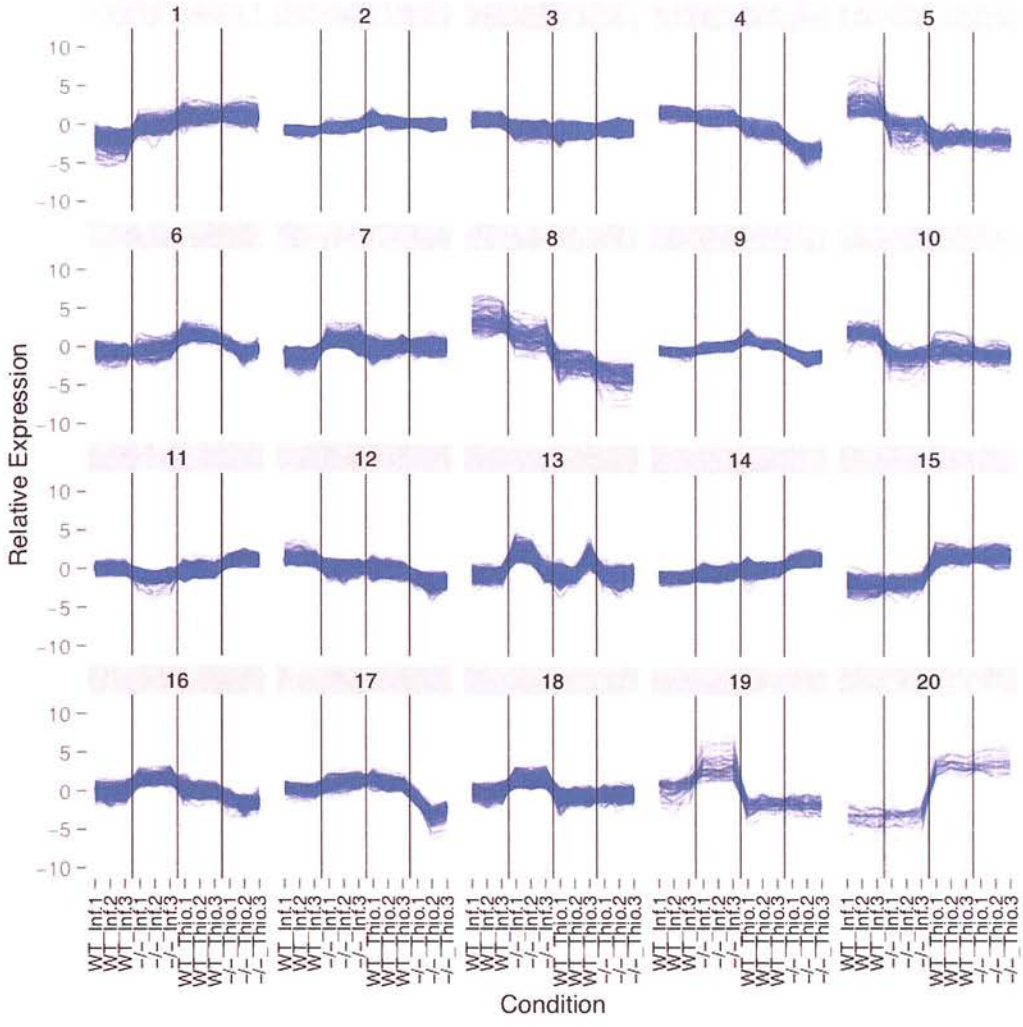
A1-s5. Bar chart showing the median absolute deviation between predicted TSS and the nearest Ensembl TSS across all loci on chromosome 2. This is shown for both forward (red) and reverse (blue) trimming policies at a variety of x cutoffs ranging from 1% to 60%.



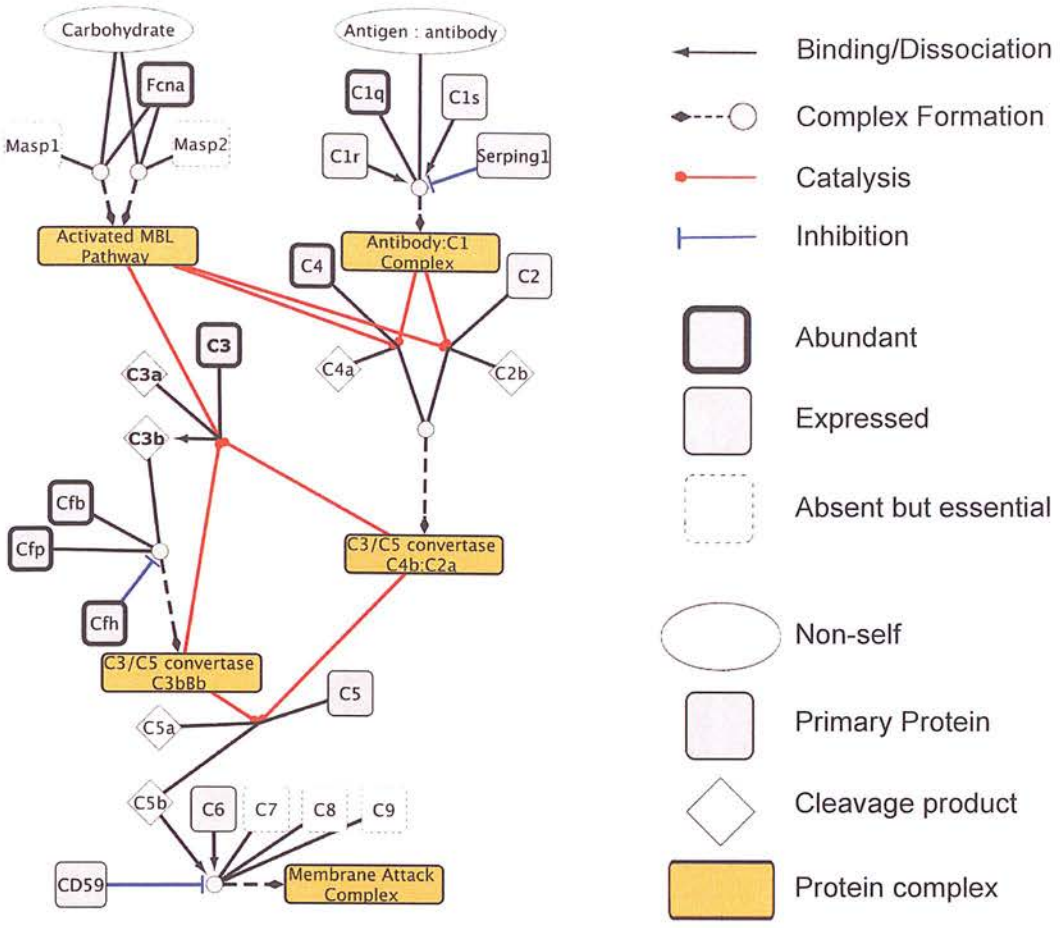
A1-s6. MAD plot showing the median absolute deviation between predicted TSS and nearest TSS as a function of expression level for all expressed genes. Genes were ranked by expression level and 20 bins containing equal numbers genes were created. For each bin the MAD from the nearest Ensembl TSS is shown.



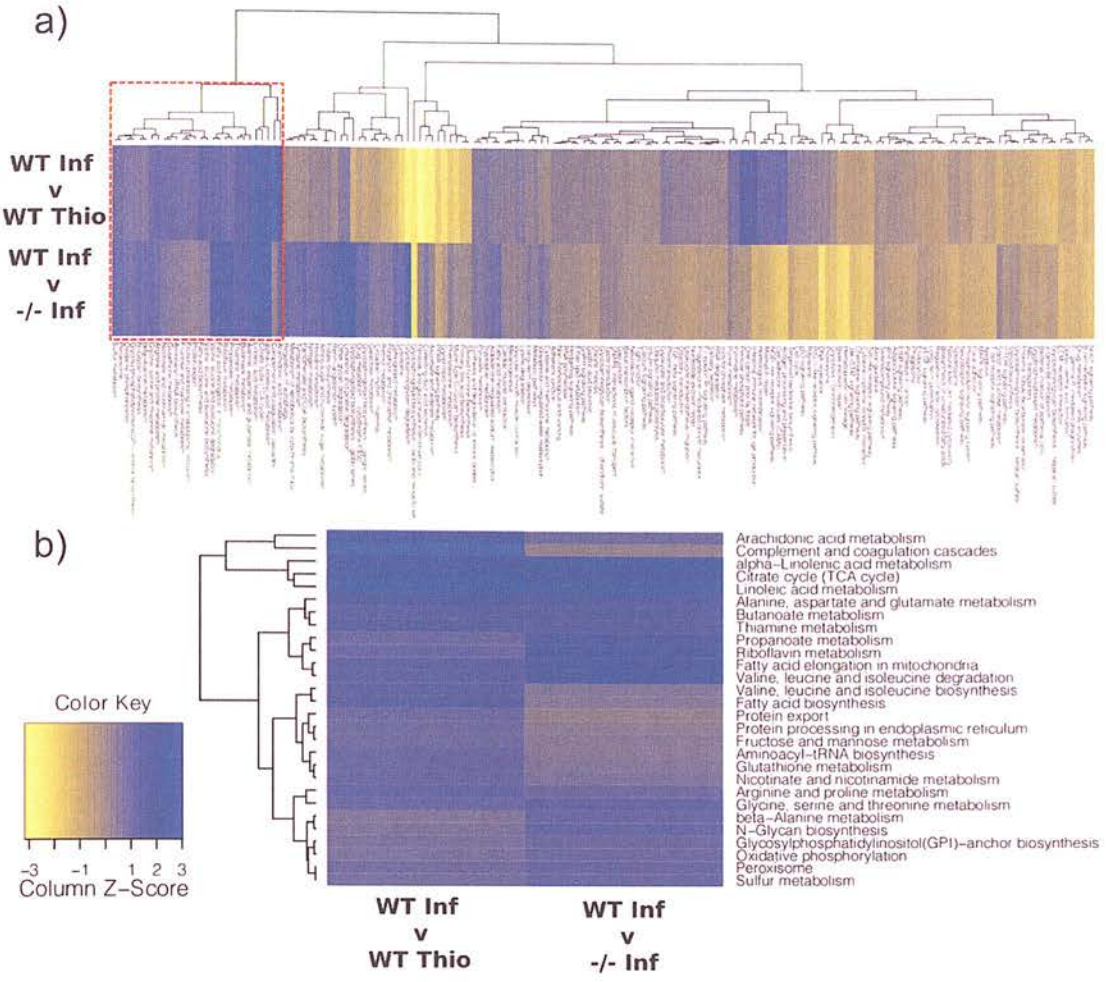
A1-s7. MvA plots demonstrating the differences in the magnitude of differential expression between the 3 key comparisons, WT-NeMϕ vs WT-ThioMϕ (top left), WT NeMϕ vs IL-4rα$^{-/-}$-NeMϕ (top right) and WT-ThioMϕ vs IL-4rα$^{-/-}$-ThioMϕ (bottom left).



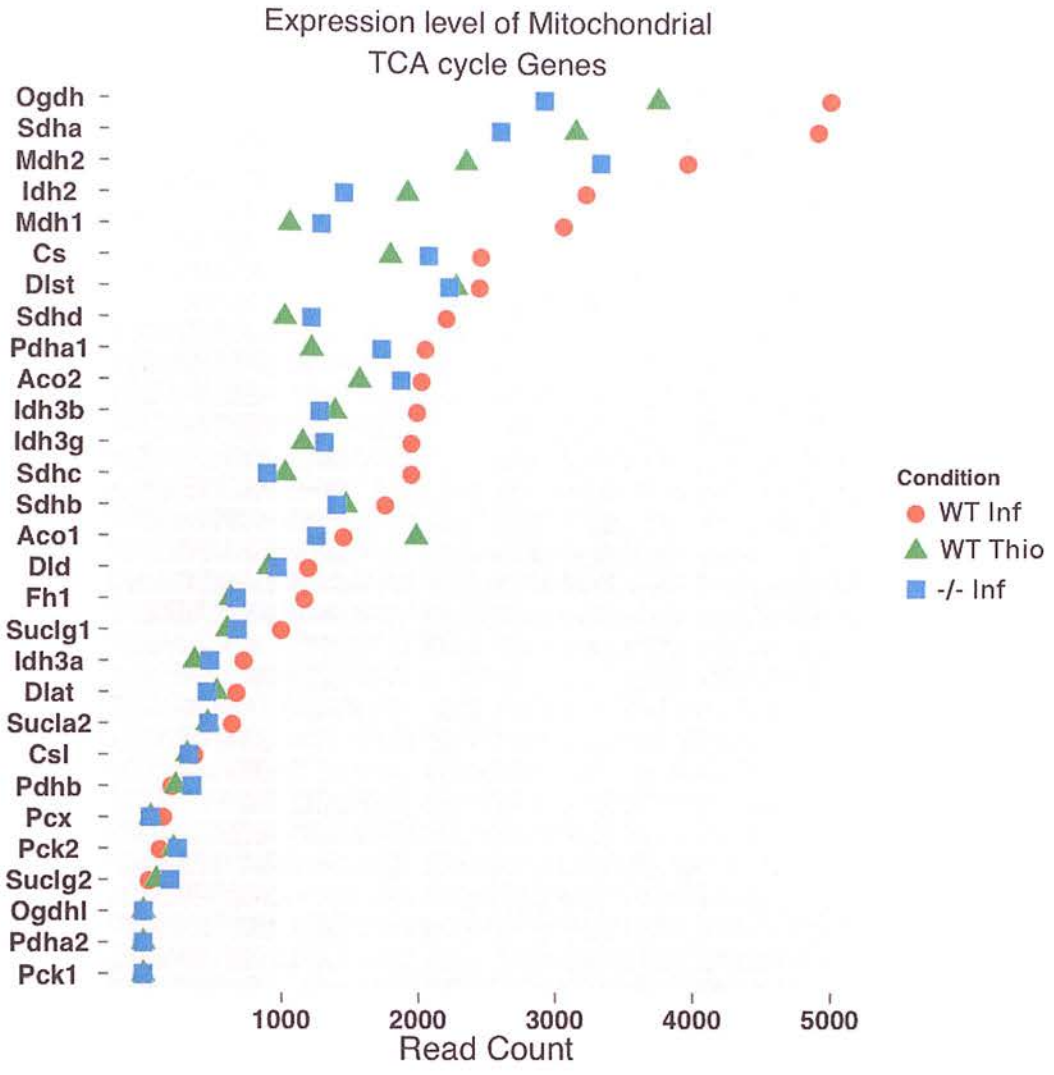
A1-s8. Expression profile of all genes assigned to clusters using hierarchical agglomerative clustering. The following clusters were assigned as AAM-associated: AAM-up.1 = 5, AAM-up.2 = 8, AAM-up.3 = 10, AAM-up.4 = 12, AAM-up.5 = 3, AAM-down.1 = 1, AAM-down.2 = 7, AAM-down.3 = 14, AAM-down.4 = 2, AAM-down.5 = 9.



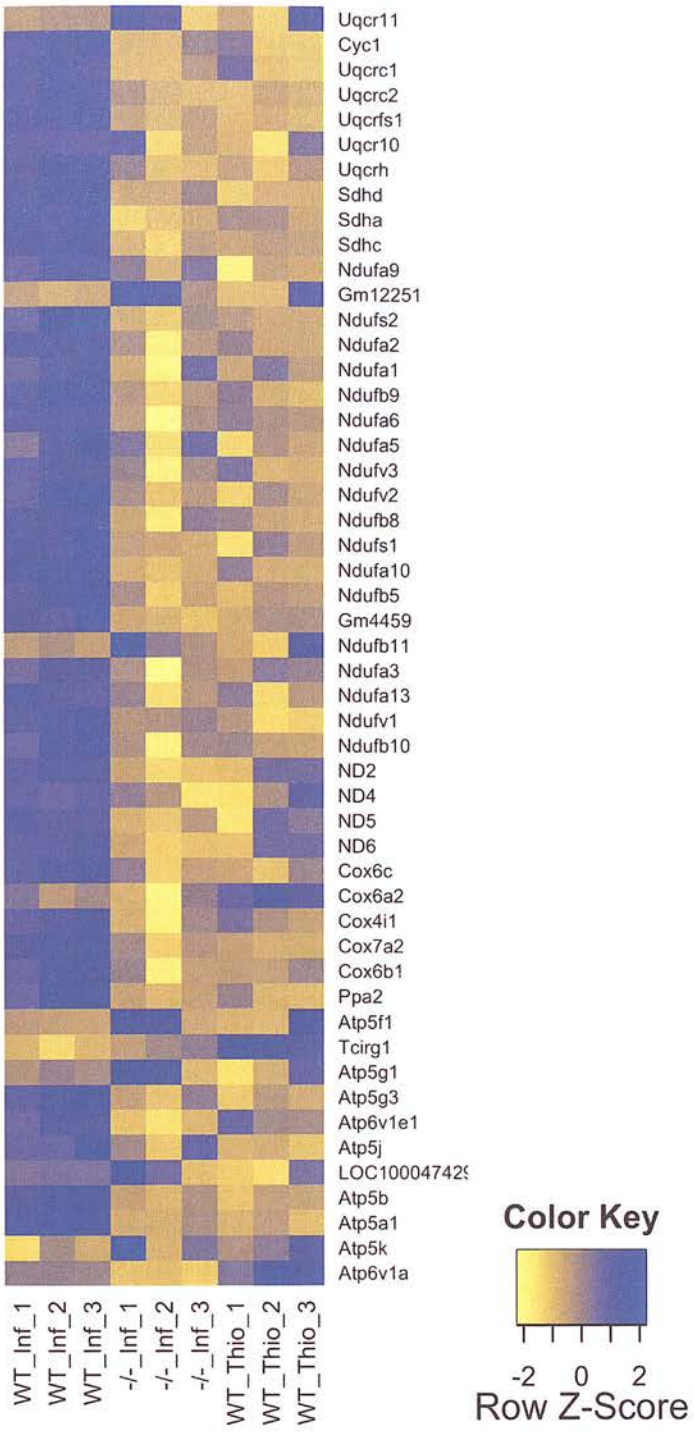
A1-s9. Graphical representation of the complement cascade highlighting genes expressed by NeMϕ. Abundantly expressed genes (thick node borders) are within the 10% most highly expressed WT-NeMϕ transcripts. This figure clearly shows that AAMϕ produce C3bbb C3/5 convertase complex components as well as FicolinA (*Fcna*) and C1q complex components.



