The effects of development and virulence on gene expression in the protozoan parasite *Neospora caninum*

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

Kittichai Unjit

February 2018

AUTHOR'S DECLARATION

Apart from help and advice acknowledged, this thesis represents the unaided work of the author

…….

Kittichai Unjit

February 2018

This research was carried out in the Department of Infection Biology and School of Veterinary Science, University of Liverpool

DEDICATION

I want to dedicate this thesis to my parents, brother and sister who have always supported me all my entire life especially during my PhD study. All of you are the greatest. Thank you very much for being with me all the times.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to Thai Government for providing me a scholarship to study aboard. I would like to thank my supervisors, Dr. Andrew Jackson and Professor Jonathan Wastling for accepting me to be a student. They always support and guide me during my PhD study to achieve my goal. Additionally, I would like to thank Dr. Nadine Randle and Dr. Dong Xia for their kindly help and support throughout my study period. Special Thanks to Dr. Nick Evans for accepting me to be his student for a certain time. I also thank to an advisory teams as Dr. Ben Makepeace and Dr.Janine Coombes for their guidance and suggestions in a research study. I would like to deeply appreciate Professor Luis Ortega-Mora for providing *Neospora* samples in this thesis and a good collaboration. Thanks to Prof. Andrew Hemphil, University of Bern, Switzerland for providing BAG1 antibody and Dr. Olivier Touzelet for providing a secondary antibody Alexafuor 488 for IFA test. Special thanks to Mrs.Catherine Harley, Dr. Nathifa Moyo and Mrs.Jenna Dawson for their help and support for the laboratory work. I would also like to thank Dr. Stuart Armstrong, Dr.Vicky Hunt, Dr.Gareth Weedall, Dr.Luca Lenzi, Dr. Sam Haldenby, Ross Low and Dr. Pisut Pongchaikul for suggesting me about bioinformatics. Many thanks to colleagues in Department of Infection Biology for having fun and supporting; Tonk, Poomie, Dr. Juriah, Dr. Dashty, Dr.Mariwan, Dr.Sarah, Dr.Corrado, Dr.Patrick, Dr.Fazila, Zuliza, Intan, Sanaria and the others. Finally, I would like to thank a group of Thai friends in Thai society in Liverpool and Chaba Chaba Thai restaurant for hospitality and friendships which make me feel like at home.

ABSTRACT

The effects of development and virulence on gene expression in the protozoan parasite *Neospora caninum*

Kittichai Unjit

Neospora caninum is an obligate intracellular protozoan parasite which causes abortion in cattle and neuromuscular disease in dogs. Neosporosis causes substantial economic losses in both dairy and beef cattle industries worldwide. This thesis explores changes in genes and proteins expression associated with development and virulence of *N. caninum* in order to gain more biological information in this parasite. To achieve the global quantification of genes and proteins expression under different phenotypic conditions, transcriptomics and proteomics approaches are used to investigate these phenomenons.

Chapter 2 examines the differential protein expression during tachyzoites-bradyzoites stage conversion using a label-free proteomics approach and finds that the differential protein expression of most apical secretory proteins decrease in abundance in bradyzoites stage. In addition, most of proteins associated with parasite motility were also reduced in abundance in bradyzoites. This indicates that proteins associated with host cell adhesion, invasion and gliding motility are down-regulated in the quiescent cyst-forming stage.

Chapter 3 compares the transcriptomic profiles of low- and high-virulence *N. caninum* using RNA-Seq and finds that protein phosphatase 2C (PP2C) was preferentially expressed in the high-virulence strain, while a group of SRS family proteins was preferentially expressed in low-virulence conditions. This finding suggests that PP2C might play a role in virulence, while a group of SRS proteins might be associated with to limiting *N. caninum* virulence.

Chapter 4 compares protein expression in low- and high-virulence *N. caninum* strains at three different time points using a label-free quantitative proteomic approach and

finds that Rop24 was preferentially expressed in high-virulence *N. caninum* at all time points. Proteins associated with host cell attachment such as SAG3, TgSRS35A and cathepsin C2 also showed preferentially expression throughout the time course. This finding indicates that Rop24 might be associated with the *N. caninum* virulence, while the discovery of proteins associated with host cell attachment in low-virulence conditions might suggest that such strains are more efficient in invading host cells than high-virulence strains.

Overall, this thesis identifies genes and proteins such as the rhoptry associated proteins PP2C and Rop24, which might be associated with virulence in neosporosis. It also shows that many proteins involved in host cell adhesion, invasion and parasite movement decrease in abundance in the cyst-forming bradyzoite stage. These discoveries enhance our understanding of parasite biology, providing a basis for future research into novel ways to prevent and control the disease by inhibiting parasite survival and transmission.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BP	Biological process
bp	Base pair
cDNA	Complementary deoxyribonucleotide
CID	Collision induced dissociation
HPLC	High performance liquid chromatography
hr	hour
FC	Fold change
FDR	False discovery rate
Hp	Heat shock protein
IFA	Immunofluorescence assay
DMDM	Dulbecco's Modified Media
LC	Liquid Chromatography
M2AP	Microneme associated protein
MIC	Microneme
mRNA	messenger ribonucleic acid
MS	Mass spectrometry
MyoA	Myosin A
m/z	mass to charge ratio
N. caninum	*Neospora caninum*
NO	Nitric oxide
PBS	Phosphate buffered saline
P. falciparum	*Plasmodium falciparum*
p.i.	post infection
PV	parasitophorous vacuole
PVM	parasitophorous vacuole membrane
qPCR	quantitative polymerase chain reaction
RNA	ribonucleic acid
RNA-Seq	RNA sequencing
RON	Rhoptry neck protein
ROP	Rhoptry protein
RPKM	Read per kilobase

RPM	Revolution per minutes
RT	Room temperature
SAG	Surface antigen glycoprotein
SNP	Sodium nitroprusside
SNPs	Single Nucleotide polymorphism
SRS	Surface antigen glycoprotein-related sequence
STAT	Signal transducer and activator of transcription
TEM	Transmission electron microscopy
T. gondii	*Toxoplasma gondii*
Tg	*Toxoplasma gondii*
v/v	volume/volume
w/v	weight/volume
1D	One dimensional
2D	Two dimensional

Chapter 1: Introduction

1.1 Neosporosis

Neosporosis is a parasitic disease caused by *Neospora caninum*, an obligate intracellular protozoan belonging to phylum Apicomplexa (Dubey et al., 1988a). *N. caninum* is a member in the same phylum as three well known parasites: *Toxoplasma gondii* (causing toxoplasmosis in humans and animals) (Dardé et al., 1998), *Plasmodium falciparum* (causing malaria in humans) (Nabet et al., 2016), and *Eimeria tenella* (causing intestinal diseases in diverse animals) (Jordan et al., 2011). Apicomplexan parasites have been characterized by an apical complex situated at one end of their structure. The apical complex is important as a unique evolutionary adaptive for host cell invasion (Goodswen et al., 2013).

N. caninum causes neosporosis in cattle (*Bos taurus*) and domestic dogs (*Canis familiaris*). Neosporosis is a major cause of abortion in both dairy and beef cattle worldwide and neuromuscular disease in infected dog. It has been found in cattle in every country where the disease has been sought (Al-Qassab et al., 2010). The loss to cattle typically happens after primary infection during pregnancy period and the recrudescence of the persistent infection is common during gestation (Dubey et al., 2006). *N. caninum* is closely related to *T.gondii* and they share many common biological and morphological characteristics, until 1988, *N. caninum* was misdiagnosed as *T. gondii* (Bjerkas and Dubey, 1991; Dubey et al., 1988a; Dubey and Lindsay, 1996). Transplacental (vertical) transmission is the major route of transmission for neosporosis. This route of transmission play a significant role in maintaining the infection within the herd (Rojo-Montejo et al., 2009a). Once infected, cattle and dogs can remain infected for their entire lives (Barber and Trees, 1998).

Abortion is the main clinical sign in cattle infected with *Neospora*. Pregnant cows may abort from three months of gestation up to term but mostly between 5-7 months of pregnancy. Infected foetuses are mostly born clinically normal but persistently infected, however some may die, be reabsorbed, mummified, autolyzed, or stillborn. Aborted foetuses typically present lesions as multifocal necrosis surrounded by inflammatory cells affecting organs such as brain, spinal cord, heart, lung, and

placenta (Barr et al., 1991). Approximately 80-90% of infected calves might be *Neospora* carriers; most of these are clinically normal.

In dogs, *N. caninum* can cause severe disease to dogs of all ages but it is mostly found in puppies (Nishimura et al., 2015). Pathology in dogs typically comprises pelvic limb paralysis and rigid hyperextension (Dubey and Schares, 2011; Reichel et al., 2007). Dogs may show encephalomyelitis and myositis which results in paralysis and early death in puppies (Bjerkas et al., 1984). However, the neurological signs which are seen in adult dogs are variable depending on the site of parasite multiplication in central nervous system (Dubey, 2003).

N. caninum is a heteroxenous parasite requiring more than one host to complete its life cycle; sexual replication occurs in definitive, canid hosts and asexual replication occurring in intermediate, bovine hosts (Donahoe et al., 2015; Dubey et al., 2006). Dogs (*Canis familiaris*) are typically the definitive host of *N. caninum* (McAllister et al., 1998) although other canids such as coyotes (*Canis latrans*) (Gondim et al., 2004), gray wolves (*Canis lupus*) (Dubey et al., 2011) and Australian dingoes (*Canis lupus dingo*) (King et al., 2010) have been observed to excrete oocysts after feeding on infected tissue cyst. *N. caninum* is a haploid organism which displays sexual reproduction in its canine host. Although, dogs infected with *N. caninum* are not always sero-positive (Lindsay et al., 1999a; McAllister et al., 1998), all dogs should be considered as an important source of *N. caninum* transmission whether they have seroconverted or not (Reichel et al., 2014).

Several animal species can be infected with *N. caninum*: parasites isolated from dogs have been classified as Nc-1 (Dubey et al., 1988b), Nc-Bahia (Sawada et al., 2000), and Nc Liverpool (Barber et al., 1993). Nc BPA1, BPA2 (Conrad et al., 1993), Nc KBA1, KBA2 (Kim et al., 2000), Nc-Spain 1H (Rojo-Montejo et al., 2009b), Nc-Spain 2H, 3H, 4H, and 5H as well as Nc-Spain 6, 7, 8, 9, and 10 (Regidor-Cerrillo et al., 2008) were isolated from cattle; Nc-Sheep from sheep (Koyama et al., 2001), and NcBrBuf 1, 2, 3, 4, and 5 from water buffalo (Rodrigues et al., 2004). In addition there are reports antibodies against *N. caninum* being detected in white-tailed deer (*Odocoileus virginianus*) (Lindsay et al., 2002), crows (*Corvus corax*) (Molina-

López et al., 2012), hares (*Lepus europaeus*), zebra (*Equus burchelli*), Thorold's deer (*Cervus albirostris*), sika deer (*Cerus Nippon pseudodaxis*), fallow deer (*Dama dama*), moose (*Alces alces*), mouflon (*Ovis ammon*), European bison (*Bison Bonasus*), sea otter (*Enhydra lutris*) and dolphin (*Tursiops truncatus*) (Dubey et al., 2007). These observations suggest that wild animals also play an important role in the epidemiology of neosporosis.

1.2 Life cycle of *N. caninum*

The life cycle of *N. caninum* comprises three distinct asexual infectious stages known as sporozoites, tachyzoites and bradyzoites (Dubey, 1999b) (Figure 1.1). Unsporulated oocysts are excreted by definitive host into the environment within 5 to 17 days after ingesting tissue cysts (Lindsay et al., 1999a). Sporulated oocysts are active after 3 days in the environment. Each sporulated oocyst contains two sporocysts and each sporocyst cantains four sporozoites (Dubey et al., 2002). The oocyst is resistant to environmental degradation, with a wall that is tolerant to freezing and drying, allowing infective sporozoites to survive in the external environment for several weeks. The intermediate host as well as the definitive host can be infected with *N. caninum* by ingesting sporozoite containing oocysts. Sporozoites are released from ingested oocyst into the gut of the host, these then invade the gut wall and differentiate to tachyzoites. Cell invasion by tachyzoites results in the creation of a parasitophorous vacuolar membrane (PVM) that is followed by rapid asexual replication by endodyogeny. The tachyzoites can be transmitted to the developing foetus via the placenta. In response to host immune reponses, the rapidly dividing tachyzoites can convert into slowly replicating bradyzoites which can survive for the life of the intermediate host (Dubey and Lindsay, 1996). The bradyzoites residing within the quiescent tissue cysts can lead to recrudescent infection if the host immune is suppressed, for example during pregnancy or concomitant disease (Haddad et al., 2005). When the same tissue cysts are ingested by dogs, often via contaminated meat, the bradyzoites are released from the cyst and infect epithelial cells of the small intestine, ultimately leading to the differentiation of gametocytes. The fusion of gametocytes to produce unsporulated oocysts in the lumen of canine gut completes the life cycle.

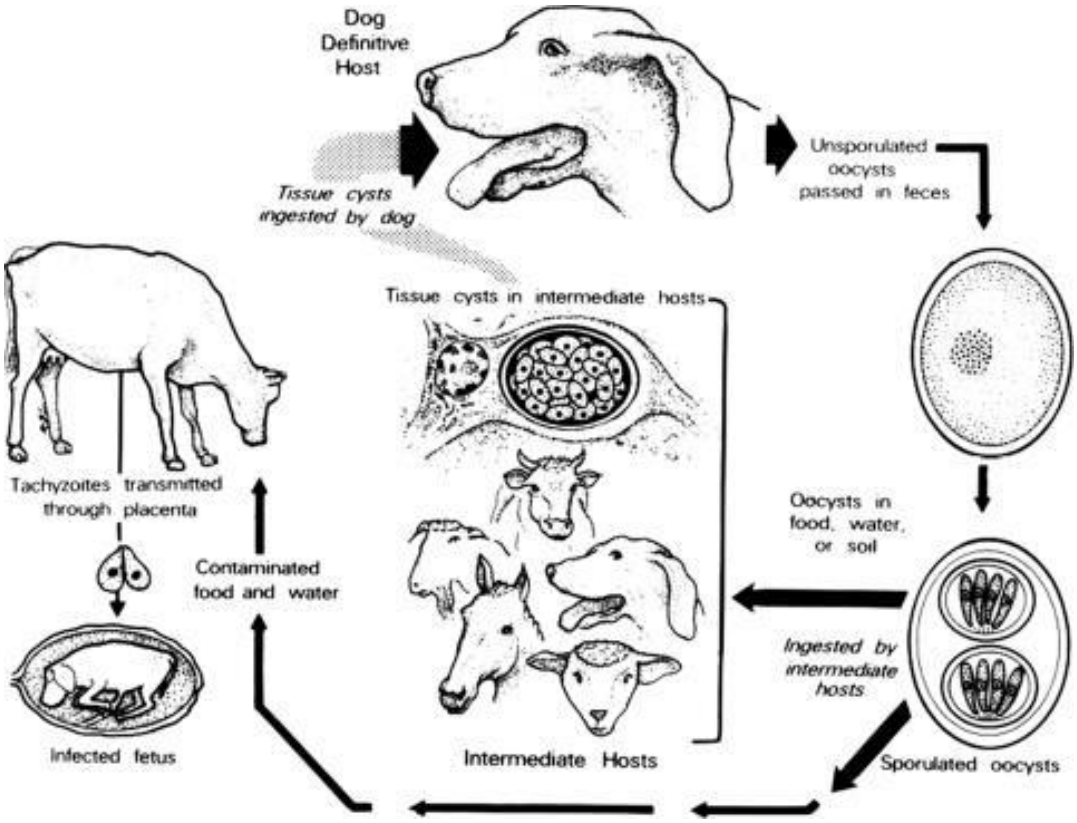

Figure 1.1 Life cycle of Neospora caninum (reprinted from Dubey (1999a)). The dog, a definitive host, ingests tissue cysts and passes unsporulated oocysts into the environment in its faeces. Intermediate hosts, such as cattle, ingest sporulated oocyst, then sporozoites release and form tachyzoites in cell lining of the gut. The definitive host can be infected by ingesting tissue cyst or the tachyzoites will transmit through the placenta and infect the foetuses. The definitive host is infected by ingesting infected meat containing tissue cysts.

1.3 Mode of transmission in cattle

Both horizontal and vertical transmission play an important role in the transmission of *N. caninum* among cattle. Horizontal transmission occurs by cattle ingesting sporolated oocysts distributed in the environment. To date, there is no report of cow-to-cow transmission nor evidence that cows can shed oocysts to the environment. There are two main routes of vertical transmission that transmit parasites from the

dam to their offspring. Exogenous transmission occurs after the dam is infected by ingesting sporulated oocysts, while endogenous transmission occurs when the infection is recrudescent following a previously infected pregnancy (Trees and Williams, 2005). Vertical transmission is very efficient in cattle and it is the major route of transmission (Barr et al., 1993; Hietala and Thurmond, 1999; Schares et al., 2005). The infected cow might remain infection for a whole life (Trees et al., 1999) and they may transfer the parasites to their fetuses in several next gestation or intermittently (Boulton et al., 1995; Piergili Fioretti et al., 2003; Wouda et al., 1998). Moreover, transplacental transmission can be re-occurred in the same animal (Barr et al., 1993; Bjerkas et al., 1984; Dubey et al., 1988b). Thus, lifelong infections of animals contribute to persistent infection of the herd through this mechanism (Beck et al., 2009). However, French et al. (1999) examined the transmission and control of *N. caninum* using deterministic and stochastic models. They found that vertical transmission alone was not sustained the infection of *N. caninum* in the cattle herds even though the high probability of vertical transmission was shown. Molecular analysis of *N. caninum* populations suggested that a clonal population is typical where vertical transmission is efficient (Perez-Zaballos et al., 2005).

Neosporosis remains a challenging disease to control. *N. caninum* can be killed by drugs but these are not adequate for disease control. The most effective way to control neosporosis is vaccination. However, at the present there are no effective vaccines against neosporosis available (Reichel and Ellis, 2009). While the analysis from Reichel & Ellis (2009) showed that vaccination is the most cost effective approach to control neosporosis. In the absence of effective drugs and vaccines, the most appropriate ways to control bovine neosporosis are employing good and effective management systems, for example by avoiding contact between dogs and grazing cattle.

1.4 *N. caninum* cellular structure and function

There are three infectious stages of *N. caninum* known as tachyzoites, bradyzoites, and sporozoites. Tachyzoites and bradyzoites are found in the tissues of intermediate and definitive hosts, while sporozoites are found in sporocysts inside the oocysts,

which are formed in the intestine of definitive host and excreted via their faeces (Dubey et al., 2006).

1.4.1 The tachyzoite stage

Tachyzoite is a fast replicating form of *N. caninum* associated with acute infection. The size of *N. caninum* tachyzoite is 3-7 x 1-5 µm (Dubey et al., 2002) which is slightly larger than *T. gondii* and depending on the stage of replication (Dubey &Lindsay, 1996). The *N. caninum* tachyzoite has a crescent shape, pointed at the anterior end and rounded at the posterior end (Figure 1.2). It contains two apical rings, 22 subpellicular microtubules, conoid, cytoskeleton, micropore, subpellicular microtubule, secretory organelles (such as rhoptries, micronemes and dense granules), mitochondria, apicoplast, endoplasmic reticulum, nucleus, Golgi complex and ribosome (Lindsay et al., 1993; Speer and Dubey, 1989). The outer membrane is covered with a pellicle which consists of three layers of plasma membrane.

Tachyzoites divide into two zoites by endodyogeny and they can be found in neural cells, macrophages, fibroblasts, vascular endothelial cells, hepatocytes, myocytes, and renal tubular epithelial cells (Bjerkas and Presthus, 1989; Dubey et al., 1988a; Speer and Dubey, 1989). Tachyzoites invade host cells by active invasion, localise within a parasitophorous vacuole (PV), and become intracellular within 5 minutes after attach with host cells (Hemphill et al., 1996). They may be located in more than one PV per host cell.

1.4.2 The bradyzoite stage

Bradyzoite is a slowly replicating form of *N. caninum* that represents the chronic stage of the disease. Bradyzoite form cysts with surrounding walls up to 4 µm thick; cysts vary in size depending on the number of bradyzoites contained within (Dubey et al., 2006). The shape of tissue cysts are often round to oval and up to 107 µm long. They are mostly found in neural tissues such as brain spinal cord, nerves and retina (Dubey et al., 1988a; Dubey et al., 1990) and they can be found in skeletal muscles (Dubey et al., 1989). Bradyzoite is slender and 6-8 x 1-1.8 µm in size. They contain

the same organelles as found in tachyzoites except there are fewer of rhoptries and more amylopectin granules than tachyzoites (Dubey and Lindsay, 1996).

1.4.3 The sporozoite stage

The sporozoite stage of *N. caninum* is found within the oocyst. Each oocyst contains two sporocysts and each sporocyst contains four sporozoites (Lindsay et al., 1999b). The size of oocysts is roughly 10 x 12 µm with smooth colourless wall of 0.6 - 0.8 µm thickness. The oocysts are spherical to sub-spherical and excreted with the faeces of the canid host in unsporulated form. Sporocysts are ellipsoidal with size of 8 x 6 µm and their wall thickness are 0.5 -0.6 µm. Sporozoites are characterized by the position of the nucleus, which is located centrally or slightly posteriorly (Lindsay et al., 1999b).

Although little is known about the biology of the *N. caninum* oocyst, it is a key factor in neosporosis epidemiology in cattle (Dubey et al., 2007). The large number of contaminated oocysts in the environment leads to the high chance of infection (Goodswen et al., 2013). The prepatent period before the shedding of oocysts by definitive host varies from 5- 10 days after *N. caninum* infection (Al-Qassab et al., 2010) and they might be continually excreted for 10 days (McAllister et al., 1998; McCann et al., 2007). The age of the host, type of feeding tissues, and immunity status of host affect the rate of oocyst shedding (Dubey et al., 2007). For example, the number of oocysts shedded was higher in dogs that ingest infected calf tissue than in those that ingested mouse tissue cysts. It has also been shown that infected puppies shed oocysts at a higher rate than adult dogs (Gondim et al., 2002).

1.5 Stage conversion from tachyzoites to bradyzoites stage of *N. caninum*

The *Neospora* tachyzoites can convert into bradyzoites in response to host immunity for survival in the host. Tachyzoites differentiate to bradyzoites after triggering by the host immune response leading to cyst forming in host tissues. Recrudescence can occur in pregnant cattle caused by the suppression of host immunity and incur transplacental transmission. The vertical transmission is an important route for the

persistent infection from generation to generation. The tachyzoite-bradyzoite conversion process is central to pathogenesis, establishing a chronic infection, and disease reactivation (Lyons et al., 2002). The knowledge of protein expression involved in *N. caninum* stage conversion will provide a better understanding of pathogenesis for neosporosis.

In vitro tachyzoite-bradyzoite conversion can be induced in *T. gondii* using inhibitors of mitochondrial function and inducers of oxidative stress (Bohne et al., 1994; Soete et al., 1994). Nitric oxide (NO) has been shown to induce the stage conversion as it is an inhibitor of mitochondrial function and an indirect stressor of the parasites (Lyons et al., 2002). The bradyzoites-tachyzoites stage conversion could also be induced using alkaline medium (Soete et al., 1993). In *N. caninum*, Vonlaufen et al. (2002) successfully developed the stage conversion assay from tachyzoites to bradyzoites using 70 µM sodium nitroprusside (SNP) and the parasites can survive for up to 8 days. The murine epidermal keratinocytes was used as host model which was infected with tachyzoites of *N. caninum* Liverpool strain. Their result revealed significant reduction in parasites multiplication. The immunofluorescence assay was used to detect BAG1 antibodies which could be firstly detected at day 4 (35% BAG1 expression), then at day 8 (60% BAG1 expression) in parasitophorous vacuole (PV) and those PV expressing BAG1 exhibited strongly reduced in expression of tachyzoites marker antigens Nc SAG1. The other *N. caninum* stage conversion development was also performed by Vonlaufen et al. (2004) using another host cell type as Vero cells for the development. They found that the optimal concentration of SNP was 17 µM. This concentration did not visibly destroy the structural integrity of the Vero cells monolayer. This developed procedure has been applied in this study.

1.6 Host cell invasion and *N. caninum* apical complex

Invasion process of host cells and subsequent intracellular development is vital for *N. caninum* to survive and proliferate (Hemphill et al., 1999; Hemphill et al., 2004). The invasion process is similar to *T. gondii.* A low-affinity contact between tachyzoites and the host cell surface membrane is established, followed by the attachment process which requires *N. caninum* metabolic energy for invading the host cell

(Hemphill et al., 1996). The tachyzoite invades the host cell by reorientating itself perpendicularly to the surface of the host cells and advancing at the anterior end. Once in the host cell cytoplasm it is enclosed by the PVM acquired from host cell membrane (Hemphill and Gottstein, 2006). The tachyzoite immediately starts to divide by endodyogeny which can be visualised at approximately 6 hours after invasion. Host cell lysis occurs after 72-80 hours leading to parasite to egress and infect the neighbouring cells (Hemphill, 1999). The unique secretory organelles that comprise the apical complex (micronemes, rhoptries and dense granules) are shown in Figure 1.2 and play a key role in this process.

1.6.1 Micronemes

Micronemes are rod-like shaped organelles located at the anterior region of apicomplexan parasites (Dubey et al., 1998). Microneme proteins (MICs) play an important role in host cell invasion and are the first proteins to be released upon attachment to the host cell (Carruthers and Sibley, 1997b; Sonda et al., 2000). The micronemes seccrete adhesins which interact with actin-myosin motors to contribute driving movement essential for host cell invasion (Dobrowolski and Sibley, 1996; Keeley and Soldati, 2004; Morisaki et al., 1995), while several microneme proteins are associated with the gliding movement of apicomplexa (Tomley and Soldati, 2001). *N. caninum* and *T. gondii* contain 38 and 33 genes respectively that encode microneme proteins (Reid et al., 2012). MIC1 has adhesive domains helping the parasite to attach to host cells (Garnett et al., 2009). MIC1 forms a complex with MIC4 and MIC6 which is important for the invasion process of the parasite (Reiss et al., 2001a; Saouros et al., 2005a). TgMIC2 contains transmembrane adhesion which play a vital role in gliding motility and host cell invasion (Huynh and Carruthers, 2006). NcMIC2 contains integrin- and TSP-like domains which are a member of the TSP family of adhesive proteins (Naitza et al., 1998). A complex of MIC2 and MIC2-associated protein (M2AP) is essential for host cell entry and attachment (Jewett and Sibley, 2004). NcMIC3 has also been shown to interact with the host cell surface proteoglycan during host cell binding (Naguleswaran et al., 2002).

1.6.2 Rhoptries

Rhoptries are club shaped organelles with a narrow anterior neck extending into the conoid, which act as ducts to release their contents (Bannister et al., 2003; Bradley et al., 2005). Rhoptries secrete rhoptry neck proteins (RONs) such as RON2, RON4, RON5 and RON8, shortly after microneme proteins. These form a complex with microneme protein apical membrane antigen 1 (AMA1) to effect a moving junction with the host cell membrane during invasion (Alexander et al., 2005; Besteiro et al., 2009; Tyler and Boothroyd, 2011). The moving junction is the mechanism by which *N. caninum* enters host cells and establishes a PV (Alexander et al., 2005). Rhoptry proteins (ROP) are released slightly after RONs and the invasion process. ROP play a significant role in early and later stage of invasion as well as localisation and formation of the PV (Bradley et al., 2005). Soluble ROP1 is released into the PV after its formation (Saffer et al., 1992) while ROP2 is secreted into the PVM where it attaches to the host mitochondria and may enable the parasite to obtain host nutrients (Beckers et al., 1994; Sinai and Joiner, 2001).

Some ROP proteins have been shown to be associated with virulence in *T. gondii* such as ROP5, ROP16 and ROP18. TgROP5 can reduce interferon- γ (IFN-γ)-induced and immunity–related GTPases (IRGs), which promote parasite killing by disrupting the PVM (Niedelman et al., 2012). TgROP16 is secreted into the host cytosol during host cell invasion where it phosphorylates STAT3/6 and so modulates the expression of inflammatory mediators such as IL-10 and IL-23 (Jensen et al., 2011). TgROP18 abundance correlates with parasite multiplication rate with kinase activity in *T. gondii* (El Hajj et al., 2007). However, Reid et al. (2012) found that ROP18 in *N. caninum* is pseudogenised and so does not play the same role. Consequently, *N. caninum* is unable to phosphorylate host immunity-related GTPases via this mechanism.

1.6.3 Dense granules

Dense granules are located at the posterior end of *N. caninum* and are similar to micronemes in their appearance (Bannister et al., 2003; Speer et al., 1999). Dense granule (GRA) proteins are secretory proteins that contain transmembrane domains.

In general, GRA proteins are secreted into the PV; GRA1 is released into the PV (Sibley et al., 1995) and GRA2, GRA4 and GRA6 are involved in intravacuolar network membrane (Labruyere et al., 1999). They are involved in PV maturation and are essential for parasite growth (Carruthers and Sibley, 1997b; Dubremetz et al., 1993). Most GRA proteins are secreted within one hour of PV formation (Dubremetz et al., 1993). GRA proteins are associated with many important processes in host invasion, for example providing nutrient acquisition, parasite egression and modulating the initial immune response (Braun et al., 2013; Cesbron-Delauw, 1994; Coppens et al., 2006; Okada et al., 2013; Travier et al., 2008).