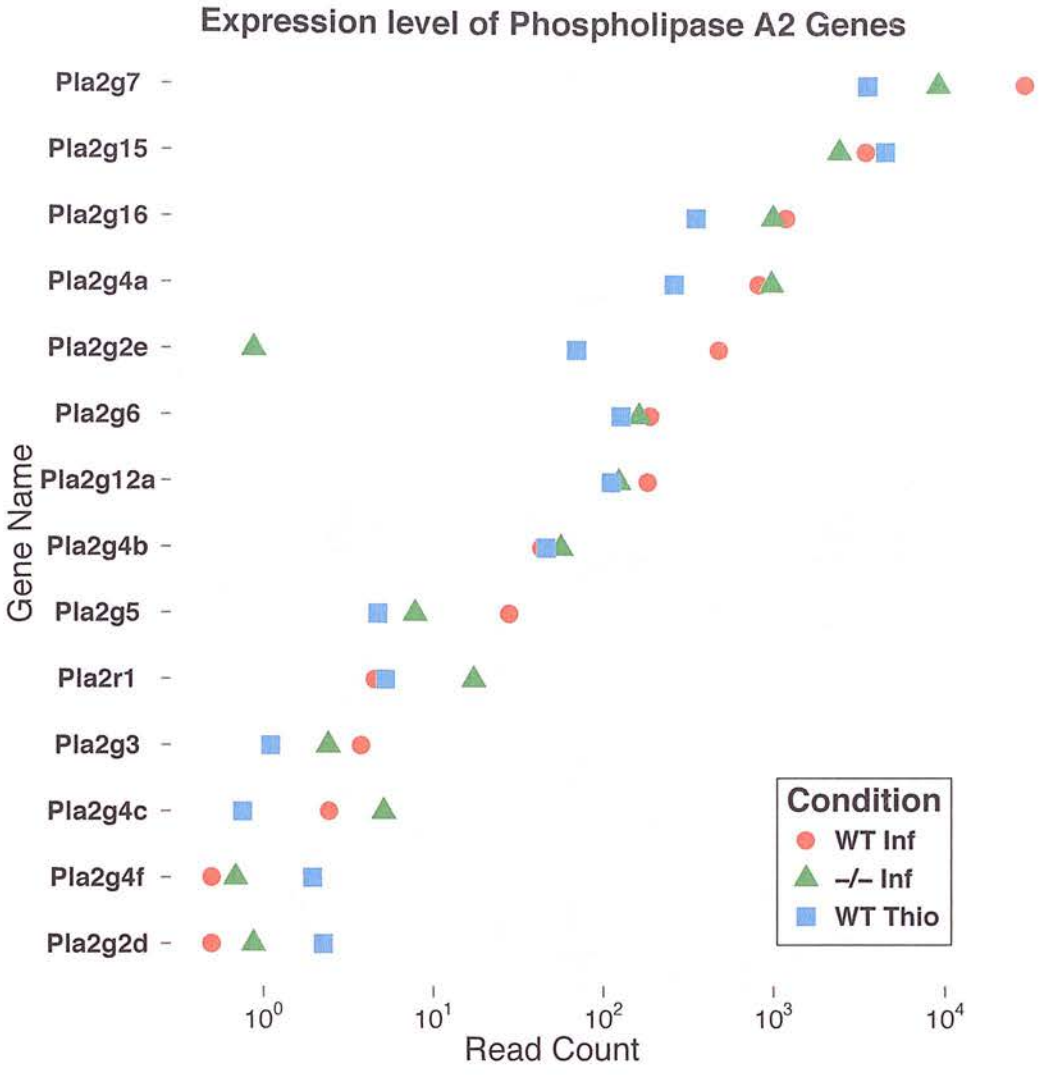
A1-s10 a) Heatmap showing GSEA profiles of all KEGG pathways for the comparisons WT-NeMϕ vs WT-ThioMϕ and WT-NeMϕ vs. IL-4rα$^{-/-}$-NeMϕ. B) Close-up view of the red box in a) showing the over-representation of KEGG metabolic pathways in WT-NeMϕ relative to both control conditions.



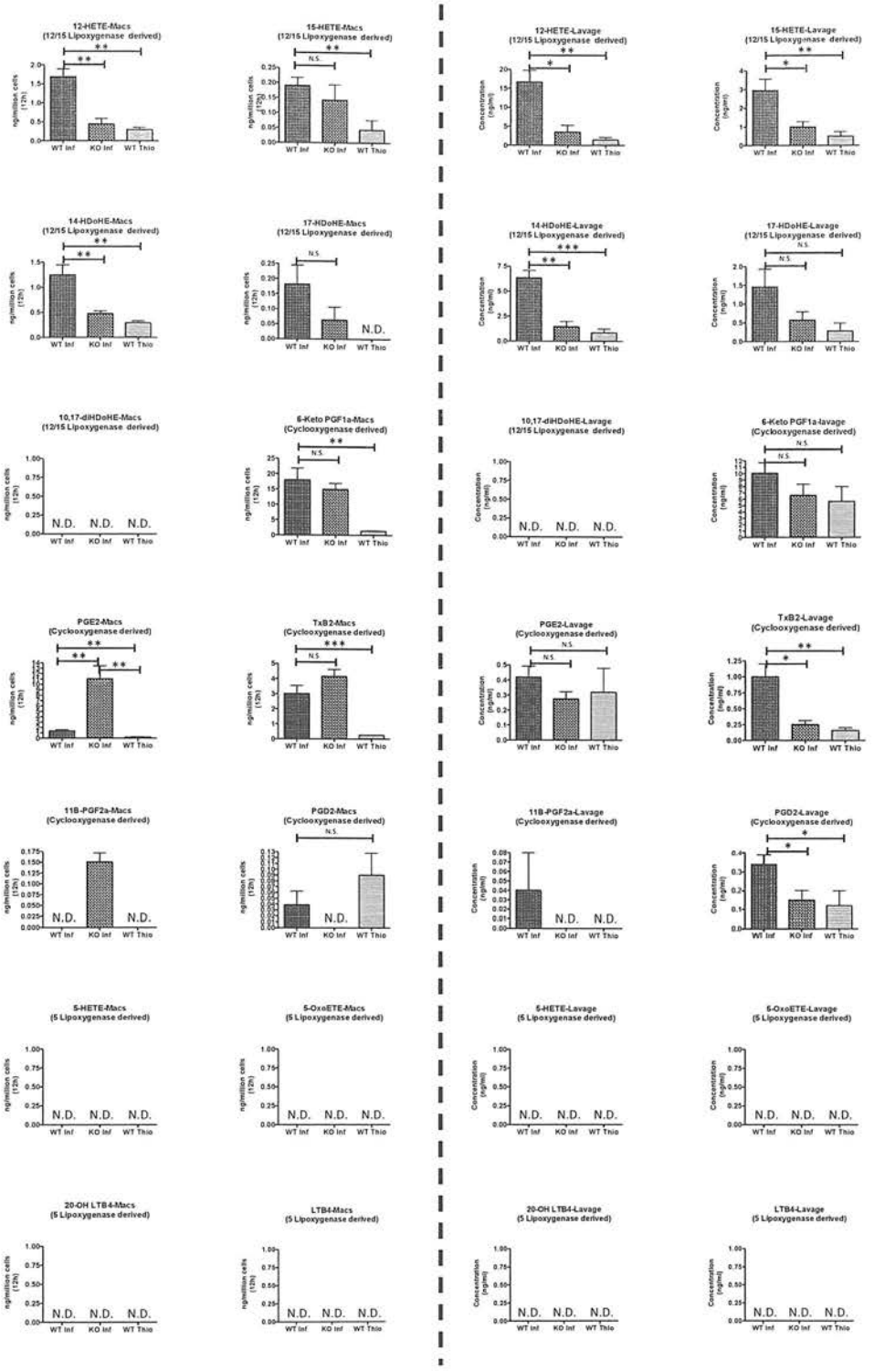
A1-s11. The expression level of mitochondrial TCA cycle genes.



A1-s12. Heatmap showing relative expression of mitochondrial electron transport chain (ETC) components between WT-NeMϕ, IL-4rα$^{-/-}$-NeMϕ and WT-ThioMϕ. The majority of ETC components are most highly expressed in WT-NeMϕ.

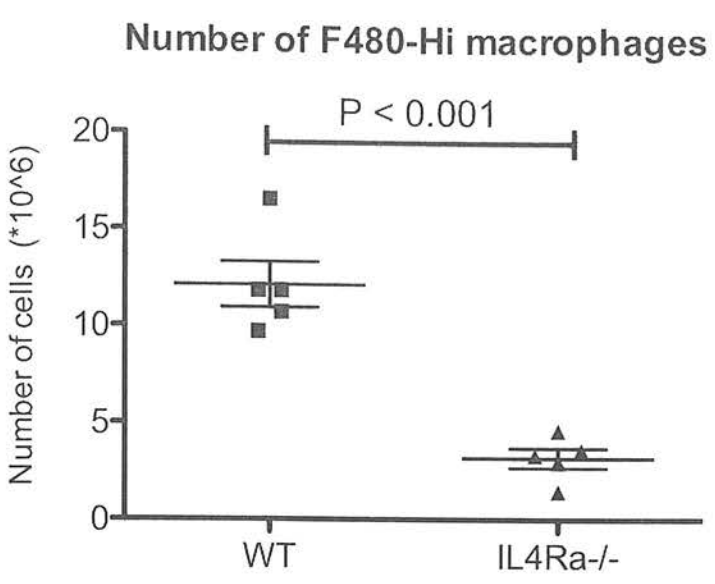


A1-s13. Expression levels of phospholipase a2 enzymes in macrophage populations show no association with alternative activation.

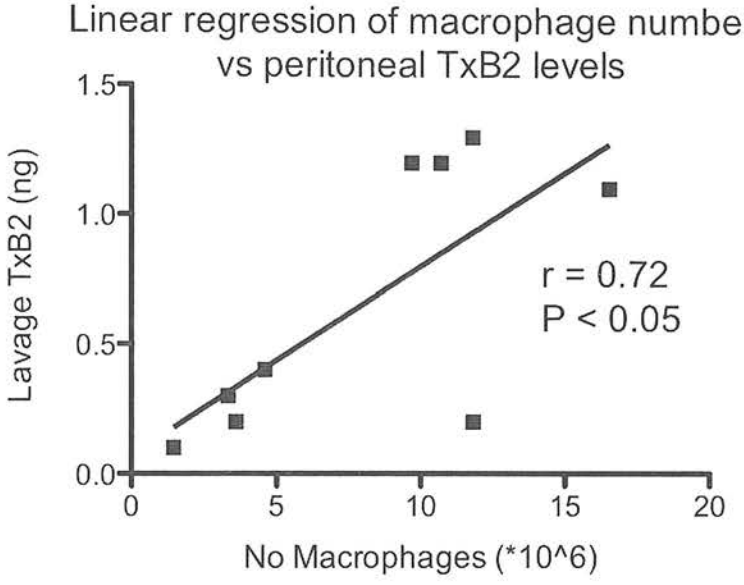


A1-s14. Concentration of measured eicosanoids in 12 hour cultures of adherence-purified peritoneal macrophages *ex vivo* (left panel) and peritoneal lavage supernatant (right panel) 21 days post implantation with *B. malayi* or 3 days post thioglycollate injection in BALB/c (WT) and IL-4rα$^{-/-}$ (KO) mice. N.D. = not detectable.

a)



b)



A1-s15 a) Counts of the total number of F4/80 positive macrophages in the peritoneal cavities of WT and IL-4rα$^{-/-}$ *B. malayi* implanted mice. B) Linear regression showing the significant positive correlation between lavage macrophage number and peritoneal TxB2 concentration.

APPENDIX 1B - SUPPLEMENTARY TABLES FOR CHAPTER 2

Group	library ID	% Purity (library)
WT-Thio	BTG1	98.5
WT-Thio	BTG2	98.1
WT-Thio	BTG3	98.3
WT-Infected	BNe1	96.5
WT-Infected	BNe2	99.4
WT-Infected	BNe3	99.3
KO-Thio	4TG1	97.2
KO-Thio	4TG2	96.6
KO-Thio	4TG3	97
KO-Infected	4Ne1	94.9
KO-Infected	4Ne2	96.1
KO-Infected	4Ne3	96.1

Supplementary Table A1-ST-1. F4/80^{hi} macrophage purities in FACS sorted macrophage popultaions used for RNA-Seq gene expression profiling.

Subset marker	Lineage	KNe1	KNe2	KNe3	KTG1	KTG2	KTG3	BNe1	BNe2	BNe3	BTG1	BTG2	BTG3
Emr1 (F4/80)	Macrophage	1650	1158	1237	691	858.	696	3910	4237	3900	470	844	779
Lyz2 (LyzM)	Myeloid	11995	12322	7315	24302	31760	29994	2952	1151	2125	9665	1253	2209
Prg3	Eosinophils	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Prg2	Eosinophils	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.07	0.00	0.00	0.00	0.00
Epx	Eosinophils	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Siglec5 (SiglecF)	Eoinophils	0.90	0.84	0.67	1.02	0.87	0.38	1.80	1.07	1.18	1.11	0.45	0.65
Mpo	Neutrophils	0.00	0.00	0.00	0.15	0.14	0.00	0.07	0.00	0.14	0.07	0.00	0.12
Ly6g (Gr-1)	Neutrophils	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00
Cd19	B-Cells	22.02	16.24	21.49	2.45	1.45	1.15	2.35	1.35	2.72	0.81	1.02	0.89

Supplementary Table A1-ST-2. Read-per million normalized (Read count per gene/ (total mapped reads/1000000)) gene expression values for marker proteins of potential contaminants in RNA-Seq lanes. BNe = WT Infected, KNe = IL4Rα$^{-/-}$ Infected, BTG = WT Thioglycollate, KTG = IL4Rα$^{-/-}$ Thioglycollate.

Gene Symbol	WT Inf Read Count	WT Thio Read Count	KO Inf Read Count	KO Thio Read Count	WT Inf v WT Thio P val	WT Inf v KO Inf P val	WT Thio v KO Thio P val
Ccl6	40318	65555	54537	192247	7.05E-02	1.86E-01	1.42E-03
Ccl9	16563	19075	12169	42256	5.60E-01	8.85E-02	1.53E-03
Ccl24	3718	438	987	1041	2.90E-37	4.52E-24	4.52E-04
Ccl8	176	7	28	49	6.35E-21	4.12E-15	5.97E-04
Cxcl12	9	2	1	2	5.11E-02	2.06E-04	1.00E+00
Cxcl13	40	1	13	0	2.08E-10	2.23E-03	7.20E-01
Ccl3	9	592	279	603	7.67E-133	1.48E-68	1.00E+00
Ccl4	6	308	162	419	3.51E-78	2.39E-42	3.38E-01
Ccl2	6	373	169	618	1.74E-95	2.27E-45	1.88E-01
Ccl7	4	109	50	185	4.28E-30	1.01E-13	2.41E-01
Cxcl14	164	239	673	412	2.96E-02	6.90E-34	6.17E-02
Cxcr7	7401	369	1225	341	3.45E-53	5.36E-39	8.99E-01
Ccr1	4166	2264	2448	5464	1.41E-05	1.56E-05	1.66E-04
Ccr2	332	2523	4029	5557	2.94E-82	1.36E-124	1.87E-03
Cxcr1	0	3	1	2	7.03E-03	3.81E-01	6.62E-01
Cxcr2	3	12	11	27	4.55E-03	1.19E-02	1.68E-01
Cxcr3	0	5	11	11	9.95E-03	3.44E-06	4.71E-01
Ccr5	2421	2991	5162	3798	1.52E-01	6.58E-12	4.76E-01
Ccrl2	194	874	1244	1094	1.00E-38	5.36E-66	7.18E-01
Il1rl1	2097	40	17	30	1.97E-69	2.87E-130	5.98E-01
Il27ra	1139	121	460	57	2.77E-35	6.08E-12	5.19E-02
Gfra2	1059	74	13	21	3.50E-41	1.33E-99	4.31E-03
Lepr	443	1	30	2	7.01E-56	1.19E-43	8.73E-01
Il6	157	2	76	3	4.11E-24	3.12E-04	9.30E-01
Wnt2	2275	9	114	3	1.77E-90	9.57E-79	2.19E-01
Bmp6	315	1	2	0	2.19E-44	1.04E-60	7.20E-01
Il10	1	1	4	2	1.00E+00	4.92E-02	8.79E-01
Tnf	9	93	211	92	2.47E-19	9.48E-52	9.00E-01
Lif	1	7	10	2	5.12E-03	7.37E-05	1.96E-01
Il1b	20	58	869	160	8.57E-05	5.69E-156	1.20E-02
Il16	165	391	279	402	7.69E-10	7.02E-04	1.00E+00
Tnfsf9	1	15	50	17	8.03E-06	7.20E-20	7.70E-01
Ebi3	16	86	131	83	2.12E-11	5.39E-22	1.00E+00
Pparg	18	632	20	927	2.63E-119	1.00E+00	2.33E-01
Ppard	547	538	724	100	1.00E+00	3.73E-02	1.15E-09
Ppara	1	1	0	0	9.50E-01	7.65E-01	7.20E-01

Supplementary Table A1-ST-3. List of genes, and gene expression values (average number of reads mapped per condition after adjustment for library size in DESeq) for all discussed genes in Chapter 2.