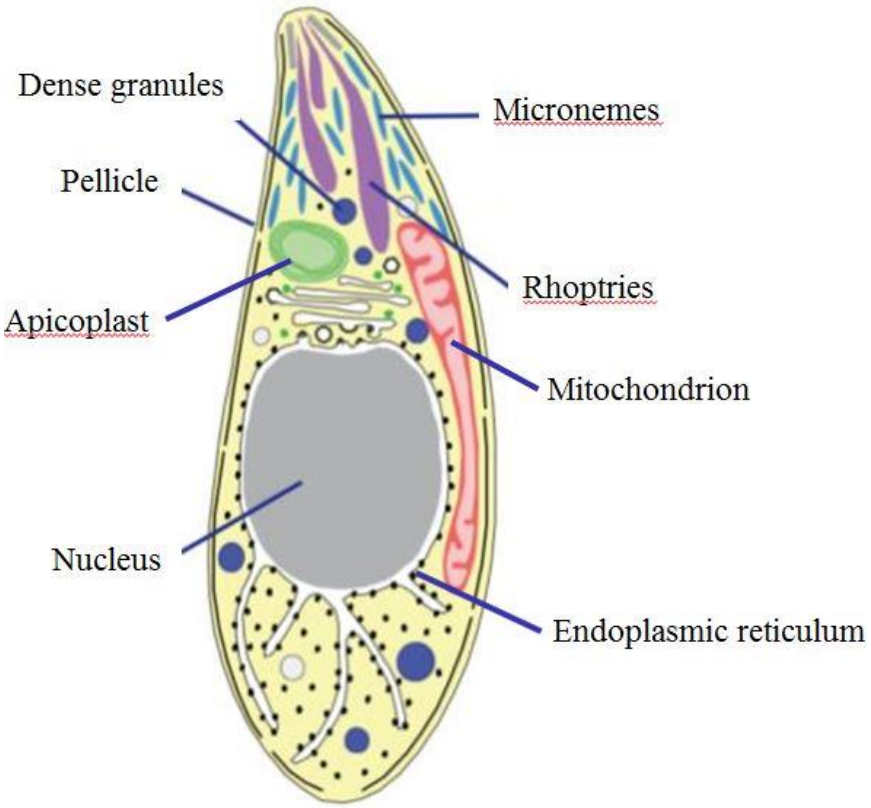

Figure 1.2 Diagram of *N. caninum* tachyzoite which organelles are labelled. (Reprinted from (Reid et al., 2012))

1.7 The virulence of *Toxoplasma gondii*

The closest parasite related to *N. caninum* is *T. gondii* which show huge amount of studies. The crucial of invasion process of *T. gondii* is rhoptry organelle. The clinical

symptoms of toxoplasmosis in human vary from asymptomatic to debilitating or even life threatening. While, *T. gondii* strains differ in their virulence in animals1. Sibley & Boothroyd, 1992 analysed 28 strains of *T. gondii* from many of hosts from five continents and found that the ten virulent strains have potential same genotype, whereas the nonvirulent strains were moderately polymorphic. These data highly suggest that virulent strains of *T. gondii* originated from a single lineage. Previous studies identified three clonal lineages of *T. gondii* that exhibit different virulence phenotypes in mice, with type I strains are lethal in mice, whereas type II show intermediate virulence that varies with mouse strain and type III have been considered avirulent 26 (Melo et al., 2011; Robert-Gangneux and Darde, 2012; Saeij et al., 2005; Xiao and Yolken, 2015), there are many factors influenced the severity of toxoplasmosis such as the age and immune status of the infected host (Petersen and Dubey, 2001), and stage of the life cycle causing infection (Munoz-Zanzi et al., 2010), parasite migratory capacity, the growth rate *in vitro* and the parasite loads in mice (Saeij et al., 2005), as well as polymorphism and differential expression of the dense granule and rhoptry proteins (Melo et al., 2011). The other different phenotypes among the three major lineages include growth rate (Radke et al., 2001), frequency of differentiation, motility and ability to cross biological barriers (Barragan and Sibley, 2002), disturbance of host cell signalling, intestinal pathology induction during acute stage (Liesenfeld, 2002) and CNS pathology development during chronic stage in the mouse (Suzuki and Joh, 1994). ROP proteins play an important role in modulating of host cell signaling and parasite virulence (Bradley and Sibley, 2007). The study of genetic basis for virulence in type I lineage was performed by analysis of progeny from genetic cross with type III. It showed that ROP18 was designated as the major determinant of virulence in type I lineage while this gene was shown type III strain with extremely low level (Taylor et al., 2006). ROP18, an active serine/threonine (S/T) protein kinase, is a major factor that contributes to strain-specific differences in virulence (Saeij et al., 2006; Taylor et al., 2006). It is discharged into the host cell and decorates the surface of the parasitophorous vacuole membrane (PVM). The expression levels of ROP18 vary from high abundance in type I and II strains to a low level in type III strains. Transgenic modification of ROP18 from type I or type II strains in the type III strain of *T. gondii* increase virulence approving that this rhoptry protein play role in

virulence between strains (Saeij et al., 2006; Taylor et al., 2006). The increased virulence of parasites with type I allele of ROP18 is related to increase replication of *T. gondii* (Saeij et al., 2006). A comparison of ROP18 region in *T. gondii* with the related parasite *N. caninum* suggests that it went through a rearrangement in the ancestral progenitor to type I and II strains (Khan et al., 2009), leading to high abundance of ROP18 expression in both type I and II lineages and enhance virulence. Both virulence of types I and II and avirulence of type III alleles are constant in the parasite population revealing evidence of long term stability (Khan et al., 2009). The other rhoptry kinase is ROP16 which is key virulence determinant. ROP 16 acts to sustain the phosphorylation of STAT3 which is a negative regulator of Th1 immune response. It is a regulator of host cell transcription of type I and III during host cell invasion by direct tyrosine phosphorylation of STAT pathways and target to the host nucleus (Saeij et al., 2007a). The ROP16 allele shared by type I and III which is effective in mediating sustained phosphorylation of STAT3. Type I and III strains may be able to avoid detection by immune system by modulating host gene expression. This protein not only induced the phosphorylation but also induced nuclear translocation of STAT5 to generate protective immunity (Jensen et al., 2013; Rosowski and Saeij, 2012). Chang et al. (2015) found that ROP16 can induce SH-SY5Y apoptosis via mitochondria-dependent p53 pathway by increasing Bax expression and decreasing the Bcl-2 expression. *T. gondii* rhoptry protein5 (ROP5) has been determined as a key virulence factor that diminishes the accumulation of immunity-related GTPases (IRG) in PVM. IRG has role in PVM integrity and evades IFNγ-mediated killing by intracellular parasites (Etheridge et al., 2014). ROP5 locus is the quantitative trait locus (QTL) found in a genetic cross between type I and II strains (Jensen et al., 2013). The genetic disturbance of the ROP5 in a type I strain revealed an attenuated of *T. gondii* (Jensen et al., 2013; Rosowski and Saeij, 2012). The comparison of the three pair-wise cross among three different type of *T. gondii* suggested that ROP5 and ROP18 interact together showing about 90% of the differences in acute virulence between these three lineages.

Table 1.1 Summary of *T. gondii* virulence

***T. gondii* (strain)**	**Virulence**	**High abundance of ROP protein**
Type I	high	ROP18 ROP16 ROP5
Type II	intermediate	ROP5
Type III	low	ROP16

1.8 The virulence of *Neospora caninum*

Virulence in *N.caninum* is the ability of the parasite to induce pathology in mice and of the parasite to transmit via the placenta (Al-Qassab et al., 2010). Virulence can vary among *N. caninum* strains and ranges between asymptomatic infections to abortion. For example, the Nc Spain 1H strain, which was isolated from the brain of infected healthy calf, is considered a low-virulence strain. While Nc Liverpool, which was collected from the cerebrum of a congenitally infected puppy, has displayed high virulence (Barber et al., 1993).

.

There are several experimental studies concerning virulence in animal models such as pregnant mice and pregnant cattle. They reveal marked phenotypic differences between low- and high-virulence strains, for example in invasion rate and tachyzoite yield in cell culture (Pereira Garcia-Melo et al., 2010). The *in vitro* invasion efficiency and intracellular proliferation rate associated with virulence in *N. caninum* study was performed by (Regidor-Cerrillo et al. (2011)). Nc-Liverpool and ten Spanish *N. caninum* isolates (Nc-Spain 1H –Nc-Spain 10) were used for the investigation. They found that the Nc Spain 4H and Nc- Liverpool revealed the significant highest of invasion rate than Nc-Spain 3H and Nc-Spain 1H at 4 hr pi by Dunn's test. Nc Spain 7 has shown high virulence in a mouse model, similar to Nc-1 (Pereira Garcia-Melo et al., 2010). Infection with Nc Spain 7 is associated with greater severe lesions in the brain compared to controls, as well as higher parasite burdens in late phase infection and increased parasite dissemination (Pereira Garcia-Melo et al., 2010).

Conversely, low-virulence strains such as Nc Spain 1H (Rojo-Montejo et al., 2009a) and Nc Nowra (Miller et al., 2002) often do not cause any clinical signs in cattle.

Rojo-Montejo et al. (2009b) examined pregnant and non-pregnant BALB/c mice infected with Nc Spain 1H. The result showed that no clinical signs and foetal death occurred, and no parasites could be detected in mouse brain tissue during the chronic infection stage. However, pregnant mice infected with Nc-1 did show high placental transmission leading to high neonatal mortality rate. The other virulent study to visualise the immunome among three biologically different isolates as Nc Spain 7 and Nc-Liv (high virulence) and Nc Spain 1H (low virulence) was performed by Regidor-Cerrillo et al. (2015). The sera from experimentally infected mice from three isolates were determined by 2-DE immunoblot. They found that all isolates exhibited a highly similar antigenic pattern which most of the spots located in acidic region (pH 3-7) and located in 3 antigenic areas as 250-70, 45-37 and 35-15KDa. The sera from Nc-Spain 7 and Nc-Liv infected mice showed the highest number of antigens. Chryssafidis et al. (2014) studied the pathogenicity of Nc-Bahia (low virulence) and Nc-1 (high virulence) by intravenous inoculation of $5x10^8$ *N. caninum* tachyzoites on the 70^{th} day of gestation in pregnant cows and buffaloes. The result revealed that abortion was occurred in only 1 from 5 cows on 42 days post infection (dpi) with Nc-Bahia and no abortion was found in buffaloes (n=3) while all animals (n=5), two cows and three buffaloes infected with Nc-1 aborted on day 35 dpi. *N. caninum* can be detected by semi-nested PCR in maternal, foetal and placental tissues which revealed the evident of vertical transmission in both strains. They suggested that Nc-Bahia had shown lower pathogenicity than Nc-1 as showing fewer abortions and milder histological lesions.

Schock et al. (2001) analyzed biological and genetic variation of six *N. caninum* isolates obtaining from bovine and canine origin by measuring growth rate *in vitro*, Western blotting and random amplification of polymorphic DNA (RAPD). The intra-species diversity of *N. caninum* analysis showed that heterogeneity exists within species. The growth rate *in vitro* revealed significant variation among isolates by determining [H] uracil uptake. The multiplication of Nc-Liverpool showed the fastest rate than others and at more than two time of Nc-1 multiplication rate. On the other hand, Nc-SweB1 multiplied significantly slower rate which were 10 times less than Nc-Liverpool. The isolates of Nc-LivB1, BPA-1, JPA-2, and Nc-1 are intermediate in multiplication rate. The result of western blotting revealed no antigenic differences

detected between any of 6 isolates. Lastly, the result of RAPD illustrated that it was able to differentiate DNA polymorphisms between *N. caninum* isolates for example the pattern of band for Nc BPA-1, JPA-2, NC-1, NC LivB1 and Nc-SeB1 showed identical band but NC-Liverpool revealed clear polymorphic band at around 360, 480 and 630 bp. These could be indicated intra-species diversity in *N. caninum* and exhibits heterogeneity exists within species. The pathogenicity study of NC-Liverpool and NC-SweB1 was performed by Atkinson et al. (1999). They inoculated the tachyzoites to mice and found more necrosis in brain lesion of mice infected with NC-Liverpool than NC-SweB1. The pathogenicity in cattle depended on timing of infection, immune status of the host. Due to the vertical transmission is predominant in *N. caninum* infection, these might limit the opportunity of sexual recombination of the parasite leading to have a similar clonal population structure (French et al., 1999). Knowledge from above experiment revealed clinical signs related to virulence among *N. caninum* strains. Thus, the genetic diversity of neosporosis will be useful for understanding the pathogenicity and epidemiology of the parasite.

Table 1.2 Summary of *N. caninum* virulence

Isolates	**Animal Origin**	**Virulence**	**Country**	**Reference**
Nc-Liverpool	Dog	High	United Kingdom	(Barber et al., 1993)
Nc Spain 7	Bovine	High	Spain	(Regidor-Cerrillo et al., 2008)
Nc-1	Dog	High	USA	(Dubey et al., 1988b)
BPA1	Bovine	High	USA	(Conrad et al., 1993)
Nc Spain 1H	Bovine	Low	Spain	(Rojo-Montejo et al., 2009b)
Nc Nowra	Bovine	Low	Australia	(Miller et al., 2002)
Nc Bahia	Dog	Low	Brazil	(Gondim et al., 2001)
Nc JAP1	Bovine	Low	Japan	(Yamane et al., 1996)

1.9 Economic impact of bovine neosporosis

Neosporosis causes economic losses to both dairy and beef cattle production (Larson et al., 2004). This is due to reproductive failures including stillbirths, infertility, foetal death in early stage and resorption but also reduction in productivity, for example in milk production (Hobson et al., 2002) weight gain (Barling et al., 2001) growth rate and feeding efficiency (Trees et al., 1999). Moreover, neosporosis results in prolonged calving interval, increased culling rate (Hall et al., 2005), increased diagnosis cost and lactation time as well as decreased breeding stock value (Haddad et al., 2005), all of which impact on the cost of farm management. One study of *N. caninum* sero-prevalence in adult cattle in south west England revealed that 94% of herds were infected (i.e. at least one animal in the herd was sero-positive) (Woodbine et al., 2008). In England and Wales, neosporosis is the main cause of abortion in cattle estimated to cause 12.5% or 6000 abortions per year (Davison et al., 1999). Overall, neosporosis may represent an economic loss of 2-5% per year to each farm however, some farms can incur losses up to 20% due to this disease. Globally, losses to livestock systems caused by neosporosis are approximately one million dollars per year (Reichel et al., 2013). The estimated median losses caused by abortion were annually accounted for US$ 1,298.3 million which could be divided by continent as US$ 852.4 million (65.7%) in North America, US$ 239.7 million (18.5%) in South America, US$ 137.5 million (10.6%) in Australia and US$ 67.8 million (5.3%) in Europe. There were 46.5 million cows of dairy industry at annual risk of abortion (pregnant cows) in the 10 countries which can be estimated to be US$ 18.16 per individual cow on average. In addition, there were 102.2 million beef cattle at risk (pregnant) in eight countries which was estimated to be US$ 4.46 individually. At farm level, the median loss per farm was approximately accounted for US$ 1,600.0 in dairy industry and only US$ 150.0 in beef industry (Reichel et al., 2013).

1.10 Omics Approaches to investigate gene and protein expression

1.10.1 Transcriptomics

Nowadays, transcriptomics has been rapidly developed because of the advance in next-generation sequencing technology. Transcriptomics is a technique used to investigate mRNA molecules or transcriptomes in cells or organisms of interest. Transcriptome is a collection of all RNA existing in cells at any given time point. The level of RNA transcripts is varying in different developmental stage and conditions. The study of transcriptomes provides the information of change happening in the cell at the transcriptional level. In addition, it helps understanding the function of gene, revealing the molecular composition of cells and detecting development of disease (Wang et al., 2009). There are three methods which are commonly performed to investigate the transcriptome expression profile; real-time quantitative PCR (qPCR), microarrays and RNA-Sequencing (RNA-Seq).

1.10.1.1 Real-time quantitative PCR (qPCR)

qPCR provides highly quantitative and sensitive data. It is generally appropriate for a relatively small number of transcripts in a large set of samples. In this method, reverse transcriptase transcribes RNA into complementary DNA (cDNA) which act as template for qPCR reaction.

The quantification is performed by measuring the change of expression levels of mRNA interpreted by cDNA. The relative quantification is easy to perform as the amount of genes are directly compared to the amount of the control reference genes.

1.10.1.2 Microarrays

Microarray can explore tens of thousands of transcripts which has led to determine a wide range of biological problems. (Zhao et al., 2014). This technique can identify genes that are preferentially expressed between two conditions, for example, diseased and non-diseased, developmental processes and/or drug responses as well as the evolution of gene regulation in different species (Kerr et al., 2008; Passador-

Gurgel et al., 2007). This technique quantifies a set of predetermined sequences and are still a widely used for transcriptomic profiling. However, there are some limitations of this technique, for instance the accuracy of expression measurements is limited by background hybridization, especially transcripts present in low abundance. Furthermore, probes differ considerably in their hybridization properties, and arrays are restricted to interrogate only those genes for which probes are designed (Zhao et al., 2014). RNA-Seq have been developed to overcome these limitations.

1.10.1.3 RNA-Sequencing (RNA-Seq)

RNA-Seq involves direct sequencing of transcripts or complementary DNAs (cDNAs) using high-throughput sequencing technologies with computational method to quantify transcripts in RNA extraction sample. There are several advantages of RNA-Seq over the other methods. RNA-Seq data can be mapped to an existing genome sequence to reveal sequence variation and strand-specific expression. Unlike microarrays, RNA-Seq does not have an upper limit for quantification, it has larger dynamic range (> 9000-fold range) of expression levels over which transcripts can be detected than microarrays. In addition, RNA-Seq has shown high accuracy for quantifying level of transcript and also show high levels of reproducibility (Wang et al., 2009). Hence, RNA-Seq generally performs with greater reliability, reproducibility and precision than microarrays and are now the preferred method for whole transcriptome profiling (Beyer et al., 2012; Montgomery et al., 2010; Mortazavi et al., 2008; Nagalakshmi et al., 2010; Wang et al., 2009). The process of RNA-Seq starts when the mRNA is converted to cDNA fragments with adaptors at one end or both ends. Individual cDNAs are sequenced in a high-throughput manner to acquire short sequences from one end (single-end sequencing) or both ends (paired-end sequencing). The reads range is size between 30 and 400 bp. Many high-throughput sequencing technologies can be used with RNA-Seq such as the Illiumina IG (Mortazavi et al., 2008; Nagalakshmi et al., 2008; Wilhelm et al., 2008), Applied Biosystems SOLiD (Cloonan et al., 2008) and Roche 454 Life Science system (Barbazuk et al., 2007; Emrich et al., 2007; Vera et al., 2008)

1.10.2 Proteomics method

Proteins are an important part in the organisms with many functional roles**.** Proteomics is a study of proteins (Anderson and Anderson, 1998) which is a subsequent step in the biological study after genomics and transcriptomics**.** Proteomics is the systematic quantification of all peptides within a cell or organism**.** Mass spectrometry plays a vital role in proteomic analysis (Cravatt et al., 2007)**.**

1.10.2.1 Gel-based proteomics

Gel-based proteomics is a very popular technique to separate and quantify protein. This approach can investigate the protein expression in large scale which is cheaper than gel-free proteomics.

Two-dimensional electrophoresis (2DE) is a widely used gel-based method which based on two different biochemical characteristics of proteins. First, proteins are separated according to their isoelectric point. Then SDS-PAGE separates proteins according to their molecular mass. The spots are then visualized, evaluated and finally identified proteins by mass spectrometry. 2D Fluorescence Difference Gel Electrophoresis **(**2D-DIGE**)** has been developed which provides more accurate and reliability in protein quantification**.** Therefore, samples in comparison are run together in the same gel as a consequence of reducing the potential gel-to-gel variation **(**Kondo, 2008**).** However, this technique requires some skills, time consuming, laborious, low sensitivity and has some technical restrictions associated with protein resolving mass (Old et al., 2005; Zhu et al., 2010)**.** In addition, this technique is limited in dynamic range, and often struggles to identify hydrophobic proteins and those with high molecular weight and pI values (Zhu et al., 2010)**.**

1.10.2.2 Non-Gel-based proteomics

The non-gel-based methods such as label-based quantitative technique and label-free quantitative technique were developed to reduce the gel-based method limitations. These methods provide potential tools to investigate large-scale protein expression

and characterisation in biological samples (Domon and Aebersold, 2006; Motoyama and Yates, 2008).

1.10.2.2.1 Label-based quantitative proteomics

The label - based quantitative technique as isotope labelling method has been incorporated into mass spectrometry as internal standards or relative references (Zhu et al., 2010). This technique differentiates the proteins in the sample by using chemical and physical properties of isotope labelled compounds that match with their natural compounds of the samples except their masses. This method offers valuable flexibility to study protein expression change using quantitative proteomics tools. Though, most labelling quantification methods also have the limitation such as increased time and complex sample preparation, require higher concentration of samples, expensive, incomplete labelling and require specific software for quantification. These restriction of label –based quantitative method lead to increase the interest of label-free proteomics techniques which provide quicker, cleaner and simpler quantification outcome (America and Cordewener, 2008; Chen and Yates Iii, 2007; Old et al., 2005; Patel et al., 2009) .

1.10.2.2.2 Label-free quantitative proteomics

The method of label-free quantitative proteomic was used to analyse the proteins expression throughout this thesis. The samples preparation for these techniques were individually performed then each injected to LC-MS/MS (Figure 1.3). There are 2 categories of measurement used for protein quantification as measurement of ion intensity changes for example peptides peak area or peak heights in chromatography and measurement of the spectral counting of identified proteins. The peptide peak intensity (or spectral count) is measured for each LC-MS/MS run and the change in protein expression are calculated by comparing between different conditions (Zhu et al., 2010).

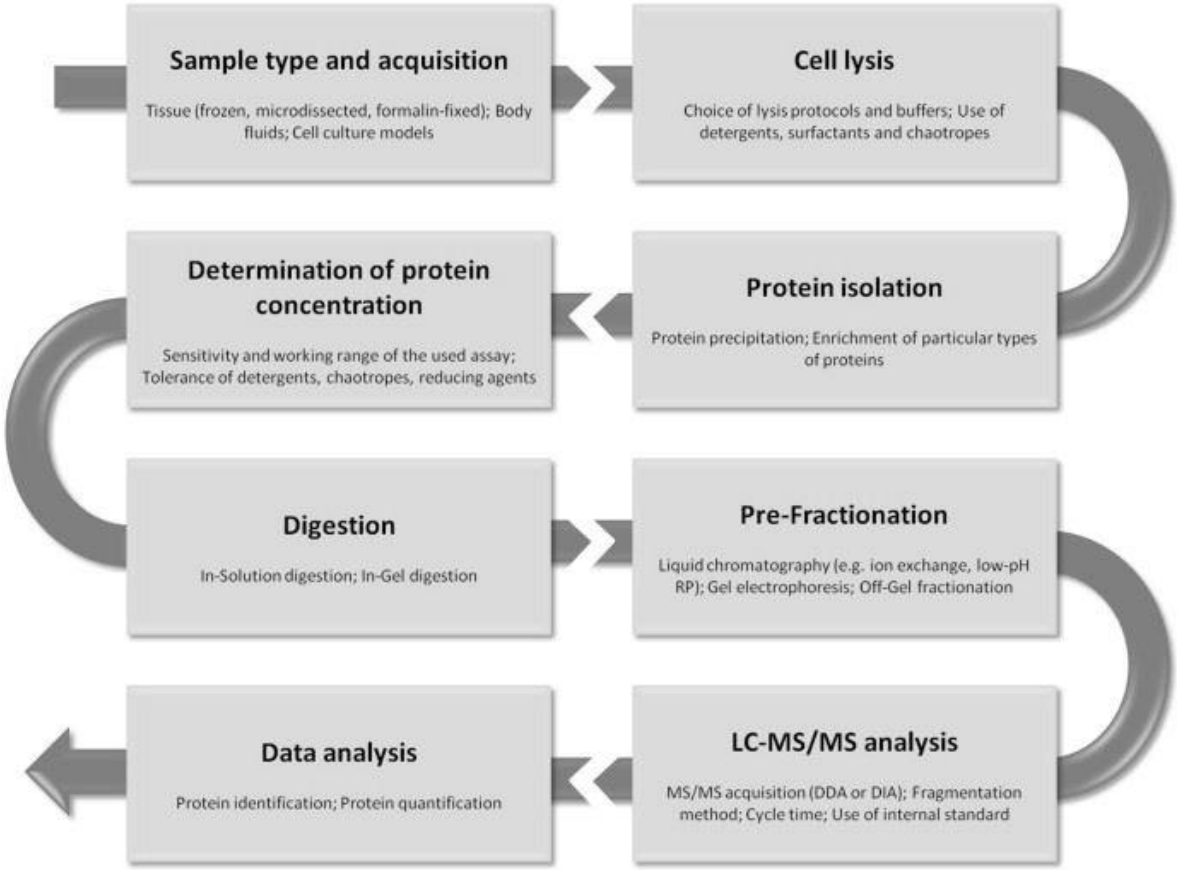

Figure 1.3 Diagram of label-free quantitative proteome analysis**.** **(**Reprinted from Megger et al. (2013)**)** The workflow start from samples cells are lysed using lysis detergents followed by protein extraction and digestion**.** Then the samples are put through LC-MS/MS followed by protein identification and quantification by determining ion intensity generated by spectral counting**.** Data analysis is performed using bioinformatics tool**.**

1.10.3 Reverse-phase high performance liquid chromatography (LC)

This technique is used to separate a complex mixture of peptides in order to achieve the better resolution of the peptides containing in the samples**.** High performance liquid chromatography **(**HPLC**)** is an analytical process in which the sample mixtures as peptides contained in liquid mobile phase**.** The protein samples are forced under the pressure to pass through a stationary phase within a tightly-packed column**.** Then compounds bind to the column based on their chemical properties in the stationary phase**.** In reverse-phase **(**RP**)** HPLC, non-polar molecules are recollected whereas

polar molecules are prior to elute**.** The mobile phase solvent buffer is adjusted along the gradient, by slowly increasing the acetonitrile content, in order to elute more polar molecules by increasing hydrophobicity (Simpson 2003).

1.10.4 Tandem mass spectrometry (MS/MS)

Mass spectrometry estimates the molecular mass of the samples by measuring the mass to charge ratio **(***m/z***)** of ions in gas phase**.** Peptides in each sample are changed to the gas phase by an ionisation source for example electrospray ionisation source **(**ESI**).** The measurement of peptides mass **(***m/z***)** is performed by the detector followed by fragmentation of peptides via collision with an inert gas**.** The three most abundance peptides are chosen for fragmentation using collision induced dissociation **(**CID**).** The *m/z* of the fragmented ion is measured**.** The data of each sample from mass spectra are matched to known peptides as peptide spectrum match **(**PSM**).** The data are then analysed using bioinformatics tools**.**

1.10.5 Protein identification and bioinformatics analysis

Mass spectral data are searched against an *in silico* database to identify proteins**.** This process is performed to obtain all possible peptides that are cleaved by the enzyme**.** For *T. gondii*, *N. caninum* and related parasite species, ToxoDB is the resource that contains the most effective *in silico* database used for searching**.** A false discovery rate **(**FDR**)** scoring system is used as a cut off to determine identified peptides**.** The higher the number of significant peptides, leads to the higher the confidence of protein identifications**.** Many algorithms have been developed for protein identification using mass spectral data as an input, for example MASCOT (Perkins et al., 1999), SEQUEST (Eng et al., 1994), X**!**tandem (Craig and Beavis, 2004) and PEAKs (Zhang et al., 2012)**.** These search engines identify peptides by using algorithms which is identical to the precursor and MS/MS spectral data against the translated genome database**.** Then, the score unique to the spectra searched against theoretical spectra of the organism of interest is obtained. In order to confirm that peptides identified do not include any false positives, so the false discovery rate (FDR) is required. It is done by re-searching the data with a false database by using

the reverse amino acid sequences of the actual database used for the protein search. The acceptable value should be around 1% or less.The identified proteins obtaining from LC-MS/MS analysis with less than 2 peptides are excluded for further analysis**.**

1.10.6 Omics and apicomplexan parasites

Nowadays many researchers have been working with -omics technologies to understand the biology of organisms as well as the interactions between host and parasites (Goodswen et al., 2013). These techniques composed of genomics, transcriptomics, proteomics and metabolomics. The recent genomics research comparing *N. caninum* and *T. gondii* provides vast knowledge about the pathogen which delivers the information of the evolution of both species (Reid et al., 2012). These -omics research does not only provide the understanding of the organisms itself, but can also be identify the drug targets and vaccine against the disease (Goodswen et al., 2013). The variations occurring at molecular level of nucleotides, genes and proteins can be detected by comparing the -omics data between species or strains of organisms.

The purpose of system biology study is to incorporate several biological data through different levels of structure and scale which represent an initial paradigm in the scientific process (Reid et al., 2012). The system is important for recognising the biological process and events occurring at any given time or in response to infection (Jovanovic et al., 2015). The -omics approach has driven the movement of the studies to boarder a biological knowledge and overview all information in system.

There are several studies of apicomplexan parasite virulence related to -omics study. Xia et al. (2008) used 2D-electrophoresis, gel liquid chromatography coupled with tandem mass spectrometry and MudPIT to analyse the proteome of *T. gondii*. The draft genome of *T. gondii* predicted approximately 8,000 genes with varying degrees of confidence. Their data revealed how proteomics can perform the predictions which can discover new genes. Approximately one-third (2,252) of predicted proteins together with 2,477 intron-spanning peptides contributed supporting evident for correct splice site annotation. Another proteomics study was done by Regidor-

Cerrillo et al. (2012a). They studied the proteomics profile between virulent and attenuated *N. caninum* isolates using DIGE and MALDI-TOF MS method. The results revealed that changed in 60 protein spots; 39 spots were preferentially expressed in Nc Spain7 and 21 spots were abundant in Nc Spain 1H. The highest number of differentially expressed spots were shown in Nc Spain7. They found NcMIC1 decreased in abundance in Nc Spain 1H than Nc Spain 7 which might due to reduce invasiveness of this isolate. The ROP40 was found in Nc Spain7 which might play different role in both strains. The additional study of a large scale proteogenomics in *T. gondii* and *N. caninum* were performed by Krishna et al. (2015). They queried proteomics data against RNASeq derived gene models; a panel of official and alternate models. Additionally, several newly generated and many previously published MS datasets were used for proteomics analysis. Their result revealed the identification of loci that were not existing in current genome annotations. It could be indicated that RNA-seq can be used as a tool for validation of genome annotation models. Yang et al. (2013) identified *T. gondii* genes candidate which might represent for phenotypic variation among type I strains using comparative genomics and transcriptomics approaches. There were 1,394 single nucleotide polymorphisms (SNPs) and insertions/deletions (dels) between RH and GT1from whole genome sequencing. The achievement of gene candidates could represent for type I phenotypic differences by analysing SNPs/indels combined with differentially expressed of genes between RH and GT1 strains. The result revealed a polymorphism of GRA2 differences in the invasion of the interferon gamma response between RH and GT1 strains. Moreover, only RH strain recruited phosphorylated IkBa to PVM which was partially dependent on GRA2. However, the other strains in type I cannot recruit. The comparative genomics and transcriptomics study between *T. gondii* and *N. caninum* was performed by Reid et al. (2012). They sequenced the tachyzoites stage of *N. caninum* genome and transcriptome. The result showed an expansion of surface antigen gene families in *N. caninum* and the rhoptry kinase ROP18 is pseudogenised in *N. caninum* which leads *N. caninum* unable to phosphorylate host immunity-related GTPases like how *T. gondii* does.