		AAM up 1		
5430407P10Rik	Cd209d	Ggct	Olfml2a	Sdsl
5730528L13Rik	Cd209e	Glrp1	Omp	Sec16b
6230427J02Rik	Cd55	Gm10369	Parm1	Serpinb1a
A530064D06Rik	Chac1	Gm684	Pax3	Sfrp1
Abi3bp	Chi3l3	Gnao1	Pcdh7	Sla
Acy1	Chi3l4	Gpr55	Pcolce2	Slamf1
Acy3	Chn2	Gpr56	Pdia5	Slc16a9
Adcyap1r1	Clec4g	Grhl3	Pdlim1	Slc34a2
AF529169	Cpne8	Hfe	Pgcp	Slc4a4
Aif1l	Cxcr7	Hoxa1	Plin4	Sorcs2
Aldh1l2	Dapk2	Il1rl1	Plin5	St8sia6
Alox12e	Ddx4	Isyna1	Pmepa1	Sult1a1
Alox5	Dixdc1	Itga7	Prdm16	Svs2
Amotl1	Dmrt2	Kif21a	Prelp	Syt7
Ank	Dnmt3l	LOC100502940	Prkaa2	Tdo2
Antxr1	Dtna	LOC100504608	Prodh	Thrb
Aqp9	Dzip1	Mgat3	Prr15	Tiam2
Arhgap29	Ear11	Mgl2	Ptk6	Tmem176a
Arl4d	Eltd1	Mogat1	Pvrl2	Tmem176b
Asb4	Entpd3	Msx3	Rap1gap	Tmem39b
AU021092	Esr1	Mtmr7	Rbms3	Tmprss13
Bai1	Fam110c	Muc20	Retnla	Tnfsf10
Batf3	Fam190a	Muc4	Rhou	Tnfsf14
BC051142	Fam3b	Naip1	Rpgr	Tnip3
Bmp6	Fkbp14	Nap1l3	Rxrg	Ttc25
C6	Gatm	Nfil3	Ryk	Vsig4
Camk2a	Gbe1	Nipal1	S1pr4	X99384
Car12	Gfi1b	Nos1	Scara5	Zfp57
Car4	Gfpt2	Nphp1	Scn1b	
Cd209a	Gfra2	Ofcc1	Scube1	
		AAM up 2		
1100001G20Rik	Cmah	H2-DMb1	Nhsl2	Selp
1600021P15Rik	Cst7	H2-Eb1	Nlrp4f	Serpinb2
2310046K01Rik	Cxcl13	Hdc	Nt5e	Serpinb7
9130008F23Rik	Cyp26b1	Hspa12a	Ntng2	She
A430105I19Rik	Dab2ip	Htr2a	Ocln	Slc22a23
Ablim1	Dock5	Il18bp	Padi4	Slc36a2
Acpp	Efnb2	Il27ra	Pdcd1lg2	Slc7a2
Adamts2	Enpp4	Il6	Pde4b	Slco2a1
Aldh1l1	Enpp5	Insr	Pecam1	Spint2
Alox15	Epha2	Irs1	Pelp1	Srgap3
Amigo2	Epha4	Itih1	Pi16	Svep1

Arhgef16	Ephx1	Itln1	Plekhg1	Sytl4
Aspa	Erbb2	Itlnb	Plin1	Tanc1
Bcam	Erc2	Kank2	Plxdc2	Tenc1
Btbd11	Evi5	Kcnn3	Ppl	Tex14
C2	F5	Lbp	Ppp1r14c	Tgfb2
C3	Fasn	Lepr	Prg4	Tgfbr3
C4b	Fgfr1	LOC100047937	Prickle1	Tjp1
C77370	Flnb	LOC100048759	Prickle2	Tln2
Cadps2	Flt1	LOC100505003	Prkcq	Trim72
Capn5	Fscn1	LOC675521	Prss35	Tsku
Cd244	Fxyd6	LOC677008	Prtn3	Uaca
Cd38	Garnl3	Loxl2	Ptgis	Ucp1
Cd74	Gata6	Lpar1	Ptgs1	Ust
Cd81	Gbp3	Lrg1	Ptprm	Vill
Cdkn2a	Gbp5	Ltbp1	Ramp3	Vmn2r26
Cish	Gm4951	Mall	S1pr5	Wnt2
Cldn10	Gpr116	Mst1r	Saa3	
Clmn	Grlf1	Nbl1	Sardh	

		AAM up 3		
1700084C01Rik	Chi3l7	Gm10384	Maml2	Scamp5
1810011O10Rik	Chit1	Gm7120	Mapk12	Sema3b
2010110P09Rik	Col13a1	Gm826	Mkx	Sh3d19
4931408A02Rik	Csn3	Gng4	Mmp28	Slc15a5
5033411D12Rik	Cst13	Grrp1	Mog	Slc16a1
A4galt	Ctsf	Gstt3	Mpp6	Slc2a5
A930005I04Rik	Cux2	Hc	Nova1	Smtnl2
Abcd2	Cyp2ab1	Hecw1	Npdc1	Sprr2a1
Acss1	Cyp2u1	Hormad1	Npl	Sprr2a2
Adck3	D6Wsu116e	Hrct1	Nrap	Srms
Adcy3	Dhrs9	Hs3st2	Nrg4	St3gal4
Afap1l1	Dhtkd1	Hydin	Odz4	Stap2
Agmo	Dnaic1	Igf1	Olfr1033	Sult2b1
Aldh1a2	Dnm1	Il1rl2	Pcyox1l	Tbc1d30
Aldh5a1	Dpp6	Il20rb	Pde4d	Thbs4
Amhr2	E030010A14Rik	Irf4	Pde4d	Tjp2
Arhgap12	Ear12	Itgax	Pde4dip	Tmem119
Armc2	Ear2	Itgb3	Pdgfc	Tmem52
Asb11	Ear3	Itgb4	Pgm5	Tmem64
Asz1	Efs	Jup	Pla2g2e	Tmigd1
BC048546	Ehd2	Kcnk13	Pla2g7	Tmprss5
Bend6	Eid2	Kcnt2	Plxna2	Tppp
Bfsp1	Epas1	Klhl13	Prkcdbp	Trf
Bzw2	Fam115c	Kpna6	Prnp	Trim12a

Cacna1a	Fam198b	Lgi3	Prox1	Trim43a
Calml4	Fam19a5	Lhx8	Ptprg	Usp9y
Ccl24	Fcgr2b	Lipc	Ptrf	Wtip
Ccl8	Fcna	LOC100044656	Rasgef1a	Zfyve28
Cd209f	Fcrls	LOC100503254	Rem1	
Cd209g	Fgd5	LOC100504580	Ren1	
Cdh1	Fhl1	Lrp3	Ren2	
Cdr2l	Galnt14	Lypd6	Ripk4	
Chchd10	Glb1	Lyplal1	Ripply3	
		AAM up 4		
1110067D22Rik	Cpeb3	Ibtk	Nr1d1	Slc26a2
1300001I01Rik	Crim1	Ids	Nr3c2	Slc2a8
1300010F03Rik	Crtap	Igf2	Ntn4	Slc35d1
1600029D21Rik	Csf1r	Il4ra	Nucb1	Slc37a1
1700020L24Rik	Cttn	Il6st	Numbl	Slc9a3r2
1700040L02Rik	Cux1	Inadl	Nup153	Smap2
2310022B05Rik	Cyb561	Ip6k3	Nupr1	Smc1b
2310028H24Rik	Cyp1b1	Iqck	Nynrin	Smg5
2410016O06Rik	Cyp46a1	Itch	Ogdh	Smurf1
2610507B11Rik	D18Ertd653e	Itfg3	Otud7b	Snd1
4931406P16Rik	Dach1	Itsn1	Pacs1	Sort1
4932438A13Rik	Deptor	Jrk	Pcbp3	Sox13
5430411K18Rik	Dip2c	Kcnab3	Pcx	Sp1
5730419I09Rik	Dnahc2	Kcnb1	Pdcd4	Spats2
9030418K01Rik	Dnaja4	Kcnk6	Pde10a	Spint1
9030625A04Rik	Dopey1	Kctd9	Pde2a	Src
9430020K01Rik	Dstyk	Khk	Pde3b	Ssh2
Acad10	Dusp10	Klc1	Pdk2	St3gal2
Acot11	Dym	Klf12	Pdpr	St7
Adam15	Dysf	Klf4	Peli3	Steap3
Add1	Efnb1	Klf9	Pex26	Ston1
Agfg2	Eif2ak4	Klhl21	Pex6	Sun1
Agtrap	Ell	Krba1	Pfkl	Syt3
AI314180	Epb4.1l1	Krt7	Pfkm	Tarm1
Ak1	Epb4.9	Laptm5	Pigq	Tbc1d2b
Ak8	Epn2	Layn	Pik3r3	Tbl1x
Aldh18a1	Espl1	Ldhd	Pik3r4	Tctn1
Als2cl	Espn	Ldlr	Pim3	Tctn2
Angel1	Etohi1	Lgals3bp	Pitpnc1	Tctn3
Ankrd13b	Etv3	Lgr4	Plcd1	Tek
Ano6	Ext2	LOC100046899	Pld1	Tep1
Apln	F10	LOC100504473	Plekha6	Tha1
Arhgef12	Fam102a	LOC100504500	Plekhf1	Thada

Arl4c	Fam117a	LOC100505126	Plekhg3	Thsd4
Arvcf	Fam13a	LOC100505278	Podxl2	Tmcc2
Asah2	Fam167a	Lpar5	Pofut1	Tmem190
Asb10	Fam20a	Lpin1	Polr1a	Tmem20
Asgr2	Fam43a	Lrig3	Pou6f1	Tmem201
Ash1l	Fam53b	Lrp6	Prkcb	Tnfrsf1a
Atm	Fam73b	Lrrc1	Prkdc	Tnik
Atp8a1	Fam92a	Lrrc36	Ptger4	Tns4
Axin2	Fbxl20	Lrrc66	Ptpn13	Tom1l2
Bank1	Fbxo40	Lrrc8b	Ptpn3	Tprn
Baz2b	Fbxw7	Luzp1	Ptprh	Traf4
Bbx	Fchsd2	Lyst	Pygl	Trim30c
BC016423	Fcrlb	Magi1	Pyroxd2	Trim46
BC051628	Fermt3	Man2b1	Rab11fip3	Trim65
BC057079	Fgd6	Man2b2	Rab11fip5	Trp53inp2
BC068157	Fgf11	Mansc1	Rab44	Trpv4
Bcas3	Fgfrl1	Map3k6	Rad54b	Tshz1
Bhlhe40	Fgr	Mapk11	Ralgapa2	Tspan18
Bmp2	Ficd	Mavs	Ralgps1	Txndc11
Bmpr1a	Fkbp9	Mbd5	Rarb	Ubr1
Braf	Foxc2	Meis3	Rarg	Uggt1
Brwd1	Foxo1	Mertk	Rassf5	Unc5b
Btbd2	Frk	Mfsd7a	Rbpms	Urgcp
Btg2	Frmd4b	Mga	Reep6	Utp20
C1qc	Fto	Mical1	Rfx2	Vash1
C2cd3	Fzd4	Mical2	Rhoq	Vgll4
C330019G07Rik	Gaa	Micalcl	Rhpn2	Vopp1
C85492	Gabpb2	Mlec	Ric3	Vps13a
Cables1	Gemin5	Msi2	Rin1	Wfikkn2
Cacna1c	Gm16515	Mtmr4	Rnf216	Wipi1
Cbx5	Gm608	Mtr	Rp1	Wrn
Cbx7	Gng7	Mtss1	Sbf2	Zadh2
Cc2d2a	Gnmt	Mxd4	Sec14l1	Zbed3
Ccdc19	Golim4	Myo6	Sec24a	Zbtb20
Ccdc50	Gphn	Myo9a	Sec24d	Zbtb38
Ccrn4l	Gprc5a	Nacc2	Sec31a	Zbtb46
Cd209c	Gprc5b	Nbas	Sel1l	Zc3h12a
Cd5l	Grk5	ND2	Sema4b	Zc3h7b
Cdc42bpa	Grk6	ND4	Sema6b	Zcchc11
Cds2	Gtf2i	ND5	Sep-11	Zdhhc14
Cdsn	H6pd	ND6	Setx	Zfp106
Celsr2	Hdlbp	Nedd4	Sf3b3	Zfp192
Cgnl1	Heatr5a	Neo1	Sh2d4b	Zfp219
Chn1	Heatr5b	Net1	Sh3bp5l	Zfp395

Chst13	Hhipl1	Nf1	Shpk	Zfp629
Clcn5	Hk1	Nfe2	Slc10a6	Zfp652
Col20a1	Homer3	Nipal3	Slc25a23	Zfyve20
Copz2	Hyou1	Nol6	Slc25a42	

		AAM up 5		
0610007P08Rik	Cdk5rap3	Glul	Ndufab1	Sirt5
0610009O20Rik	Cdkl4	Gm10191	Ndufaf1	Slamf9
0610010O12Rik	Cela1	Gm11992	Ndufaf4	Slc10a3
0610031J06Rik	Cenpv	Gm13004	Ndufs1	Slc22a18
1110008L16Rik	Cep57	Gm19444	Nek1	Slc25a12
1110031I02Rik	Cept1	Gm4459	Neu3	Slc25a39
1190002H23Rik	Cfp	Gm5150	Nfs1	Slc30a4
1500012F01Rik	Cisd1	Gm6762	Nln	Slc35b1
1700026L06Rik	Clcn3	Gm7689	Nlrp10	Slc39a12
1700034H14Rik	Clec10a	Gm8759	Nop2	Slc39a8
1700037H04Rik	Clec2d	Gmppb	Nop58	Slc39a9
1810012P15Rik	Clec4a1	Gna11	Nphp3	Slc41a2
1810030O07Rik	Clec4a3	Gna15	Nradd	Slc46a1
1810033B17Rik	Clec4b1	Gnb5	Nsl1	Smad5
2010011I20Rik	Clk4	Gne	Nucb2	Smagp
2010106G01Rik	Clrn3	Gnl3	Nudt17	Smyd4
2200002D01Rik	Clybl	Golga1	Nudt19	Snap29
2210411K11Rik	Cmtm6	Golm1	Nudt5	Snap47
2310008H09Rik	Cnot6l	Gorasp1	Nudt8	Snx10
2310046O06Rik	Coasy	Gpd1l	Numb	Snx13
2410002O22Rik	Cog4	Gpx7	Oat	Snx22
2410066E13Rik	Cog5	Grinl1a	Ogfod1	Snx29
2610301B20Rik	Cog6	Gsg1	Olfm1	Socs5
2810055F11Rik	Comt	Gsta4	Optn	Spaca1
4732418C07Rik	Copg	Gstm5	Oxnad1	Spcs2
4833442J19Rik	Copz1	Gsto1	Oxsm	Sphk1
4921513D23Rik	Cpne3	Gtf2e1	P2ry1	Srgn
4930572J05Rik	Cpsf3	Gtf3c4	P2ry14	Srm
4933403F05Rik	Cpt2	Gyk	P2ry2	Srp72
4933421E11Rik	Cr1l	Hbegf	P4hb	Srxn1
6030446N20Rik	Crat	Heatr3	Paics	Ssr1
6330578E17Rik	Crebl2	Hebp2	Parn	Ssr3
6530418L21Rik	Creld2	Hectd2	Pcca	Ssr4
6720456H20Rik	Cry1	Herc4	Pcmtd2	St6gal1
A430107O13Rik	Cryl1	Herpud1	Pde8b	Stambp
A530054K11Rik	Cse1l	Hibadh	Pdha1	Stbd1
A630007B06Rik	Ctps	Hibch	Pdia3	Stim2
A830007P12Rik	Ctsa	Hist1h2bc	Pdia4	Stt3a

A930001N09Rik	Ctso	Hist1h2be	Pdia6	Stxbp3a
A930038C07Rik	Cxcl12	Hist1h2bg	Pdk1	Suox
Abcb8	Cxcl16	Hist2h2bb	Pdk3	Supt3h
Abce1	Cyb5d2	Hmg20a	Pdlim2	Supv3l1
Abcf3	Cyc1	Hn1l	Pdp2	Susd1
Abhd11	Cyld	Hpn	Pdxdc1	Syp
Abhd14a	Cysltr2	Hps5	Pecr	Tars2
Abhd14b	Cytip	Hpse	Pex11a	Tbc1d10a
Abhd4	D10Wsu102e	Hsf2	Pex12	Tbc1d23
Acad12	D15Ertd621e	Hsp90b1	Pfkfb4	Tbc1d8b
Acat1	D3Ertd751e	Hspa13	Pgm2	Tbccd1
Acer2	D630037F22Rik	Hspa5	Pgm3	Tbrg4
Acox1	Dapp1	Hspa9	Phf6	Tcfec
Acsl5	Dars2	Hyal2	Phkb	Tcn2
Adal	Dbndd2	Hyal3	Phlda3	Tcp11l2
Adam10	Dbt	I830077J02Rik	Phyh	Tef
Adc	Dcaf4	Iars2	Piga	Tfdp2
Adck4	Dcaf6	Ica1	Pign	Tfg
Add3	Dck	Ide	Pigz	Tgfbr2
Adhfe1	Dclre1b	Idh3a	Pitrm1	Thap11
Adipor2	Dcps	Ikbkg	Pja2	Ticam1
Adora2b	Dcun1d4	Il18r1	Pla1a	Timp1
Adora3	Ddost	Impad1	Pla2g15	Timp2
Adss	Dennd2c	Inpp5a	Pla2g16	Tiparp
Afg3l2	Dfna5	Ipmk	Pla2g5	Tlr4
Agl	Dgka	Ipo11	Plekha3	Tm7sf3
Agpat9	Dguok	Irak4	Plscr4	Tm9sf2
Ahcyl2	Dhcr7	Isoc1	Plvap	Tmed3
AI467606	Dhrs7	Ispd	Pm20d1	Tmem106a
Aifm1	Dhx40	Itgam	Pold4	Tmem129
Aifm2	Dkc1	Itgb1	Polr3gl	Tmem135
Ak2	Dlat	Kars	Pomt2	Tmem143
Ak3	Dnaja3	Kcnn1	Pon3	Tmem177
Akap1	Dnajb9	Kcnn4	Pop4	Tmem186
Akap7	Dnajc10	Kctd21	Pot1a	Tmem22
Alas1	Dnajc28	Kdelc1	Ppapdc1b	Tmem39a
Aldh4a1	Dnajc3	Kdelc2	Ppat	Tmem41a
Aldh6a1	Dnd1	Kdelr1	Ppfibp2	Tmem53
Aldoa	Donson	Kdelr2	Ppib	Tmem69
Aldoart1	Dpagt1	Kdelr3	Ppif	Tnfaip8
Alg12	Dse	Kif9	Ppm1k	Tnfrsf26
Alg2	Dtd1	Klf10	Pramef8	Tnfsf13b
Amacr	Dusp19	Klhdc1	Prdx3	Tomm70a
Ang	Dusp23	Klhl10	Prdx4	Trap1