Some of -omics studies related to *N. caninum* stage conversions were also performed. Marugan-Hernandez et al. (2010) investigated the differentially

expression proteins of tachyzoites and bradyzoites stage using DIGE couple with MS analysis. They found 72 spots were differentially expressed in which 53 spots were more abundant in bradyzoites and 19 spots were high abundant in tachyzoites. The 26 proteins were obtained and identified from MS in which 20 proteins were from bradyzoites and 6 proteins were from tachyzoites. The key proteins found in bradyzoites stage included glycolysis related proteins such as enolase and glyceraldehyde-3-phosphatedehydrogenase, stress response related such as HSP70 and HSP90 and GRA9. The more abundant protein related to krebs cycle as isocitrate dehydrogenase 2 was found in tachyzoites stage. The other transcriptomic study of stage conversion was performed by Vermont (2012). The RNA-Seq technology was employed to investigate the mRNA expression during tachzoites-bradyzoites conversion process. The key finding had shown that rhoptry proteins reduced abundance of transcripts expression in bradyzoites stage which might due to a less importance of invasive proteins in bradyzoites stage.

1.11 Aims and objectives

The overall aim of this thesis is to advance the understanding of *N. caninum* biology by investigating the systematic view of specific genes and proteins related to the parasites virulence as well as the key proteins during stage conversion from tachyzoites to bradyzoites. The biological information of *N. caninum* virulence will provide the divergent evolutionary biology between two different isolates. Moreover, the investigation of proteome involved in *N. caninum* stage conversion will provide more biological knowledge in the parasite life cycle.

In chapter two, the proteomics profile of *N. caninum* stage conversion from tachyzoites to bradyzoites were analysed using label-free quantitative techniques. The differential protein expression between the two phenotypically stages were compared to identify the key genes involved in each stage.

In chapter three, the transcriptomics analysis was used to compare differential transcript expression between low- and high- virulent of *N. caninum* using RNA-Seq approach. Two different isolates causing differential pathogenicity were investigated to visualise the similarity and different in term of transcripts expression.

In chapter four, label-free proteomics technique was used to compare the differential protein expression between low and high virulence isolates of *N. caninum.* The proteomics profile of both isolates were then analysed using bioinformatics tools such as progenesis.

Chapter 2: Analysis of protein expression during *N. caninum* tachyzoites to bradyzoites stage conversion

2.1 Introduction

There are three invasive stages in the life cycle of *N. caninum*: tachyzoites, bradyzoite-containing tissue cysts and sporozoite-containing oocysts. Tachyzoites are responsible for acute infection while the slowly replicating bradyzoites are responsible for establishing chronic infection (Buxton et al., 2002). *N. caninum* has to survive in a fluctuating environment, withstanding physiological and immunological changes from the host, and it has evolved a diverse set of regulatory mechanisms. (Eastick and Elsheikha, 2010). The mechanisms stimulating tachyzoite to bradyzoite stage conversion can be viewed as one such adaptive response (Marugan-Hernandez et al., 2010). Moreover, the reactivation of tissue cysts in an immunocompromised animal, especially during pregnancy, may cause bradyzoite to tachyzoite retransformation followed by placental infection and congenital fetal infection (Quinn et al., 2002). Knowledge of tachyzoite-bradyzoite stage conversion will help to clarify mechanisms of persistent *N. caninum* infection by determining key proteins that show high abundance in the bradyzoite stage.

2.1.1 Stage conversion in Apicomplexan parasites

The stage conversion phenomenon is common in organisms with complex life cycles in order to survive in the host, such as apicomplexan parasites. (Ferreira da Silva Mda et al., 2008; Schwan and Hinnebusch, 1998). Not only does the characteristic shape of tachyzoites differ from bradyzoites but the function of each stage is different also (Vonlaufen et al., 2004). Tachyzoites replicate rapidly while bradyzoites replicate slowly. Once the rapid replicating tachyzoites become stressed by the host immune response they change into the dormant cysts which are tolerant to host immunity. Stage conversion can be characterized with several techniques. Morphological changes affecting cellular ultrastructure can be compared using transmission electron microscopy (TEM) (Jardine, 1996; Kang et al., 2008; Speer et al., 1999). For example, bradyzoites are slimmer than tachyzoites and the nucleus of bradyzoite is located at the posterior end while it is located at the middle of tachyzoites. Moreover, the bradyzoite contains more rhoptries, amylopectins, and micronemes, while dense granules are numerous in tachyzoites stage (Kang et al.,

2008). Immunohistochemistry and immunofluorescent assays are able to differentiate between stages by labelling the parasites with *N. caninum* stage-specific antibodies, for example, *N. caninum* tachyzoite surface antigens NcSAG1 and NcSRS2 (Fuchs et al., 1998; Vonlaufen et al., 2004) and *T. gondii* bradyzoite antigen1(TgBAG1). TgBAG1 have shown cross reactivity to *N. caninum* bradyzoite antigen 1 (McAllister et al., 1996). There are three more proteins determined as bradyzoite specific antigens including NcSAG4, NcBSR4 (Fernandez-Garcia et al., 2006; Risco-Castillo et al., 2007) and SRS13 (Reid et al., 2012). Additionally, there are some proteins identified as tachyzoite specific proteins such as surface antigens SAG1 and SRS2 (Hemphill et al., 1997b; Howe et al., 1998), MIC1 (Keller et al., 2002) and GRA1, GRA2, GRA3, GRA4, GRA5 and GRA6 (Fuchs et al., 1998; Tilley et al., 1997).

To study stage conversion process, there have been many techniques developed to induce stage conversion and encystation *in vitro.* For example, changing the pH of growing medium and increasing growing temperature (Weiss et al., 1999). It can also achieve by stressing the culture with chemical agents such as nitric oxide (NO) and NO donor sodium nitroprusside (SNP) (Vonlaufen et al., 2004; Vonlaufen et al., 2002) as well as inducing bradyzoites specific marker (BAG1) expression (McAllister et al., 1996; Risco-Castillo et al., 2004; Vonlaufen et al., 2002).

2.1.2 The proteomics studies of stage conversion in *N. caninum*

Proteomics is one of valuable tool to study protein expression in tachyzoites and bradyzoites of *N. caninum* (Fuchs et al., 1998; Kang et al., 2008). Vonlaufen et al., (2004) developed the technique for initiating *Neospora stage* conversion from tachyzoites to bradyzoites stage by using SNP. They found that 17 µM SNP could induce the stage conversion while not harming the culture. Sodium dodecyl sulfate-polyacrilamide gel electrophoresis (SDS-PAGE) and immunoblotting were used to investigate the protein differential expression in each parasite stage. Their results showed that bradyzoite specific protein BAG1 preferentially increased in expression while the expression of tachyzoites surface antigens NcSAG1 and NcSRS2 were reduced during stage conversion. *Neospora dense* granule proteins such as GRA1,

GRA2 and GRA7 were shown incoporated in cyst wall of bradyzoites. Another study was done by Kang et al. (2008), who investigated the differential protein expression of *Neospora tachyzoites* and bradyzoites stage in different time point including day 3, 5, 7 and 9 post induction with 70 µM SNP. Proteomics analysis was carried out using SDS-PAGE and western blot. They found that the tachyzoite specific antigens were identified at ~ 40 kDa MW in the tachyzoite stage and were not seen at the bradyzoite stage. While rabbit polyclonal antibodies against bradyzoites detected the significant increase in expression of bradyzoites stage at MW ~25 kDa by western blotting. BAG1 was used as a marker to determine stage conversion by IFA. Marugan-Hernandez et al. (2010) identified differentially expressed proteins during stage conversion using difference gel electrophoresis (DIGE) technology followed by MS analysis. They found 72 spots were differentially expressed ($p < 0.05$ and fold change >1.5). Among these protein spots, 53 spots showed high abundant in bradyzoites and 19 spots were more abundant in tachyzoites. There were 26 proteins identified by MS analysis in which 20 proteins were of higher abundance in bradyzoites and 6 proteins showed preferentially expression in tachyzoites. Proteins with high abundance in bradyzoites include enolase, glyceraldehyde-3-phosphate dehydrogenase, heat shock protein HSP70 and 90, and dense granule GRA9. Proteins with higher abundance in tachyzoites was isocitrate dehydrogenase2. Another study investigating stage conversion of *N. caninum* was done by (Sarah, 2012) using transcriptomics apporach. They investigated mRNA level during tachyzoites-bradyzoites interconversion using Illumina Hi-Seq2000 next generation sequencing technology. Stage conversion was induced using 50 µM SNP and the samples were collected at day 0, 1, 3 and 6 post induction. The result focused on on three groups of apical proteins which showed that the rhoptry gene expression decreased over the experimental time course; microneme genes expression displayed both increases and decreases, while the dense granule genes were not significantly differentially expressed.

2.1.3 Aims and Objectives

The aim of this study is to investigate changes in protein expression during the stage conversion from *N. caninum* tachyzoites to bradyzoites, over a six-day experimental course.

Objective of this study:

1. To undertake *in vitro* stage conversion of *N. caninum* using SNP with confirmation by IFA
2. To identify and quantify differentially-expressed proteins during stage conversion using a label-free quantitative proteomics approach

2.2 Materials and Methods

2.2.1 Cell culture

Cultivation of Vero cells (African green monkey kidney fibroblasts) were maintained in DMEM (Dulbecco's Modified Eagle's medium, Sigma-Aldrich®) containing 10% foetal bovine serum (Sigma-Aldrich®), 100 U/ml penicillin and 100 µg/ml streptomycin sulphate (Sigma-Aldrich®). The medium was passed through a filter prior to use. The amount of $1x10^5$ Vero cells in 5 mL medium was added into 25 cm^2 flasks (T25). The flasks were placed in a 37 ^oC incubator with 5% CO_2 and observed daily.

2.2.2 *Neospora* tachyzoites infection

The purified *N. caninum* Liverpool strain tachyzoites $1x10^5/cm^2$ were added to the Vero cells culture flask which had shown 10-20% confluent (about 24 hours after passaging). The infected Vero cells with *N. caninum* were placed at 37^oC in a 5% CO_2 incubator.

2.2.3 Isolation of *N. caninum* tachyzoites

Briefly, the parasite infected Vero cells were scraped and transferred to 50 ml centrifuge tube. The culture medium containing *N. caninum* and Vero cells were suspended through a 20-gauge needle and put through a 47mm diameter 3 µm nucleopore™ polycarbonate membranes (Whatman) in order to purify the parasite. The parasite was washed with cold phosphate buffered saline (PBS) and centrifuged twice at 1500 x g for 10 minutes at 4 ^oC. Then parasite pellets were re-suspended in 2 mL of cold PBS. The purified parasites were centrifuged at 13,000 x g for 10 minutes at 4 ^oC. The supernatant was discarded. The parasite pellets were kept at -80 ^oC prior to use.

2.2.4 *Neospora* stage conversion induction using sodium nitroprusside

The *Neospora* stage conversion was done by adding sodium nitroprusside (SNP) (Sigma-Aldrich®) at a final concentration of 17 µM (stock SNP solution 10 mg/ml in distilled water) at the time of infection. The infected Vero cells culture with *N. caninum* were incubated in 37 ºC with 5% CO_2. The fresh medium containing 17 µM SNP were replaced every day up to 6 days. The tachyzoites sample of *N. caninum* were collected at three different times as day 0 of induction (day 3 post infection without SNP treatment), day 3 (day 6 post infection) and day 6 (day 9 post infection) respectively. The parasite were harvested using the same protocol as in section 2.2.2. Three biological replicates were collected for each time point (Table 2.1).

In addition, the parallel experiment was established using a SNP concentration of 70 µM to induce the stage conversion. However, the Vero cells started to detach after 3 days of SNP treatment and Vero cells were unhealthy. Moreover, the additional stage conversion of day 8 post SNP treatment (day 11 post infection) were also attempted in order to visualise more bradyzoite forming but the Vero cells were mostly detached at day 6 post SNP treatment.

Table 2.1 Description of samples of tachyzoites/bradyzoites stage conversion collected at Day 0, 3 and 6 post SNP treatment

Samples	Days duration after SNP treatment	Days duration after *N. caninum* infection
1	0	3
2	3	6
3	6	9

2.2.5 Immunofluorescence assay (IFA) to determine stage conversion

The IFA was used to confirm the stage conversion by observing the presence of bradyzoite-specific marker BAG1. The IFA slide was coated with poly-L lysine (molecular weight 70,000-150,000, Sigma Aldrich®) for 20 mins at room temperature. Vero cells was cultured on the slides (in parallel with the culture in the flasks) at the concentration 1×10^4 cells per well and kept in the incubator at 37 ^oC with 5% CO_2. The Vero cell culture was observed daily until 80-90% confluent. Between $5x10^3$-$1x10^4$ tachzoites of the *N. caninum* Liverpool strain were added to each well containing Vero cells. The induction buffer with 17µM SNP, 10% FBS and 1% PS in DMEM was made and added to the culture on day 3 post induction. Same medium without SNP was added to Vero cells as a control. The medium was changed daily.

The process of IFA were carried out as follows. Parasite-infected cells were washed three times with PBS, then fixed and permeablised with 4% (v/v) formaldehyde and 0.5% (v/v) TritonX-100 in PBS. The slides were incubated at room temperature for 15 minutes, followed by three washes with PBS. The slides were blocked with 5% (v/v) normal goat serum twice for ten minutes. After that, all slides were washed with 1% (v/v) FBS/PBS. The BAG1 primary antibody polyclonal rabbit antiserum (kindly provided by Prof. Andrew Hemphil, University of Bern, Switzerland) was diluted into 1:200 with 1% (v/v) FBS/PBS and put on the slides, then incubated at room temperature for one hour. The slides were quickly washed six times with PBS and a secondary antibody Alexafuor 488 was added (kindly provided by Dr. Olivier Touzelet, University of Liverpool) which was diluted into 1:1000 with 1% (v/v) FBS/PBS and incubated at room temperature for one hour. The slides were briefly washed for six times with 1% (v/v) FBS/PBS and then with PBS followed by six times washeswith ddH_2O. The slides were mounted with Vectashield containing DAPI and covered with foil, left to dry and kept in the fridge until observing under fluorescence microscopy.

2.2.6 Sample digestion for proteomics analysis

This step was performed by Dr. Dong Xia, Department of Infection Biology, University of Liverpool. A sample was diluted in 25 mM ammonium bicarbonate to a volume of 160 µl, then added 10 µl of 1% Rapigest (Waters) (final concentration 0.05% (w/v). Samples were incubated at 80 ^{o}C for 10 mins. A total of 10 µl of 9.2 mg/kg ditiothreitol reagent (DTT) was applied to each sample (at a final concentration of 3 mM) and incubated at 60 ^{o}C for 10 minutes. Then a total of 10 µl of 33 mg/ml iodoacetamide was added (at a final concentration of 9 mM) to each sample and incubated at room temperature in the dark. After that a total 10 µl of trypsin (200 µg/ml of 25 mM ammonium bicarbonate) was added to sample at the ratio of 50:1 protein: trypsin and incubated at 37 ^{o}C overnight. Trifluoroacetic acid (TFA) was added to all samples at a final concentration of 0.5% (v/v)) and incubated at 37 ^{o}C for 45 minutes. The samples were centrifuged at 13,000 x g at 4 ^{o}C for 20 minutes. The supernatant of each sample was collected and kept in fridge (4 ^{o}C) prior to mass spectrometry analysis. Digest was loaded onto a 12% SDS-PAGE to determine the complete digestion of proteins as a quality control

2.2.7 Liquid chromatography mass spectrometry LC-MS/MS

This procedure was performed by Dr. Dong Xia, University of Liverpool. Separated peptides in samples were analysed by on-line nanoflow LC using the Thermo Ultimate 3000 nano system (Thermo Fisher Scientific) attached with Q-Exactive mass spectrometer (Thermo Fisher Scientific). Examined samples were put on a 50 cm Easy-Spray column with 75 µm internal diameter contained 2µm C18 particles attached to a silica nano-electrospray emitter (Thermo Fisher Scientific). The column was performed under 35^{o}C. Chromatography was done with a buffer containing 0.1% formic acid (buffer A) and 80% ACN in 0.1% formic acid (buffer B). The separation of peptides was performed by a linear gradient of 5-50% buffer B of 120 min at flow rate 300 nl/min. The operation of Q-Exactive were achieved in data-dependent mode with survey scans required at a resolution of 70,000. The first top ten abundant isotope pattern with charge conditions +2, +3 and/or +4 from survey scan were chosen with an isolation window of 2.0. They were fragmented by higher energy

collisional dissociation and normalized collision energies of 30. The highest ion injection times for the survey scan was 250 ms and the MS/MS was 100 ms. The ion target value was set for survey scans and MS/MS at 1e6 and 1e4, respectively. Repetitive sequencing of peptides was reduced to 20 s through dynamic exclusion of peptides sequence.

2.2.8 MS data processing by Progenesis

The raw data obtained from LC-MS/MS were imported to Progenesis^{TM} QI software (Version 3.0, Nonlinear Dynamics). Data were divided into three groups following the induction time points as day 0, 3 and 6 respectively. The data of MS scans were transformed to peak lists. Then sample reference was chosen among tested samples by the software after assessing the 2-D mapping (*m/z* against retention time) for LC variability. The rest of samples were aligned to the reference by retention time. After the peak picking, spectra were exported for peptide identification using Mascot (version 2.3.02, Matrix Science) against the sequence database. The identified peptides were imported back to Progenesis and the data were normalised against all identified peptides. Protein abundance (iBAQ) was calculated as the sum of all the peak intensities (from Progenesis output) divided by the number of theoretically observable tryptic peptides (Schwanhausser et al 2011). Protein abundance was normalised by dividing the protein iBAQ value by the summed iBAQ values for that sample. The reported abundance is the mean of the biological replicates.

Q values are the adjusted p-value which are using an optimised FDR)False Discovery Rate(approach. FDR is the proportion of false positive expected to get from a test. P- value present the probability of false positive from one test while q – value is a measure of significance in term of FDR associated with each tested feature. Q value is needed when thousands of variables)e.g. gene expression levels(have to measure. So an adjustment of p-value is necessary when the multiple tests are making on the same (Storey, 2003).

2.2.9 Protein identification by Mascot

Mascot (Version 2.3.02, Matrix Science) is a search engine to identify protein from MS spectra. The spectra were also compared to a decoy database containing reversed peptide sequences to obtain a false discovery rate (FDR). Protein sequences were downloaded from ToxoDB database of *N. caninum* Liverpool strain (ToxoDB, Version 11). The search parameters of Mascot were set with precursor mass tolerance of 10 ppm and fragment mass tolerance of 0.01 Da. Carbamidomethylation of cysteine (C) was set as a fixed modification and oxidation of methionine (M) was established as a variable modification. One missed tryptic cleavage site was allowed with charge states +1, +2 and +3. The protein identifications were made at FDR < 0.01 with a minimum of two unique peptides per protein.

2.2.10 Cluster analysis

This method was done in order to visualise the trend of protein expressions over the stage induction time course. The GProX software package version 1.1.12 (Rigbolt et al., 2011) was employed to generate the cluster analysis. The clustering is based on fuzzy c-means algorithm as a fuzzification value of 2 for 100 iterations (Futschik and Carlisle, 2005). The similar protein expression trends were grouped together into 6 clusters.

2.2.11 Gene Ontology enrichment analysis

GO enrichment analysis was performed to interpret the biological relevance of the protein clusters. topGO package (Alexa and Rahnenfuhrer, 2016) was used for obtaining GO terms enriched for each cluster. The R scripts for topGO was kindly provided by Dr. Vicky Hunt, University of Liverpool.

2.3 Results

IFA was used to verify the stage conversion of *N. caninum*. The BAG1 bradyzoite-specific antibody was detected in *N. caninum* infected Vero cells at day 5 post induction while this antibody was not detected in the control samples (Figure 2.1).

2.3.1 Differential protein expression and cluster analyses

A total number of 1,475 proteins were identified and quantified by Progenesis ^{TM}. There were 299 proteins showing preferential expression at different point during the stage conversion time course with statistical significance (unique peptides ≥ 2, adjusted p value < 0.05 and fold change ≥ 2). All 229 proteins were clustered into 6 groups following their expression trend using GProX (Figure 2.2). A majority of proteins showed a downward trend in expression over the time course (Cluster1 and 6, n = 143). Proteins in cluster 5 (n=23) showed an upward trend from day 0 to day 3 and then dropped from day 3 to day 6. On the other hand, cluster 3 (n=42) and cluster 4 (n=47) exhibited an upward trend throughout the time course while cluster 2 (n=44) showed a downward trend from day 0 to day 3 and a slightly upward from day 3 to day 6.

As expected, most of microneme and dense granule proteins showed a downward trend in abundance throughout the time course for example one unidentified microneme protein (cluster 6), MIC7 (cluster1), MIC2, MIC3, MIC6, MIC10 and MIC11 (cluster5) as well as GRA1, GRA3 (cluster1), GRA6 and GRA7 (cluster 6).

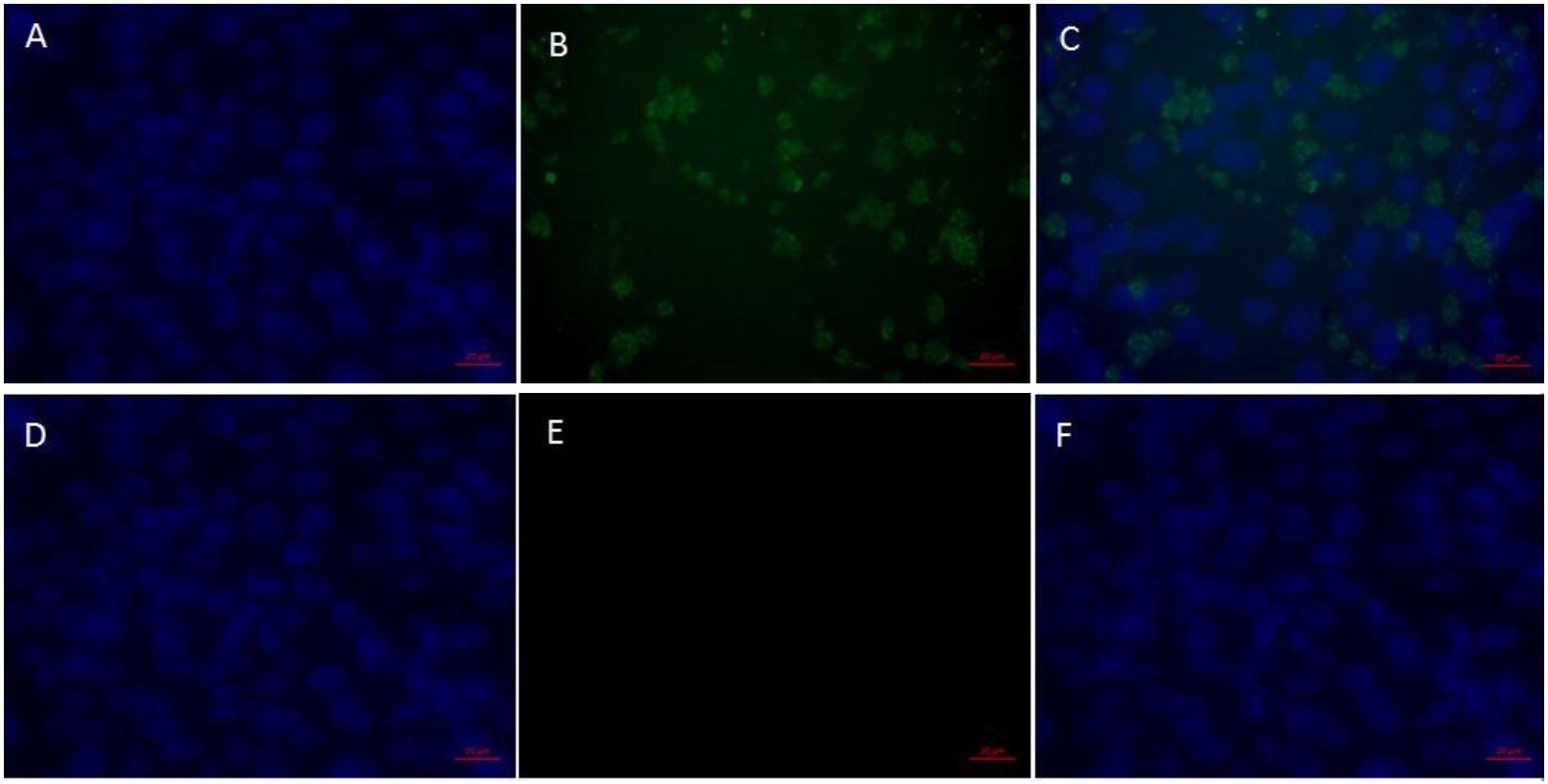

Figure 2.1 Immunofluorescence image of *N. caninum* infected Vero cells inducing stage conversion with 17μM SNP at day 5 post induction A and D) Vero cells staining with DAPI; B and E) *N. caninum* staining with BAG1 bradyzoite-specific antibody (Alexafuor 488-green) C and F) Combining image of Vero cells infected with *N. caninum* Images A,B,C are a stage induction with SNP and Images D,E,F are control without SNP

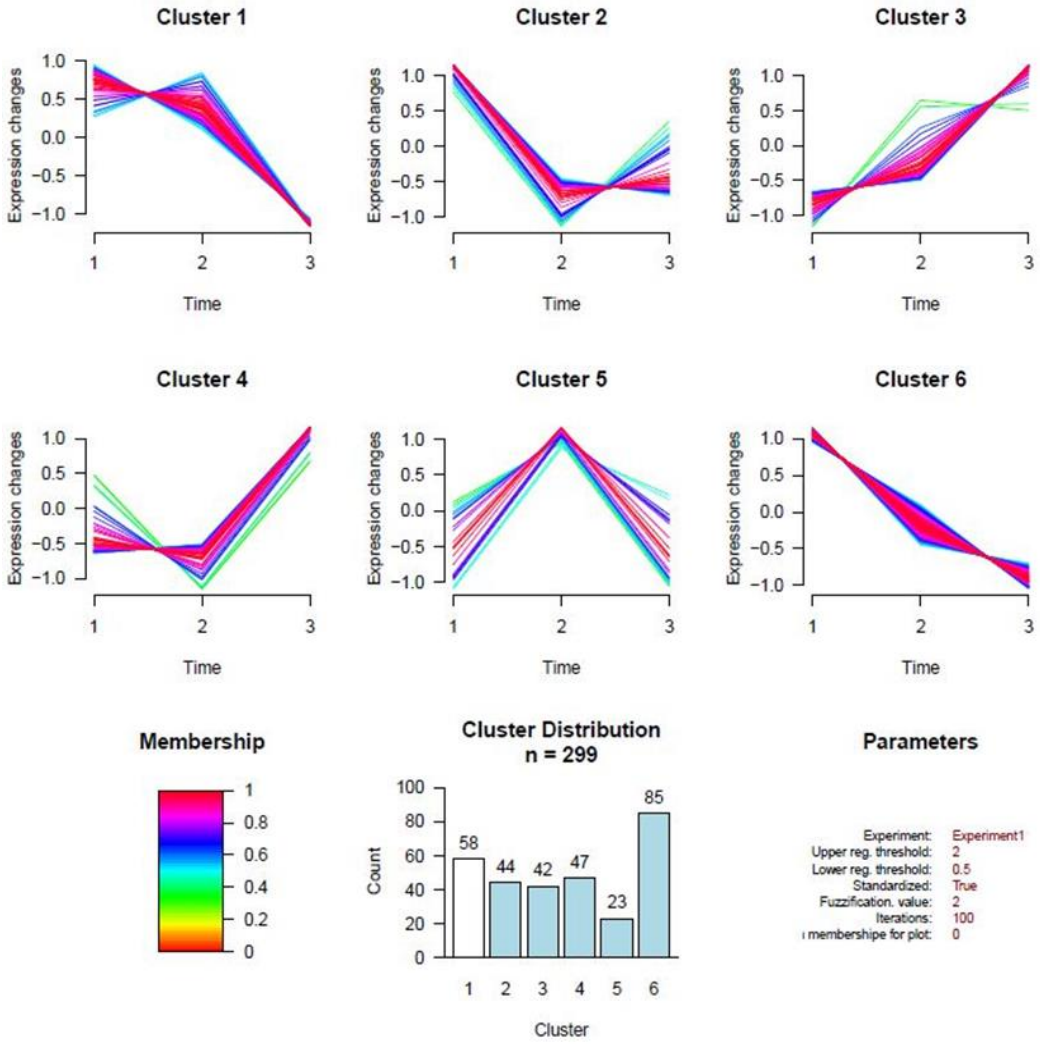

Figure 2.2 Cluster analyses of *N. caninum* stage conversion from tachyzoites to bradyzoites stage using SNP treatment. The X-axis shows the experimental time point as 1 = Day 0, 2 = Day 3 and 3 = Day 6. The Y-axis reveals the protein expression changes of *N. caninum.*

Proteins in cluster 1 (Table 2.2, Appendix Table I and Figure 2.2) revealed no change in expression from day 0 to day 3 followed by decreasing expression to day 6. The proteins with the highest fold changes include a putative calmodulin CAM2, ubiquitin family protein and a GAP40 protein, which is associated with parasite movement and host cell invasion. A putative cyclophilin that has roles in parasite biochemistry and host parasite interaction also shows downward trend. Apical proteins found in this cluster include secretory proteins such as microneme protein MIC7, dense granule protein GRA3 and GRA1 as well as rhoptry protein ROP17. This finding indicates that stage conversion of tachyzoite to bradyzoites was successful, as the apical secretory proteins are expected to be less abundant by day 6.