Ang2	Dusp3	Klhl12	Prdx5	Trex1
Ankrd46	Ears2	Klhl22	Preb	Trim30a
Ankrd55	Echdc1	Klhl24	Prepl	Trim35
Ap4b1	Eci2	Klhl7	Prkd3	Trmt2b
Aplp2	Edem2	Klraq1	Prkrir	Trpm3
Appl2	Ednrb	Krt18	Prorsd1	Ttc15
Arf4	Eef1b2	Lancl1	Pros1	Ttc21b
Arfgap3	Egln3	Lap3	Prr13	Ttc23
Arfip1	Ehd1	Laptm4b	Prrg1	Ttc3
Arg1	Eif2ak1	Lbx2	Psme4	Ttc30b
Arhgef3	Eif3e	Letmd1	Pspc1	Tubb3
Armc8	Eif3f	Lipn	Ptp4a3	Txndc5
Armc9	Eif4a2	Lman1	Ptplad2	Txnip
Arsg	Eif4e3	Lmbr1	Ptpn4	Txnrd3
Asap2	Elmod2	LOC100045212	Ptpn7	Uap1
Asb13	Elp3	LOC100045796	Pxk	Ube2e2
Asnsd1	Eml2	LOC100046166	Pxmp4	Ube2t
Aste1	Eml6	LOC100046302	Qrsl1	Ube3c
Ate1	Emr1	LOC100504754	Qtrtd1	Ubxn2a
Atf5	Entpd1	LOC100504962	Rab32	Uhmk1
Atg13	Entpd4	LOC100505031	Rab33a	Uhrf1bp1l
Atp5a1	Entpd5	LOC100505081	Rab38	Uqcrh
Atp5b	Epm2a	LOC631033	Rab4a	Uso1
Atxn3	Eprs	LOC636901	Rac2	Usp3
Auh	Eps8	LOC636969	Rai14	Uty
Avil	Erap1	LOC638399	Ranbp6	Vkorc1
B230219D22Rik	Ergic1	Lpcat3	Rasgrp2	Vmac
B2m	Erlec1	Lrch3	Rasl10a	Vprbp
B3galtl	Erp44	Lrp8	Rassf4	Vps37a
Bag2	Esrra	Lrpprc	Rbl2	Vps53
Bag5	Esyt2	Lrrc34	Rcl1	Wdpcp
Baiap2	Etf1	Lrrc40	Rd3	Wdr3
Bbs12	Etfdh	Lysmd4	Rdh13	Wdr36
BC017612	Exd1	Magt1	Reep3	Wdr75
BC026585	Extl2	Man1a	Retsat	Xbp1
BC026590	Fahd2a	Manba	Rfk	Xpo7
Bcap29	Fam114a2	Manf	Rhoh	Xrcc6
Bcat2	Fam135a	Map1lc3a	Rint1	Yme1l1
Bcs1l	Fam162a	Mapk1	Ripk2	Ypel2
Bet1	Fam20b	Mapk9	Rlim	Zbtb6
Bfar	Fam63b	Mar-01	Rmnd5a	Zbtb8a
Bin1	Fam81a	Mar-07	Rnasel	Zdhhc2
Birc2	Fas	Mbtps1	Rnf144b	Zfp187
Bola3	Fastkd1	Mccc2	Rnf167	Zfp213

Btk	Fbp1	Mcm8	Rnf185	Zfp235
C030006K11Rik	Fbxo18	Mdh1	Rpa1	Zfp260
C130050O18Rik	Fbxo22	Mdh2	Rpl31	Zfp296
C1qa	Fech	Med22	Rpn1	Zfp322a
C1qb	Fem1c	Med27	Rpn2	Zfp324
C1qbp	Fert2	Meis1	Rpp14	Zfp397
Cacnb4	Fgf9	Mfsd5	Rprd1a	Zfp456
Cage1	Fkbp5	Mfsd9	Rpsa	Zfp518a
Calhm2	Fkbp7	Mgmt	Rras	Zfp560
Calr	Fktn	Mill2	Rras2	Zfp672
Camk2d	Fmo5	Mlh3	Rtn3	Zfp719
Car6	Foxred1	Mlph	Sacm1l	Zfp748
Casp1	Fpgt	Mmachc	Scamp1	Zfp809
Casp7	Ftsjd1	Mobkl1a	Scfd1	Zfp866
Catsper2	Fut8	Morc4	Scfd2	Zfp869
Ccbl1	Fyttd1	Mr1	Scrn3	Zfp87
Ccdc122	G3bp2	Mras	Sdf2l1	Zfp931
Ccdc69	G6pc3	Ms4a8a	Sdf4	Zfp939
Ccng1	G6pdx	Mst1	Sdha	Zfp943
Ccng2	Galnt7	Mthfd1	Sdr42e1	Zfp945
Ccpg1	Ganc	Mtm1	Sec22b	Zfp958
Cd164	Gars	Mttp	Sec23b	Zfp963
Cd300lg	Gatsl3	Mtus2	Selplg	Zfr
Cd80	Gba	Mut	Senp8	Zkscan1
Cd82	Gca	Mxi1	Sepsecs	Zkscan5
Cd86	Gclm	N4bp2l1	Serpinb9	Zkscan6
Cd9	Gda	Naglu	Sft2d2	Zmat1
Cdc14b	Ggcx	Naip2	Sfxn1	Znrd1as
Cdc40	Ghr	Nars2	Sgms1	Znrf2
Cdc42ep3	Glce	Nbr1	Sh3bgrl	Zrsr1
Cdk2ap2	Glipr2	Ndrg3	Sil1	Zyg11b
		AAM down 1		
1110007C09Rik	Clec1a	Gpr85	N4bp3	Sat1
2310022A10Rik	Cnn2	Gpr97	Ndrg1	Sbsn
2610018G03Rik	Coro1a	Gpx3	Nek10	Sdc1
4930402H24Rik	Coro1c	Gstm1	Nlrc3	Sestd1
4930503L19Rik	Coro2a	Gusb	Nr1h3	Sft2d1
Abcb1a	Cox6a2	Havcr2	Nrp1	Sgms2
Abcb1b	Creg2	Hey1	Nt5dc2	Sgpl1
Abcb4	Cryba4	Hgsnat	Ogfrl1	Sh2d1b1
Acvrl1	Crybb1	Hmga2	Osbpl8	Sh2d3c
Adap2	Cttnbp2nl	Hopx	P2ry6	Sh3bp5
Adrb2	Cx3cr1	Hsd11b1	Pabpc1l	Slc15a3

Adrbk2	Cxcr1	Htra1	Palm3	Slc16a10
AF251705	Cxcr2	Ifi27l2a	Papss2	Slc26a11
Ambp	Cyp27a1	Igf2bp2	Parvg	Slc38a4
Amz1	Cyp4f39	Il1f9	Pcyt1a	Slc38a7
Ankrd34a	D1Ertd622e	Insig1	Pde7b	Slc3a2
Ap1s3	Dlg3	Iqcd	Penk	Slc40a1
Apbb1	Dock10	Irf8	Pid1	Slc7a11
Apobec1	Dtx3	Itgb5	Pira1	Slc9a7
Apoc2	Dtx4	Kcne3	Pira11	Slc9a9
Apoe	E230008N13Rik	Kcnk12	Pira4	Slfn10-ps
Aprt	E2f1	Kdr	Pira6	Slfn5
Armcx2	Eepd1	Kif5c	Pira7	Smad7
Armcx6	Egr2	Klf6	Plagl1	Smpd3
Arpc5	Elovl6	Klk9	Plekho1	Snx27
Atp8b2	Emb	Lass4	Plin2	Soat1
Atp8b4	Endod1	Lat	Plk2	Sox4
B230120H23Rik	Eno2	Lat2	Plp2	Sparc
Bambi	Evl	Lcp1	Por	Sqle
Bcl2l1	Fads1	Ldlrap1	Prdm1	St3gal6
Bhlhe41	Fam105a	Lilra5	Prkch	St6galnac6
Bri3	Fam107b	Lilra6	Procr	St8sia4
Bri3bp	Fam129a	Lmna	Prokr1	Stat4
Bst1	Fam134b	Lmo2	Psd2	Stom
Btbd17	Fam71f2	LOC100504922	Psen2	Stx4a
Card11	Fam83f	LOC674895	Pstpip1	Tcfeb
Ccdc48	Fcgr1	Lonrf1	Ptafr	Tie1
Ccl2	Fcrl1	Lpar6	Ptchd1	Tle1
Ccl3	Fgf1	Lpl	Ptms	Tle6
Ccl4	Fkbp1a	Lrfn4	Ptpla	Tlr5
Ccl7	Fos	Lsr	Ptprd	Tlr7
Ccr2	Ftl1	Lst1	Rab15	Tmem158
Cd22	Ftl2	Lyl1	Rab37	Tnfrsf11a
Cd276	Gcnt2	Lyz2	Rab7	Tnfrsf13b
Cd300a	Gdpd5	Mapk3	Rap2a	Tpcn2
Cd300ld	Ggta1	Marcks	Rasa4	Tpd52
Cd300lf	Glis3	Masp1	Rassf2	Tpm4
Cd36	Glp2r	Matn2	Rftn1	Trpv2
Cdk14	Gm11710	Mdk	Rgs2	Tsc22d3
Cdk15	Gm11711	Mef2c	Rnf149	Ube2v1
Cdk5r1	Gm12854	Met	Rnft2	Usf2
Cdkn1c	Gm14548	Mfng	Rph3a	Usp2
Ceacam19	Gm15448	Mmp8	Rsph9	Vegfa
Cebpa	Gm15448	Mpeg1	S100a11	Vim
Cenpb	Gm6377	Ms4a14	S100a4	Vsig8

Chic1	Gnb4	Ms4a6b	S100a6	Zranb3
Chst11	Gpr157	Ms4a6c	S1pr2	
Clec12a	Gpr34	Mtap7d3	Samd4	
		AAM down 2		
1110004F10Rik	Dnajc30	H3f3c	Mrpl48	Rsl24d1
1110038F14Rik	Dnajc9	Haghl	Mrpl51	Rtp4
1700025G04Rik	Dok3	Hck	Mrps10	Sap18
1810037I17Rik	Dpysl2	Hcls1	Mrps21	Satb1
2410127L17Rik	Dr1	Hdgf	Mrps26	Scarb1
2500003M10Rik	Dusp16	Hif1a	Ms4a4c	Scel
2610318N02Rik	Dusp2	Hlcs	Msh5	Scml4
3110003A17Rik	Dynll1	Hmgb1	Mxd1	Sdc4
6530409C15Rik	E130303B06Rik	Hmgcr	Myd88	Sema4d
A630001G21Rik	Ebi3	Hmgn2	Myl12b	Senp3
Acaa1a	Edf1	Hmgn2-ps1	Nab1	Sep-06
Acbd6	Egln1	Hmgn5	Nab2	Setd3
Acot1	Eif1	Hnrnpa2b1	Nampt	Setd8
Acsl1	Eif2s1	Hnrnpab	Napsa	Sfpi1
Actr2	Eif5a	Hnrpdl	Ncaph2	Sgk1
Adamts6	Epb4.1l3	Hsh2d	Nck2	Siah2
Aen	F11r	Hspa8	Neurl3	Skil
Afmid	F3	Hspe1	Nfu1	Slamf8
Ager	Fam103a1	Htra4	Nhp2l1	Slbp
Aif1	Fam76a	Hus1b	Nrg1	Slc17a7
Aim1	Fam82a2	Id3	Nsfl1c	Slc1a5
Akirin1	Fbxl5	Ifi204	Nup93	Slc20a1
Akt3	Fbxw17	Ifi205	Oas1a	Slc25a10
Amica1	Fdft1	Ifi30	Obfc2a	Slc25a25
Anp32b	Fen1	Ifi44	Ogfr	Slc25a5
Aoah	Fnbp1l	Ifit1	Osbpl3	Slc44a2
Ap1ar	Fnbp4	Ifit2	Oscp1	Smox
Apitd1	Foxd2	Ifit3	Otud6b	Snf8
Aplp1	Foxn2	Ifitm2	P2ry13	Snrnp48
Arf1	Fpr1	Igj	P4ha1	Snrpa
Arhgap28	Fut7	Il10ra	Pa2g4	Snrpd1
Arhgef7	Galnt3	Il12rb1	Padi3	Sntb1
Arid3b	Gdi2	Il1b	Parp8	Snx18
Arl1	Glrx3	Il9r	Pcgf1	Socs6
Arl5a	Gm10094	Immp1l	Pck2	Sod1
Arl5c	Gm10146	Irak2	Pdcd5	Sord
Armc10	Gm10224	Irf7	Pde12	Sp100
Arpc1b	Gm10282	Itga1	Pdlim4	Spin1
Arrb2	Gm10713	Itpk1	Pes1	Spn

Atf1	Gm10767	Jarid2	Pfkfb3	Spop
Atp5f1	Gm11964	Kcna3	Pfkp	Spred1
Atp7a	Gm12657	Kctd3	Pgap1	Srsf10
AU040320	Gm12952	Klhdc3	Pgs1	Srsf3
Azi2	Gm13342	Kynu	Phf20	Ssbp1
B430306N03Rik	Gm13880	Lbr	Phlda1	Stard10
Bex6	Gm14277	Ldhb	Pilra	Stfa3
Bhlhb9	Gm14535	Lfng	Pisd	Stk19
Bmf	Gm14698	Lif	Pkp3	Stx2
Bzrap1	Gm15427	Limd2	Plbd1	Styx
C5ar1	Gm15466	Lipg	Plekha2	Supt4h1
Cacna2d4	Gm15776	Llph	Plekha5	Syt11
Cacnb3	Gm1862	Lman2l	Plk4	Taf10
Ccdc114	Gm19945	Lmbr1l	Plscr1	Tarbp2
Ccdc115	Gm2036	LOC100039181	Pmaip1	Tbk1
Ccdc12	Gm3258	LOC100044374	Pnkp	Tceb2
Ccdc88a	Gm3837	LOC100044750	Ppfia3	Terf1
Ccr5	Gm4199	LOC100045191	Ppfia4	Tgfbi
Ccrl2	Gm4332	LOC100045999	Pphln1	Thoc4
Cd180	Gm4356	LOC100046104	Ppm1m	Tifab
Cd2	Gm4581	LOC100046297	Ppp1cc	Timm17a
Cd300e	Gm4596	LOC100046650	Ppp1r14b	Tlr1
Cd4	Gm5161	LOC100047252	Prelid1	Tlr11
Cd69	Gm5450	LOC100048410	Psat1	Tlr6
Cd72	Gm5471	LOC100048847	Psma4	Tmeff1
Cdc34	Gm5514	LOC100503218	Psmb3	Tmem173
Cdk10	Gm5616	LOC100504743	Psmd2	Tmem41b
Cdk2	Gm5637	LOC100505005	Psmd3	Tmsb10
Cdk9	Gm5963	LOC100505147	Pten	Tnf
Cep135	Gm6115	LOC100505328	Ptgs2	Tnfaip3
Chmp1a	Gm6181	LOC631287	Ptma	Tnfrsf1b
Chst7	Gm6314	LOC631966	Ptplb	Tnfsf9
Cks2	Gm6421	LOC632329	Ptpn6	Tomm5
Clec4n	Gm6428	LOC634785	Ptprc	Tpi1
Clec5a	Gm6594	LOC637733	Qsox1	Tra2b
Clec7a	Gm6749	Lpcat4	Rab5a	Traf3ip3
Cpn1	Gm6750	Lrrc16a	Raet1c	Trappc6b
Cpne2	Gm7194	Lrrc42	Raet1e	Triobp
Cpsf7	Gm7227	Lrwd1	Ran	Tsg101
Cpt1a	Gm7589	Lsm7	Rangap1	Tspan13
Csda	Gm7931	Ltb	Rbbp5	Ttc5
Csnk1g3	Gm8069	Ltb4r1	Rbbp7	Ttll3
Ctnnd2	Gm9525	M6pr	Rbm17	Tubb2b
Ctsc	Gm9733	Maff	Rbm38	Twistnb

Ctu2	Gm9830	Mag	Rbm4b	Twsg1
Cwc15	Gm9840	Malt1	Rbm8a	Ube2d3
Cxcl14	Gnb3	Map3k8	Rbmxrt	Ube2j2
Cxcr3	Gng5	Map4k1	Rbx1	Uck2
Cyp4f16	Golga7	Map4k4	Rdh12	Ufm1
D11Wsu99e	Gpank1	Mapk1ip1l	Rel	Usf1
D3Bwg0562e	Gpc1	Mapkapk2	Rmnd1	Usp12
Daxx	Gpc2	Mar-05	Rnase6	Vangl2
Dazap1	Gpd2	Marcksl1	Rnd3	Vav1
Dbn1	Gpr141	Mdm1	Rnf126	Vcan
Dcaf15	Gpr65	Mdm2	Romo1	Was
Dcun1d1	Gps2	Med29	Rpl10a	Whsc1l1
Ddx19a	Gpsm3	Mef2a	Rpl15	Wnt6
Ddx50	Grb2	Megf9	Rpl18a	Wsb2
Dedd2	Gtf2a1	Metap2	Rps20	Ywhaz
Dek	Gxylt1	Mgst2	Rps21	Zbtb2
Dexi	H28	Minpp1	Rps26	Zc3h12d
Dgkg	H2afy	Mlf1ip	Rps9	Zc3h15
Dna2	H2afz	Mmel1	Rrad	Zdhhc21
Dnahc17	H3f3a	Mmp14	Rrs1	Zfand5
Dnajc2	H3f3b	Mrpl21	Rsad2	Zfp28
		AAM down 3		
1200009I06Rik	Ccdc99	Gm12166	Necap2	Sc5d
2010002N04Rik	Cd101	Gm12693	Nfkbib	Scand1
2010007H12Rik	Cd14	Gm12844	Nme1	Sdcbp
2010111I01Rik	Cd300lb	Gm13540	Nme2	Selm
2210404J11Rik	Cd302	Gm13736	Nras	Sep-10
2310004N24Rik	Cd99l2	Gm14680	Nt5c	Sepw1
2310011J03Rik	Cdc37	Gm16517	Nt5c2	Serinc3
2310035K24Rik	Cdk20	Gm1673	Ntpcr	Sertad1
2310036O22Rik	Cdkn2aipnl	Gm4705	Nubp1	Sft2d1
2400001E08Rik	Cenpa	Gm6334	Nus1	Shfm1
2410001C21Rik	Cgrrf1	Gm6941	Oaf	Siva1
2410004B18Rik	Chaf1b	Gm885	Ocel1	Slc16a7
2410042D21Rik	Chmp4b	Gng11	ORF19	Slc31a1
2610029G23Rik	Cklf	Gpt	Ostm1	Slc9a3r1
2900073G15Rik	Cldnd1	Gpx1	P2rx4	Slfn2
4930412F15Rik	Clec4a2	Gtf2h2	Pafah1b3	Sms
4930420K17Rik	Clec4d	H2afj	Pdap1	Snrpa1
5430437P03Rik	Clptm1	Haao	Pdcd2l	Snrpf
6330409N04Rik	Clta	Hebp1	Pdcl3	Snx12
8430410A17Rik	Cmpk1	Hes6	Pepd	Snx24
9430023L20Rik	Cndp2	Hmox2	Pgk1	Snx3

A630033H20Rik	Cno	Hras1	Pgrmc1	Snx8
Abcg2	Cox4nb	Hsd3b7	Phf23	Spg20
Abcg3	Crlf3	Hsdl1	Pkib	Srp9
Abhd12	Csrp1	Hspbap1	Pkig	Ssbp2
Abt1	Csrp2	Ifi202b	Plau	Steap4
Acss2	Ctnnbl1	Ifitm3	Plekha1	Stk17b
Actl6a	Ctsb	Ifngr1	Plekhj1	Stx11
Ada	Ctss	Ifngr2	Plin3	Stx3
Adam17	Ctsz	Ifrd1	Pls3	Suds3
Adipor1	Cuedc2	Igfbp6	Pno1	Sumf1
Adssl1	Cyb5r1	Igsf6	Pomp	Susd3
Agpat2	Cyp39a1	Il1r2	Ppa1	Svip
Aim2	Cyp4v3	Ilvbl	Ppm1a	Sypl
Alcam	Cyp51	Ing4	Prim1	Tbcc
Aldh3b1	D230037D09Rik	Ino80b	Prmt2	Tctex1d2
Aldh9a1	Daglb	Isy1	Prrg2	Tgfbr1
Alkbh7	Dcxr	Jdp2	Psma6	Tifa
Anxa2	Dffa	Kank3	Psmb6	Timm13
Ap3s1	Dhps	Kcmf1	Psmd8	Tmco3
Arf5	Dhrs1	Kitl	Psmd9	Tmem120a
Arf6	Dhrs3	Klhl9	Psmg4	Tmem144
Arhgap24	Dirc2	Lage3	Ptger2	Tmem165
Arhgap9	Dmkn	Lair1	Ptpro	Tmem19
Arl10	Dnajc14	Lass5	Rab1	Tmem192
Arl3	Dnalc4	Leprotl1	Rab13	Tmem205
Arl6ip4	Dnase1l1	Lin7c	Rab21	Tmem38b
Arpc2	Dok1	Lmo4	Rabac1	Tmem86a
Arpc3	Dpy19l4	LOC100038947	Rala	Tnfaip8l2
Asl	Echdc3	LOC100044068	Ralb	Tnfrsf12a
Asna1	Efhd2	LOC100048137	Raly	Tnfrsf8
Atp1b3	Eif1ax	LOC675577	Ranbp1	Tnni2
Atp5k	Eif2b2	Lsm4	Rap2b	Tor2a
Atp6ap2	Eif4a3	Lsp1	Rbfa	Tppp3
Atp6v0b	Eif5a2	Lta4h	Rbm7	Tpst2
Atp6v1b2	Eif6	Ly86	Rdbp	Tradd
Atp6v1f	Emp1	Maged2	Rdh11	Trappc3
Atp6v1g1	Emp3	Mapkapk5	Rdh5	Trim13
Atxn10	Epsti1	Mbd2	Reep5	Tspan4
Aup1	Erh	Mea1	Rexo2	Tspan5
B020018G12Rik	Exosc3	Med10	Rftn2	Tst
B3gat3	Fabp7	Metap1	Rhoa	Ttc12
B3gnt2	Faim	Metrnl	Rilpl2	Tuba4a
B4galt7	Fam105b	Mfsd1	Rnaseh2b	Tubb2a
B630005N14Rik	Fam158a	Mfsd11	Rnaseh2c	Tusc3

Bax	Fam173a	Mgst3	Rnf115	Twf2
BC003266	Fam73a	Mlf2	Rnf130	Ubtd1
BC031181	Fam96b	Mocos	Rnf180	Ubxn6
BC096441	Fbxo30	Mpdu1	Rnh1	Ugdh
Bloc1s2	Fibp	Mpnd	Rpl34	Ugt1a6a
Bola2	Fnip1	Mpp1	Rpl34-ps1	Ugt1a6b
Brms1	Fth1	Mpst	Rps19	Umps
Btbd10	Fuca2	Mrpl19	Rps19-ps4	Vapb
C3ar1	Fxn	Mrps18c	Rps19-ps6	Vdac2
Cab39	Gabarapl1	Ms4a6d	Rps19bp1	Vma21
Cacybp	Galnt9	Msr1	Rraga	Vmp1
Calm1	Gch1	Msrb2	Rragc	Wdfy1
Calm2	Get4	Mthfd2l	Rsu1	Wdr1
Calm3	Gins2	Mtpn	Rwdd1	Wsb1
Camk1	Gipc1	Myoz1	Ryr1	Ydjc
Capg	Gm10086	Naa20	S100a9	Ypel3
Car5b	Gm10094	Nasp	Sap18	Zfp771
Ccdc124	Gm10131	Ndufb11	Sap30	

		AAM down 4		
1110012D08Rik	Dnajb14	Jag2	Plxnb3	Slc2a3
1110014N23Rik	Dok2	Klhdc4	Pnpla7	Slc30a1
1190002N15Rik	Dtx2	Kras	Poc1b	Slc39a3
1600027N09Rik	Dusp5	Lass2	Ppap2c	Slc41a3
1700019L03Rik	Dusp6	Lca5	Ppp1r12c	Snrnp35
2510009E07Rik	Dvl2	Lemd2	Ppp2r5a	Snrpc
4930471M23Rik	Dync1li2	Lilrb3	Pqlc2	Snx20
Acbd5	Eaf1	LOC100045021	Prdx6	Snx32
Acp2	Eea1	LOC100047702	Prkar2b	Specc1
Actn4	Eed	LOC100504508	Prkcd	Spnb4
Adck5	Efcab4b	LOC677654	Psap	Sptlc2
Ado	Ehd4	Loxl3	Psmc3ip	Ssh3
Agap3	Eif4ebp1	Lpcat1	Ptpra	Stard3
Agpat6	Elk1	Lrrfip1	Pwwp2b	Stat6
Ahrr	Elk3	Lrrfip2	Pygb	Stk11
AI846148	Elovl1	Maf	Rab20	Stk25
AK010878	Epb4.1l2	Mafk	Rab22a	Stk30
Ankrd13a	Ezr	Map2k1	Rab24	Stub1
Ankrd13d	Fam110b	Map2k2	Rab34	Swap70
Ankrd40	Fam134a	Mapk6	Rabgef1	Taf6l
Arl8a	Fam187b	Mapre2	Rabggta	Tal1
Asb6	Fam3c	Mbd3	Ranbp3	Tatdn2
Asph	Fam40a	Mcl1	Rapgef6	Tbc1d14
Atg16l2	Fam46c	Mcoln1	Rbms1	Tbc1d22a

Baiap2l1	Fam50a	Mettl17	Rcbtb1	Tcf4
Banp	Fam89a	Mex3b	Rcor3	Tex264
Bbc3	Fance	Mex3d	Reep4	Tgif1
BC085271	Fbxw4	Mfsd6	Rfxap	Thap3
Bcl7c	Fem1b	Mfsd7b	Rgag4	Thsd1
Cap1	Fkrp	Mgrn1	Rhog	Tjp3
Capn2	Flcn	Mid1ip1	Rnaseh1	Tlr13
Cbfb	Frat2	Mknk1	Rnasek	Tm6sf1
Ccdc125	Fut4	Mkrn1	Rnf166	Tmem109
Ccdc163	Fxyd5	Mmp19	Rnf220	Tmem181a
Cd2bp2	G0s2	Mobkl2a	Rpl26	Tmem43
Cd300lh	Gab3	Mtdh	Rrp1	Tmem66
Cd44	Galns	Mtmr11	Rtn4	Tmem8
Cdc42se2	Galt	Naa50	Rufy1	Tollip
Cdk17	Gas8	Ncf1	Rufy3	Trim36
Cds1	Ghdc	Neu1	Samd8	Trim45
Cebpg	Glud1	Neurl2	Sap30l	Trip10
Chfr	Gm9531	Nfe2l2	Scamp3	Trp53i13
Chka	Gnl1	Nfkbil1	Scamp4	Ttl
Chmp6	Gpatch3	Nfya	Scpep1	Ttpal
Cited2	Gpr146	Nkiras2	Scyl2	Tubb4
Clk3	Gtf2f1	Npepl1	Sec62	Tubb6
Cln3	Habp4	Nr2f6	Senp2	Tuft1
Cln8	Herc3	Nudc	Set	Tulp3
Cnr2	Hhex	Nup50	Sfmbt1	Tusc1
Col4a3bp	Hk3	Pabpn1	Sgta	Ube2j1
Coq10a	Hmg20b	Paqr7	Sh3bgrl3	Ube2m
Crip2	Hnrpll	Parp1	Sh3gl1	Uhrf2
Csnk1e	Hs1bp3	Pbx2	Shf	Upf3a
Ctnnbip1	Hspbp1	Pcyox1	Shisa5	Usp4
Ctsd	Hvcn1	Pcyt2	Sirt1	Vsig10l
Cwc27	Ier2	Pde5a	Sirt2	Wbp4
Cyhr1	Ifnar1	Pde8a	Slc11a2	Wdr11
Cyth1	Igsf8	Pfn1	Slc15a4	Zbtb17
Cyth2	Il11ra1	Pgm2l1	Slc16a6	Zbtb22
Cyth4	Il16	Pgp	Slc22a4	Zfp238
Dcun1d3	Il3ra	Phka2	Slc24a6	Zfp444
Dda1	Ing1	Pi4k2a	Slc25a24	Zfp787
Ddit3	Inpp1	Pik3cb	Slc25a28	Zfp821
Def6	Ipo5	Pip4k2a	Slc25a43	Zfp90
Dgke	Irf5	Pld4	Slc29a1	Zfyve9
		AAM down 5		
1500002O20Rik	Clip1	Gtpbp2	Myo1g	Smtn

1700020C07Rik	Clip2	Gtpbp6	Nr4a2	Snrk
2310014H01Rik	Clk2	Hdac5	Nt5dc3	Snrnp70
2310035C23Rik	Cnksr3	Herpud2	ORF61	Sntb2
2410131K14Rik	Cnnm3	Hlx	Otud1	Sphk2
2510012J08Rik	Cnnm4	Hmha1	Pacs2	Srebf2
2510039O18Rik	Crebzf	Hnrnpm	Pacsin2	Srf
5031439G07Rik	Csnk2a2	Hnrnpul2	Palm	Srrt
Acvr1b	Ctdp1	Hps6	Papd5	Stim1
Adamts10	Cxxc1	Hs6st1	Pbxip1	Sun2
Adat3	Ddx23	Igsf9	Pdgfb	Sykb
Adcy7	Ddx54	Incenp	Phldb3	Taf1c
Aebp1	Dennd1c	Inpp4a	Pias4	Tbc1d1
Aebp2	Dgkh	Inppl1	Pik3r5	Tbc1d10b
Afap1	Dgkq	Iqgap1	Pip4k2c	Tbc1d17
Agpat1	Dhx16	Irf2bp1	Pip5k1c	Tbc1d8
Ano8	Dip2a	Itga5	Pkn1	Tchp
Anxa11	Dusp7	Itprip	Pld2	Tesk1
Ap2a1	Dusp8	Itsn2	Plekhm1	Tfip11
Arhgef40	Dvl1	Jag1	Plekho2	Tgfb1
Arih1	Eif3a	Jund	Plxna3	Tjap1
Arrdc2	Epn1	Kcnab2	Ppm1h	Tmem115
Atp10a	Ets2	Kif1c	Pvrl4	Tmem164
Atp13a2	Exoc8	Klhdc8b	Rab43	Tmem184b
Auts2	Fam160a2	Kri1	Rabep2	Tob2
Axin1	Fam20c	Lasp1	Ralbp1	Trim25
B3gnt3	Fbxl19	Lats2	Rasa3	Trim8
B3gnt7	Fchsd1	Lemd3	Rgl2	Trpm4
B4galnt1	Frmd8	Limd1	Rgmb	Tsc22d1
Bach1	Furin	Lmf2	Rhbdl1	Unc45a
Bach2	Fxr2	Lnx2	Rnf31	Unk
BC005537	Fzd5	LOC677502	Rps6ka1	Usp10
Bcl6	Fzd7	Lrch4	Runx1	Vac14
Bicd2	Gab1	Lrrc45	Safb	Yy1
Brf1	Galnt10	Lrrc8d	Samd1	Zbtb42
Brpf1	Gan	Mafb	Sap25	Zbtb7b
C330006K01Rik	Gas2l1	Maml3	Sart1	Zcchc24
Cacnb1	Gcc1	Map3k11	Scly	Zeb2
Capn1	Gcfc1	Map3k3	Scyl1	Zfp276
Cc2d1a	Gdf15	Mapk7	Sepn1	Zfp295
Cc2d1b	Gga1	Mapk8ip3	Sh2b3	Zfp316
Ccdc142	Git2	Mesdc1	Sirpa	Zfp408
Ccdc9	Gm2260	Mfhas1	Slc12a4	Zfp608
Cd97	Gm2274	Mgat4b	Slc12a9	Zfp668
Cdk11b	Gm962	Mier2	Slc16a14	Zfyve1

Cep170	Gpr124	Mtap4	Slc26a10	Zgpat
Chkb	Gramd1b	Mtap7d1	Slc38a2	Zrsr2
Ckap4	Gramd4	Mum1	Slc4a2	Zswim4
Clasrp	Gripap1	Myh10	Slc6a6	
Clcn7	Gtf2ird2	Myo1c	Smo	

Supplementary Table A1-ST-4. List of gene symbol IDs for genes in AAMΦ-up and AAMΦ-down clusters identified in Chapter 2. As described in the methods section, all differentially expressed genes between WT-NeMΦ and either IL4Rα$^{-/-}$-NeMΦ or WT-ThioMΦ were clustered using heierarchical agglomerative clustering and subdivided into 20 clusters using Ward's method. The resulting gene expression profiles were assessed and clusters deemed to represent AAMΦ-up and AAMΦ-down gene expression profiles were identified. Gene symbols in this table are for the genes in those clusters.

APPENDIX 2 – SOFTWARE USED FOR TSS-PREDICTOR

Software/Annotation	Version
TopHat	2.0.0
Cufflinks	2.0.2
Ensembl	67
BEDTools	2.17.0
SAMTools	0.1.8

Table A2-1. List of software and versions used in Chapter 3.

APPENDIX 3 – HELP DOCUMENTATION FOR TSS-PREDICTOR SOFTWARE

TSS-Predictor

Version 1.0 by Graham Thomas

Contents

1.0 Introduction

TSS-Predictor has been developed to identify promoter usage for genes in RNA-Seq experiments. It is designed to be used when both a high quality reference genome and annotation set are available, for example human or mouse. The output of TSS-Predictor is a BED file containing predicted transcription start sites for each reference gene, these may then be used for downstream applications such as *cis*-regulatory analysis.

TSS-Predictor is written in perl and is available at GitHub `https://github.com/GrahamThomas/TSS-Predictor` and as an online tool at Galaxy (Not Yet!!!) and GeneProf (Not Yet!!!).

2.0. Getting Started

2.1. Dependencies

In order to use TSS-Predictor you must have working versions of the following in your $PATH:-

1. A short read mapper compatible with Cufflinks. I use TopHat `http://tophat.cbcb.umd.edu/`.
2. Cufflinks `http://cufflinks.cbcb.umd.edu/`
3. BEDTools `http://code.google.com/p/bedtools/` and
4. R, with the libraries ggplot2 and reshape2 installed `http://cran.r-project.org/`

To follow this guide you will need to merge and sort some BAM files. For this you will also need SAMtools `http://samtools.sourceforge.net/`.