Table 2.2 Top twenty most differentially expressed proteins of cluster 1 showing downward expression trend throughout three time points

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
NCLIV_025600	putative calmodulin	calmodulin CAM2 (CAM2)	265739.36	263341.31	25032.09	10.61595	0.00139
NCLIV_043870	conserved hypothetical protein	hypothetical protein	632804.95	491318.85	83413.24	7.58639	0.00066
NCLIV_036200	putative cyclophilin	peptidyl-prolyl cis-trans isomerase family 1	440164.70	311944.18	61153.88	7.19766	0.00194
NCLIV_053590	WD-40 repeat protein, related	transducin beta-like protein TBL1 (TBL1)	1078432.05	1120640.79	167584.27	6.68703	0.00339
NCLIV_043320	rna recognition motif (RRM)-containing protein,related	U2 snRNP auxilliary factor, large subunit, splicing factor subfamily protein	999741.20	910921.59	149615.12	6.68209	0.00293
NCLIV_017040	Phosphofructokinase, related	phosphofructokinase PFKII (PFKII)	846932.48	613245.67	156661.14	5.40614	0.00459
NCLIV_047600	hypothetical protein	hypothetical protein	245166.59	265982.91	50582.39	5.25841	0.00485
NCLIV_064340	putative prefoldin subunit	prefoldin subunit 6, putative	252216.14	225237.83	49038.51	5.14323	0.00198
NCLIV_042770	hypothetical protein	hydrolase, NUDIX family protein	191186.70	246180.35	48626.41	5.06269	0.02926
NCLIV_048800	hypothetical protein	ubiquitin family protein	1616946.72	1243199.11	319526.01	5.06045	0.00152
NCLIV_019600	hypothetical protein	clathrin adaptor complex small chain subfamily protein	584126.80	516189.27	119401.46	4.89212	0.00175
NCLIV_053360	conserved hypothetical protein	transcription elongation factor SPT6 (SPT6)	133160.44	107140.26	30173.92	4.41310	0.00156
NCLIV_001770	putative DNAK family domain containing protein	tetratricopeptide repeat-containing protein	1353555.83	938777.78	321447.71	4.21081	0.00267
NCLIV_056590	putative phosphatidylinositol-4-phosphate 5-kinase	MORN repeat-containing protein	1058324.45	896339.46	251841.30	4.20235	0.00751
NCLIV_024440	hypothetical protein	vacuolar protein sorting-associated protein 26, putative	265255.35	220355.64	66567.25	3.98477	0.00371
NCLIV_056360	Eukaryotic initiation factor, related	DEAD (Asp-Glu-Ala-Asp) box polypeptide DDX6 (DDX6)	512884.26	586998.49	151636.41	3.87109	0.02395
NCLIV_054550	Cytochrome c oxidase subunit 2, related	cytochrome C oxidase subunit IIb, putative	3972133.77	2765119.30	1059769.01	3.74811	0.00058
NCLIV_019450	hypothetical protein	signal peptidase	551941.50	669909.50	192573.35	3.47872	0.00073
NCLIV_025710	unspecified product	microneme protein MIC7 (MIC7)	4010693.96	3108615.04	1165075.86	3.44243	0.00528
NCLIV_005190	putative transcription elongation factor FACT 140 kDa	transcriptional elongation factor FACT140 (FACT140)	1102959.20	893566.69	329756.63	3.34477	0.00230

Cluster 2 (Table 2.3, Appendix Table II and Figure 2.2) revealed decrease in expression from day 0 to day 3 followed by minimal change until day 6. Most of the proteins in the top twenty differentially-expressed proteins were hypothetical proteins. Some notable proteins in this cluster are a putative U4/U6 small nuclear ribonucleoprotein, which is required to assemble an active spliceosome, and catalase, which is responsible for degradation of hydrogen peroxide to water and oxygen. Moreover, some key proteins involved in host cell invasion are found in this cluster, for example, two rhoptry proteins as ROP6 and ROP9, rhoptry neck protein RON10 and dense granule protein DG32, two myosin A proteins, rhomboid protease ROM4, glideosome-associated protein with multiple-membrane spans GAPM3 (NCLIV_035190) and glideosome-associated protein with multiple-membrane (NCLIV_044350). This finding suggests that lower expression of these invasion protein is required after the differentiation of the parasite to quiescent stage.

Table 2.3 Top twenty most differentially expressed proteins of cluster 2 showing downward expression trend and then not change in expression

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
NCLIV_021510	conserved hypothetical protein	DUF3228 domain-containing protein	396479.64	66940.64	20525.99	19.31598	0.01330
NCLIV_021410	conserved hypothetical protein	hypothetical protein	1934838.90	217867.71	638107.78	8.88080	0.00297
NCLIV_056730	conserved hypothetical protein	hypothetical protein	588892.91	92014.19	80675.27	7.29955	0.01047
NCLIV_020430	hypothetical protein	cleft lip and palate transmembrane protein 1 (clptm1) protein	714046.52	151961.38	122149.83	5.84566	0.04728
NCLIV_018260	putative U4/U6 small nuclear ribonucleoprotein	WD domain, G-beta repeat-containing protein	155932.92	44142.10	27703.00	5.62874	0.03875
NCLIV_040600	hypothetical protein	signal recognition particle receptor beta subunit protein	5691004.50	1028281.60	1039360.97	5.53448	0.03800
NCLIV_008860	hypothetical protein	UDP-galactose transporter family protein	671184.72	203690.15	142488.38	4.71045	0.00485
NCLIV_065570	putative translation initiation factor SUI1	translation initiation factor SUI1, putative	1995635.11	558339.68	505226.56	3.94998	0.00066
NCLIV_046970	conserved hypothetical protein	hypothetical protein	1103666.88	301124.81	411043.97	3.66515	0.00993
NCLIV_048040	conserved hypothetical protein	hypothetical protein	1109215.06	316639.53	401233.78	3.50308	0.01931
NCLIV_051590	putative GTP binding protein	GTP-binding protein, putative	8497994.99	3170482.10	2547632.15	3.33564	0.00293
NCLIV_036630	putative 14-3-3 protein	14-3-3 superfamily protein	153305.83	46317.57	114537.63	3.30989	0.02400
NCLIV_004510	putative kelch motif domain-containing protein	kelch repeat-containing protein	136045.70	41180.82	88555.08	3.30362	0.04414
NCLIV_005050	putative coatomer delta subunit	adaptor complexes medium subunit family protein	3716585.02	1425511.92	1139790.55	3.26076	0.00156
NCLIV_005040	conserved hypothetical protein	hypothetical protein	2968418.05	943944.08	1068584.01	3.14470	0.00273
NCLIV_032240	Catalase (EC 1.11.1.6), related	catalase	66004502.11	21314140.59	23758910.72	3.09675	0.00282
NCLIV_058300	hypothetical protein	hypothetical protein	2576396.59	840942.93	1649767.24	3.06370	0.03922
NCLIV_035190	conserved hypothetical protein	glideosome-associated protein with multiple-membrane spans GAPM3	3676957.38	1267464.45	2785751.42	2.90103	0.00854
NCLIV_045490	conserved hypothetical protein	hypothetical protein	1081874.99	394781.00	617225.23	2.74044	0.00548
NCLIV_058000	putative alanine dehydrogenase	NAD(P) transhydrogenase subunit beta, putative	12759821.55	4755062.00	5570964.71	2.68342	0.02397

Proteins in cluster 3 (Table 2.4, Appendix Table III and Figure 2.2) showed an upward expression trend throughout the time course. The proteins found in this cluster include a ubiquitin family domain containing protein, HEAT repeat-containing protein, and stress indicator proteins such as heat shock proteins HSP60 and HSP90. Proteins involved with parasite movement such as microtubule associated protein SPM2 (SPM2) and putative kinesin heavy chain, the enzyme create energy storage such as ATPase synthase subunit alpha and ATP synthase beta subunit ATP-B (ATPB) were also in this cluster. However, some proteins potentially involved in host cell invasion were also found in this cluster such as ROP35, myosin I and camoludin. This may reflect the fact that some tachyzoites remain in the culture at day 6, as not all parasites have converted into bradyzoites.

Table 2.4 Top twenty most differentially expressed proteins of cluster 3 showing upward expression trend throughout three time points

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
NCLIV_012950	putative ubiquitin family domain-containing protein	ubiquitin family protein	32183.17	352394.63	1089634.23	33.85726	0.00250
NCLIV_001290	conserved hypothetical protein	HEAT repeat-containing protein	385271.78	820139.66	4696425.55	12.18990	0.00139
NCLIV_035590	conserved hypothetical protein	hypothetical protein	1513363.22	9352111.17	12779680.14	8.44456	0.01333
NCLIV_033910	hypothetical protein	WD domain, G-beta repeat-containing protein	665310.80	1515968.93	5159820.96	7.75550	0.00714
NCLIV_012600	hypothetical protein	UBX domain-containing protein	1216519.76	3110979.71	9384912.85	7.71456	0.01076
NCLIV_031460	conserved hypothetical protein	hypothetical protein	4253666.60	7494357.69	28251210.99	6.64161	0.00156
NCLIV_066970	putative enoyl-acyl carrier reductase	enoyl-acyl carrier reductase ENR (ENR)	7574215.54	22602986.54	41683983.44	5.50341	0.04438
NCLIV_015620	60S ribosomal protein L36, related	ribosomal protein RPL36 (RPL36)	3902566.58	7980362.45	18483879.84	4.73634	0.00329
NCLIV_044410	unspecified product	rhoptry kinase family protein ROP35 (ROP35)	20917373.66	28776282.42	95433729.32	4.56241	0.00762
NCLIV_038590	conserved hypothetical protein	hypothetical protein	150718.32	429752.79	655892.18	4.35177	0.00589
NCLIV_065440	hypothetical protein	calmodulin, putative	8454788.39	15033666.83	34716268.40	4.10611	0.00415
NCLIV_065210	KLLA0F09449p, related	heat shock protein HSP60 (HSP60)	36358177.28	54507306.08	140030918.30	3.85143	0.00569
NCLIV_014050	conserved hypothetical protein	microtubule associated protein SPM2 (SPM2)	522829.03	827953.23	1913522.53	3.65994	0.01407
NCLIV_030650	putative 26S protease regulatory subunit 6b	26S proteasome regulatory subunit 6b, putative	1195117.95	1771237.37	4252425.39	3.55816	0.00085
NCLIV_024860	Proteasome/cyclosome repeat family protein,related	Proteasome/cyclosome repeat-containing protein	2564313.44	3917997.98	9027531.44	3.52045	0.00419
NCLIV_006060	conserved hypothetical protein	hypothetical protein	728248.46	1356145.96	2366549.83	3.24965	0.00655
NCLIV_034090	putative kinesin heavy chain	kinesin heavy chain, putative	2048222.47	4914383.62	6617486.51	3.23084	0.01873
NCLIV_022370	Ribophorin I, related	hypothetical protein	125163.22	218550.97	383389.85	3.06312	0.02062
NCLIV_002190	conserved hypothetical protein	hypothetical protein	1368076.06	2540517.20	4097414.88	2.99502	0.01794

Cluster 4 (Table 2.5, Appendix Table IV and Figure 2.2) exhibited unchanged protein expression from day 0 to day 3, followed by increasing expression from day 3 to day 6. Proteins involved in parasite growth are seen in this cluster, such as serine/threonine-protein phosphatase 2A and serine/threonine-protein phosphatase PP1. This might due to the bradyzoite formation. Enzymes related to energy metabolism such as isocitrate dehydrogenase as well as the proteasome (Prosome, macropain) subunit, beta type, 1, which is related to degrade the unnecessary proteins by proteolysis, were also seen in this cluster. An activator of HSP90 ATPase 1 family protein was identified, which is related to the stress condition as from the parasite pressure from the treatment. Moreover, surface antigen such as SAG-related sequence SRS35A, rhoptry kinase family protein, rhoptry protein and actin, which are associated with host cell invasion, were found in this cluster. This could be caused by some tachyzoites still being in the culture at day 6.

Table 2.5 Top twenty most differentially expressed proteins of cluster 4 showing not change or slightly down expression trend at the beginning and then upward expression trend

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
NCLIV_020970	conserved hypothetical protein	GYF domain-containing protein	451657.79	625660.82	9665977.74	21.40111	0.00032
NCLIV_019580	SRS domain-containing protein	SAG-related sequence SRS35A (SRS35A)	303766.07	373644.29	3658692.90	12.04444	0.00114
NCLIV_004250	putative nuclear RNA binding protein	hypothetical protein	2796716.17	4026163.48	23359224.10	8.35238	0.00066
NCLIV_060500	conserved hypothetical protein	hypothetical protein	37674990.75	53448716.98	297628121.50	7.89989	0.00417
NCLIV_016250	putative cytidine deaminase	cytidine and deoxycytidylate deaminase zinc-binding region domain-containing protein	1710510.18	2416250.81	12384871.19	7.24045	0.00066
NCLIV_007800	unspecified product	hypothetical protein	57927434.60	69343600.66	343683829.60	5.93301	0.00133
NCLIV_031700	DEHA2G03652p, related	LsmAD domain-containing protein	528306.96	101558.13	579518.71	5.70628	0.04211
NCLIV_048330	hypothetical protein	serine/threonine-protein phosphatase 2A catalytic subunit beta isoform, putative	354085.24	232071.42	1157635.38	4.98827	0.00392
NCLIV_007510	conserved hypothetical protein	hypothetical protein	2835411.64	2798676.06	13455993.20	4.80799	0.00473
NCLIV_039280	isocitrate dehydrogenase-like protein, related	isocitrate dehydrogenase	3402418.68	4394418.36	16146865.25	4.74570	0.00196
NCLIV_017840	conserved hypothetical protein	membrane protein	3742394.05	3674093.12	15954504.80	4.34243	0.00170
NCLIV_016330	putative phosphomannomutase 2	phosphomannomutase	78653.89	66253.03	273498.16	4.12809	0.02291
NCLIV_031970	hypothetical protein	pre-mRNA processing splicing factor PRP8 (PRP8)	4616679.00	5443197.93	18867085.16	4.08672	0.00499
NCLIV_041270	hypothetical protein	RuvB family 2 protein	957256.25	658961.86	2433646.21	3.69315	0.00118
NCLIV_055690	hypothetical protein	small GTPase Rab2, putative	2685377.41	2725196.60	9531330.85	3.54934	0.00298
NCLIV_063230	hypothetical protein	CRAL/TRIO domain-containing protein	989286.38	817158.05	2592223.66	3.17224	0.02502
NCLIV_051820	hypothetical protein	small GTP binding protein rab1a, putative	7441716.63	7966699.96	23471753.59	3.15408	0.00367
NCLIV_054750	hypothetical protein	serine/threonine phosophatase PP1	6688114.92	5946345.40	18412003.87	3.09636	0.00332
NCLIV_009170	proteasome (Prosome, macropain) subunit, beta type, 1, related	proteasome subunit beta type 2, putative	1264325.79	808412.90	2493083.11	3.08392	0.01410
NCLIV_052510	hypothetical protein	hypothetical protein	5436994.54	5025834.72	14992101.37	2.98301	0.00937

Cluster 5 (Table 2.6, Appendix Table V and Figure 2.2) revealed upward expression trend from day 0 to day 3 and downward trend from day 3 to day 6. The majority of proteins found in this chapter was microneme proteins such as MIC2, MIC 3, MIC 6, MIC 10 and MIC 11 and two surface antigen found as SAG-related sequence SRS16E(SRS6) and surface antigen repeat-containing protein.

Table 2.6 Top twenty most differentially expressed proteins of cluster 5 showing upward expression and then downward expression trend

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
NCLIV_042540	putative PBS lyase HEAT-like repeat domain-containing protein	HEAT repeat-containing protein	120685.78	233588.55	25276.21	9.24144	0.00337
NCLIV_001550	putative ribonucleoside-diphosphate reductase, large subunit	ribonucleoside-diphosphate reductase large chain	87340.33	583783.56	418757.21	6.68401	0.00638
NCLIV_051450	putative centromere/microtubule binding protein	rRNA pseudouridine synthase	223142.17	1220448.03	675565.49	5.46937	0.00778
NCLIV_033680	Solute carrier family 25 (Mitochondrial carrier,dicarboxylate transporter), member 10, related	2-oxoglutarate/malate translocase OMT (OMT)	329636.97	1525846.96	793626.42	4.62887	0.00066
NCLIV_013700	putative CHCH domain-containing protein	CHCH domain-containing protein	856631.00	3695217.99	2024348.95	4.31366	0.01140
NCLIV_032940	putative adenosine transporter	nucleoside transporter protein	891615.53	2919848.11	767821.02	3.80277	0.00226
NCLIV_037590	conserved hypothetical protein	hypothetical protein	1270323.18	2177673.77	611453.73	3.56147	0.04333
NCLIV_032950	putative deoxyuridine 5'-triphosphate nucleotidohydrolase	deoxyuridine 5'-triphosphate nucleotidohydrolase, putative	564550.47	1609532.21	768416.03	2.85100	0.00751
NCLIV_010050	srs domain-containing protein	SAG-related sequence SRS16E (SRS6)	1285383.80	2211278.86	835739.05	2.64590	0.00175
NCLIV_032910	hypothetical protein	SPFH domain / Band 7 family protein	1312561.54	3427174.12	2072315.96	2.61106	0.00802
NCLIV_061760	putative microneme protein MIC6	microneme protein MIC6 (MIC6)	99847150.72	145373153.80	57170746.58	2.54279	0.00214
NCLIV_051890	unspecified product	AP2 domain transcription factor AP2X-7 (AP2X7)	512977.59	796420.85	316090.41	2.51960	0.01159
NCLIV_022970	unspecified product	microneme protein MIC2 (MIC2)	19596185.13	34068901.48	13797268.62	2.46925	0.00674
NCLIV_001620	conserved hypothetical protein	hypothetical protein	197451.04	472436.90	366920.74	2.39268	0.00720
NCLIV_010600	putative microneme protein MIC3	microneme protein MIC3 (MIC3)	460862170.3	597623331.90	263553637.7	2.26756	0.00066
NCLIV_012130	eukaryotic translation initiation factor 3 subunit 7-like protein, related	eukaryotic translation initiation factor 3 subunit 7, putative	3058435.49	4282803.69	1901774.7	2.25200	0.00720
NCLIV_060470	hypothetical protein	cysteine-tRNA synthetase (CysRS)	1168807.61	2187167.46	993506.64	2.20146	0.04842
NCLIV_046540	putative oligoendopeptidase F	peptidase family M3 protein	559748.68	1227000.02	598720.08	2.19206	0.00556
NCLIV_013360	putative plectin	surface antigen repeat-containing protein	161159.37	212122.31	97005.97	2.18669	0.00740
NCLIV_016150	Serine--pyruvate transaminase,related	alanine-glyoxylate aminotransferase	604348.13	1232329.87	570792.00	2.15898	0.01293

Proteins in cluster 6 (Table 2.7, Appendix Table VI and Figure 2.2) displayed decreasing expression throughout the time course. Proteins in this cluster included those associated with energy metabolism such as formate/nitrite transporter protein and UDP-glucose 4-epimerase. Also proteins associated with apical secretory organelles such as dense granule protein GRA7 and GRA 6, rhoptry protein ROP15 and ROP24 and a microneme protein. Some enzymes such as acid phosphatase that localise in the rhoptries and NTPase I that localises in dense granules, as well as protease inhibitors such as protease inhibitor PI1 and protease inhibitor PI2, which are mostly expressed in tachyzoites, were also in this cluster.

Table 2.7 Top twenty most highly expressed proteins of cluster 6 showing downward expression trend throughout three time points

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
NCLIV_004050	formate/nitrite transporter family protein,related	formate/nitrite transporter protein	526974.15	182506.23	8223.52	64.08132	0.00499
NCLIV_0275	peptidyl-prolyl cis-trans isomerase NIMA-interacting 1	peptidylprolyl isomerase	1247517.81	221748.16	49890.12	25.00531	0.00332
NCLIV_045400	hypothetical protein	splicing factor 3A subunit 2, putative	369688.77	177245.27	33584.01	11.00788	0.01333
NCLIV_056130	conserved hypothetical protein	hypothetical protein	1531949.74	740896.97	178443.55	8.58507	0.00078
NCLIV_057250	conserved hypothetical protein	hypothetical protein	3146278.19	1580093.20	376648.63	8.35335	0.00236
NCLIV_046950	UDP-glucose 4-epimerase, related	UDP-glucose 4-epimerase	206692.55	85457.45	26501.99	7.79913	0.00739
NCLIV_039080	adaptin N terminal region family protein,related	beta-COP	2967373.03	1361368.99	393575.40	7.53953	0.00259
NCLIV_050780	conserved hypothetical protein	hypothetical protein	1641546.33	736822.23	226831.19	7.23686	0.00282
NCLIV_048930	hypothetical protein	importin-beta N-terminal domain-containing protein	1152760.57	366793.78	161921.32	7.11926	0.00409
NCLIV_060900	conserved hypothetical protein	hypothetical protein	343669.12	126206.43	49841.19	6.89528	0.00655
NCLIV_007120	hypothetical protein	microneme protein, putative	1298651.02	635904.54	206501.33	6.28883	0.00467
NCLIV_055980	conserved hypothetical protein	hypothetical protein	5287578.25	2117710.36	880652.61	6.00416	0.00066
NCLIV_063240	DnaJ homologue, related	DnaJ domain-containing protein	379609.76	206332.62	66489.75	5.70930	0.00876
NCLIV_044100	conserved hypothetical protein	hypothetical protein	1075443.92	373675.67	193536.67	5.55680	0.01201
NCLIV_057280	Peptidylprolyl isomerase D (Cyclophilin D),related	tetratricopeptide repeat-containing protein	389988.54	192061.67	70593.74	5.52441	0.01568
NCLIV_011690	unspecified product	rhoptry protein ROP15 (ROP15)	3193011.07	1200586.02	608179.47	5.25011	0.04398
NCLIV_058260	hypothetical protein	N/A	6073953.30	3533331.29	1282677.41	4.73537	0.00145
NCLIV_010110	hypothetical protein	ubiquinol-cytochrome c reductase hinge protein, putative	1529043.10	938260.82	333063.12	4.59085	0.00372
NCLIV_006620	trehalose-6-phosphate synthase of likely plant origin, related	trehalose-phosphatase	157243.48	69697.86	34737.52	4.52662	0.04629
NCLIV_019240	conserved hypothetical protein	hypothetical protein	2088544.07	1098747.25	465879.65	4.48301	0.00241

2.3.2 GO enrichment analysis

GO term enrichment analysis was carried out using topGO for each cluster. However, not all clusters provided the statistically enriched GO categories.

Proteins in cluster 1 have shown decreased expression during stage conversion. Enriched GO biological processes terms are shown in Table 2.8. These include peptide, cellular amide and organonitrogen compound metabolic process, peptide and amide biosynthetic process, translation and gene expression. These GO terms suggest that during stage conversion less energy is required to form a complex molecule in living cells. All proteins involved in GO term are related to protein synthesis such as putative isoleucine-tRNAsynthetase, putative 40S ribosomal protein S29, Eukaryotic translation initiation factor 5, related; Ribosomal protein S21-maize (ISS), related and Glutaminyl-tRNA synthetase, related (Table 2.8).

The GO molecular function terms enriched in cluster 3, which showed an upward expression trend, mainly associated with the binding activities that involved in the regulation of nucleic acid synthesis. Proteins involved in this molecular function were hsp90, related; putative TCP-1/cpn60 chaperonin family protein; ATP synthase subunit beta, related; putative 26S protease regulatory subunit 6b; putative kinesin heavy chain; KLLA0F09449p, related and Myosin, related (Table 2.8).

The GO terms enriched in cluster 4 revealed biological regulation processes comprising of regulation of cellular process, biological regulation and response to stimulus and signalling transduction processes comprising of cell communication, signal transduction, signalling, intracellular signal transduction, single organism signalling and cellular response to stimulus. However, most of the categories share the same set of proteins such as adp-ribosylation factor 4, Adenylate/guanylate cyclase with GAF sensor and FHA domain and two hypothetical proteins. Glutathione reductase, related was only associated to biological regulation.

Molecular function terms enriched in cluster 4 comprised hydrolase activity, GTP binding, guanyl nucleotide binding, and guanyl ribonucleotide binding. There are 13

proteins involved in hydrolase activity as proteasome (Prosome, macropain) subunit, beta type, 1, mgc81714 protein, putative cytidine deaminase, tubulin alpha chain, Family T1, proteasome beta subunit, threonine peptidase, related, hypothetical protein and 5 proteins related to GTP binding, guanyl nucleotide binding and guanyl ribonucleotide binding as adp-ribosylation factor 4, related, tubulin alpha chain and hypothetical protein.

Biological process enriched in cluster 5, which are preferentially expressed from day 0 to day 3 and then drop in abundance from day3 to day 6, included metabolic processes involved in nucleobase-containing compounds, cellular aromatic compounds, heterocycle metabolic process, organic cyclic compounds and cellular nitrogen compounds. The cellular metabolic process is related to nucleobases, nucleosides, nucleotides and nucleic acids. Significantly enriched GO terms related to cluster 5 are associated with the same set of proteins such as putative ribonucleoside-diphosphate reductase, large subunit, putative deoxyuridine 5'-triphosphate nucleotidohydrolase, putative ATP synthase, putative centromere/microtubule binding protein and hypothetical protein.

Table 2.8 Gene Ontology: enriched biological process and molecular function terms obtained from cluster 1, 3, 4 and 5 of *N. caninum* stage conversion from tachyzoite to bradyzoite using topGO

Cluster	GO categories	GO.ID	Term	Identify Proteins	Annotated	Significant	Expected	Fisher exact test
1	Biological process	GO:0006518	peptide metabolic process	putative isoleucine-tRNAsynthetase, hypothetical protein, putative 40S ribosomal protein S29, Eukaryotic translation initiation factor 5,related, Ribosomal protein S21-maize (ISS), related, Glutaminyl-tRNA synthetase,related,	17	7	3.34	0.023
		GO:0043603	cellular amide metabolic process		17	7	3.34	0.023
		GO:1901564	organonitrogen compound metabolic process		26	9	5.1	0.031
		GO:0006412	Translation		15	6	2.94	0.043
		GO:0043043	peptide biosynthetic process		15	6	2.94	0.043
		GO:0043604	amide biosynthetic process		15	6	2.94	0.043
		GO:0010467	gene expression		19	7	3.73	0.044
3	Molecular function	GO:0005524	ATP binding	hsp90, unspecified product, putative TCP-1/cpn60 chaperonin family protein hypothetical protein , ATP synthase subunit beta, related, putative 26S protease regulatory subunit 6b, putative kinesin heavy chain, hypothetical protein, KLLA0F09449p, related, Myosin, related	30	11	5.03	0.0027
		GO:0030554	adenyl nucleotide binding		30	11	5.03	0.0027
		GO:0032559	adenyl ribonucleotide binding		30	11	5.03	0.0027
		GO:0001882	nucleoside binding		41	12	6.88	0.0152
		GO:0001883	purine nucleoside binding		41	12	6.88	0.0152
		GO:0017076	purine nucleotide binding		41	12	6.88	0.0152
		GO:0032549	ribonucleoside binding		41	12	6.88	0.0152
		GO:0032550	purine ribonucleoside binding		41	12	6.88	0.0152
		GO:0032553	ribonucleotide binding		41	12	6.88	0.0152
		GO:0032555	purine ribonucleotide binding		41	12	6.88	0.0152
		GO:0035639	purine ribonucleoside triphosphate binding		41	12	6.88	0.0152
		GO:0097367	carbohydrate derivative binding		41	12	6.88	0.0152
		GO:0000166	nucleotide binding		42	12	7.04	0.0189
		GO:0036094	small molecule binding		42	12	7.04	0.0189
		GO:0043168	anion binding		42	12	7.04	0.0189
		GO:1901265	nucleoside phosphate binding		42	12	7.04	0.0189

4	Biological process	GO:0050794	regulation of cellular process	adp-ribosylation factor 4, Adenylate/guanylate cyclase with GAF sensor and FHA domain, and two hypothetical protein	8	5	1.79	0.013
		GO:0050789	regulation of biological process		9	5	2.02	0.026
		GO:0065007	biological regulation		9	5	2.02	0.026
		GO:0050896	response to stimulus		10	5	2.24	0.043
		GO:0007154	cell communication		7	4	1.57	0.043
		GO:0007165	signal transduction		7	4	1.57	0.043
		GO:0023052	signalling		7	4	1.57	0.043
		GO:0035556	intracellular signal transduction		7	4	1.57	0.043
		GO:0044700	single organism signaling		7	4	1.57	0.043
		GO:0051716	cellular response to stimulus		7	4	1.57	0.043
	Molecular function	GO:0016787	hydrolase activity	adp-ribosylation factor 4, related, tubulin alpha c, three hypothetical protein	41	13	7.64	0.014
		GO:0005525	GTP binding		11	5	2.05	0.033
		GO:0019001	guanyl nucleotide binding		11	5	2.05	0.033
		GO:0032561	guanyl ribonucleotide binding		11	5	2.05	0.033
5	Biological process	GO:0006139	nucleobase-containing compound metabolic process	putative ribonucleoside-diphosphate reductase, large subunit putative deoxyuridine 5'-triphosphate nucleotidohydrolase, putative ATP synthase, putative centromere/microtubule binding protein, hypothetical protein	19	5	1.6	0.0084
		GO:0006725	cellular aromatic compound metabolic process		19	5	1.6	0.0084
		GO:0046483	heterocycle metabolic process		19	5	1.6	0.0084
		GO:1901360	organic cyclic compound metabolic proces		19	5	1.6	0.0084
		GO:0006807	nitrogen compound metabolic process		32	5	2.69	0.0199
		GO:0034641	cellular nitrogen compound metabolic process		32	5	2.69	0.0199

2.4 Discussion

The aim of this chapter was to investigate the protein expression changes during the *N. caninum* differentiation from tachzoites to bradyzoites stage. Due to the close similarity of the two species and the more advanced status of *T. gondii* annotation, many of the *N. caninum* protein names referred to in this study for unspecified product and hypothetical proteins were based on their orthologue proteins in *T. gondii.*

The BAG1 antigen was detected in *N. caninum* at day 5 post induction with SNP using IFA technique indicated the successfully inducing stage conversion. The analysis of GO term enrichment was performed for all cluster sets using topGO in order to interpret functional involvement of these proteins. However, the results mainly returned high-level, inclusive GO categories that are not particularly associated with stage conversion. This has lead to the focus of data interpretation on individual proteins, which was discussed below.

2.4.1 Proteins involved in stage conversion

From the current study, many proteins involved in host cell invasion showed a downward trend in expression, such as several microneme, dense granule and rhoptry proteins. This suggested that the expression of these secretory proteins is required less in quiescent bradyzoite-containing cysts than actively replicating tachyzoites.

2.4.1.1 Microneme proteins

Several microneme proteins had shown downward trend in abundance throughout all three time points such as MIC7 (Reid et al., 2012), whereas , MIC2, MIC3, MIC6, MIC10 and MIC11 had higher abundance at day 3 comparing to day 0 before decreased in abundant from day 3 to day 6 (cluster 5). MICs are released from microneme organelle after attaching to host cell. They play an important role in host cell recognition and invasion process in *T. gondii*. (Carruthers and Sibley, 1999). The observation that many MICs were found more abundant at day 3 post induction

comparing to day 0 would have coincided with the second round of infections where parasites were egressing from the host cells and start to infect the new cells. The results also suggested that microneme proteins were not required in chronic bradyzoites forming stage as they all showed lowest in abundance on day 6 post induction. The results agrees with previous study which indicated that microneme proteins can be expressed in both tachyzoites and bradyzoites but mostly found in tachyzoites (Sonda et al., 2000).

2.4.1.2 Rhoptries proteins

In this study, most of rhoptry proteins decreased in expression throughout the time course such as ROP6, ROP 9, ROP15, ROP17, ROP24 and RON10. Rhoptry proteins play a critical role in host cell invasion and are also associated with virulence in *T. gondii* (Peixoto et al., 2010; Reese and Boothroyd, 2011; Saeij et al., 2007b).The finding more abundant of ROP 9 in tachyzoites stage in this study agrees with the study from Reichmann et al. (2002), which found that ROP 9 is a tachyzoites- specific rhoptry proteins in *T. gondii*. However, this finding contradicts with the study of Marugan-Hernandez et al. (2010), where they found ROP9 was overexpressed in bradyzoites stage. In this study, ROP6 was in high abundance in tachyzoites, which is determined as an intrinsically disordered protein. ROP6 have no kinase activities in *T. gondii* and its biochemical and physical properties are still unknown as 60% of ROP6 is predicted to be intrinsically disorderdered (Lee et al., 2014). ROP15 is thought to have a role in parasite virulence as the expression of ROP15 protein decreased in a mutant temperature-sensitive *N. caninum* (Pollo-Oliveira et al., 2013). The decreasing abundance of ROP15 and the other ROP proteins in the bradyzoite stage in this study correlates with the diminished activity of bradyzoites invasion of the host cells. It seems like the low abundance of rhoptry proteins in bradyzoites is because this life stage is slower invade host cells than tachyzoite stage. However, ROP35 did show increased abundance throughout the time course (cluster3) which suggests an important role in the bradyzoite stage. This finding is consistent with the study of Fox et al. (2016). They found that some of ROP proteins including ROP35 were associated with chronic infection. Moreover,

they showed that the deletion of ROP35 resulted in substantially reduced cyst burden.

2.4.1.3 Dense Granule Proteins

Most of the dense granule proteins (GRA) exhibited downward expression trend during the time course. GRA proteins are released at the sub-apical area of the parasite within 1 hour of *N. caninum* penetration and formation of PV (Carruthers and Sibley, 1997a; Dubremetz et al., 1993). In this study, GRA1, GRA3, GRA6, GRA7 and dense granule protein DG32 were observed in high abundance in tachyzoites at day 0 and they were not seen at day 6. GRA1, although did not observe in this study at day 6, might be cross-reactive with bradyzoites (Cesbron-Delauw et al., 1989), which can be found in both tachyzoites and bradyzoites stage (Nam, 2009). GRA3 is exocytosed after invasion into PV of host cells and associated with the PVM (Achbarou et al., 1991). GRA6 is localised in dense granules of tachyzoites as well as in the PV. The function of GRA6 is thought to have a role in the antigenicity and pathogenicity of *T. gondii*. This protein stabilises the tubular network with the support of GRA2 after invasion host cells (Mercier et al., 2002). Vonlaufen et al. (2004) found that dense granule NcGRA1, NcGRA2 and NcGRA7 suspected to be involved in tissue cyst wall formation because these dense granule protein were secreted and accumulated in the cyst wall during the stage conversion. Their finding suggested that NcGRA1 and NcGRA2 might be used as *N. caninum* markers to identify chronic infection by serological methods (Atkinson et al., 2001; Ellis et al., 2000). Their finding was contradicted with the current study. This might be GRA7 can also be found in tachyzoite stage. In this study, all GRA proteins showed decreased abundant over the experimental time course. This study found that several invasion proteins were reduced in expression during the induction of bradyzoite formation, which is in agreement with previous studies (Bohne et al., 1999; Kang et al., 2008; Lekutis et al., 2001; Manger et al., 1998).

2.4.1.4 Rhomboid proteins

Rhomboid 4 proteins (ROM4) showed decreased expression during conversion stage (cluster 2). A rhomboid protease is sheded by intramembrane proteolysis to support host cell invasion and motility of the parasite. TgROM4 associates in surface adhesins with MIC2, AMA1 and MIC3. Buguliskis et al. (2010) suppressed the TgROM4 causing the disruption of gliding motility, decreased MIC2 secretion and failure to invade host cells. The down-regulation of ROM 4 interrupts the normal apical-posterior gradient of adhesins leading to less efficient of host cell invasion and motility in *T. gondii*. Brossier et al. (2005) stated that ROM1, ROM4 and ROM5 are expressed in *T. gondii* tachyzoites. In this study, ROM4 decreased in expression during the stage differentiation, which might indicate that ROM4 and its associated invasion mechanism is required less into a bradyzoite stage.

2.4.1.5 Calmodulin proteins

Two calmodulin proteins including calmodulin CAM2 and calmodulin, putative showed decreased expression (cluster 1) throughout time course. Moreover, CAM2 revealed the highest fold change at 10.61. In *T. gondii*, intracellular calcium ions interact with calmodulin to induce cell motility, cell division, secretion, cytoskeletal activity and protein synthesis (Song et al., 2004). The formation of a Ca^{2+}-calmodulin complex is associated with signal transduction pathways (Cheung, 1980). Their role is associated with host cell attachment and invasion (Kieschnick et al., 2001). Once *N. caninum* converts to the bradyzoite stage, calmodulin may be less important in protein expression concerning to host cell invasion.

2.4.2 Protein associated with parasite motility

The proteins associated with parasite motility and host cell invasion such as gliding associated proteins GAP40, glideosome-associated protein with multiple-membrane spans GAPM3, glideosome-associated protein with multiple-membrane, and two proteins of myosin A have shown reduced expression throughout the induction time course (cluster 1 and 2). The gliding associated proteins GAP40 have critical roles in

intracellular replication and they are also important for parasite stability (Harding et al., 2016). Harding et al. (2016) found that lacking of GAP40 or GAP 50 in *T. gondii,* is capable to replicate but could not be completed. In this study, these proteins associated with gliding activity and intracellular replication have revealed deceased abundance during stage conversion. This reflects the slow multiplication rate of bradyzoites within tissue cysts.

2.4.3 Metabolic enzyme

Fructose-1,6-biphosphate is a glycolytic enzyme that was differentially expressed in bradyzoites (day 6) in this study. This finding is comparable to the study of Tomavo (2001). They found that *Toxoplasma* bradyzoites stage might depend on anaerobic glycolysis. Moreover this finding is consistent with the study of Marugan-Hernandez et al. (2010) who found that fructose-1,6 biphosphate was more abundant in *N. caninum* bradyzoites.

2.4.4 Heat shock proteins

The development of bradyzoites using chemical treatment such as SNP can induce a stress condition that is associated with heat shock proteins (hsp) (Weiss et al., 1998a). Weiss et al. (1998b) found that hsp70 family is induced during bradyzoite differentiation. Marugan-Hernandez et al. (2010) also found hsp 90 to be differentially abundant in bradyzoites after inducing stage conversion using SNP. These studies are consistent with the present data where hsp90 and hsp60 were most abundant in bradyzoiotes (cluster 3). However, hsp40 showed the opposite trend (cluster 1). Echeverria et al. (2005) studied the role of hsp90 in *T. gondii* during stage differentiation under stress. They suggested that hsp90 plays an important role in stage conversion and this protein is suitable for using as a potential drug target against *T. gondii*. One consequence of stressing bradyzoites may be to introduce DNA damage in leading to inhibition of cell replication (Tomavo, 2001).

2.4.5 Bradyzoites associated proteins

Two bradyzoites associated proteins such as SAG-related sequence SRS35A (NCLIV_019580) and putative high molecular mass nuclear antigen (NCLIV_030890) have both revealed increase in expression during the time course. This finding is associated that bradyzoites specific marker (BAG1) antibody revealed positive at day 5 post induction by IFA and is consistent with the study of Vonlaufen et al. (2004). However, dense granule protein GRA6 (NCLIV_052880) has exhibited decrease in expression over the time course. This might due to GRA6 can be found in both tachyzoites and bradyzoites stage.

2.4.6 *In vitro* stage conversion

The study of Vonlaufen et al. (2004) has shown that SNP could trigger the stage conversion from tachyzoites to bradyzoites stage which determined by increasing in expression of *N. caninum* bradyzoites marker such as NcBAG1 and cyst wall antigen as MAbCC2 and decreasing in expression of tachyzoites surface antigens such as NcSAG1 and NcSRS2.The SDS-PAGE and immunoblotting were used for detection. However, this technique is not appropriate to obtain higher amount of bradyzoites for further studies which require high amount of the samples such as biochemical, molecular and functional investigation. The current study has shown successful induction of stage conversion from tachyzoites to bradyzoites stage with the same SNP concentration at 17 µM. However, the SNP concentration of 70 µM has destroyed the Vero cells after a few days of application. One possible reasons might be the length of cell culture, i.e. 6 days post induction equals to 9 days post infection which might be too long for infected Vero cells to survive. The other method that can use to measure the bradyzoite formation such as transmissible electron microscope, immunohistochemistry, conventional PCR, quantitative realtime PCR, mass spectrometry. In this study, real time PCR is suggested to visualised the bradyzoite formation as this method can quantify the expression level of BAG1.

2.4.7 Differences between *Toxoplasma* and *Neospora* in their propensity to differentiate to bradyzoites

Similarly, to *T. gondii, N. caninum* can be differentiated from tachyzoites to bradyzoites stage. It seems likely that similar mechanisms such as stress induced stage conversion underlie developmental transition in both organisms. However, not all of methods used to induce stage conversion for *T. gondii*, can be used for inducing the differentiation process for *N. caninum* (Weiss et al., 1999). It might be due to the difference in its biology such as natural host range (Dubey et al., 2002), antigenicity (Bjorkman and Hemphill, 1998; Hemphill et al., 1997a; Howe and Sibley, 1999), some ultra-structural features (Hemphill et al., 2004; Speer et al., 1999) and its host cell recognition (Naguleswaran et al., 2002; Naguleswaran et al., 2003). Weiss et al. (1999) used protocol that successfully developed the stage conversion for *T. gondii* by increasing the pH of the medium for culturing *T. gondii* infected human fibroblasts with tylosine to develop stage conversion for *N. caninum*. The result showed that the yield of obtaining parasites was low during the stage conversion, indicating that the method offered lower efficiency to differentiate the conversion process *in vitro* than *T. gondii*. In addition, (Dubey and Lindsay, 1996) found that most of cyst formation of *T. gondii* forms in several organs but *N. caninum* form cyst mostly in neural tissue. However, the spontaneous rate of cyst formation for low virulent ME49 strains of *T. gondii* in culture is higher than virulent RH strains (Dubey et al., 1998; Soete et al., 1994). It is consistent with the cyst formation of *N. caninum*, a development of thick refractile cyst wall was occurred more common with high virulent Nc-Liv isolate than with lower virulent NC-2 isolate (Vonlaufen et al., 2004). This study showed the visualise of BAG1 by IFA but the vero cells was not look good at day 6 post induction. The comparison study between *T. gondii* and *N. caninum* suggested to perform to see the difference.

2.4.8 Limitations of the study and suggestions for future prospectives

In this study, cells infected with tachyzoites at day 0 after treated with SNP was used as control. The evidence of not seeing the BAG1 reaction by IFA at day 0 has been proved for the stage of parasite that was not yet convert to bradyzoites stage.

However, cells were not treated with SNP should be included for experimental control, so the stage conversion process will provide more reliable result. The validation of experiment need to be considered such as proper parasite: host ratio. In this study, the parasite to host ratio was used as 1:1 and it showed that most of Vero cells were destroyed at day 6 post induction. The proper ratio might be less parasite to host such as 1: 2-4 (parasite: host) ratio. This would allow an extension of the stage conversion experiment upto 8-9 days post induction (Vonlaufen et al., 2004). Extension of the duration of stage conversion experiment might yield purer bradyzoites productions. However, it might contain a lot of dead cells which would interfere with the experimental results. The IFA method which was used to confirm the conversion stage in this experiment was performed at day 5 post induction, which was not consistent with the experimental design. This is due to less amount of BAG1 antibody. It would be better to set the IFA method following the time point in the experiment time course, which would provide more reliable result on the stage conversion progression and . Moreover, tachyzoite specific antigen can also be added to IFA method to give more view of stage conversion synchronisation.

Furthermore, most of the proteins were annotated as hypothetical proteins and unspecified products with no known functions. It limits the depth of data interpretation. With more activities in the field of *Neospora* research, it is expected that increasing number of proteins will be annotated in the nearly future. It is also to be considered that the synchronization of *N. caninum* stage conversion should be regulated, where all parasites should be brought to same phase. This is important for studying cell cycle of the parasite for example extending the duration of SNP induction to obtain more bradyzoites. One more limitation is this study is not reflecting the real situation like in animal body as in cell culture have no immunity in cell culture

The proteomics study of tachyzoites-bradyzoites stage of *N. caninum* is important to know the biological information of this conversion stage as it is important in pathogenesis. More approach such as metabolomics method may be applied for further investigation in order to deeply know about the melobolism related to the stage conversion.

2.4.9 Comparative study of proteomics and transcriptomics of *N. caninum* tachyzoites to bradyzoites stage conversion

Vermont (2012) studied the *N. caninum* stage conversion using RNA-Seq method to investigate the transcripts expression trend during the experimental time course at day 0, 1, 3 and 6 post SNP induction with the concentration of 50 µM. Their study was focused on the apical genes as micronemes, rhoptry and dense granules genes which involved in host cells invasion process. So the comparative analysis between both methods are performed only in these three protein families. The transcriptomics result found that rhoptry genes such as ROP40, ROP14 and Toxofilin have shown decreased expression during the stage conversion time course which is consistent with proteomics study. Their results indicated that the decrease in abundant of rhoptry proteins because these proteins are less important in invasion process of bradyzoites However, ROP35 has shown increase in abundant in current study but this ROP proteins was associated to the chronic infection and cyst formation (Fox et al., 2016).

The transcriptomics finding of microneme genes has shown both increase and decrease in abundance during the time course. This finding is inconsistent with the proteomics study as most micronemes proteins as MIC2, MIC3, MIC6, MIC10 and MIC11 have revealed increase in abundant from day0 to day 3 post induction and then decrease in protein expression at day6 post induction. In addition, MIC7 has shown decrease in expression during the stage conversion time course and no MICs have shown high abundant at day 6 post induction.

No dense granule genes were expressed in the transcriptomics study. This might indicate the transcriptional process of these genes was remarkably robust or there might have been some other elements controlling the translation (Vermont, 2012). Their finding is not consistent with the proteomics results presented in this study as several dense granule proteins quantified in the study exhibited decrease in expression throughout the time course. The dense granule are known to function after micronemes and rhoptry proteins in the invasion process, which establishment in PV (Carruthers and Sibley, 1997a) which is supported by the proteomics finding.

The correlation between transcriptomics and proteomics data has been proved to be low. The mRNA transcripts abundances might be only partially correlated with the protein abundances depending on the type of organism. For example the study of human DAOY medulliblastoma cell line showing that approximately 30-40% of protein abundance variance correlated to mRNA abundance (Vogel and Marcotte, 2012). There are many factors involved in this process such as post transcriptional regulation of mRNA and posttranslational modification of protein (Perco et al., 2010; Yang et al., 2015). As a consequence, gene expression abundance might not correlate with protein expression abundance.

Many previous studies revealed the comparable between transcriptomic and proteomic data are indicated weak correlation in protozoan parasites such as Plasmodium, *Toxoplasma* and other Apicomplexan parasites. The differences between mRNA and protein expression can be explained by biological mechanisms including, firstly, regulation of expression such as selective protein degradation, and variations in protein turnover rates. Secondly, post-translation regulation such as mRNA decay, and translational repression. Moreover, the previous studies in apicomplexan parasites, it has evidences that might be occured from stage-specific adaptation that induce transcriptional changes and will reflect protein changes. In addition, technical factor such as lacks of collected sample concurrently transcriptomic and proteomic analysis, and lacks of statistical models to prove the errors in each transcriptomic and proteomic data (Wastling et al., 2009)

The experiments used to study the relation between transcriptomic and proteomic data are Expressed Sequence Tags (EST) or Microarray expression for attain transcriptomic data and 1 gel-based analysis of parasite protein (one- or two-dimensional gel electrophoresis) followed by mass spectrometry of trypsin digested bands or spots or whole shotgun proteome analysis (Wastling and Xia, 2016). These will be resolved 30-40% of proteome of protozoan parasite (Wastling et al., 2009).

2.5 Conclusions

In summary, most of the apical secretary proteins such as micronemes, rhoptries and dense granules have shown reduced expression in the bradyzoites-like stage during the stage conversion and most bradyzoites associated proteins have shown increase in expression. This indicates that when the parasites are preparing to convert stage, the apical secretory proteins might be less important in bradyzoites stage.

Chapter 3: Comparative transcriptomics of low- and high-virulence strains of *N. caninum*

3.1 Introduction

3.1.1 *Neospora* virulence

Parasite virulence can be in simple terms as destruction or the ability to form disease, or as a complex feature including parasitic infection, invasion, rate of reproduction, and pathogenicity (Garnick, 1992). The virulence of *T. gondii*, which is closely related to *N. caninum*, is defined by the multiplication rate of the parasite and the stimulation of various kinds of immune response in mice (Dubremetz and Lebrun, 2012). The factors involved in multiplication rate seem to be involved in virulence factors of *T. gondii*. Al-Qassab et al. (2010) defined virulence in *N. caninum* as the ability of the organism to induce clinical signs of neosporosis in mice, transplacental transmission and survival of offspring in infected animal. There are some *N. caninum* isolates defined as low virulence such as Nc JAP1 (Yamane et al., 1996; Yamane et al., 1997), Nc-Sheep (Koyama et al., 2001), Nc-Nowra (Miller et al., 2002), Nc-Spain 1H (Rojo-Montejo et al., 2009a) and Nc-Goias (García-Melo et al., 2009). All low virulence strains were isolated from asymtomatic calves or sheep. Conversely, some of *N. caninum* strains determined are as high-virulence, for example Nc-1 (Dubey et al., 1988b), Nc-Liverpool (Barber et al., 1993; Barber et al., 1995) and Nc Spain 7 (Regidor-Cerrillo et al., 2008).

The low and high-virulence strains of *N. caninum* used in this study were Nc Spain 1H and Nc Spain 7, respectively. Nc Spain 1 was originally isolated from a congenitally infected calf brain and was shown to be less virulent than Nc1 both in pregnant and non-pregnant mouse models, with a low tachyzoite yield. Moreover, the viability rate in *in vitro* cell culture of Nc Spain 1H was decreased (Rojo-Montejo et al., 2009a). The high-virulence Nc Spain 7 strain was isolated from naturally asymptomatic infected calf's brain (Regidor-Cerrillo et al., 2008). It causes foetal death and abortion in pregnant heifers while immunological responses could be detected after day 14 post infection (Caspe et al., 2012).

Intraspecific variation of *N. caninum* is important as the high-virulence of the parasites can cause reproductive failure in cattle and neuromuscular disorders in dogs. On the other hand, low virulence strains cause very mild clinical signs. These

variations also depend on the infected host response; for example Nc-1 has been shown to be more virulent than Nc-Liverpool in sheep, while Nc-Liverpool displayed greater virulence than Nc1 in BALB/c mice with the former being more abundant in the brain (Collantes-Fernandez et al., 2006). Moreover, host susceptibility plays a role in *N. caninum* infection; for example, inbred BALB/c mice, immunodeficient mice, INF-γ KO mice and gerbils are susceptible to neosporosis (Dubey and Lindsay, 1996; Dubey et al., 1998). Indeed, disease phenotype among *N. caninum* is not strictly conserved within the species due to significant variation levels within the population. The factors contributing to the diversity of *N. caninum* may involve broad intermediate host range, geographical distribution, and the parasite capability to reproduce sexually (Al-Qassab et al., 2010).

3.1.2 The difference in pathogenicity between low- and high-virulence *N. caninum*

Pathogenicity is the ability of an organism to cause disease in an infected host. The pathogenicity of *N. caninum* has been investigated in many approaches such as *in vivo* and *in vitro* experiments, pathological observations, and molecular techniques (Caspe et al., 2012; Cavalcante et al., 2012; Regidor-Cerrillo et al., 2014; Rojo-Montejo et al., 2009a). For example, the low virulence Nc Spain 1H strain does not cause foetopathy in infected cattle at day 70 of the gestation period but fetal death occurred in three out of five dams infected with the high-virulence Nc-1 strain (Rojo-Montejo et al. (2009a). A comparison of the *in vitro* tachyzoites yield and viability rate between these strains using plaque assays found that Nc Spain 1H exhibited lower tachyzoites yield and lower viability rate than Nc-1 (Rojo-Montejo et al. (2009a). Moreover, in the pregnant mouse model almost 100% of offspring survived in Nc Spain 1H infected group and *Neospora* DNA was detected from only 1 pup. In contrast, the high mortality rate was demonstrated in Nc-1 infected group.

The study of pathogenicity in high-virulence Nc Spain 7 was performed by Caspe et al. (2012). The result showed that four out of seven pregnant cattle aborted after inoculation of dams with high-virulence Nc Spain 7 at day 65 of pregnancy. The abortion occured at 3 to 5 weeks post infection. On the other hand, the cattle in the control group inoculated with PBS did not show any abortion. Similarly, foetal death

occurred after injecting with Nc-Liverpool (Williams et al., 2000) and Nc-1 (Macaldowie et al., 2004) tachyzoites to early pregnancy of cattle. The previous experimental study associated to virulence of *N. caninum* revealed the significant difference of the clinical evidence between low and high-virulence strain of *N. caninum*. However, the genetic factors influencing the variation of virulence in *N. caninum* are still unclear. A transcriptomics approach can be used to explore gene expression in strains of *N. caninum* with different virulence phenotypes to gain a better understanding of biological information related to virulence in *N. caninum*.

3.1.3 Transcriptomics and RNA-Seq

The transcriptome is the complete set of all the messenger RNA (mRNA) in a cell (Wang et al., 2009). The transcriptome is important for identifying the functional elements of the genome and for quantifying genome-wide gene expression. The regulation of gene expression is a basic mechanism in all cells that maintains development, homeostasis and adaptation to environments (Gomez et al., 2010). Transcriptomic data from a sample of interest can be obtained by several methods such as microarrays, expressed sequence tag (EST) collections, serial analysis of gene expression (SAGE), massively parallel signature sequencing (MPSS) tags, Cap analysis of gene expression (CAGE) and RNA-Seq (Wang et al., 2009; Wastling et al., 2009). Microarrays can detect nucleic acids in a sample by hybridization to probes on microchips. This method is high throughput and not too expensive. However, microarrays depend on present genome sequence knowledge, high background toward cross hybridization and restrict the range of detection due to interfering from background and saturation of signals. Even though the sequence-based methods such as Sanger sequencing of cDNA or EST libraries provide the cDNA sequence determination, these are relatively low throughput, expensive and not quantitative (Wang et al., 2009). The tag-based sequencing methods for example SAGE, CAGE and MPSS were developed to overcome these limitations. They provided more precise estimation. However, these techniques are still expensive and a specific portion of short tags cannot be perfectly mapped to the reference genome. In contrast, RNA-Seq provides a method for mapping and precisely quantifying mRNA with high-throughput (Chu and Corey, 2012).

3.1.4 RNA-Seq

The RNA-Sequencing (RNA-Seq) method was developed to investigate gene expression (Mortazavi et al., 2008; Nagalakshmi et al., 2008; Wang et al., 2009). It is normally used for comparative experiments where transcript abundance is determined from multiple samples (Li et al., 2012). RNA-Seq has many advantages over existing methods such as greater accuracy, sensitivity and specificity, a broader dynamic range, and it can be applied to all organisms even when there is no sequence reference. Microarrays use hybridization to measure gene expression but the measurement is limited by the background; RNA-Seq quantifies discrete, digital sequencing read counts providing a broader dynamic range with no upper limit for quantification enabling a larger range of detection of transcript expression level.

In RNA-Seq, messenger RNA (mRNA) are randomly fragmented into small pieces and the total or fractionated mRNA are converted to cDNA fragment library attached with adaptors to one or both ends using random primers. Then cDNA library is amplified by PCR and each molecule is then sequenced in a high-throughput method, resulting in millions of short sequence reads. The size-range of reads is normally 30-400 bp depending on the techniques used. The reads are then either aligned to the reference (genome or transcripts) or *de novo* assembled without requiring the genomic reference . The number of reads mapped to a specific regions are used as a measure of expression (Li et al., 2012). Reads density can be used to measure transcript and gene expression (Trapnell et al., 2012). In comparison to previous methods, RNA-seq represents notable advances. First, the detection of transcripts by RNA-Seq is not limited by the current reference genome sequence. Second, RNA-Seq has a low signal background compared to DNA microarrays, meaning a higher accuracy rate for quantifying gene expression as well as a lower requirement for starting material (Wang et al., 2009). This method can explore the whole transcriptome in a high throughput and quantitative ways as well as providing lower cost compared to arrays and large scale Sanger EST sequencing.

There are three categories involved in RNA-Seq analysis tools, namely read alignment, transcript assembly or genome annotation, and transcript or gene quantification (Trapnell et al., 2012). The quality assessment for sequence reads is

the first pipeline for differential expression analysis such as assessment a number of reads quality, trimming bases or reads below the requiring cut off and trimming residual adapter from reads sequence. Then mapping reads to a reference genome sequence is allows the determination of exon/gene expression values and the identification of differential gene expression between two conditions. Finally, functional terms associated with cohorts of differentially expressed transcripts, such as gene ontology (GO) terms, can be analyzed to interpret the biological consequences of changes in gene expression.

3.1.5 Application of transcriptomics to the study of virulence in apicomplexan parasites

RNA-seq has been applied to the transcriptomic profiles of apicomplexan parasites, to examine specific developmental stages, specific phenotypes such as the virulence of the parasites (Pedroni et al., 2013). Pedroni et al. (2013) compared transcriptomic profiles of geographically distinct virulent and attenuated *Babesia bovis* strains using microarray and RNA-sequencing methods. They found that *ves* gene expression was up-regulated in the virulent strain, consistent with the role of VESA1 proteins in cytoadherence of infected cells to endothelial cells. While the spherical body protein 2 gene family was up-regulated in an attenuated strain. The study concluded that these two gene families may play an important role in regulating virulence in *B. bovis*.

In *Plasmodium* spp., Rovira-Graells et al. (2012) found that the transcriptional diversity within clonal *Plasmodium falciparum* populations were involved in gene families associated with host parasite interactions and genes relevant to environmental pressure adaptation. Almelli et al. (2014) compared transcriptomic patterns of *Plasmodium falciparum* in children with cerebral malaria with those in asymptomatic patients using microarray and quantitative rtPCR. They found a high abundance of genes involved in pathogenesis, host cells attachment and erythrocyte aggregation in parasites taken from patients with cerebral malaria. However, genes encoding exported proteins, Mauere's cleft proteins, transcriptional factor proteins, proteins implicated in protein transport and hypothetical proteins were more abundant in parasites from asymtomatic children. Another transcriptome study of

avian malaria was performed by Lauron et al. (2014), in which erythrocytes containing *P. gallinaceum* were inoculated into white leghorn chicken, identified transcripts related to intra-erythrocytic survival and erythrocyte invasion, such as apical membrane antigen-1 (ama-1) and rhoptry neck protein 2 (RON2) were identified.

The variation in clinical presentation of neosporosis is associated with intra-specific diversity of the parasite (Regidor-Cerrillo et al., 2013). Most of the virulence studies in *N. caninum* have been investigated from the clinical appearance and pathology while the factors related to virulence are still not known. Moreover, there has been no previous analysis of genome-wide gene expression in *N. caninum* strains with contrasting virulence phenotypes using an RNA-Seq approach. A comprehensive analysis of *N. caninum* virulence by transcriptomics is needed because this technique can identify those transcripts related to virulence and might provide information on the possible function of genes associated with virulence.

3.1.6 Aims and Objectives

The aim of this chapter is to compare the transcriptomic profile of low-virulence (Nc Spain 1H) and high-virulence (Nc Spain 7) strains of *N. caninum* and identify transcripts displaying significant differential expression between the two strains. These transcripts may play a role in virulence.

Objective of this study:

1. To estimate and compare transcript abundance between low- and high-virulent strains of *N. caninum*
2. To identify differentially expressed transcripts associated with low- and high- virulence phenotypes
3. To identify enriched functional terms consistent with a role in virulence

3.2 Materials and Methods

3.2.1 *N. caninum* tachyzoites preparation

Infection of cell culture with *N. caninum* was carried out by the Ortega-Mora's group, Complutense University of Madrid, Spain. Tachyzoites pellets of low-virulent (Nc Spain 1H) and high-virulent (Nc Spain 7) strains were prepared at $1x10^8$ tachyzoites (three biological replicates each) and sent to University of Liverpool for subsequent procedure. Total RNA was extracted by the author and measured for RNA quantitation using Qubit (Life technologies) and NanoDrop 1000 spectrophotometer (Thermo Scientific).