2.2. Installation

TSS-Predictor is implemented in Perl and R. To get up and running all you need to do is:-

1. Add TSS-Predictor/bin to your $PATH variable:-
   $PATH=$PATH:/<Path to installation directory>/TSS-Predictor/bin
2. And give executable permission to All scripts in the directory.
   cd <Path to installation directory>/TSS-Predictor/bin
   chmod +x *

3.0. Using TSS-Predictor

3.1. Preparing your data

Below is a step-by-step guide on how to use TSS-Predictor beginning with raw illumina reads. In this example we use 51 base paired-end Illumina RNA-Seq data from a macrophage transcriptomics project (REF). Raw data and BAM files for this project are available from the SRA (ACCESSION... `http://www.ebi.ac.uk/ena/`). All of the intermediate files generated in the working example are available at our FTP site .

```
cd ~/TSS-Predictor-Example/RawData/
ls
4Ne_1_F.fastq  4Ne_3_R.fastq  4TG_3_F.fastq  BNe_2_R.fastq  BTG_2_F.fastq
4Ne_1_R.fastq  4TG_1_F.fastq  4TG_3_R.fastq  BNe_3_F.fastq  BTG_2_R.fastq
4Ne_2_F.fastq  4TG_1_R.fastq  BNe_1_F.fastq  BNe_3_R.fastq  BTG_3_F.fastq
4Ne_2_R.fastq  4TG_2_F.fastq  BNe_1_R.fastq  BTG_1_F.fastq  BTG_3_R.fastq
4Ne_3_F.fastq  4TG_2_R.fastq  BNe_2_F.fastq  BTG_1_R.fastq
```

A pre-built Bowtie index for mouse (mm9) was downloaded from `http://bowtie-bio.sourceforge.net/index.shtml`.

```
cd ~/TSS-Predictor-Example/BowtieIndexes/
wget ftp://ftp.cbcb.umd.edu/pub/data/bowtie_indexes/mm9.ebwt.zip
unzip mm9.zip
```

Reconstitute a reference genome from the bowtie index.

```
cd ~/TSS-Predictor-Example/BowtieIndexes/
bowtie-inspect mm9 > mm9.fa
```

An Ensembl GTF file containing reference annotations was downloaded from the Ensembl FTP site `ftp://ftp.ensembl.org/pub/release-67/gtf/mus_musculus/`. Ensembl annotations are preferable as, unlike UCSC annotations, a many-to-one relationship exists between gene 'gene_id' and 'transcript_id' fields of the GTF file. This relationship is required to combine per-transcript expression levels to per-gene expression levels.

```
cd ~/TSS-Predictor-Example/Annotations/
wget ftp://ftp.ensembl.org/pub/release-67/gtf/\
mus_musculus/Mus_musculus.NCBIM37.67.gtf.gz
```

Ensembl chromosomes need converting from NCBIM37 to mm9 style. This can be done with Ensembl2UCSC.pl. This script prints the annotations in the same order as the FASTA headers in the supplied genome and removes annotations belonging to the 'random' chromosomes. An alternative approach would be to map reads directly to the Ensembl genome build, thereby ommitting this step.

```
Ensembl2UCSC.pl -r ../BowtieIndexes/mm9.fa -e Mus_musculus.NCBIM37.67.gtf \
-o Mus_musculus.mm9.67.gtf
```

3.2. Mapping reads to the reference genome

Map reads to the reference genome using TopHat (version v2.0.0, Bowtie version 0.1.18.0 were used in this example). Multi-mapping reads are allowed by using the default parameters. Multi-mapped reads are important as they facilitate the assembly of transcripts containing repetitive sequence. In downstream analysis it may be wise to disregard transcripts containing a large proportion of multi-mapped reads as they may be unreliable.

Transcription start site prediction is more effective with greater sequencing depth. My preferred strategy is to map reads for each lane individually, and then merge the alignments prior to transcript assembly with Cufflinks.

Map the reads with TopHat:-

```
cd ~/TSS-Predictor-Example/TopHat
tophat -p 5 -o KNe_1_multi_out ../BowtieIndexes/mm9 \
../RawData/4Ne_1_F.fastq ../RawData/4Ne_1_R.fastq
...
tophat -p 5 -o BTG_3_multi_out ../BowtieIndexes/mm9 \
../RawData/BTG_3_F.fastq ../RawData/BTG_3_R.fastq
```

Merge and sort the alignments. I use SAMtools http://samtools.sourceforge.net/:-

```
cd ~/TSS-Predictor-Example/TopHat/

###Reconstitute a SAM header from one of the TopHat alignments
samtools view -H BNe_1_multi_out/accepted_hits.bam > mm9.header.sam


samtools cat -h mm9.header.sam -o Macrophage.cat.bam \
BNe_1_multi_out/accepted_hits.bam BNe_2_multi_out/accepted_hits.bam \
BNe_3_multi_out/accepted_hits.bam KNe_1_multi_out/accepted_hits.bam \
KNe_2_multi_out/accepted_hits.bam KNe_3_multi_out/accepted_hits.bam \
BTG_1_multi_out/accepted_hits.bam BTG_2_multi_out/accepted_hits.bam \
BTG_3_multi_out/accepted_hits.bam KTG_1_multi_out/accepted_hits.bam \
KTG_2_multi_out/accepted_hits.bam KTG_3_multi_out/accepted_hits.bam

samtools sort Macrophage.cat.bam Macrophage.sorted
```

3.3. Running Cufflinks and Cuffcompare

Assemble the mapped reads into transcribed fragments 'transfrags' using Cufflinks (here we have used version 2.0.2). In this example we have not provided reference annotations to guide transcript assembly. You may do so if you wish.

```
cd ~/TSS-Predictor-Example
cufflinks -p 8 -L Macrophage-Cufflinks -o Cufflinks TopHat/Macrophage.sorted.bam
```

Assign the transfrags to reference gene models using Cuffcompare

```
cd ~/TSS-Predictor-Example/Cufflinks
cuffcompare -r ../Annotations/Mus_musculus.mm9.67.sorted.gtf -M \
-s ../mm9.fa -o mm9-EnsGene transcripts.gtf
```

3.4. Generating a .coverage file with GetCoverage.pl

GetCoverage.pl makes use of BEDTools to calculate the per-base read coverage at the genomic locations reported in the combined.gtf file produced by Cuffcompare. The script takes the following arguments:-

```
GetCoverage.pl -c <cuffcompare.combined.gtf> -b <mapped_reads.bam> \
-g <.genome file> -o <output>
```

Options	Description
-c	Path to the cuffcompare combined.gtf file
-b	Path to the mapped reads BAM file
-g	Path to the .genome file
-o	Name of the output .coverage file

The .genome file is a two column tab-delimited file stating each chromosome/scaffold name and it's size in base pairs. BEDTools is shipped with .genome files for mouse and human reference genomes. If you are working with a different species you will have to create one yourself, see the BEDTools manual (`http://code.google.com/p/bedtools/`) for more information.

Making the .coverage file can take some time. We make the .coverage file for this example as follows:-

```
cd ~/TSS-Predictor-Example/CoverageFiles
cp <Path To BEDTools>/genomes/mouse.mm9.genome .

GetCoverage.pl -c ../Cufflinks/mm9-EnsGene.combined.gtf \
-b ../TopHat/Macrophage.sorted.bam -g mouse.mm9.genome \
-o mm9-EnsGene.coverage
```

3.5. Using TSS-Optimizer.pl to identify optimal cutoffs

TSS-Optimizer.pl is a wrapper script for the main TSS prediction algorithm. For a description of how the main algorithm works see section 5 or the supplementary information of (REF MY PAPER). TSS-Optimizer.pl takes one chromosome, defined by the user and predicts transcription start sites over a range of cutoff fractions. TSS-Optimizer.pl reports, for each gene at each cutoff fraction, the distance between the predicted TSS and nearest annotated TSS for the same gene. TSS-Optimizer also produces a graph showing the median absolute deviation (MAD) between the predicted and nearest annotated transcription start sites at each cutoff fraction. We choose the cutoff fraction that gives the lowest MAD for use with TSS-Predictor.

```
TSS-Optimizer.pl -a <Annotations.gtf> -cd <Cufflinks_directory> \
-cc <Cuffcompare_prefix> -t <tmap_file> -co <Coverage_file> -ch <chromosome>
```

Options	Default	Description
-a	[]	The main reference annotation file i.e. the one used with Cuffcompare
-cd	[]	Path to the Cufflinks directory
-cc	[]	Prefix used when running Cuffcompare
-t	[]	Path to the Cuffcompare transcripts.gtf.tmap file
-co	[]	Path to the .coverage file produced by GetCoverage.pl
-ch	[]	Name of chromosome for optimization
-s	[0.01]	Coverage cutoff fraction to begin trimming
-e	[0.20]	Coverage cutoff fraction to end trimming
-i	[0.01]	Increment, or step size, between lower and upper coverage cutoff fractions
-o	[./Optimizer]	Output directory (this will be suffixed with the selected chromosome)

Running TSS-Optimizer for our data set is simple. We use the following command:-

```
cd ~/TSS-Predictor-Example

TSS-Optimizer.pl -a Annotations/Mus_musculus.mm9.67.gtf \
-cd Cufflinks/ -cc mm9-EnsGene -co CoverageFiles/mm9-EnsGene.coverage \
-ch chr1 -o Optimizer-EnsGene -t Cufflinks/mm9-EnsGene.transcripts.gtf.tmap
```

The results of this analysis can be found in the directory ./Optimizer-EnsGene-chr1/ as the Optimizer script concatenates the chromosome used to the specified output directory. TSS-Optimizer produces a number of files including a chromosome.coverage file, which is simply a reduced version of the .coverage file. TSS-Optimizer.pl produces all of the files produced by TSS-Predictor.pl (see section 4). Multiple copies of isotigs.trimmed.gtf, distances.txt and tss.prediction.bed are produced, one for each cutoff fraction interrogated.

The most informative output of TSS-Optimizer.pl is the graph Optimizer-MADPlot.jpg (Figure 1). This barchart shows the MAD between predicted TSSs and the nearest annotated TSS for each gene over all cutoff fractions assessed. The cutoff fraction that gives the lowest MAD is shown in red, this should be given to TSS-Predictor.pl to identify TSS usage genome wide. The actual MAD calculated by TSS-Optimizer contains no post-filtering (see section 3.7) and is typically much greater than the value obtained in the final predictions.

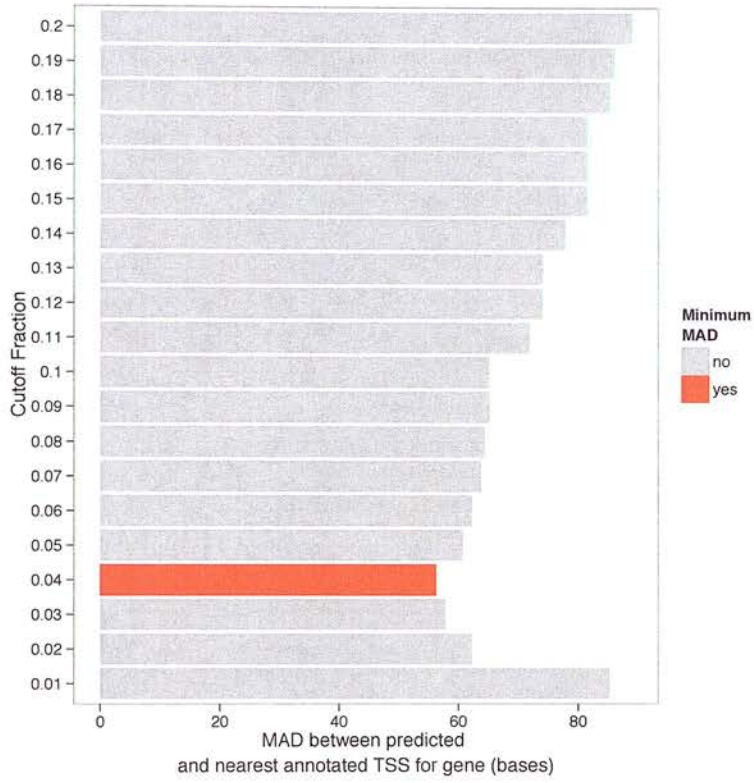


Figure 1: Barchart showing the median absolute deviation between predicted transcription start sites and the nearest annotated TSS for the gene over the coverage cutoff fractions assessed by TSS-Optimizer.pl. The cutoff fraction with the minimum MAD is shown in red (0.04).

3.6. Running TSS-Predictor.pl

In this next step we run TSS-Predictor to predict transcription start sites genome wide. This script attempts to identify the major transcription start site for each reference gene assigned to a a Cufflinks gene_id by cuffcompare. The output will include erroneous TSS calls due to a number of factors, these include incorrect transcript assembly and selection of the wrong Cufflinks transfrag for trimming. Strategies exist to get high confidence TSS predictions, see next section.

```
TSS-Predictor.pl -a <Annotations.gtf> -cd <Cufflinks_directory> \
-cc <Cuffcompare_prefix> -t <tmap_file> -co <Coverage_file>
```

Options	Default	Description
-a	[]	The main reference annotation file i.e. the one used with Cuffcompare
-cd	[]	Path to the Cufflinks directory
-cc	[]	Cuffcompare prefix
-t	[]	Path to Cuffcompare transcripts.gtf.tmap
-co	[]	Path to the .coverage file produced by GetCoverage.pl
-f	[0.1]	Cutoff fraction (of expected per-base coverage) to trim annotations
-u	[1]	Number of bases upstream of predicted TSS to report
-d	[0]	Number of bases downstream of predicted TSS to report

Next, predict TSS usage genome-wide with a coverage cutoff of 0.04:-

```
cd ~/TSS-Predictor-Example

TSS-Predictor.pl -f 0.04 -a Annotations/Mus_musculus.mm9.67.gtf \
-cd Cufflinks/ -cc mm9-EnsGene -co CoverageFiles/mm9-EnsGene.coverage \
-t Cufflinks/mm9-EnsGene.transcripts.gtf.tmap
```

3.7. Post-filtering for sensible results

TSS-Predictor.pl produces a few plots to help you choose an appropriate method for cleaning up your data. The final step in the TSS-Predictor pipeline is SelectTSS.pl. This takes user-defined cutoffs and reports only transcription start sites that pass these filter(s). For SelectTSS.pl usage see the next section.

Expression-based filtering

Transcripts that are expressed at a low level are less likely to be assembled correctly, and hence TSS predictions for the least abundant transcripts are less likely to be accurate. Expression-based filtering involves disregarding genes expressed below a user-defined level, optimal cutoffs vary between experiments and depend on a number of factors including sequencing depth and read length. In some cases expression-based filtering may not be suitable at all. TSS-Predictor uses R to produce two plots to aid the selection of appropriate cutoffs.

Figure 2 shows the MAD expression plot produced by TSS-Optimizer for our sample dataset, this clearly demonstrates the effect of gene expression on TSS prediction accuracy. We assume that one of the annotated transcription start sites for each expressed gene is correct. For high quality, well annotated genomes such as mouse and human this appears to be a relatively safe assumption. The MAD expression plot shows the median absolute deviation for genes, grouped into bins based on the per-base coverage (expression). We observe a much higher variation in the distances between predicted and annotated TSS for genes expressed at a lower level.

The dashed black line in Figure 2 shows the MAD for all genes in the dataset. TSS-Predictor advises an expression cutoff as the median per-base gene expression of the first bin below the line. The effect of trimming at this cutoff can be seen in the ExprCutoffPlot.jpg file produced by TSS-Predictor.pl, which for our example dataset is shown in Figure 3. This shows that a large number of lowly expressed outliers will be removed by disreagarding genes expressed below the red line. In principle this approach facilitates the discovery of novel TSSs by identifying highly expressed genes with predicted TSS a large distance from the nearest predicted TSS for the gene. In practise however, these 'novel' TSSs most often arise from incorrect transcript assembly, or selection of an inappropriate reference transfrag to trim. For some datasets expression-based filtering does not work well, usually because the distinction between low expression (and high variance) and high expression (and low variance) is not particularly clear. In this case an alternative approach to filtering is distance based, see below.

Distance based filtering

Distance based filtering assumes explicitly that one of the annotated transcription start sites for each gene is correct. Genes with a TSS greater than a given distance from the nearest predicted TSS are disregarded. 500 bases often appears to be a good cutoff. The DistanceCutoffPlot (Figure 4) shows the effect of removing TSS greater than 500bp from the nearest annotated TSS for the gene.

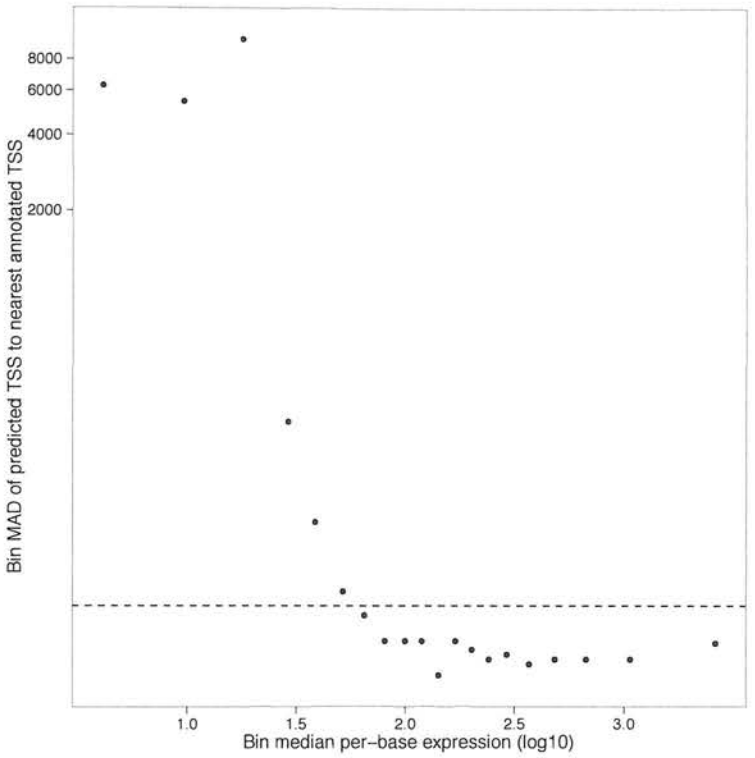


Figure 2: The MAD expression plot.

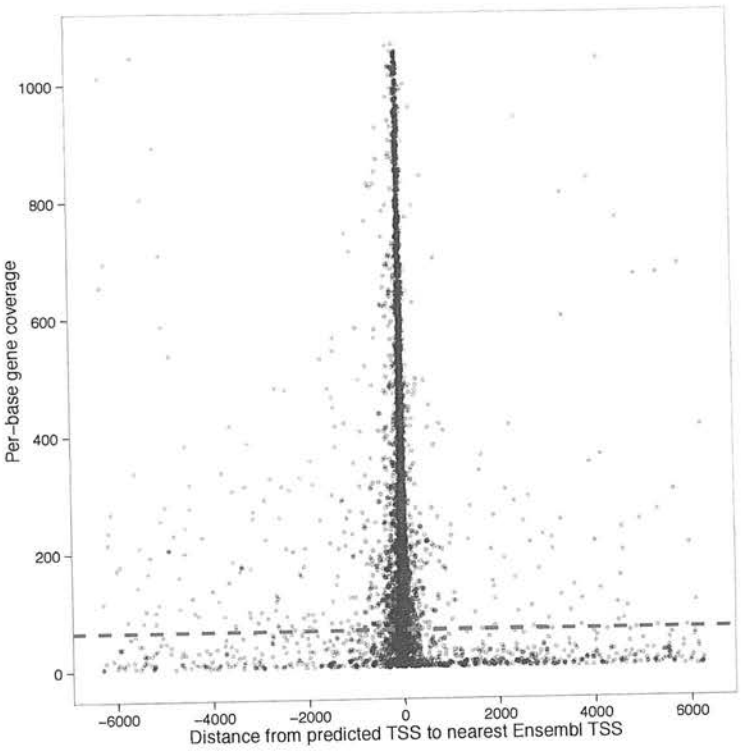


Figure 3: The expression cutoff plot. TSS-Predictor reports the value of the expression cutoff to the command line. In this instance it is 65.09, this value is printed to the terminal during TSS-Predictor analysis.