3.2.2 RNA Extraction

The total RNA of both *N. caninum* strains was extracted by using RNeasy mini kit (Qiagen). Briefly, samples were suspended with 600 µl RLT buffer in order to disrupt the tachyzoites and then the lysate was transferred to QIAshredder columns (Qiagen) to homogenize the tachyzoites and reduce viscosity, affected by high molecular weight cellular components and cell debris. Samples were centrifuged at full speed for 2 minutes. The flow-through fluids were kept and 600 µl of 70% molecular grade ethanol was added to the fluid. The sample mixture was transferred to RNeasy spin column and centrifuged at 10,000 rpm for 30 seconds. The flow-through was discarded and 700 µl of buffer RW1 was added into the column to wash the filter. The columns were centrifuged at 10,000 rpm for 30 seconds and the flowthrough was discarded. Five hundred µl of RPE was added to the column and centrifuged at 10,000 rpm for 30 seconds. The previous step was repeated by adding 500 µl of RPE and centrifuged at 10,000 rpm for 2 minutes. The columns were then inserted to the other collection tube and centrifuged at 10,000 rpm for 1 minute to remove the remaining fluid. The columns were placed into new collection tubes and the total RNA was eluted using 30 µl RNase-free water. The columns were incubated at room temperature for 1 minute and then centrifuged at 10,000 rpm for 1 minute. The flow-through was collected and stored at -80 ^oC.

3.2.3 RNA Quantification

The concentration of total RNA was measured using Qubit assay (Life technologies) and the purity was measured using NanoDrop 1000 spectrophotometer (Thermo Scientific). These procedures were performed according to the manufacturer's protocols.

3.2.4 Poly-A selection and RNA-Seq libraries preparation

The isolation of messenger RNA (mRNA) was done by the Centre for Genomic Research (CGR), at the University of Liverpool. mRNA was isolated from total RNA by selecting Poly-Adenylated RNA using two rounds of selection with the Poly-A selection kit and Dynabeads mRNA Purification kit (Life Technology, Catalogue No. 61006). Preparation of RNA-Seq libraries were performed from the Poly-A selected material using the Epicentre ScriptSeq v2 RNA-Seq Library Preparation kit (Catalogue No. SSV21106, Epicentre). 50 ng of the poly-A selected material was used as input into the library preparations, followed by 10 cycles of amplification. Then libraries were purified using AMPure XP beads. Qubit was employed for library quantification and the Agilent 2100 Bioanalyzer was used for size determination. Finally, the libraries were pooled in equimolar amount using the Qubit and Bioanalyzer data. Each pool was evaluated for their quantity and quality using Bioanalyzer and later by qPCR using the Illumina Library Quantification kit from Kapa on a Roche Light Cycler LC480II following manufacturer in protocol. The sequenced of pool of libraries was performed on one lane of HiSeq 2500 at 2x100bp by paired- end sequencing with rapid-run mode chemistry.

3.2.5 Processing and Quality assessment of the RNA-sequence data

The initial processing and the quality assessment of RNA-sequence data was done by the CGR using an in-house pipeline. Base-calling and de-multiplexing of indexed reads of 6 RNA-Seq samples was prepared using CASAVA version 1.8.2 (Illumina) in fastq format. The Illumina adapter sequences were removed from raw fastq files by trimming the sequence using Cutadapt version 1.2.1 (Dubey et al., 2011). The option "-O,3" was used to trim the 3' end of any reads which matched the adapter

sequence more than 3 bp. Later the low quality reads were trimmed using Sickle version 1.200 and the minimum window quality score of 20 was set. Then reads that were shorter than 10 bp were removed. Finally, the R1 (forward reads) and R2 (reverse reads) were identified if both reads from a pair passed existing filter, while the R0 (unpaired reads) were recognized if only one of a read pair passed the filter.

3.2.6 Reads mapped to *N. caninum* reference genome sequence

Read mapping was carried out by the CGR. Each of the 6 samples of RNA-Seq library was aligned to *N. caninum* genome sequence obtained from ToxoDB website (Version 11.0) using TopHat2 (Braun et al., 2013). Reads that mapped to the reference in more than one position were rejected.

3.2.7 Differential gene expression analysis

Differential gene expression was identified using the Cufflinks package (Trapnell et al., 2010) (http://cufflinks.cbcb.umd.edu/). Cufflinks is a free software used to assemble the transcripts in an RNA-Seq sample (Trapnell et al., 2012a). The package is also used for gene discovery and gene expression analysis of transcriptome (mRNA) sequencing data. This program can calculate the abundance of transcripts with or without a reference annotation. However, Cufflinks needs a reference genome and RNA-Seq reads alignments. It can be used for data normalisation in order to remove artifacts and biases which affect the quantification. The mapped reads against reference genome (as done by CGR using TopHat 2 aligner (http://tophat.cbcb.umd.edu/)) were used as input files which the reads of each biological replicate of *N. caninum* were mapped individually. The assembly files were combined with the reference transcriptome annotation which produced a united annotation for further analysis. Then Cuffdiff, a part of Cufflinks, was employed for quantifying the merged annotation and provided expression data in a tabular file. Cuffdiff examines expression log-fold-change in contrast to the null hypothesis of no change for example true log fold-change equal to zero. It can detect significant differences in transcript abundance between multiple RNA-seq data sets, allowing differential or preferential expression to be identified among multiple conditions

(Trapnell et al., 2012a). fold-changefold-changeTranscript abundance in expressed as Fragments per kilobase of transcript per million fragments mapped (FPKM). These data were then used for further analysis such as gene identification, gene differentially expression and gene ontologies.

3.2.8 Gene ontology (GO) enrichment analysis

Transcript sets with significantly greater abundance in in low- and high-virulence strains of *N. caninum* respectively were separately analysed for their ontologies and biological function. The GO enrichment analysis was performed using ToxoDB website (http://www.toxodb.org/toxo/) in order to explore the gene representatio either over-expression or under -xpression of ontological terms in each of gene set by comparing to the abundance of these terms across the genomic background. An adjusted-P value (q value) of less than 0.05 was used to determine significant enrichment of GO terms. The enrichment analysis was divided into three categories: biological process, molecular function and cellular component ontologies. Fold enrichment was calculated by using the proportion of annotated genes possessing a given GO term divided by the proportion of genes with the same term in the background (Ashburner et al., 2000).

The other approach for obtaining GO term analysis from each *Neospora* strain was performed using topGO (Alexa and Rahnenfuhrer, 2016). This method provides tools for analysing GO terms using the R package to run the analysis. The R script for generating the topGo was kindly provided by Dr. Vicky Hunt, Department of Infection Biology, University of Liverpool.

3.3 Results

3.3.1 Transcriptomic profiles of low and high virulence *N. caninum*

The transcriptomes of the low and high-virulence strains Nc Spain 1H and Nc Spain 7, respectively, were sequenced at the Centre for Genomic Research, University of Liverpool. The sequence data was of high quality, as shown by the total number of RNA-Seq read counts obtained and the distribution of read length after trimming

The percentage of reads mapped to the *N. caninum* reference genome ranged from 62-77 % (Table 3.1). The percentage of the reads not mapped to reference genome is likely to be contamination from the host cells.

Table 3.1 Summary of number and percentage of total reads mapped to the *N. caninum* reference genome using TopHat Version 2.0.10.

Samples	No. of reads to map	Reads mapped to genome	Percentage of total reads mapped
Nc-1H_1	69,200,330	45,999,810	66.5%
Nc-1H_2	49,185,566	30,667,545	62.4%
Nc-1H_3	53,434,568	33,061,233	61.9%
Nc-7_1	41,148,538	29,866,305	72.6%
Nc-7_2	50,510,474	39,176,661	77.6%
Nc-7_3	59,062,862	38,018,906	64.4%

3.3.2 Differentially expressed transcripts between low and high virulence strains of *N. caninum*

A Total of 192 transcripts display differential expression between low and high-virulence *N. caninum* with statistically significance (q value < 0.05). Of these, 132 genes show differential expression with fold-change more than two (log 2 fold-change < -1 and > 1). Among these genes, 106 genes showed a higher level of expression in Nc Spain 1H in comparison to Nc Spain 7 (Table 3.2 and Appendix

Table VII). A further 26 genes showed lower levels of expression in Nc Spain 1H in comparison to Nc Spain 7 as shown in Table 3.3.

Table 3.2 Top twenty most preferentially expressed genes in Nc Spain 1H by FPKM (Fragments per kilobase of exon per million fragments mapped)

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold-change	q value
NCLIV_005760	putative SRS13	10.8762	1.94971	5.578368	0.005436
NCLIV_060015	unspecified product	44.0336	9.91198	4.442463	0.005436
NCLIV_027990	putative WD domain, G-beta repeat-containing protein	4.49324	1.0373	4.331669	0.049833
NCLIV_046230	putative cGMP-inhibited 3',5'-cyclic phosphodiesterase	2.83636	0.664509	4.268355	0.005436
NCLIV_067430	conserved hypothetical protein	5.3933	1.28166	4.208058	0.005436
NCLIV_016130	conserved hypothetical protein	10.8443	2.5969	4.175864	0.026189
NCLIV_042840	hypothetical protein	23.6494	5.82076	4.062940	0.005436
NCLIV_069900	zinc finger (CCCH type) protein, putative	2.86731	0.729169	3.932298	0.034852
NCLIV_042910	malate dehydrogenase, related	118.622	30.3059	3.914155	0.005436

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold-change	q value
NCLIV_003260	putative NcMCP3	95.7119	24.7783	3.862731	0.005436
NCLIV_038930	srs domain-containing protein	19.956	5.29338	3.769992	0.017441
NCLIV_000750	conserved hypothetical protein	75.5329	20.0536	3.766551	0.005436
NCLIV_017300	putative 3', 5'-cyclic nucleotide phosphodiesterase	49.9496	13.4756	3.706670	0.005436
NCLIV_031960	conserved hypothetical protein	13.7958	3.78912	3.640898	0.005436
NCLIV_015120	hypothetical protein	16.5114	4.63039	3.565877	0.005436
NCLIV_013310	hypothetical protein	6.41478	1.81093	3.542257	0.011354
NCLIV_068830	PAN domain-containing protein (EC 3.4.21.27), related	13.9061	4.03794	3.443860	0.014103
NCLIV_019580	SRS domain-containing protein	1086.73	319.311	3.403359	0.005436
NCLIV_009520	conserved hypothetical protein	11.335	3.34599	3.387637	0.016162
NCLIV_021430	hypothetical protein	8.56093	2.60186	3.290312	0.005436

106 transcripts were significant preferentially expressed in Nc Spain 1H, the greatest difference between low and high-virulent strains was a putative SRS13 (NCLIV_005760). This is a bradyzoite-specific gene showing 5.58 fold-change (adjust p-value < 0.05). There are high abundance transcripts of the other six SRS domain-containing transcripts (NCLIV_038930, NCLIV_019580, NCLIV_020100, NCLIV_043600, NCLIV_034731 and NCLIV_049510) as well as some metabolic enzymes. For example, a putative 3', 5'-cyclic nucleotide phosphodiesterase (NCLIV_017300), which has a role in the breakdown of cyclic nucleotides and malate dehydrogenase (NCLIV_042910), which has a role in energy metabolism.

By contrast, there are 26 genes in Nc Spain 1H that are significantly less abundant than in Nc Spain 7 (Table 3.4 and Appendix Table VIII). Most of these genes of Nc Spain 1 H comprise many hypothetical proteins and unspecified products (19 from 26 genes).The seven genes with verified identifications are putative retinitis pigmentosa GTPase regulator (NCLIV_015810), two putative protein phosphatase 2C (NCLIV_019010 and NCLIV_022340), putative actin (NCLIV_004950), putative adenosine transporter (NCLIV_019000), putative zinc finger (CCCH type) protein (NCLIV_064250) and putative PPR repeat-containing protein (NCLIV_066300**).**

Table 3.3 Top twenty most preferentially expressed genes in Nc Spain 7 by FPKM

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold-change	q value
NCLIV_047150	conserved hypothetical protein	64.2754	513.257	-7.9852	0.0054
NCLIV_015810	putative retinitis pigmentosa GTPase regulator	0.377197	1.69162	-4.4847	0.0092
NCLIV_038025	unspecified product	0.716453	2.61539	-3.6504	0.0092
NCLIV_047420	conserved hypothetical protein	0.528506	1.84384	-3.4887	0.0092
NCLIV_019010	putative protein phosphatase 2C	1.47862	4.63615	-3.1354	0.0054
NCLIV_006200	hypothetical protein	5.4881	16.8742	-3.0746	0.0054
NCLIV_069070	hypothetical protein	110.042	310.791	-2.8242	0.0141
NCLIV_011060	conserved hypothetical protein	10.8528	29.419	-2.7107	0.0113
NCLIV_058960	conserved hypothetical protein	2.43305	6.50768	-2.6747	0.0054
NCLIV_038021	unspecified product	3.4118	8.72897	-2.5584	0.0054
NCLIV_050550	hypothetical protein	6.37614	15.8601	-2.4874	0.0113

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold-change	q value
NCLIV_006110	conserved hypothetical protein	8.66654	21.1521	-2.4406	0.0054
NCLIV_004950	putative actin	17.1389	41.1343	-2.4000	0.0054
NCLIV_046220	hypothetical protein	2.65258	6.27052	-2.3639	0.0113
NCLIV_019000	putative adenosine transporter	273.568	626.164	-2.2888	0.0135
NCLIV_027970	conserved hypothetical protein	243.123	555.452	-2.2846	0.0249
NCLIV_053760	unspecified product	19.3859	44.275	-2.2838	0.0054
NCLIV_012010	hypothetical protein	64.0725	144.574	-2.2564	0.0054
NCLIV_064250	putative zinc finger (CCCH type) protein	9.50535	21.2026	-2.2305	0.0054
NCLIV_066300	putative PPR repeat-containing protein	5.46844	11.6782	-2.1355	0.0092

3.3.3 GO enrichment analysis of differentially expressed transcripts

After identifying the transcripts preferentially expressed in Nc Spain 1H (n = 106) and Nc Spain 7 (n = 26), gene ontology enrichment analysis was performed. The biological functions of the differentially expressed transcripts were interogated using a GO-term enrichment analysis generated by the ToxoDB website and topGO. However, topGO did not produce any significant GO enrichment; this may be due to the different parameters used by topGO and ToxoDB to generate GO terms. Only the transcripts that were preferentially expressed in the low-virulence strain (Nc Spain 1H) provided a GO-term enrichment while those that showed an increase in expression in Nc Spain 7 did not provided any significant GO term.

The GO term enrichment of high abundance Nc Spain 1H transcripts comprised 37 significant biological process GO terms (Benjamini < 0.05), as shown in Table 3.5 and Appendix Table III. The GO enrichment of biological process has several functions but most of biological process terms contribute a single transcript and some terms are associated with the same transcripts leading to widespread redundancy among the results.

The significant and non-redundant biological process terms involved parasite movement such as microtubule-based movement (GO:0007018), cellular component movement (GO:0006928), microtubule-based process (GO:0007017), and carbohydrate metabolic process (GO:0005975). In addition, the significant and non-redundant cellular component terms also involved parasite movement such as dynein complex (GO:0030286), microtubule associated complex (GO:0005875), microtubule cytoskeleton (GO:0015630), cytoskeletal part (GO:0044430) and cytoskeleton (GO:0005856).

Moreover, the significant molecular function terms comprised of microtubule motor activity (GO:0003777), motor activity (GO:0003774), ATPase activity (GO:0016887) catalytic activity (GO:0003824), oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor (GO:0016616), and cofactor binding (GO:0048037).

Table 3.4 Gene Ontology enrichment: Biological Process, Cellular Component and Molecular Function of 106 preferentially expressed genes in Nc Spain 1H obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process**:**								
GO:0007018	microtubule-based movement	29	5	17.2	10.04	10.89	0.00022	0.00539
GO:0006928	cellular component movement	31	5	16.1	9.39	10.18	0.00029	0.00539
GO:0007017	microtubule-based process	41	5	12.2	7.1	7.67	0.00093	0.01149
GO:0005975	carbohydrate metabolic process	86	5	5.8	3.38	3.61	0.01775	0.04983
Cellular Component**:**								
GO:0030286	dynein complex	10	5	50	29.1	31.75	3.05E-06	2.13E-05
GO:0005875	microtubule associated complex	17	5	29.4	17.12	18.64	2.44E-05	8.54E-05
GO:0015630	microtubule cytoskeleton	22	5	22.7	13.23	14.38	7.01E-05	0.00016
GO:0044430	cytoskeletal part	35	5	14.3	8.32	9.01	0.00048	0.00084
GO:0005856	cytoskeleton	40	5	12.5	7.28	7.87	0.0008	0.00117
Molecular Function**:**								
GO:0003777	microtubule motor activity	29	5	17.2	10.04	10.89	0.00022	0.00529
GO:0003774	motor activity	40	5	12.5	7.28	7.87	0.00084	0.01009
GO:0016887	ATPase activity	111	7	6.3	3.67	4.04	0.003326	0.019487
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	18	3	16.7	9.7	10.18	0.00490	0.01948

Table 3.5 Identified genes involved in GO term: biological process, cellular component and molecular function of preferentially expressed transcripts of Nc Spain 1 H obtained from ToxoDB database

Gene ID	Description	*T. gondii* Ortholog	Biological process	Cellular component	Molecular function
NCLIV_050200	GF18580, related	Dynein heavy chain family protein	microtubule-based movement	dynein complex	ATP binding, ATPase activity, microtubule motor activity, nucleoside-triphosphatase activity, nucleotide binding
NCLIV_054090	hypothetical protein	Dynein heavy chain family protein			
NCLIV_065960	Dynein heavy chain 1, axonemal, related	dynein heavy chain family protein			
NCLIV_026260	GA26239, related	Dynein heavy chain family protein			
NCLIV_034170	hypothetical protein	Dynein heavy chain family protein			

Gene ID	Description	*T. gondii* Ortholog	Biological process	Cellular component	Molecular function
NCLIV_042910	malate dehydrogenase, related	malate dehydrogenase	carbohydrate metabolic process, cellular carbohydrate metabolic process, malate metabolic process, oxidation-reduction process	null	L-malate dehydrogenase activity, catalytic activity, oxidoreductase activity
NCLIV_043550	glycerol-3-phosphate dehydrogenase, related	glycerol-3-phosphate dehydrogenase, related	carbohydrate metabolic process, glycerol-3-phosphate catabolic process, oxidation-reduction process	cytoplasm	NAD binding, glycerol-3-phosphate dehydrogenase [NAD+] activity, protein homodimerization activity, oxidoreductase activity

Gene ID	Description	*T. gondii* Ortholog	Biological process	Cellular component	Molecular function
NCLIV_062520	3-ketoacyl-(Acyl-carrier-protein) reductase, related	3-ketoacyl-(Acyl-carrier-protein) reductase, related	null	null	oxidoreductase activity
NCLIV_015830	hypothetical protein	ABC transporter transmembrane region domain-containing protein	transmembrane transport, transport	integral to membrane	ATP binding, ATPase activity, nucleoside-triphosphatase activity, nucleotide binding
NCLIV_022240	ATPase, related	ATPase, related	ATP biosynthetic process, ATP catabolic process, cation transport, metabolic process	integral to membrane, membrane	ATP binding, ATPase activity, catalytic activity, metal ion binding, nucleotide binding

The transcripts preferentially expressed in Nc Spain 1H that were identified as contributing to each GO term are shown in Table 3.6. There are five genes which are involved in microtubule and parasite movement such as GF18580, related (*T. gondii* Ortholog: Dynein heavy chain family protein) (NCLIV_050200), hypothetical protein (*T. gondii* Ortholog: Dynein heavy chain family protein) (NCLIV_054090), Dynein heavy chain 1, axonemal, related (NCLIV_065960), GA26239, related (Ortholog *T. gondii*: Dynein heavy chain family protein) (NCLIV_026260), and hypothetical protein (*T. gondii* Ortholog: Dynein heavy chain family protein) (NCLIV_034170).

Moreover, five metabolic enzymes have been identified; malate dehydrogenase related (NCLIV_042910), glycerol-3-phosphate dehydrogenase related (NCLIV_043550), 3-ketoacyl-(Acyl-carrier-protein) reductase related (NCLIV_062520) involved with oxidoreductase activity, ATPase related (NCLIV_022240) and a hypothetical protein (NCLIV_015830) involved ATPase activity.

3.4 Discussion

3.4.1 The SRS gene family may be associated with limiting *N. caninum* virulence

The aim of this study was to identify differentially expressed transcripts between low and high-virulence strains of *N. caninum*. The results have shown that a putative SRS13 gene was the most significantly differentially expressed transcripts with the highest fold-change and is more abundant in Nc Spain 1H than Nc Spain 7. The SRS13 gene has been described as a bradyzoite-specific gene (Tomita et al., 2013). The bradyzoites are the slowly-multiplying, encysted, quiescent form that are thought to be less important in the host cell invasion. A further six SRS domain-containing transcripts were identified as being preferentially expressed in the low-virulence Nc Spain 1 H strain. All of these genes belong to the SRS surface antigen gene family which localises to the cell surface of *N. caninum* (Reid et al., 2012) . SRS genes play an important role in attaching to the host cells, stimulating the host immunity and regulating parasite virulence (Kim & Boothroyd, 2005). It was suggested that the large number of surface antigen genes (SRSs) is important for infecting a wide range of host in *T. gondii* (Howe et al., 1998; Boothroyd, 2009). However, (Reid et al., 2012) found that SRS genes were more significantly expanded in *N. caninum* than *T. gondii*, although *N. caninum* has narrower host range than *T. gondii*. A study of SRS genes in *T. gondii* revealed that some of these genes reduce the virulence of the parasite. Wasmuth et al. (2012) found that overexpression of SRS29C attenuated the virulence of the *T. gondii* WT strain in mice and they discovered that SRS29C is an important negative regulator of acute *T. gondii* in murine models. Moreover, Lekutis et al. (2001) stated that the SRS genes could play a significant role in invading the host cell, modulating immune response and/or virulence attenuation. In this study, the SRS genes were preferentially expressed in the low-virulence strain (Nc Spain 1 H); hence, it may be that SRS genes are involved in limiting the virulence of *N. caninum* through a moderation of its host interaction not yet understood.

3.4.2 Malate dehydrogenase

The malate dehydrogenase related gene was found to be more abundant in Nc Spain 1H. This enzyme plays a vital role as a part of the energy production cycle of *T. gondii.* Malate dehydrogenase (MDH) is associated with tricarboxylic acid cycle (TCA) which provides the energy sources for the activities of parasites such as motility, host cell invasion and rapid multiplication (Hassan et al., 2014). This energy production enzyme is important in parasites for maintaining their life cycles (Nwagwu and Opperdoes, 1982). Liu et al. (2016) suggested that malate dehydrogenase in *T. gondii* (TgMDH) could be a potential vaccine candidate as this enzyme has been shown to be an immunogenic compound. The finding of malate dehydrogenase in low virulence strains reflect the general biology of the parasite for survival.

3.4.3 Microneme adhesive repeat domain-containing proteins (MCPs)

Two microneme adhesive repeat domain-containing proteins (MCPs) such as putative NcMCP3 and NcMCP4 were more abundant in Nc Spain 1H than Nc Spain 7. The MCPs are a subset of the micronemes (MICs) which have been found to bind a component of glycoproteins called sialic acid, and glycolipids from the host cell membrane. It targets host cells and is common in moving junction proteins. The studied of *T. gondii* by (Friedrich et al., 2010)stated that MCPs are unique and conserved in coccidian protozoa such as *Toxoplasma, Neospora , Sarcocystis* and *Eimeria* parasites. MCPs are not present in *Plasmodium, Theileria* and *Cryptosporidium*. Gong et al. (2014) found that TgMCP4 localised near the rhoptry of parasite and can be detected in PV. Three type of heat shock protein can interact with TgMCP4 and disruption this proteins was affected theoir growth. The finding of these proteins might indicate the better growth rate of Nc Spain 1H.

3.4.4 Protein phosphatase 2C may be associated with *N. caninum* virulence

In contrast, there are 26 genes showing preferential expression in Nc Spain 7. Most of them encode hypothetical proteins and unspecified products (19 from 26 genes) of which the function is not known. However, there are two putative protein phosphatase 2C genes found to be more abundant in high-virulence Nc Spain 7 strain. This is consistent with a study by Regidor-Cerrillo et al. (2015) who examined immunome expression in *N. caninum* tachyzoites of three strains (Nc Liv, Nc Spain 1H and Nc Spain 7). They found that serine-threonine phosphatase 2C could not be detected in the Nc Spain 1H pooled sera, while it was present in sera from animals infected with Nc Liv and Nc Spain 7.

The protein phosphatase 2C (PP2C) is one of four main classes of serine/threonine specific protein phosphatase. They have many cellular functions such as cell signalling, cell splicing, the cell cycle and cell development (Luan, 2003). PP2C is a rhoptry-localized protein that is secreted during *T. gondii* invasion to transport a parasite-derived protein phosphatase 2C direct into the nucleus of infected cell (Gilbert et al., 2007). Then PP2C will be in the host nucleus, but the protein may have targets within the rhoptries. An identified cytoplasmic *Toxoplasma* PP2C has been found interact with actin binding protein Toxophilin in parasite lysate (Delorme et al., 2003). It is suspected that PP2C has no function as a phosphatase, alternatively it may bind to host substrates resulting in them not being available for orientation by host or parasite phosphatase (Gilbert et al., 2007). The PP2C delivery into the host nucleus illustrates that the rhoptry protein can be entered to the host cytoplasm during the burst by secreting rhoptry proteins. It is indicated that this mechanism is very important to deliver parasite protein into the host cells. PP2C disruption caused mild growth appearance that can be complemented by the wild-type gene indicating that it is important in parasite's lytic cycle. Thus, it may be that PP2C has a role in determining *N. caninum* virulence as these proteins found high abundant in Nc Spain 7 and they can transport the parasite-derived molecule directly to the nucleus of host cells.

3.4.5 Validation of identified gene

The transcriptomic study will come out with the interested gene which shows high expression level. The method used for validate the gene is Real-time quantitative PCR (RT-qPCR)(Zhang et al., 2010). RT-qPCR is a reliable and sensitive method for measuring gene expression level. This technique can quantify gene expression and characterization of gene regulation. The internal reference genes are used to remove sample-to-sample variations which may increase because of test variation.

3.4.6 GO terms enrichment of genes preferentially expressed in the low virulence strain are involved in parasite movement

The biological process (BP) GO terms of low-virulence Nc Spain 1H has revealed several functions involved in movement such as microtubule-based movement, cellular component movement and microtubule-based process. This is consistent with the cellular component (CC) GO terms which are also involved in movement such as, dynein complex, microtubule associated complex and microtubule cytoskeleton. Furthermore, movement associated GO terms are enriched in molecular function (MF) term such as microtubule motor activity and motor activity.

The microtubules in eukaryotes are essential for chromosome segregation during multiplication, elongate shape of invasive zoites and motility (Morrissette, 2015). There are five genes such as Dynein heavy chain 1, GF18580, GA26239, and two hypothetical proteins which are orthologous to the Dynein heavy chain family protein in *T. gondii.* All of these genes made the same contribution to the GO term as microtubule-based movement, dynein complex and microtubule motor activity.

The biological GO term of microtubule-based movement (GO: 0007018) is associated with motor proteins and are involved in the movement of organelles, other microtubules or other particles along microtubules. Microtubules are critically important in apicomplexan parasites. They are responsible for many kinds of movement in eukaryotic organisms and also involved in cell division, cell replication, fertility, intracellular structure organization, intracellular transport and the movement of cilia and flagella (Morrissette, 2015). It suggests that the low

virulence Nc Spain 1H might have a better capability in movement, cell division and host cell invasion than the high-virulence Nc Spain 7 strain. Microtubules are necessary components of structures in *Toxoplasma gondii.* The microtubule cytoskeleton is mainly conserved in four "zoite" stages which role in host cell invasion and asexual replication (Morrissette, 2015). The tachyzoites of *Toxoplasma* and other apicomplexan zoites use an immobilized, the inner membrane complex (IMC) -localized myosin motor to move actin filaments which is linked to the cytoplasmic tails of secreted adhesins during the invasion process (Morrissette & Gubbels 2014).

There are five genes in microtubule-based movement GO term which have been identified as dynein heavy chain 1, axonemal, related and dynein heavy chain family. The functional GO terms of these genes are associated with microtubule-based movement (BP), microtubule motor activity (MF) and dynein complex (CC). The dynein is an extremely large motor protein that translocate along microtubules through the minus end. While the other motor protein, called kinesin, moves toward the plus end of the microtubule. Dynein can be divided into two classes as axonemal and cytoplasmics dyneins (Milisav, 1998). It comprises of heavy, intermediate and light polypeptides chain. Cytoplasmic dyneins involved in transport vesicles and organelles moves in the opposite direction with kinesin. While, axonemal dynein generates the motility of cilia and flagella by sliding of outer microtubule doublets (Cooper., 2000). The dyneins role in microtubule-based motility and their biochemistry were reviewed in many publications (Gibbons, 1995, 1996; Milisav, 1998; Ogawa and Mohri, 1996). Dynein heavy chain (DHCs) encompass a motor domain. Axonemal dynein may contain one to three DHCs which connect two microtubules to generate the sliding of flagellar movement. Three accessory subunit types as the WD40 repeat intermediate chain, axonemal assembly factor and axonemal light chain are specific to flagellar dyneins in *T. gondii* (Morrissette and Gubbels, 2014).

3.4.7 Limitations of the study

3.4.7.1 Parasite strains

In this study differentially expressed transcripts were compared between a single low- and a single high-virulent strain of *N. caninum.* This means that each virulent phenotype was represented by only one strain of *N. caninum* which is not enough to to identify the similarity and/or difference in term of transcript expression related to virulence**.** Increasing sample size can increase confidentiality of the result. However, the virulence of each strain needs to be determined by showing its ability to cause pathogenicity in the animal. Therefore, this study provided a preliminary result for further investigation of *N. caninum* virulence**.**

3.4.7.2 Most genes do not have GO terms

Most of identified transcripts were conserved hypothetical proteins and unspecified products. The hypothetical proteins have not been annotated for and have no experimental evidence to support their function. While unspecified proteins are predicted and have experimentally proven but there is no biochemical evidence for their function. So hypothetical proteins and unspecified products cannot be assigned a functional role. Many of these proteins were identified as the differentially expressed genes between the two strains of *N. caninum* and could not be included in the GO analysis. However, many hypothetical genes will be gradually annotated which might be explained more for their role in nearly future.

3.4.7.3 Transcript expression is not protein expression

RNA-Seq provides a transcript expression profile that is not equivalent to protein expression as a transcript can have multiple translational variants**.** So the transcript expression from this study is the expression of specific transcript**.**

3.5 Conclusions

In conclusion, PP2C is expected to be involed in virulence of *N. caninum*, while the SRS proteins may reduce the virulence of the parasite. These two proteins might work together to modulate the virulence of *N. caninum*. All significant GO enrichment term found in biological process, cellular component and molecular function of Nc Spain 1 H are associated with the parasite movement and parasite growth.

Chapter 4: Comparative proteomics of low- and high-virulence strains of *N. caninum*

4.1 Introduction

Chapter 3 was a study of the comparative transcriptomics of low- and high-virulent strains of *N. caninum.* This detected two sets of genes that were differentially expressed between the conditions. Those might be related to *N. caninum* virulence, such as protein phosphatase 2C (PP2C), which was significantly more abundant in the high-virulent strain, and SRS genes, which were significantly preferentially expressed in the low-virulent strain. This chapter will further analyse the proteins expressed by *N. caninum* with distinct virulence-related phenotypes and compare the differential expression of transcripts and the proteins, which could be associated with *N. caninum* virulence. Since transcript abundance can be an inaccurate indicator of gene expression for several reasons, this will produce a more robust picture of the relationship between virulence and gene expression.

It is often assumed that transcript and protein expression are well correlated, based on the Central Dogma (Haider and Pal, 2013). However, many studies have revealed that the association between mRNA and protein expression can be low due to many factors (Maier et al., 2009). The reasons for a poor correlation include post-translational modification of proteins, altered half-life of specific proteins, and the method used for protein detection (Perco et al., 2010). However, protein abundance cannot be predicted by transcript expression alone so post-translational modification of proteins and post-transcriptional regulation of mRNA should be considered together (Juschke et al., 2013; Yang et al., 2015).

This study has performed comparative and quantitative proteomic analysis of low- and high-virulence of *N. caninum* at three different time points, i.e. 12, 36 and 56 hours post infection (hour p.i.). These data will enable proteins that are preferentially expressed in the virulent strain and tachyzoites stage to be identified. The three time points chosen to measure protein abundance relate to the development tachyzoites *in vivo;* 12 hours pi. is approximately the time when tachyzoites invade and proliferate in the host cells, 36 hours p.i. is the time when tachyzoites begin to egress, and by 56 hours p.i. all tachzoites have egressed. The proteomic data in this chapter were also

compared with the transcriptomic data from Chapter 3 to visualise the similarities and differences between them in gene expression values.

4.1.1 Label-free proteomics

Label-free proteomics is a method for generating mass spectrometry data from protein samples without requiring a stable isotope-containing compound to label the protein. The raw data can be directly compared to the protein abundances in each protein run (Cutillas and Vanhaesebroeck, 2007; Wastling et al., 2012). The advantages of this technique are that it is simple, requires a relatively low concentration of sample, and does not require extra experimental steps compared to label-based quantitative technique such as gel electrophoresis and gel imaging. However, a label-free proteomics method may provide a variable result, as every sample has to be prepared separately and individually run on LC-MS/MS. There are two different quantification approaches of label-free proteomics, known as spectral counting (Liu et al., 2004) and peak intensity measurement (Bondarenko et al., 2002). Spectral counting quantifies the number of fragment ion spectra (MS/MS) for peptides of a specific protein. However, this method depends on a simple counting of spectra rather than determining any direct physical property of peptides. It assumes that the linearity of response are the same for every protein. Actually, they are different for every peptide in terms of the chromatographic profiles such as retention time and peak width, which vary for every peptide (Bantscheff et al., 2007).While, the measurement of peak intensity has been performed by calculating chromatographic peak area to measure peptide abundance. The chromatographic peak area of identified peptides depends on the concentration of peptides. If the peptide concentration of increases, the peak area will be increased.

In this study, peak intensity was employed for label-free quantification. Subsequently, mass spectrometry data were analysed using commercial software such as Progenesis QI (http://www.nonlinear.com/progenesis) and Mascot program (Version 2.3.02, Matrix Science) to identify proteins.

4.1.2 Proteomic analysis of *N. caninum* virulence

Several proteomic studies have addressed parasite virulence in *N. caninum*. Proteomic comparison *N. caninum* tachyzoites (Nc KBA-2 and Nc JPA-1) with *T.gondii* was performed by Lee et al. (2005) using 2-DE, immunoblotting and mass spectrometry. They found considerable differences between the *N. caninum* and *T. gondii* proteomes as the 2-DE spots were different pattern of both parasites. Conversely, proteomics comparisons among *N. caninum* strains, for example of Nc KBA-2 (high-virulence strain) from South Korea (Kim et al., 2000) and Nc JPA-1 (low-virulence strain) from Japan (Yamane et al., 1996) revealed great similarity, with 78% of proteins spots by 2-DE and 80% of antigen spots on immunoblotting of 2-DE being consistent (Lee et al., 2005). Another proteomic study visualised the similarity and differences between two geographically distinct *N. caninum* strains, KBA-2 from South Korea and VMDL-1 from USA (Shin et al. (2005). Both strains are considered to be virulent as they both caused abortion in cattle (Hyun et al., 2003; Kim et al., 2000). Shin et al. (2005) found that most spots on 2-DE and antigenic spots on 2-DE immunoblots had the same locations by isoelectric point and molecular weight. They concluded that these virulent strains displayed consistent proteomic and antigenic profiles, despite coming from very different locations.

The proteomic analysis of virulence specifically began with Nischik et al. (2001), who used 2D-PAGE and mass spectrometry to identify host proteins that displayed a modified expression profile after attenuation of a virulent strain in mice (*T. gondii* type I). The parasites were attenuated by continuous passage in fibroblast cell culture, and these were then compared to the parental strain inoculated in mice. Attenuated parasites induced a reduced interleukin-12 and interferon-γ response in the mouse relative to virulent *T. gondii*. Moreover, the abundance of diverse parasite proteins common secreted by tachyzoites was reduced in the attenuated *T. gondii*, such as actin, catalase, microneme protein 5, nucleoside triphosphate hydrolase 1 and dense granule proteins 1, 2, 3, 4, 5, 6, 7 and 8 (Nischik et al. (2001).

Regidor-Cerrillo et al., (2012) compared protein expression patterns in the virulent *N. caninum* strains Nc-Liv and Nc Spain 7 with the attenuated strain Nc Spain 1H using Difference Gel Electrophoresis (DIGE) coupled to mass spectrometry (MS)

(Regidor-Cerrillo et al., 2012b). The comparison of Nc Spain 7 and Nc Spain 1H showed that 39 protein spots were more abundant in Nc Spain 7, while 21 protein spots were more abundant in Nc Spain 1H. In addition, 3 Nc-Liv protein spots were significantly more abundant when compared to Nc Spain 1H (Regidor-Cerrillo et al., 2012). Protein spots were also compared between the two high virulence strains Nc Liv and Nc Spain 7. 24 spots were more abundant in Nc Spain 7 and 12 spots in Nc-Liv. Of the three strains, Nc Spain 7 was found to have the highest number of differentially expressed protein spots (Regidor-Cerrillo et al., 2012).

The functions associated with these differentially-expressed proteins were associated with known aspects of virulence, and given some indications of its mechanism. Eleven differentially expressed proteins were involved in gliding motility and lytic cycles, as well as in oxidative stress. The key proteins that they found increased in expression in Nc Spain 7 included NcNTPase, microneme protein NcMIC1 and NcROP40, which are thought to be associated with virulence.

These previous studies have shown various similarities and differences among *Neospora* strains, as well as with the related parasite as *T. gondii*. The purpose of this study is to provide the first quantitative proteomic analysis of expression profile by using label free proteomics techniques in two strains of *N. caninum* known to differ in virulence phenotype, which were shown different in term of phenotype and pathogenicity. This analysis will identify the proteins that are differentially expressed in the two conditions, and therefore, consistent with a role in regulating virulence. These new insights can advance our understanding of the molecular mechanisms of pathogenesis.

4.1.3 Aims and Objectives

The main aim of this chapter is to identify proteins preferentially expressed in either low- and high-virulence strains of *N. caninum* at different time points post-infection using label-free quantitative proteomics, and then to examine the possible role of these proteins in virulence by inferring their likely functions.

Objectives of this study:

1. To compare, identify and quantify the protein expression profile between low- and high- virulence of *N. caninum* using a label-free quantitative proteomics approach
2. To identify enriched functional terms among proteins consistent with a role in virulence.

4.2 Materials and Methods

4.2.1 Parasite preparation

The MARC-145 monkey kidney cell culture infected with the *N. caninum* isolated Nc Spain 1H (low-virulence) and Nc Spain 7 (high-virulence) was prepared by Prof. Luis Ortega-Mora's group, Complutense University of Madrid, Spain at 12, 36 and 56 hours post infection. Each of the three biological replicates of both *N. caninum* tachyzoites were purified using PD10 columns. The pellets of *N. caninum* tachyzoites were then sent to Department of Infection Biology, University of Liverpool for subsequent analysis.

4.2.2 *N. caninum* tachyzoites preparation

A total of eighteen pellets of Nc tachyzoites (three replicates in three different time points for both strains) in the range of 1.6-7×10^7 tachyzoites were prepared in the same methods described in section 2.2.6.

4.2.3 Protein identification and quantification

4.2.3.1 Liquid chromatography mass spectrometry LC-MS/MS

LC-MS/MS was performed by Dr. Dong Xia, Department of Infection Biology, University of Liverpool following the procedure described in section 2.2.7.

4.2.3.2 Peptides quantification and protein identification

The raw data acquired from LC-MS/MS were imported to Progenesis^{TM} QI software (Version 3.0, Nonlinear Dynamics) for the label-free quantitative data analysis as previously explained in section 2.2.8. The protein identification was performed following the procedure mentined in section 2.2.9. The protein expressions in three different time points were compared between two strains.

4.2.4 Gene Ontology (GO) enrichment analysis and Cluster analysis

In this study, GO enrichment analysis has taken an individually set of protein list from low- or high- virulence of *N. caninum* in different time points as the input data in order to find high representation or low representation of ontologies in comparisons between both strains. The input proteins data used for the enrichment test was the differentially expressed proteins with statistical significance at q-value < 0.05 and fold-change greater than 2. The interpretation of GO enrichment result was described in chapter 3.2.8.

4.3 Results

4.3.1 Preferential expression of proteins in low and high-virulence *N. caninum*

A total of 1,390 proteins from Nc Spain 1H and Nc Spain 7 sampled at three different time points were quantified by LC MS/MS.

At 12 hours p.i., 351 proteins showed significant ($q < 0.05$) differential expression between low (Nc Spain 1H) and high (Nc Spain 7) virulent conditions, while 77 proteins showed differential expression with fold-change greater than two. Of these 77 proteins, 56 proteins were significantly more abundant in Nc Spain 1H (Table 4.1, Appendix Table 1 and Figure 4.1), while 21 proteins were significantly more abundant in Nc Spain 7 (Table 4.2, Appendix Table1 and Figure 4.1).

Most of the proteins preferentially expressed in Nc Spain 1H at 12 hours pi were associated with host cell adhesion and invasion. The SRS domain-containing protein (NCLIV_019580), an orthologue of the SRS35A protein of *T. gondii* (TgSRS35A), showed the highest fold-change (6.53). SAG3 (NCLIV_034740), which plays an important role in host cell attachment during the infection process, was also highly abundant. SRS proteins are a family of cell surface glycoproteins in cyst-forming coccidian parasites, thought to mediate attachment to host cells.

Cathepsin C2 (CPC2) (NCLIV_007010), which are lysosomal peptidases with an important role in growth and survival of Apicomplexan protozoa, had a fold-change of 2.57. The hypothetical protein (NCLIV_069110) is an orthologue of the rhoptry protein ROP1 of *T. gondii* (TgROP1) and was highly abundant with a fold-change of 2.47.

A group of microneme proteins were all found to be highly abundant in Nc Spain 1H: NCLIV_022970 and NCLIV_033690 are hypothetical proteins homologous to microneme protein MIC2 in *T. gondii* (TgMIC2), NCLIV_062770 is orthologous to MIC9 in *T.gondii*, NCLIV_010600 is orthologous to putative microneme protein

MIC3 and NCLIV_002940 is orthologous to microneme protein MIC4 in *T.gondii* (MIC4).

Other proteins associated with cell surface and host invasion machinery that were highly abundant in Nc Spain 1H at 12 hours pi include subtilisin (SUB1; NCLIV_021050), which is involved in host cell surface processing of micronemal secreted proteins in *T.gondii* (Lagal et al., 2010); protein phosphatase 2C (NCLIV_067490) which is secreted from the rhoptry and enters the host nucleus during invasion; and apical membrane antigen 1 (AMA1) (NCLIV_028680), which has a key role in host cell invasion in *T. gondii*.

There are some metabolic enzymes that were highly abundant in Nc Spain 1H, for example NCLIV_022920, the ortholog of aspartyl protease ASP1 (ASP1) in *T. gondii*, putative histidine acid phosphatase domain containing protein (NCLIV_053400), putative orotate phosphoribosyltransferase (NCLIV_027130), putative 3',5'-cyclic phosphodiesterase (NCLIV_019740), HIT family protein, related (NCLIV_018300), Lysyl-tRNA synthetase, related (NCLIV_012460). putative DEAH-box RNA/DNA helicase (NCLIV_011110), putative Hsp20/alpha crystallin domain-containing protein (NCLIV_041850) and ubiquitin carrier protein, related (NCLIV_010370).

Besides these various characterized proteins, there were nine hypothetical proteins and 1 uncharacterized protein that were preferentially expressed in this strain at 12 hours pi.

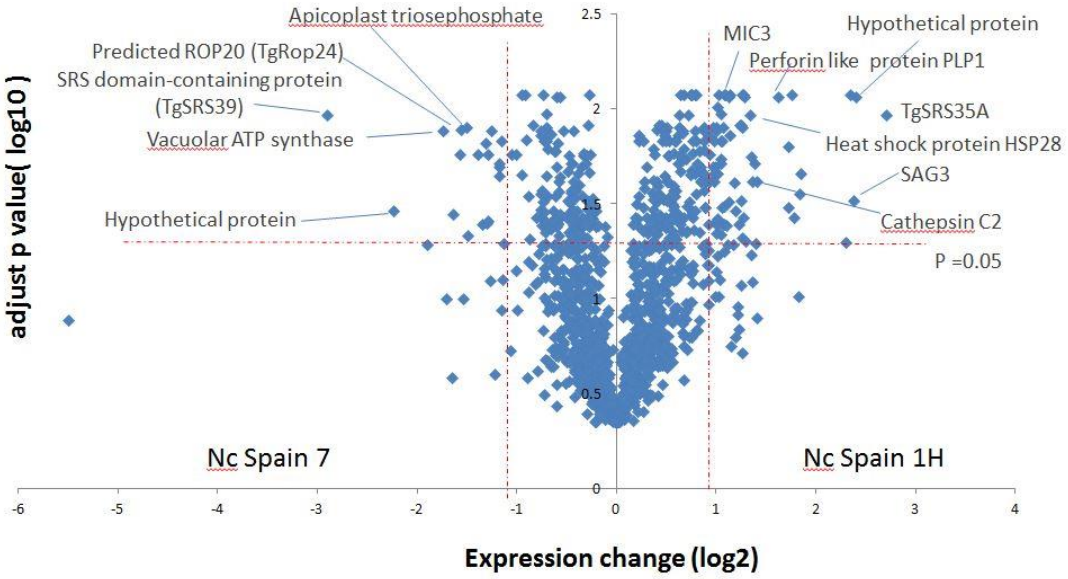

Figure 4.1 Volcano plot of quantified *Neospora* proteins of Nc Spain 1H (right) and Nc Spain 7 (left) at 12 hours pi to host cells.

The volcano plot (Figure 4.1, 4.2 and 4.3) of three time points show fold-change in protein expression of low- (Nc Spain 1H) and high-virulent (Nc Spain 7) strain. The plot compared between both strains. The y-axis is log10 of adjust p-value and the x-axis has shown the differential protein expression (measured in log2 fold change) between Nc Spain 1H and Nc spain 7. Vertical red dash lines between -1 and +1 indicate the fold-change of protein expression equal to 2 (log2 fold-change -1 and 1). While horizontal dash line on y-axis is equivalent to p-value of 0.05. Significant proteins with fold-change greater than two of Nc Spain 7 were located at the top left corner, while significant proteins with fold-change greater than two of Nc Spain 1H were located at the top right corner.

Table 4.1 Summary of the most important proteins identified as significantly differentially regulated of Nc Spain 1H at 12 hours pi.

Accession	Description	*Toxoplasma* Ortholog	Nc 1H-12hr	Nc 7-12hr	Max fold change	q Value
NCLIV_019580	SRS domain-containing protein	SAG-related sequence SRS35A	231858.78	35504.03	6.53	0.010746
NCLIV_022920	Renin, related	aspartyl protease ASP1 (ASP1)	34023.50	6502.10	5.23	0.030770
NCLIV_012090	putative ATP-dependent protease ATP-binding subunit	N/A	23024.58	6765.14	3.40	0.008507
NCLIV_018300	HIT family protein, related	Hit family protein involved in cell-cycle regulation, putative	1402055.32	421821.52	3.32	0.015779
NCLIV_012460	Lysyl-tRNA synthetase, related	hypothetical protein	2738.42	826.68	3.31	0.033022
NCLIV_037460	hypothetical protein	citrate synthase I	404581.56	152092.37	2.66	0.024397
NCLIV_007010	putative cathepsin C2	cathepsin CPC2 (CPC2)	274324.17	106677.59	2.57	0.024397
NCLIV_011110	putative DEAH-box RNA/DNA helicase	helicase associated domain (ha2) protein	20806.13	8151.25	2.55	0.017953
NCLIV_041850	putative Hsp20/alpha crystallin domain-containing protein	heat shock protein HSP28 (HSP28)	2691566.63	1056490.91	2.55	0.010746
NCLIV_069110	hypothetical protein	rhoptry protein ROP1 (ROP1)	27187.72	11024.39	2.47	0.049761
NCLIV_022970	unspecified product	microneme protein MIC2 (MIC2)	1968924.78	807753.17	2.44	0.008713
NCLIV_062770	unspecified product	microneme protein MIC9 (MIC9)	1338525.62	552301.84	2.42	0.037391
NCLIV_010600	putative microneme protein MIC3	microneme protein MIC3 (MIC3)	34648535.79	14310252.42	2.42	0.008507

On the other hand, from Table 4.2, Appendix Table1 and Figure 4.1, there were a smaller number of significant preferentially expressed proteins revealed in Nc Spain 7 at 12 hours pi comparing to Nc Spain 1H. Twenty-one proteins show most abundant in Nc Spain 7. SRS domain-containing protein (NCLIV_023620) orthologue of the SRS39 protein in *T. gondii* (TgSRS39) showed highest fold-change of 7.48 which determines as a surface protein of the parasite. The other preferentially expressed protein with fold-change 2.94 is Predicted rhoptry kinase, subfamily ROP20 (NCLIV_068850) orthologue of *T. gondii* predicted rhoptry kinase, subfamily ROP24 (TgROP24). This protein is suspected to play vital role in the early invasion process and later in invasion process. Putative serine/threonine protein phosphatase 5 (NCLIV_066900) show fold-change of 2.44, which is important for the invasion process of the parasite. There were some metabolic emzyme showing high abundant of protein expression in Nc Spain 7 at 12 hours pi which composed of transporter/ permease protein (NCLIV_039290) with fold-change 3.33, putative vacuolar ATP synthase subunit d (NCLIV_023610), putative acetyltransferase domain-containing protein (NCLIV_030730), putative CW-type zinc finger domain-containing protein (NCLIV_032350), Peroxiredoxin-2E-1, related (NCLIV_014020), putative regulator of chromosome condensation (NCLIV_013660) and high mobility group protein, related (NCLIV_043670). Eight hypothetical proteins have also found more abundant in Nc Spain 7 at 12 hours pi.

Table 4.2 Summary of the most important proteins identified as significantly differentially regulated of Nc Spain 7 at 12 hours pi.

Accession	Description	*Toxoplasma* Ortholog	Nc 1H-12hr	Nc 7-12hr	Max fold change	q Value
NCLIV_023620	SRS domain-containing protein	SAG-related sequence SRS39 (SRS39)	11502.31	86069.13	-7.48	0.010746
NCLIV_039290	hypothetical protein	transporter/permease protein	27564.48	91771.33	-3.33	0.013090
NCLIV_023610	putative vacuolar ATP synthase subunit d	V-type ATPase, D subunit protein	15790.05	46768.32	-2.96	0.017404
NCLIV_068850	Predicted rhoptry kinase, subfamily ROP20	rhoptry kinase family protein ROP24 (incomplete catalytic triad) (ROP24)	307278.32	902109.59	-2.94	0.013006
NCLIV_026210	hypothetical protein	apicoplast triosephosphate translocator APT1 (APT1)	184546.29	520631.34	-2.82	0.012491
NCLIV_066900	putative serine/threonine protein phosphatase 5	serine/threonine protein phosphatase	9279.11	22678.42	-2.44	0.017355
NCLIV_014020	Peroxiredoxin-2E-1, related	redoxin domain-containing protein	282524.95	673761.93	-2.38	0.013090
NCLIV_013660	putative regulator of chromosome condensation	regulator of chromosome condensation RCC1	6377.03	14349.80	-2.25	0.020239
NCLIV_043670	high mobility group protein, related	HMG (high mobility group) box domain-containing protein	1151938.37	2390237.21	-2.07	0.017404

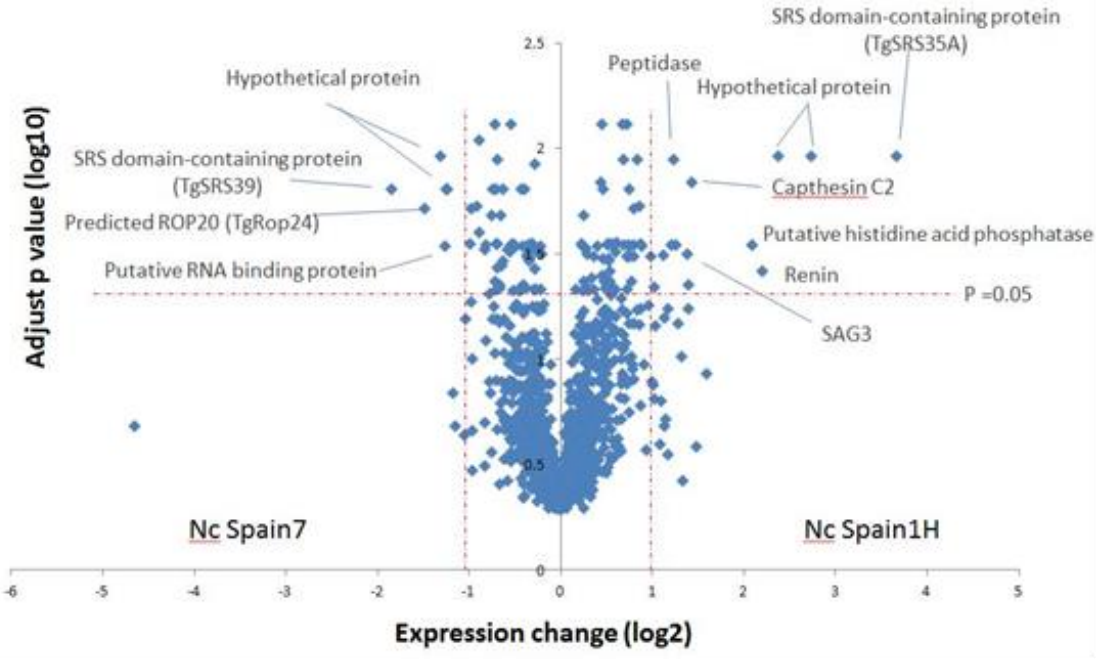

Figure 4.2 Volcano plot of quantified *Neospora* proteins of Nc Spain 1H (right) and Nc Spain 7 (left) at 36 hours pi to host cells.

At 36 hours pi, 136 proteins show differential expression between low (Nc Spain 1H) and high (Nc Spain 7) virulence strains of *N. caninum* with statistically significant difference (q value < 0.05). A much smaller number of significant proteins expression at 36 hours pi has found comparing to 12 hours pi that only 19 proteins showed differential expression with fold-change more than two. Among 19 differentaily expressed proteins, 13 proteins were preferentially expressed in Nc Spain 1H (Table 4.3, Appendix Table2 and Figure 4.2) and 6 proteins were preferentially expressed in Nc Spain 7 as shown in Table 4.4, Appendix Table2 and Figure 4.2.

Among 13 proteins which were preferentially expressed in Nc Spain 1H, there were 9 same proteins have found both at 12 hours pi and 36 hours pi (Figure 4.4). There were comprised of SRS domain-containing protein (NCLIV_019580) (TgSRS35A), SAG3 protein, related (NCLIV_034740), Renin, related (NCLIV_022920), putative histidine acid phosphatase domain containing protein (NCLIV_053400), putative cathepsin C2 (CPC2) (NCLIV_007010) , Oligoribonuclease, related (NCLIV_054180), Glucosamine-fructose-6-phosphate aminotransferase

(Isomerizing), related (NCLIV_031610) and two conserved hypothetical protein (NCLIV_036640, NCLIV_049520),

The rest of four proteins, which showed more abundant in Nc Spain 1H at 36 hours pi and were not found at 12 hours pi. included protein F23B2.12, partially confirmed by transcript evidence, related (NCLIV_008320), peptidase, S9A/B/C family, catalytic domain protein, related (NCLIV_063570), putative glycerol-3-phosphate acyltransferase (NCLIV_029980), and hypothetical protein (NCLIV_031190).

Table 4.3 Summary of the most important proteins identified as significantly differentially regulated of Nc Spain 1H at 36 hours pi.

Accession	Description	*Toxoplasma* Ortholog	Nc 1H-36hr	Nc 7-36hr	Max fold change	q Value
NCLIV_019580	SRS domain-containing protein	SAG-related sequence (SRS35A)	310976.44	24221.35	12.84	0.010883
NCLIV_022920	Renin, related	aspartyl protease ASP1 (ASP1)	32867.31	7075.07	4.65	0.038356
NCLIV_053400	putative histidine acid phosphatase domain containing protein	hypothetical protein	40155.73	9350.73	4.29	0.028609
NCLIV_007010	putative cathepsin C2 (CPC2)	cathepsin CPC2 (CPC2)	233283.80	85862.13	2.72	0.014564
NCLIV_008320	protein F23B2.12, partially confirmed by transcript evidence, related	RNA binding protein, putative	12493.30	4700.62	2.66	0.044566
NCLIV_034740	SAG3 protein, related	N/A	1041019.33	396426.25	2.63	0.031809
NCLIV_054180	Oligoribonuclease, related	small nuclease	41807.20	17333.89	2.41	0.028769
NCLIV_063570	Peptidase, S9A/B/C family, catalytic domain protein, related	acylaminoacyl-peptidase, putative	21390.60	9018.98	2.37	0.011348
NCLIV_029980	putative glycerol-3-phosphate acyltransferase	glycerol-3-phosphate-acyltransferase	3380.46	1537.80	2.20	0.031926

Six proteins were preferentially expressed in Nc Spain 7 at 36 hour pi with fold-change > 2 as shown in Table 4.4, Appendix Table2 and Figure 4.2. There were 3 proteins (from 6 proteins) found high abundant in both 12 hours pi and 36 hours pi (Figure 4.5) which composed of SRS domain-containing protein (NCLIV_023620) orthologue of the TgSRS39 with fold-change 3.57, Predicted rhoptry kinase, subfamily ROP20 (NCLIV_068850) orthologue of *T. gondii* rhoptry kinase family protein ROP24 (incomplete catalytic triad) (TgROP24) with fold-change 2.78, and Peroxiredoxin-2E-1, related (NCLIV_014020) with fold-change 2.34.

The other three more abundant proteins in Nc Spain 7 at 36 hours pi which were not found at 12 hours pi. included putative RNA binding protein (NCLIV_021440) and two conserved hypothetical protein (NCLIV_050060 and NCLIV_041790)

Table 4.4 Summary of the most important proteins identified as significantly differentially regulated of Nc Spain 7 at 36 hours pi.

Accession	Description	*Toxoplasma* Ortholog	Nc 1H-36hr	Nc 7-36hr	Max fold change	q Value
NCLIV_023620	SRS domain-containing protein	SAG-related sequence SRS39 (SRS39)	25932.55	92463.37	-3.57	0.015602
NCLIV_068850	Predicted rhoptry kinase, subfamily ROP20	rhoptry kinase family protein ROP24 (incomplete catalytic triad) (ROP24)	258639.81	718755.68	-2.78	0.019337
NCLIV_021440	putative RNA binding protein	RNA binding protein, putative	2927.82	6963.26	-2.38	0.028897
NCLIV_014020	Peroxiredoxin-2E-1, related	redoxin domain-containing protein	374177.49	877317.10	-2.34	0.015602

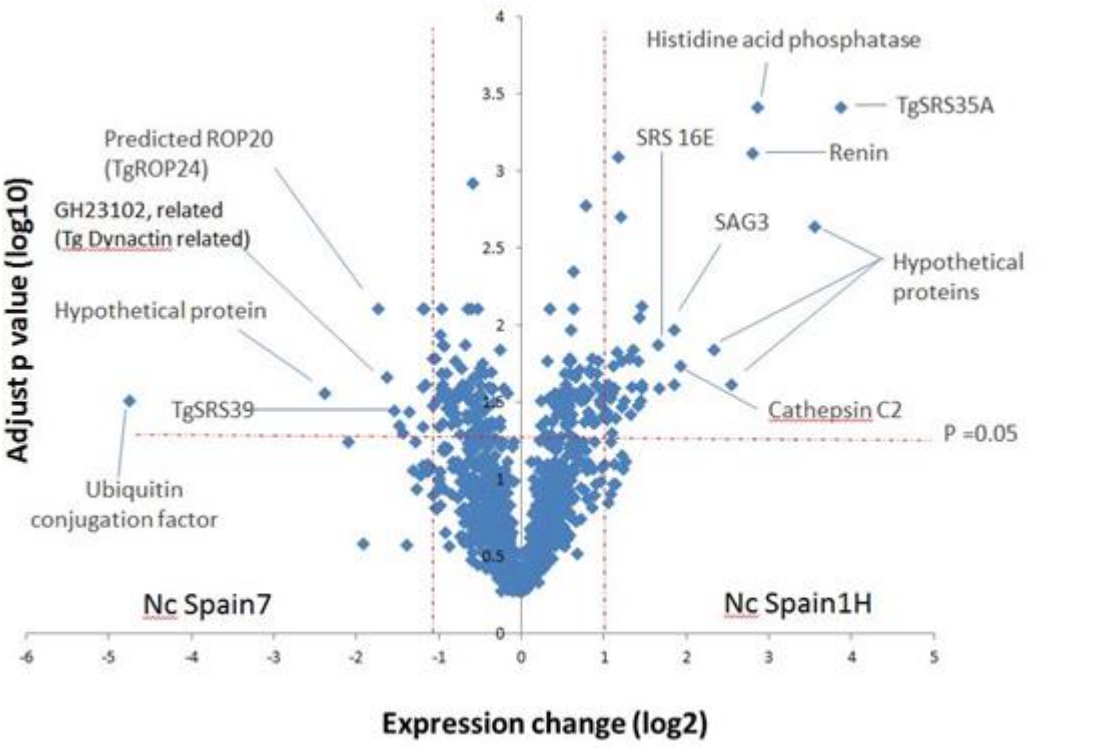

Figure 4.3 Volcano plot of quantified *Neospora* proteins of Nc Spain 1H (right) and Nc Spain 7 (left) at 56 hours pi to host cells.