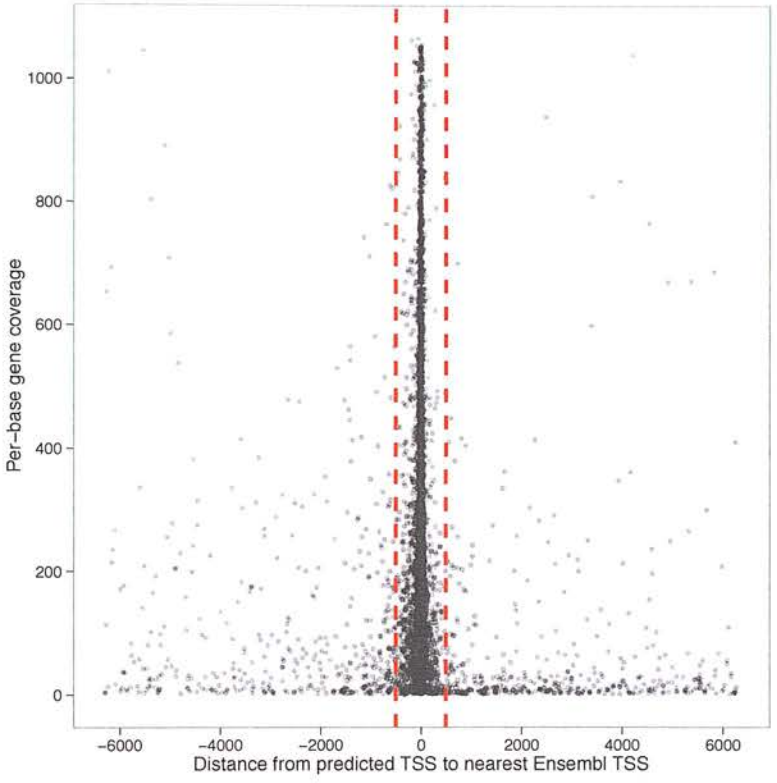


Figure 4: The distance cutoff plot.

3.8. Selecting a high confidence TSS set using SelectTSS.pl

SelectTSS.pl is used to filter TSS predictions. A tss.selected.bed file is produced which contains only transcription start sites that pass either expression-based or distance-based filtering. If both a distance and expression cutoff are supplied then only transcripts that pass both filters are reported. There is an additional option to specify whether to report the predicted TSS, or the TSS of the nearest annotated transcript for the gene.

```
SelectTSS.pl [options -d -e -r] -o <output prefix>
```

Options	Default	Description
-o	tss.selected	Output file prefix
-d	[]	TSS distance cutoff
-e	[]	Expression level cutoff
-r	[]	Report nearest reference transcription start site

We want to select only high confidence TSS's in our example. The following command will report predicted TSS's that satisfy both distance AND expression criteria:-

```
cd ~/...
SelectTSS.pl -d 500 -e 65.09
```

4.0. Output files and formats

4.1. Coverage file

The coverage file produced by GetCoverage.pl is a tab-delimited text file containing the per-base coverage of each transcribed location on the reference genome as determined by Cufflinks. The file format is as follows:-

Field position	Field name	Desription
1	Chromosome	Reference chromosome/scaffold
2	Position	Position on the reference chromosome
3	Read depth	Read coverage at position on genome

4.2. trim.combined.gtf

The trim.combined.gtf file is produced by Prep4Trim.pl. This GTF file is a reduced version of the Cuffcompare combined.gtf file and contains one transfrag record for each expressed reference gene (as determined by Cuffcompare). Specifically, the longest transfrag on the same strand as the reference gene is reported. For a detailed descripion combined.gtf format see the Cufflinks manual `http://cufflinks.cbcb.umd.edu/manual.html`.

4.3. trim.gtf.tmap

The trim.gtf.tmap file is an amended version of the transcripts.gtf.tmap file produced by Cuffcompare. This tab seperated file contains only the reference isotigs selected for trimming. There are also a couple of changes in this file relative to the Cuffcompare output. Firstly, the *ref_id* field contains the reference gene, as defined by the *gene_id* tag in the annotations GTF file, rather than the reference transcript. The second difference is the *gene.cov* field. The cuffcompare output contains a *cov* field reporting the expected transcript coverage, whereas the trim.gtf.tmap contains the expected per-base coverage for the gene calculated by summing the length-normalized coverage estimates for all transcripts in the gene (see section 5.1).

TSS-Predictor makes use of fields 2 and 10, the rest of the file is produced by Cufflinks. The fields in the trim.gtf tmap file are as follows:

Field position	Field name	Desription
1	ref_gene_id	Gene name assigned by Cuffcompare (*gene_name* attribute of GTF file)
2	ref_id	*gene_id* attribute of the the GTF file
3	class_code	Cuffcompare class code
4	cuff_gene_id	The gene id assigned by Cufflinks. Multiple *cuff_gene_id*'s may map to a single *ref_gene_id* attribute
5	cuff_id	Name of the Cufflinks transcript id selected for trimming for *ref_gene_id*
6	FMI	Expression of this isoform relative to major isoform for the gene (defined by Cufflnks, not used)
7	FPKM	FPKM for the transcript *cuff_id*. Note this is the transcript (NOT gene) expression
8	FPKM_conf_lo	Upper confidence interval for FPKM
9	FPKM_conf_hi	Lower confidence interval for FPKM
10	gene.cov	The per-base coverage estimate for the gene as used by TSS-Predictor
11	len	*cuff_id* transcript length
12	major_iso_id	*cuff_id* of the genes (*cuff_gene_id*) major isoform

4.4. isotigs.trimmed.gtf

The isotigs.trimmed.gtf file is an amended version of the trim.combined.gf file (see section 4.1). An additional flag, *contain_tss*, is added to field 9, this contains either "yes" or "no" and identifies the exons which are predicted to contain the primary TSS for the gene. For transcripts on the '+' strand with *contain_tss* = "yes" the exon start position (field 4) is amended to reflect the predicted TSS. When *contain_tss* = "yes" and the transcript is on the '-' strand the exon end position (field 5) is amended accordingly.

4.5. tss.prediction.bed

This BED (tab delimited) file contains the TSS predictions for each expressed reference gene. By default the TSS lenth is 1 base. Proximal promoter regions flanking the predicted TSS can be extracted by providing TSS-Predictor with

the *-u* and *-d* flags. Please note that changing the values of *-u* and *-d* will affect the distance estimates produced in the distances.txt file. The format of tss.predictions.bed is as follows:-

Field position	Field name	Desription
1	Chromosome	Reference chromosome/scaffold
2	Start	TSS prediction start
3	End	TSS prediction end
4	Gene name	The reference gene id
5	Score	Not used
6	Strand	Strand the reference gene is on

4.6. distances.txt

The tab-delimited distances.txt file contains a variety of information for each expressed reference gene. The format of this file is as follows:-

Field position	Field name	Desription
1	Gene	Reference gene id
2	Chr	Chromosome
3	Prediction_start	Predicted TSS
4	Nearest_transcript	Closese reference transcript to TSS prediction
5	Nearest_transcript_start	TSS of nearest annotated reference transcript
6	Distance	Distance between predicted and nearest reference TSS

4.7. tss.selected.bed

This BED file contains the subset of tss.prediction.bed containing high-confidence TSS predictions (i.e. those which have passed the filters set by SelectTSS.pl).

5.0. Overview of the algorithm

TSS-Predictor works by selecting and trimming Cufflinks gene annotations based upon the ratio between the genes expected coverage (expression level) and per-base coverage at each position on the gene. Below is an overview of the steps involved:-

1. Firstly, reads are mapped to the reference genome allowing multi-mapping reads. Because the Cufflinks output is used by TSS-Predictor I use TopHat, but this is not essential. **Depth is important for TSS prediction, for this reason I merge BAM alignments after read mapping.**
2. Cufflinks is then run with the default parameters and transcribed fragments 'transfrags' are linked to reference genes using Cuffcompare. Ensembl is my favourite annotation source, but all you need is a many-to-one transcript_id to gene_id relationship in the GTF file. Note UCSC GTF files obtained from the table browser do not maintain this relationship.
3. A coverage file is produced using *GetCoverage.pl*. This script makes use of the Cufflinks output and BEDTools to calculate the per-base read coverage for every transcribed base reported in the Cufflinks cuffcompare.combined.gtf file.
4. TSS-Optimizer uses a reduced dataset of one chromosome to determine optimal conditions (see Section 3.5 *Using TSS-Optimizer to identify optimal coverage cutoffs*). TSS-Predictor predicts major TSS utilization genome-wide once the optimal conditions have been determined. TSS-Optimizer and TSS-Predictor run the same core algorithm and are described together below:-

5. 1. Prep4Trim.pl

Prep4Trim.pl calculates the expected coverage for each reference gene and selects one representative transfrag for trimming. Expected gene expression is calculated by summing length-normalized expression (coverage) for all Cufflinks transfrags associated with each reference gene, as defined by the *gene_id* tag in the reference GTF file (Figure 5). Additionally, one representative transfrag for each reference gene is selected for trimming. Specifically, the longest Cufflinks transfrag on the same strand as the reference gene is selected.

Prep4Trim.pl reports two files - trim.combined.gtf and trim.gtf.tmap. Trim.combined.gtf is a reduced version of the Cufflinks cuffcompare.combined.gtf file, and contains only the transfrags selected for trimming in the next stage of the algorithm. Trim.gtf.tmap is an amended version of the cuffcompare .tmap output. This contains a mapping between each Cufflinks transfrag to be trimmed and the reference gene, as well as the amended coverage estimate for

$$C_{(G)} = \sum_{i=1}^{n} \frac{C_{(Ti)} L_{(Ti)}}{L_{(Tmax)}}$$

Figure 5: Gene-wise coverage calculation. Each expressed reference gene G is composed of n transfrags (T), each with coverage (C) and length (L). Thus, C_{Ti} is the coverage for the i^{th} transfrag of gene G, and L_{Ti} is the length of the i^{th} transfrag of gene G.

the gene (see section 4.3). All other fields in trim.gtf.tmap file are transfrag level information generated by Cuffcompare.

5.2. Trim_Cufflinks_GTF.pl

The core of the algorithm. Trim_Cufflinks_GTF.pl takes the expected gene-level coverage information from trim.gtf.tmap and transfrag information from trim.combined.gtf. Using the per-base coverage information calculated by *Get-Coverage.pl* each selected isotig is trimmed at a fraction of the expected coverage for the gene (X).

The algorithm iterates, one base at a time, from the last base (3') in the first exon of each transfrag (providing it is covered by $>X * C_G$) towards the first base. The GTF file is trimmed when the per-base coverage of transfrag T_{max} drops below $X * C_G$. If the last base of exon 1 is expressed at a lower level than X then the second exon is examined, so on and so forth...

The value of X is provided by the user, and determined through empirical optimization. This is what the TSS-Optimizer is for. Optimal values for X are then used to predict transcription start sites genome-wide using TSS-Predictor.pl.

Trim_Cufflinks_GTF.pl reports an isotigs.trimmed.gtf file. This is an amended version of trim.combined.gtf with revised transcription start site information. An additional variable *"contain_tss"* is added to the 9$^{\text{th}}$ field of the GTF file to identify exons predicted to contain transcription start sites.

5.3. Get-TSS.pl

This script links reports a predicted transcription start site for each reference gene in BED format. By default this is tss.prediction.bed. The *-u* and *-d* flags are used to set the number of bases upstream (u) and downstream (d) of the predicted start site to report.

APPENDIX 4 – ARRAYQUALITYMETRICS REPORT FOR *L. SIGMODONTIS* MICROARRAY TIME COURSE

arrayQualityMetrics report for BSData.vsn.rsn

Browser compatibility

This report uses recent features of HTML 5 which have not yet been implemented by all browsers. Thus, unfortunately, browser compatibility currently needs to be considered:

- Firefox 4 - tested, works well,
- Chrome 10 - tested, works well,
- Safari 5 - the interactive (SVG) plots will be missing, since this browser does not support the embedding of the <svg> tag in HTML.

- Array metadata and outlier detection overview

array	*1	*2	*3	Sample	Day	RNA.batch	type	Index
1				15	8	1	I	1
2				4	4	2	I	2
3				36	18	2	I	3
4				23	11	3	I	4
5				27	14	4	I	5
6				46	28	4	I	6
7				20	11	1	I	7
8				16	8	2	I	8
9				44	28	2	I	9
10				29	14	3	I	10
11				41	0	3	N	11
12				33	18	4	I	12
13				34	18	1	I	13
14				21	11	2	I	14
15				5	4	3	I	15
16				45	28	3	I	16
17				18	8	4	I	17
18				22	11	5	I	18
19				43	28	1	I	19
20				28	14	2	I	20
21				17	8	3	I	21
22				32	18	3	I	22
23				24	11	4	I	23
24				35	18	5	I	24

The columns named *1, *2, ... indicate the calls from the different outlier detection methods:

1. outlier detection by Distances between arrays
2. outlier detection by Boxplots
3. outlier detection by MA plots

The outlier detection criteria are explained below in the respective sections. Arrays that were called outliers by at least one criterion are marked by checkbox selection in this table, and are indicated by highlighted lines or points in some of the plots below. By clicking the checkboxes in the table, or on the corresponding points/lines in the plots, you can modify the selection. To reset the selection, reload the HTML page in your browser.

Section 1: Between array comparison

- Figure 1: Distances between arrays.

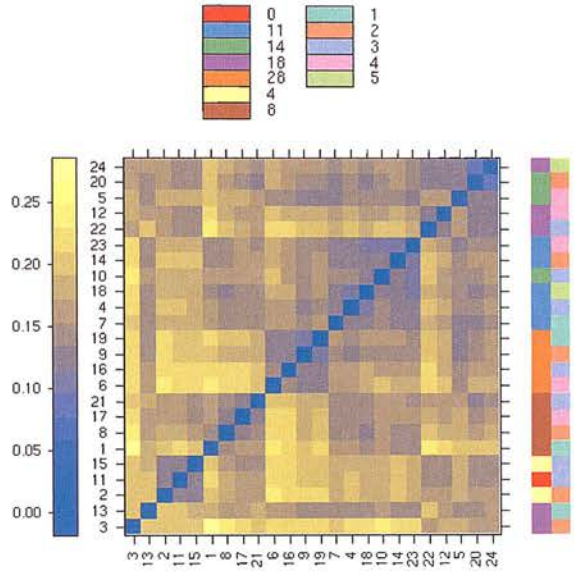


Figure 1 (PDF file) shows a false color heatmap of the distances between arrays. The color scale is chosen to cover the range of distances encountered in the dataset. Patterns in this plot can indicate clustering of the arrays either because of intended biological or unintended experimental factors (batch effects). The distance d_{ab} between two arrays a and b is computed as the mean absolute difference (L_1-distance) between the data of the arrays (using the data from all probes without filtering). In formula, d_{ab} = mean $| M_{ai} - M_{bi} |$, where M_{ai} is the value of the i-th probe on the a-th array. Outlier detection was performed by looking for arrays for which the sum of the distances to all other arrays, $S_a = \Sigma_b d_{ab}$ was exceptionally large. No such arrays were detected.

- Figure 2: Outlier detection for Distances between arrays.

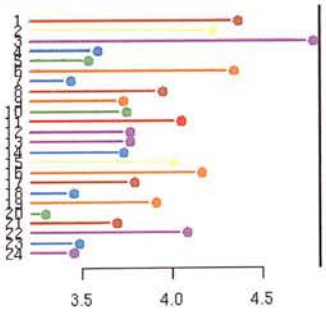


Figure 2 (PDF file) shows a bar chart of the sum of distances to other arrays S_a, the outlier detection criterion from the

previous figure. The bars are shown in the original order of the arrays. Based on the distribution of the values across all arrays, a threshold of 4.83 was determined, which is indicated by the vertical line. None of the arrays exceeded the threshold and was considered an outlier.

- Figure 3: Principal Component Analysis.

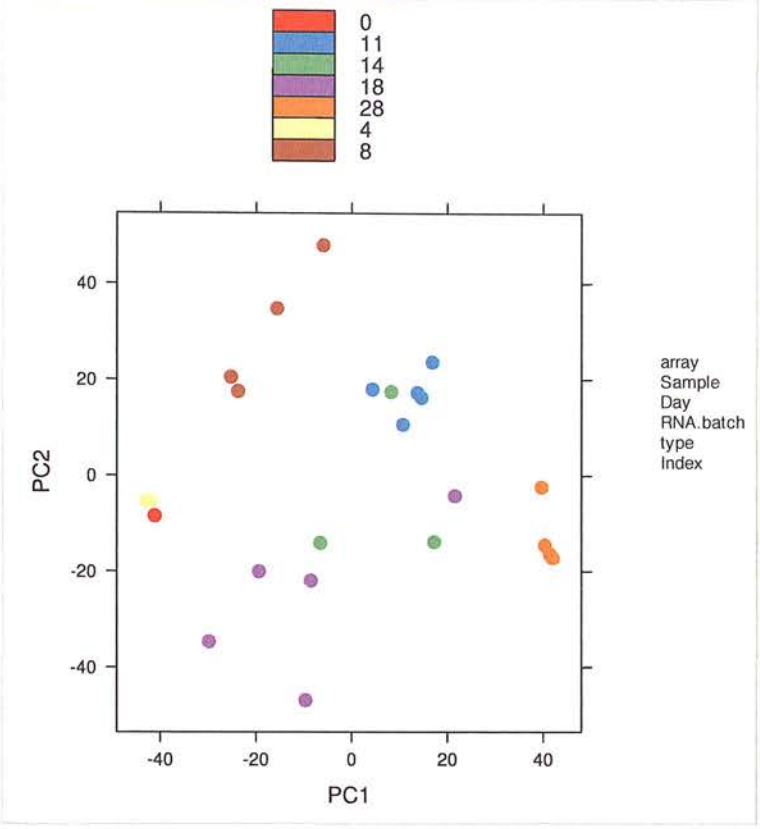


Figure 3 (PDF file) shows a scatterplot of the arrays along the first two principal components. You can use this plot to explore if the arrays cluster, and whether this is according to an intended experimental factor, or according to unintended causes such as batch effects. Move the mouse over the points to see the sample names.
Principal component analysis is a dimension reduction and visualisation technique that is here used to project the multivariate data vector of each array into a two-dimensional plot, such that the spatial arrangement of the points in the plot reflects the overall data (dis)similarity between the arrays.

Note: the figure is static - enhancement with interactive effects failed. This is either due to a version incompatibility of the 'SVGAnnotation' R package and your version of 'Cairo' or 'libcairo', or due to plot misformating. Please consult the Bioconductor mailing list, or contact the maintainer of 'arrayQualityMetrics' with a reproducible example in order to fix this problem.

Section 2: Array intensity distributions

- Figure 4: Boxplots.

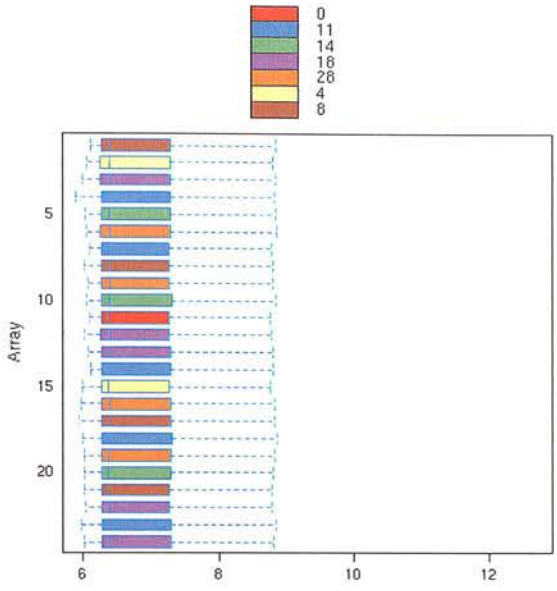


Figure 4 (PDF file) shows boxplots representing summaries of the signal intensity distributions of the arrays. Each box corresponds to one array. Typically, one expects the boxes to have similar positions and widths. If the distribution of an array is very different from the others, this may indicate an experimental problem. Outlier detection was performed by computing the Kolmogorov-Smirnov statistic K_a between each array's distribution and the distribution of the pooled data.

- Figure 5: Outlier detection for Boxplots.

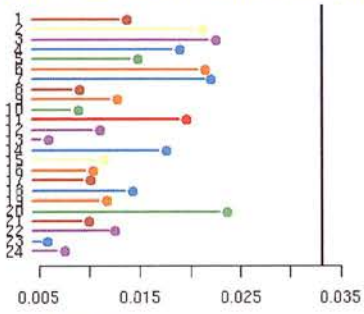


Figure 5 (PDF file) shows a bar chart of the Kolmogorov-Smirnov statistic K_a, the outlier detection criterion from the previous figure. The bars are shown in the original order of the arrays. Based on the distribution of the values across all arrays, a threshold of 0.0332 was determined, which is indicated by the vertical line. None of the arrays exceeded the threshold and was considered an outlier.

- Figure 6: Density plots.

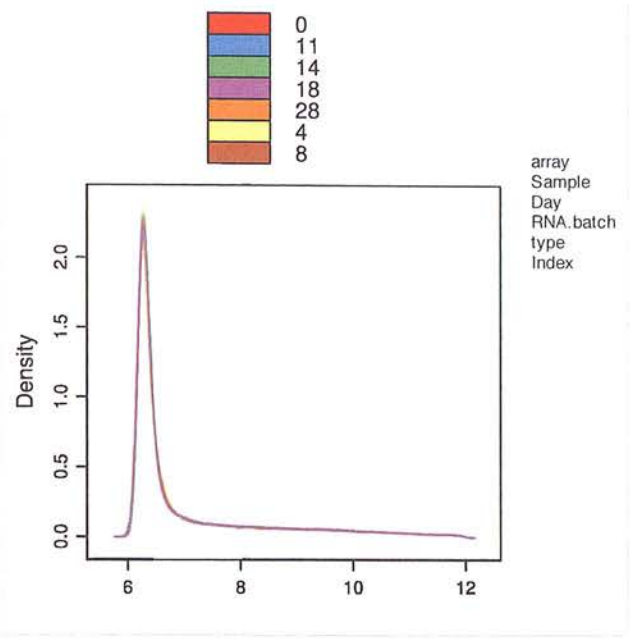


Figure 6 (PDF file) shows density estimates (smoothed histograms) of the data. Typically, the distributions of the arrays should have similar shapes and ranges. Arrays whose distributions are very different from the others should be considered for possible problems. Various features of the distributions can be indicative of quality related phenomena. For instance, high levels of background will shift an array's distribution to the right. Lack of signal diminishes its right right tail. A bulge at the upper end of the intensity range often indicates signal saturation.

Section 3: Variance mean dependence

- Figure 7: Standard deviation versus rank of the mean.

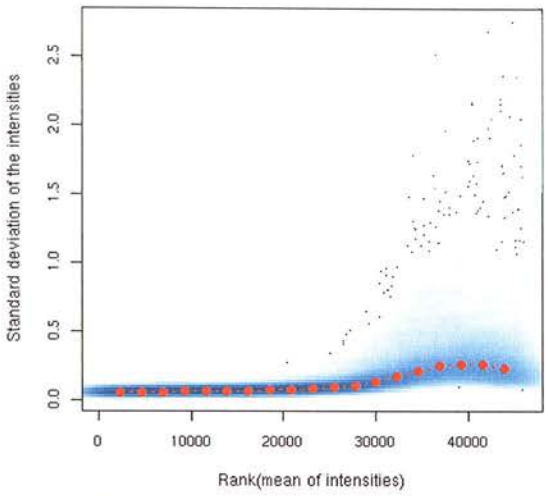


Figure 7 (PDF file) shows a density plot of the standard deviation of the intensities across arrays on the *y*-axis versus the rank of their mean on the *x*-axis. The red dots, connected by lines, show the running median of the standard deviation. After normalisation and transformation to a logarithm(-like) scale, one typically expects the red line to be approximately horizontal, that is, show no substantial trend. In some cases, a hump on the right hand of the x-axis can be observed and is symptomatic of a saturation of the intensities.

Section 4: Individual array quality

- Figure 8: MA plots.

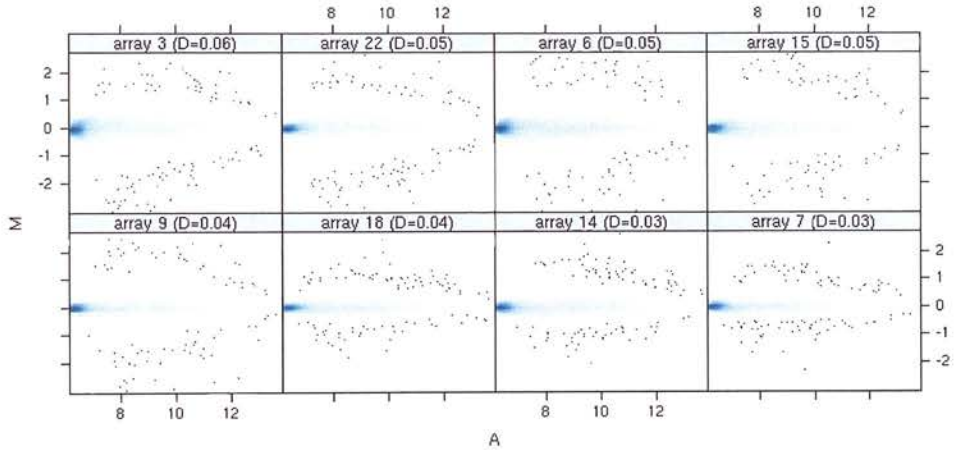


Figure 8 (PDF file) shows MA plots. M and A are defined as:

$M = \log_2(I_1) - \log_2(I_2)$

$A = 1/2\ (\log_2(I_1)+\log_2(I_2))$,

where I_1 is the intensity of the array studied, and I_2 is the intensity of a "pseudo"-array that consists of the median across arrays. Typically, we expect the mass of the distribution in an MA plot to be concentrated along the M = 0 axis, and there should be no trend in M as a function of A. If there is a trend in the lower range of A, this often indicates that the arrays have different background intensities; this may be addressed by background correction. A trend in the upper range of A can indicate saturation of the measurements; in mild cases, this may be addressed by non-linear normalisation (e.g. quantile normalisation).

Outlier detection was performed by computing Hoeffding's statistic D_a on the joint distribution of A and M for each array. Shown are the 4 arrays with the highest value of D_a (top row), and the 4 arrays with the lowest value (bottom row). The value of D_a is shown in the panel headings. 0 arrays had D_a>0.15 and were marked as outliers. For more information on Hoeffing's D-statistic, please see the manual page of the function `hoeffd` in the `Hmisc` package.

- Figure 9: Outlier detection for MA plots.

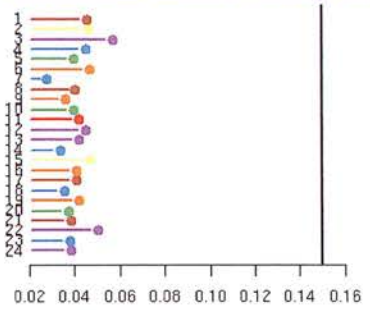


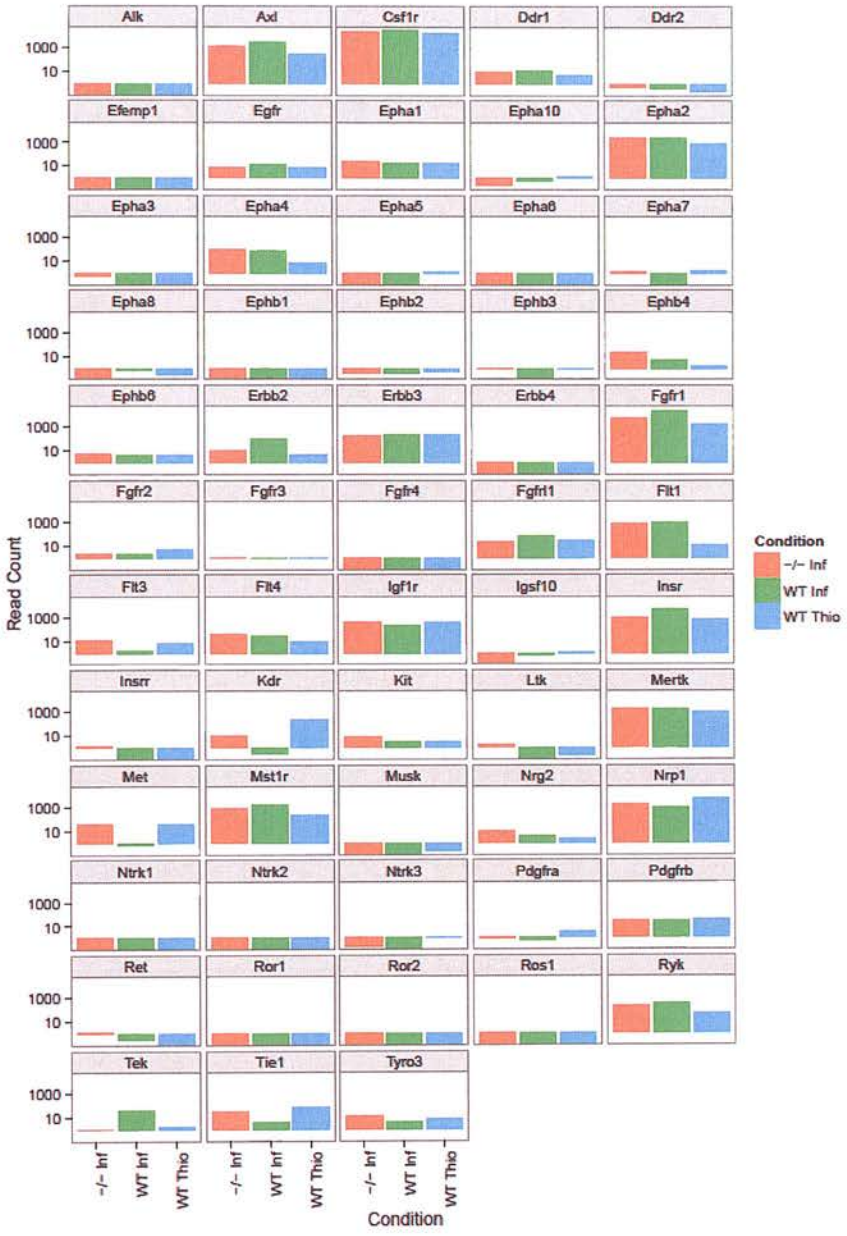
Figure 9 (PDF file) shows a bar chart of the Hoeffding's statistic D_a, the outlier detection criterion from the previous figure. The bars are shown in the original order of the arrays. A threshold of 0.15 was used, which is indicated by the vertical line. None of the arrays exceeded the threshold and was considered an outlier.

This report has been created with arrayQualityMetrics 3.8.0 under R version 2.13.0 (2011-04-13).

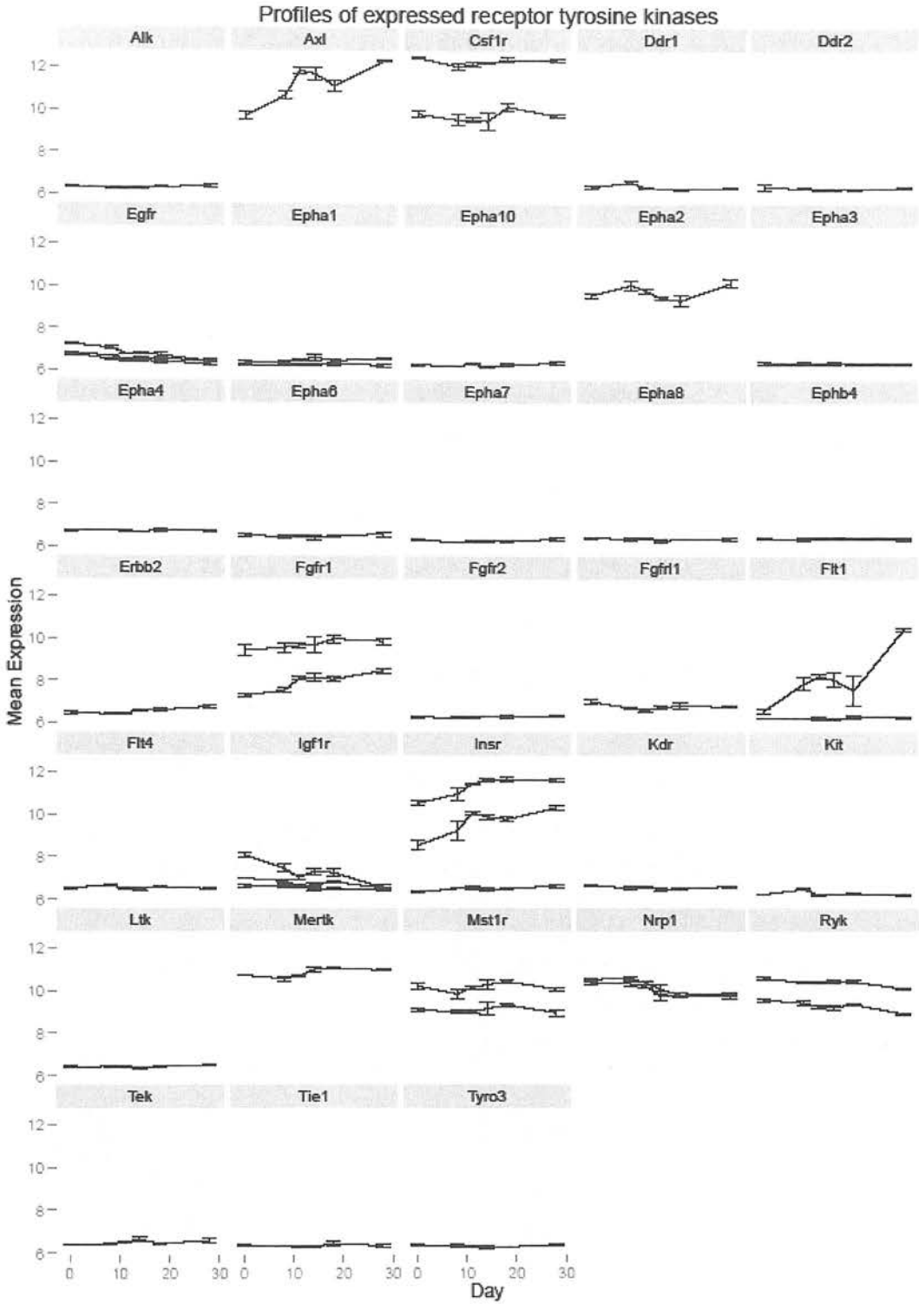
(Page generated on Fri Dec 9 13:39:42 2011 by hwriter)

13-1. arrayQualityMetrics report for normalized microarray data from the *L. sigmodontis* infectoin time course presented in chapter 7.

APPENDIX 5 – RTK GENE EXPRESSION PROFILES IN PERITONEAL AND PLEURAL CAVITY MACROPHAGE POPULATIONS



A5-1. Gene expression profiles of expressed receptor tryosine kinase proteins (GO:0004714) in the RNA-Seq dataset generated in Chapter 5.



A5-2. Gene expression profiles of expressed receptor tryosine kinase proteins (GO:0004714) in the *L. sigmodontis* microarray time course presented in Chapter 7.

APPENDIX 6 - RESEARCH OUTPUT ARISING FROM THIS THESIS AT THE TIME OF SUBMISSION

Aspects of the work within this thesis has been presented at the following conferences:

"Immunology and Metabolism" – Oral presentation. January 2011. CIML, Marseille, France.

BSI congress - Poster presentation – December 2011. Liverpool, UK.

The Biology of Cytokines – Poster presentation. February 2012. Keystone, Colorado, USA.

Chapter 2 of this thesis has been published in Blood and can be accessed under the following citation:

Thomas, G.D., Rückerl, D., Maskrey, B.H., Whitfield, P.D., Blaxter, M.L., and Allen, J.E. (2012). The biology of nematode- and IL-4Rα-dependent murine macrophage polarization in vivo as defined by RNA-Seq and targeted lipidomics. Blood.