At 56 hours pi, 214 proteins showed differential expression between Nc Spain 1H and Nc Spain 7. However, 62 proteins exhibited preferential expression between low (Nc Spain 1H) and high (Nc Spain 7) virulence strains with statistically significant difference (q value < 0.05) at fold-change more than two. Among 62 proteins, 44 proteins were more abundant in Nc Spain 1H (Table 4.5, Appendix Table3 and Figure 4.3) and 18 proteins were significantly preferential expressed in Nc Spain 7 as shown in Table 4.6 , Appendix Table3 and Figure 4.3.

Forty-four proteins showed preferentially expressed in Nc Spain 1H with statistically significant difference (q value < 0.05) and fold-change > 2 (Table 4.5, Appendix Table 3 and Figure 4.3). There were 7 proteins (from 44 proteins) found preferentially expressed in all three time points at 12, 36 and 56 hours pi respectively (Figure 4.4). Three out of seven proteins were related to host cell invasion which composed of putative cathepsin C2 (CPC2) (NCLIV_007010) and two surface antigens as SAG3 protein, related (NCLIV_034740) and SRS domain-containing protein (NCLIV_019580) orthologue of the TgSRS35A protein. The other four

proteins comprised of putative Renin, related (NCLIV_022920), histidine acid phosphatase domain containing protein (NCLIV_053400) and two conserved hypothetical proteins (NCLIV_049520 and NCLIV_036640).

There were 12 proteins found more abundant in Nc Spain 1H at both 12 and 56 hours pi ($q < 0.05$ and FC >2). Surprisingly, a group of microneme proteins which is important for host cell attachment were found in these two time points but not in 36 hours pi such as two microneme protein MIC2 (TgMIC2) (NCLIV_022970 and NCLIV_033690), putative microneme protein MIC3 (NCLIV_010600) and microneme protein MIC4 (MIC4) (NCLIV_002940). The other proteins which associated to host cell invasion were presented in these time points as subtilisin SUB1 (SUB1) (NCLIV_021050) and putative apical membrane antigen 1 (AMA1) (NCLIV_028680). Some metabolic enzyme proteins such as putative 3',5'-cyclic phosphodiesterase (NCLIV_019740), putative DEAH-box RNA/DNA helicase (NCLIV_011110), perforin-like protein PLP1 (PLP1) (NCLIV_020990), putative signal recognition particle protein (NCLIV_019540), Lysyl-tRNA synthetase, related (NCLIV_012460) and conserved hypothetical protein (NCLIV_065500) were revealed preferentially expressed in these time points.

Three high abundant proteins of Nc Spain 1H found in both 36 and 56 hours pi ($q < 0.05$ and FC >2) composed of some metabolic enzymes as Peptidase, S9A/B/C family, catalytic domain protein, related (NCLIV_063570), putative glycerol-3-phosphate acyltransferase (NCLIV_029980) and protein F23B2.12, partially confirmed by transcript evidence, related (NCLIV_008320).

There were 22 proteins which exhibited preferentially expression in Nc Spain 1 H at only 56 hours pi ($q < 0.05$ and FC >2). More surface antigen proteins had been found as SRS domain-containing protein (NCLIV_010050) ortholog of SAG-related sequence SRS16E in *T. gondii* (TgSRS16E), and hypothetical protein (NCLIV_046140) ortholog of SAG-related sequence SRS67 (TgSRS67). Moreover the other microneme proteins were found more abundant at 56 hours pi as putative microneme protein MIC1 (NCLIV_043270), putative MIC2-associated protein M2AP (NCLIV_051970), microneme protein MIC10 (MIC10) (NCLIV_066250),

putative microneme protein MIC11 (NCLIV_020720). The cathepsin L, related (NCLIV_004380) which related to host cell invasion was also found in this time points. The ALVEOLIN1, related (NCLIV_026500) was found in Nc Spain 1 H at 56 hours pi. This protein play key role in maintenance of the cell shape in *T. gondii.* The other high abundant proteins were found at 56 hours pi composed of some metabolics enzymes such as Protein-L-isoaspartate O- methyltransferase, related (NCLIV_057890), dolichol-phosphate-mannose synthase family protein (NCLIV_068040), class I chitinase, related (NCLIV_000740), putative peptidase M16 domain containing protein (NCLIV_044230), GD24343, related (NCLIV_025760), ABC transporter transmembrane region domain-containing protein (NCLIV_009810), Acyl-CoA synthetase, related (NCLIV_054250) and small GTPase Rab2, putative (NCLIV_055690), putative rRNA processing protein (NCLIV_047570) and some hypothetical proteins (NCLIV_038110, NCLIV_061620, NCLIV_060220, NCLIV_031300 and NCLIV_018530).

Table 4.5 Summary of the most important proteins identified as significantly differentially regulated of Nc Spain 1H at 56 hours pi.

Accession	Description	*Toxoplasma* Ortholog	Nc 1H-56hr	Nc 7-56hr	Max fold change	q Value
NCLIV_019580	SRS domain-containing protein	SAG-related sequence SRS35A (SRS35A)	405070.47	27750.65	14.60	0.029754
NCLIV_053400	putative histidine acid phosphatase domain containing protein	histidine acid phosphatase superfamily protein	50228.98	6897.47	7.28	0.027695
NCLIV_022920	Renin, related	aspartyl protease ASP1 (ASP1)	48104.27	6882.95	6.99	0.030872
NCLIV_038110	hypothetical protein	microneme protein MIC17C (MIC17C)	53036.03	9072.69	5.85	0.036448
NCLIV_007010	putative cathepsin C2 (CPC2)	cathepsin CPC2 (CPC2)	436309.74	114586.24	3.81	0.041009
NCLIV_034740	SAG3 protein, related	N/A	1452783.02	400652.78	3.63	0.016437
NCLIV_008320	protein F23B2.12, partially confirmed by transcript evidence, related	serine carboxypeptidase s28 protein	18642.01	5178.58	3.60	0.026346
NCLIV_057890	Protein-L-isoaspartate O-methyltransferase, related	L-isoaspartyl protein carboxyl methyltransferase family protein	43970.20	13736.82	3.20	0.024390
NCLIV_010050	SRS domain-containing protein	SAG-related sequence SRS16E (SRS6)	220825.35	70365.37	3.14	0.028942
NCLIV_033690	Hypothetical protein	microneme protein MIC2 (MIC2)	5164280.08	1867479.75	2.77	0.033339

Accession	Description	*Toxoplasma* Ortholog	Nc 1H-56hr	Nc 7-56hr	Max fold change	q Value
NCLIV_002940	putative microneme protein MIC4	microneme protein MIC4 (MIC4)	24897767.32	9017437.25	2.76	0.016437
NCLIV_019740	putative 3',5'-cyclic phosphodiesterase	3'5'-cyclic nucleotide phosphodiesterase domain-containing protein	3620.19	1317.63	2.75	0.029575
NCLIV_028680	putative apical membrane antigen 1	apical membrane antigen AMA1	6386731.22	2354879.39	2.71	0.049602
NCLIV_020990	perforin-like protein PLP1 (PLP1)	perforin-like protein PLP1 (PLP1)	311947.73	115987.59	2.69	0.027695
NCLIV_068040	dolichol-phosphate-mannose synthase family protein	dolichol-phosphate-mannose synthase family protein	29312.09	11032.95	2.66	0.025724
NCLIV_043270	putative microneme protein MIC1	microneme protein MIC1 (MIC1)	35922016.21	14004447.55	2.57	0.025376

Eighteen proteins revealed significant preferentially expressed in Nc Spain 7 at 56 hours pi as shown in Table 4.6, Appendix Table 3 and Figure 4.3. Among 18 proteins, only 2 proteins were found preferentially expressed in all time points as Predicted rhoptry kinase, subfamily ROP20 (NCLIV_068850) orthologue of *T. gondii* predicted rhoptry kinase, subfamily ROP24 (TgROP24) and SRS domain-containing protein (NCLIV_023620) orthologue of the TgSRS39 protein (TgSRS39). The other two proteins were shown preferentially expressed in 12 and 56 hours pi as putative acetyltransferase domain-containing protein (NCLIV_030730) and hypothetical protein (NCLIV_026210). Ferthermore, 2 proteins were exhibited high abundant in 36 and 56 hours pi as putative RNA binding protein (NCLIV_021440) and conserved hypothetical protein (NCLIV_050060) (Figure: 3.5).

The other 12 high abundant proteins which were seen only at 56 hours pi were as following. The putative ubiquitin conjugation factor (NCLIV_019660) was shown the highest fold-change of 26.90 and the putative ubiquitin family domain-containing protein (NCLIV_012950) was also found in this time point. The ubiquitin protein is involved in cell division and host cell invasion of the parasite. GH23102, related (NCLIV_063510) ortholog of dynactin related protein in *T. gondii* which involed in intracellular transport. Moreover, putative Rhoptry kinase family protein, truncated (incomplete catalytic triad) (NCLIV_007770) was shown in 56 hours pi. The Rhoptry kinase protein play an important role in host cell invasion in Apicomplexa parasite. Some metabolic emzymes were found in this time point such as, putative opine dehydrogenase (NCLIV_042450), putative TPR domain-containing protein (NCLIV_037480), YbaK / prolyl-tRNA synthetases associated domain containing protein (NCLIV_0292) and beta-1 tubulin, putative (NCLIV_005150). Four conserved hypothetical protein (NCLIV_014040, NCLIV_025990, NCLIV_011740 and NCLIV_046320) were also found in 56 hours pi which their function were not assigned

Table 4.6 Summary of the most important proteins identified as significantly differentially regulated of Nc Spain 7 at 56 hours pi.

Accession	Description	*Toxoplasma* Ortholog	Nc 1H-56hr	Nc 7-56hr	Max fold change	q Value
NCLIV_019660	putative ubiquitin conjugation factor	U-box domain-containing protein	137.21	3690.97	-26.90	0.030673
NCLIV_068850	Predicted rhoptry kinase, subfamily ROP20	rhoptry kinase family protein ROP24 (incomplete catalytic triad) (ROP24)	335590.44	1111069.09	-3.31	0.002306
NCLIV_063510	GH23102, related	Dynactin related protein	3411.84	10578.65	-3.10	0.000387
NCLIV_023620	SRS domain-containing protein	SAG-related sequence SRS39 (SRS39)	27474.34	80197.10	-2.92	0.000768
NCLIV_042450	putative opine dehydrogenase	NAD/NADP octopine/nopaline dehydrogenase, alpha-helical domain-containing protein	261448.45	725179.73	-2.77	0.024390
NCLIV_037480	putative TPR domain-containing protein	tetratricopeptide repeat-containing protein	7415.09	20097.99	-2.71	0.027695
NCLIV_026210	apicoplast triosephosphate translocator APT1 (APT1)	apicoplast triosephosphate translocator APT1 (APT1)	150238.98	343906.78	-2.29	0.010723
NCLIV_007770	putative Rhoptry kinase family protein, truncated (incomplete catalytic triad)	Rhoptry kinase family protein, truncated (incomplete catalytic triad)	433281.75	981688.73	-2.27	0.007757
NCLIV_012950	putative ubiquitin family domain-containing protein	ubiquitin family protein	5902.62	12357.41	-2.09	0.021485

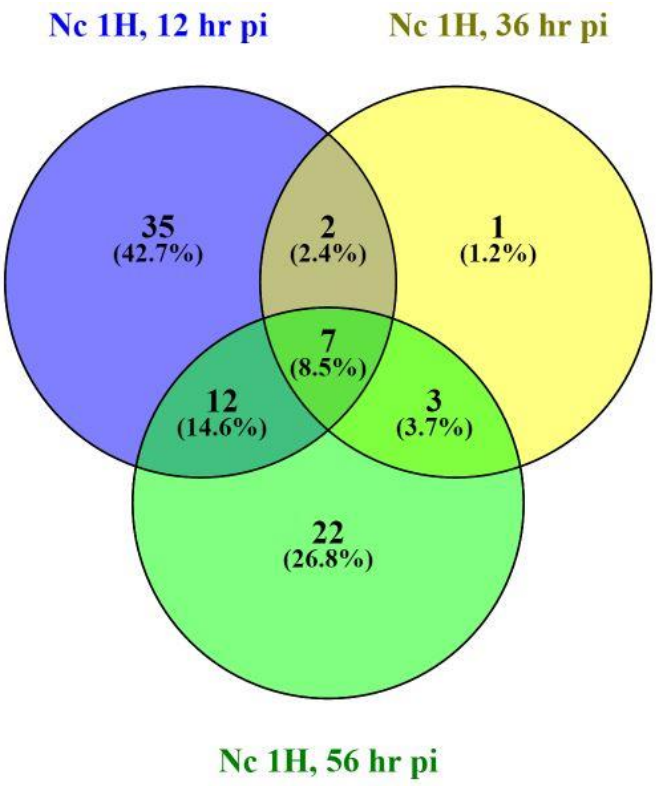

Figure 4.4 Venn diagram showing the number of significant ($q < 0.05$) preferentially expressed of proteins in Nc Spain 1H in three time points as 12, 36, 56 hours pi., respectively. Seven proteins were found more abundant in all three time points in Nc Spain 1H

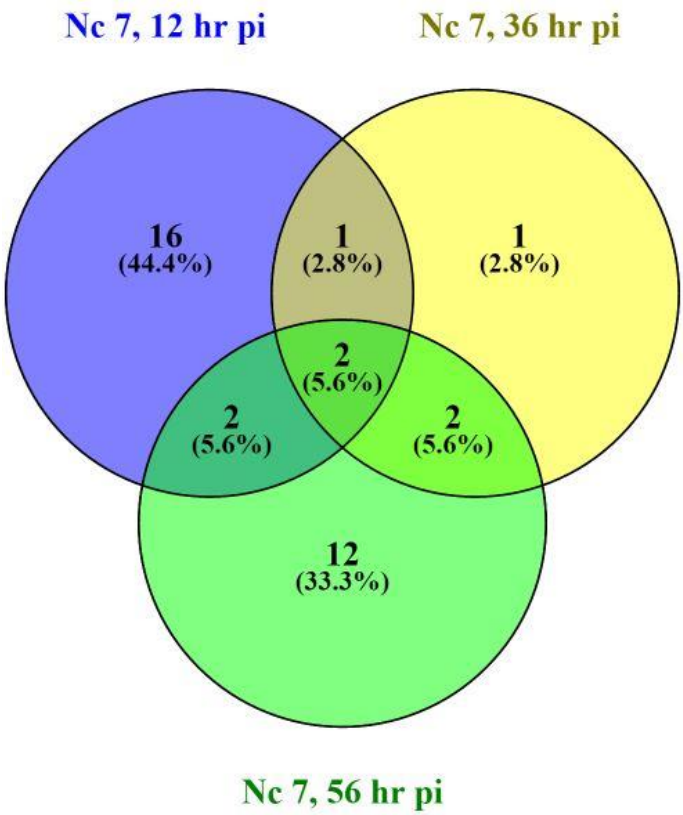

Figure 4.5 Venn diagram showing the number of significantl ($q < 0.05$) preferentially expressed of proteins in Nc Spain 7 in three time points as 12, 36, 56 pi respectively. Two proteins were found more abundant in all three time points in Nc Spain 7

The venn diagram showed the number of high abundant of proteins found in one time points or overlapping to the other time points. The most protein abundant in Nc Spain 1H which were found in all time course included SRS domain-containing protein (NCLIV_019580), conserved hypothetical protein (NCLIV_049520), conserved hypothetical protein (NCLIV_036640), Renin, related (NCLIV_022920), putative histidine acid phosphatase domain containing protein (NCLIV_053400), putative cathepsin C2 (CPC2) (NCLIV_007010) and SAG3 protein, related (NCLIV_034740) (Figure 4.4). Conversely, there were 2 most protein abundant in Nc Spain 7 which were found in all time course were SRS domain-containing protein (SRS39) (NCLIV_023620) and predicted rhoptry kinase, subfamily ROP24 NCLIV_068850.

4.3.2 GO enrichment analysis

The proteins that showed preferentially expressed in Nc Spain 1H and Nc Spain 7in three time points as 12, 36 and 56 hours pi, respectively were grouped together and individually searched for their GO term using ToxoDB database.

The GO term enrichment analysis could be obtained from only preferentially expressed proteins in Nc Spain 1H for all time points (Table 4.7, 3.8 and 3.9). However, the GO term enrichment analysis for more abundant proteins in Nc Spain 7 returned no significant GO term as there are small number of protein to be an input. The number of preferentially expressed proteins input for GO term in Nc Spain7 were 21 proteins of 12 hr pi, 6 proteins of 36 hr pi and 18 proteins of 56 hr pi, respectively.

Among 56 preferentially expressed proteins of low-virulence Nc Spain 1H, there were 14 significantly enriched biological process terms (Benjamini < 0.05), as shown in appendix Table 4. However, only carbohydrate metabolic process (GO:0005975), metabolic process (GO:0008152) and primary metabolic process (GO:0044238) of biological process GO terms revealed a statistically significant. The other biological process and molecular function terms illustrated a single or few proteins and some terms are related to the same proteins as a result in extensive redundancy among the results.

Seven proteins that found associated with carbohydrate metabolic process **(**GO**:**0005975**)** were Glucosamine--fructose-6-phosphate aminotransferase **(**Isomerizing**)**, related **(**NCLIV_031610**)**, putative phosphoglucomutase **(**NCLIV_010960**)**, Fructose-1,6-bisphosphatase, related **(**NCLIV_050070**)**, Fructose-1,6-bisphosphatase class 1, related **(**NCLIV_050080**)**, cbr-GPDH-1 protein, related **(**NCLIV_001180**)**, and two hypothetical protein **(**NCLIV_037460 and NCLIV_004200**)**

4.3.2.1 Gene Ontology (GO) enrichments of the low- and high- virulence of *N. caninum* in different time points

Table 4.7 Protein Ontology enrichment: Biological Process and Molecular Function of 56 high abundant proteins in Nc Spain 1H at 12 hours post infection obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0005975	carbohydrate metabolic process	86	8	9.3	7.48	9	0.000013	0.000186
GO:0008152	metabolic process	1317	24	1.8	1.46	2.08	0.013670	0.049091
GO:0044238	primary metabolic process	1075	20	1.9	1.5	1.95	0.024545	0.049091
Molecular Function:								
GO:0003824	catalytic activity	1365	23	1.7	1.35	1.78	0.043708	0.048274

Table 4.8 Gene Ontology enrichment: Biological Process and Molecular Function of 13 high abundant proteins in Nc Spain 1H at 36 hours post infection obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0006508	proteolysis	133	4	3	12.69	24.38	0.00016	0.00063
GO:0019538	protein metabolic process	607	4	0.7	2.78	4.56	0.03996	0.04425
GO:0008152	metabolic process	1317	6	0.5	1.92	4.69	0.04425	0.04425
Molecular Function:								
GO:0016787	hydrolase activity	537	5	0.9	3.93	8.81	0.00380	0.00945

Table 4.9 Gene Ontology enrichment: Biological Process and Molecular Function of 44 high abundant proteins in Nc Spain 1H at 56 hours post infection obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0006508	proteolysis	133	9	6.8	8.79	12.91	0.000001	0.000030
GO:0019538	protein metabolic process	607	11	1.8	2.35	3.35	0.003563	0.028737
GO:0008150	biological_process	1771	20	1.1	1.47	3.02	0.009605	0.028737
GO:0043170	macromolecule metabolic process	823	12	1.5	1.89	2.66	0.013039	0.028737
GO:0044238	primary metabolic process	1075	14	1.3	1.69	2.5	0.016877	0.028737
GO:0008152	metabolic process	1317	16	1.2	1.58	2.5	0.017164	0.028737
Molecular Function:								
GO:0070011	peptidase activity, acting on L-amino acid peptides	99	7	7.1	9.18	12.2	0.000009	0.000122
GO:0008233	peptidase activity	106	7	6.6	8.57	11.37	0.000014	0.000122

Table 4.10 Identified genes involved in GO term: biological process and molecular function of preferentially expressed proteins of Nc Spain 1 H at 12 hous pi obtained fromToxoDB

Gene ID	Description	*Toxoplasma* ortholog	Biological process	Molecular Function
NCLIV_001180	cbr-GPDH-1 protein, related	NAD-dependent glycerol-3-phosphate dehydrogenase	carbohydrate metabolic process, glycerol-3-phosphate catabolic process, oxidation-reduction process	NAD binding, glycerol-3-phosphate dehydrogenase [NAD+] activity, protein homodimerization activity
NCLIV_010960	putative phosphoglucomutase	glucosephosphate-mutase GPM2 (GPM2)	carbohydrate metabolic process	null
NCLIV_031610	Glucosamine--fructose-6-phosphate aminotransferase (Isomerizing), related	glucosamine-fructose-6-phosphate aminotransferase	carbohydrate biosynthetic process, carbohydrate metabolic process	carbohydrate binding, glutamine-fructose-6-phosphate transaminase (isomerizing) activity
NCLIV_037460	hypothetical protein	citrate synthase I	cellular carbohydrate metabolic process	null
NCLIV_050070	Fructose-1,6-bisphosphatase,related	sedoheptulose-1,7-bisphosphatase	carbohydrate metabolic process	phosphoric ester hydrolase activity
NCLIV_050080	Fructose-1,6-bisphosphatase class 1, related	sedoheptulose-1,7-bisphosphatase	carbohydrate metabolic process	phosphoric ester hydrolase activity

Gene ID	Description	*Toxoplasma* ortholog	Biological process	Molecular Function
NCLIV_004200	hypothetical protein	glycosyltransferase	biosynthetic process, carbohydrate metabolic process	catalytic activity, cation binding
NCLIV_035300	putative lipoate-protein ligase a	lipoyltransferase and lipoate-protein ligase subfamily protein	cellular protein modification process, protein lipoylation	catalytic activity, transferase activity
NCLIV_018300	HIT family protein, related	Hit family protein involved in cell-cycle regulation, putative	null	catalytic activity
NCLIV_067490	putative protein phosphatase 2C	protein phosphatase 2C domain-containing protein	null	catalytic activity

Table 4.11 Identified proteins involved in GO term: biological process of preferentially express proteins of Nc Spain 1 H at 36 hous pi obtained ToxoDB

Gene ID	Description	*Toxoplasma* ortholog	Biological process
NCLIV_007010	putative cathepsin C2 (TgCPC2)	cathepsin CPC2 (CPC2)	proteolysis
NCLIV_008320	protein F23B2.12, partially confirmed by transcript evidence, related	serine carboxypeptidase s28 protein	proteolysis
NCLIV_022920	Renin, related	aspartyl protease ASP1 (ASP1)	proteolysis
NCLIV_029980	putative glycerol-3-phosphate acyltransferase	glycerol-3-phosphate-acyltransferase	metabolic process
NCLIV_031610	Glucosamine--fructose-6-phosphate aminotransferase (Isomerizing), related	glucosamine-fructose-6-phosphate aminotransferase	carbohydrate biosynthetic process, carbohydrate metabolic process
NCLIV_063570	Peptidase, S9A/B/C family, catalytic domain protein, related	acylaminoacyl-peptidase,	proteolysis

Table 4.12 Identified genes involved in GO term: biological process and molecular function of preferentially expressed proteins of Nc Spain 1 H at 56 hous pi obtained from ToxoDB

Gene ID	Description	*Toxoplasma* ortholog	Biological process	Molecular Function
NCLIV_002940	putative microneme protein MIC4	microneme protein MIC4 (MIC4)	blood coagulation, proteolysis	protein binding
NCLIV_004380	cathepsin L, related	cathepsin CPL (CPL)	proteolysis	cysteine-type peptidase activity
NCLIV_007010	putative cathepsin C2 (TgCPC2)	cathepsin CPC2 (CPC2)	proteolysis	cysteine-type peptidase activity
NCLIV_008320	protein F23B2.12, partially confirmed by transcript evidence, related	serine carboxypeptidase s28 protein	proteolysis	serine-type peptidase activity
NCLIV_021050	unspecified product	subtilisin SUB1 (SUB1)	proteolysis	serine-type endopeptidase activity
NCLIV_022920	Renin, related	aspartyl protease ASP1 (ASP1)	proteolysis	aspartic-type endopeptidase activity
NCLIV_038110	hypothetical protein	microneme protein MIC17C (MIC17C)	blood coagulation, proteolysis	protein binding
NCLIV_044230	putative peptidase M16 domain containing	toxolysin TLN4 (TLN4)	proteolysis	catalytic activity, metal ion binding, metalloendopeptidase

Gene ID	Description	*Toxoplasma* ortholog	Biological process	Molecular Function
	protein			activity, zinc ion binding
NCLIV_063570	Peptidase, S9A/B/C family, catalytic domain protein, related	acylaminoacyl-peptidase,	proteolysis	serine-type peptidase activity

Among 44 preferential expressed proteins of low-virulence Spain 1H, 59 biological process term were enriched (Benjamini < 0.05), as shown in appendix Table 7. However, only proteolysis (GO:0006508) of biological process GO terms showed statistically significant and was not redundancy of proteins with enrichment GO terms.

Nine proteins that found related to proteolysis involved putative microneme protein MIC4 (NCLIV_002940), cathepsin L, related (NCLIV_004380), putative cathepsin C2 (TgCPC2) (NCLIV_007010), protein F23B2.12, partially confirmed by transcript evidence, related (NCLIV_008320), unspecified product (NCLIV_021050), Renin, related (NCLIV_022920), hypothetical protein (NCLIV_038110), putative peptidase M16 domain containing protein (NCLIV_044230) and eptidase, S9A/B/C family, catalytic domain protein, related (NCLIV_063570).

There were four proteins involved in **proteolysis** in Nc Spain 1 H at 36 hours pi which comprised of protease or peptidase enzymes as putative cathepsin C2 (TgCPC2) (NCLIV_007010) protein F23B2.12, partially confirmed by transcript evidence, related (NCLIV_008320) ortholog of serine carboxypeptidase s28 protein in *T. gondii*, Renin, related (NCLIV_022920) ortholog of aspartyl protease ASP1 (TgASP1) and Peptidase, S9A/B/C family, catalytic domain protein, related (NCLIV_063570) (Table 3.11)

There were nine proteins related to proteolysis in Nc Spain 1 H at 56 hour pi which comprised of two micronemes proteins as putative microneme protein MIC4 (NCLIV_002940) and hypothetical protein (NCLIV_038110) ortholog of *T. gondii* microneme protein MIC17C (TgMIC17C), seven peptidase or protease enzymes as cathepsin L related (NCLIV_004380) and putative cathepsin C2 (TgCPC2) (NCLIV_007010), protein F23B2.12, partially confirmed by transcript evidence, related (NCLIV_008320) ortholog of serine carboxypeptidase s28 protein in *T. gondii*, Renin, related (NCLIV_022920) ortholog of aspartyl protease ASP1 (TgASP1), Peptidase, S9A/B/C family, catalytic domain protein, related (NCLIV_063570), unspecified product (NCLIV_021050) ortholog of *T. gondii*

subtilisin SUB1 (TgSUB1), putative peptidase M16 domain containing protein ortholog of *T. gondii* toxolysin TLN4 (TgTLN4) (Table 3.12)

4.4 Discussion

The aim of this study is to identify key differences in the protein expression of low- (Nc Spain 1H) and high- (Nc Spain 7) virulence *N. caninum* at three different time points post-infection. In this study, all experiments have used three biological replicates of *N. caninum* rather than five because a minimum of three biological replicates was sufficient for detecting naturally biological variation of samples caused either by genetic variation or by the effect of environmental factors (Westermeier et al., 2008). Many label free quantitative proteomic studies were chosen three biological replicates in their study (Dudekula and Le Bihan, 2016; Zhang et al., 2015). The financial constraints to run samples, laborious and time consuming are major factors to keep in considering to select an appropriate amount of *N. caninum* replicates. However, more reliable proteomics profile will be increased if the experimental study includes more replicates. GO term enrichment was used to infer possible functional roles of differentially expressed proteins in each parasite strains and at each time point. GO enrichment analysis could only be performed in Nc Spain 1H at all three time points due to insufficient terms in other conditions. No GO terms were significantly enriched in Nc Spain 7 at any time point.

4.4.1 GO enrichment analysis of preferentially expressed proteins in low-virulence Nc Spain 1H at three time points

Carbohydrate metabolism (GO:0005975) was significantly enriched at 12 hours pi. This process is involved in the formation, breakdown and interconversion of carbohydrates. Seven preferentially expressed proteins in Nc Spain 1H are associated with carbohydrate metabolism, i.e. glucosamine-fructose-6-phosphate aminotransferase (isomerizing), phosphoglucomutase, fructose-1,6-bisphosphatase, fructose-1,6-bisphosphatase class 1, cbr-GPDH-1 protein and two hypothetical proteins.

Fructose-1,6-bisphosphate aldolase is important for cell invasion by *T. gondii. T. gondii* tachyzoites invade host cells using an acomysin motor connected to extracellular adhesins, in which fructose-1,6-bisphosphate aldolase acts as a bridging protein. This enzyme also has a role in glycolysis (Boucher and Bosch, 2014).

Glucosamine-fructose-6-phosphate aminotransferase (GFAT) plays a key role in amino sugar biosynthesis and glycosylation (Durand et al., 2008).

There were 13 preferentially expressed proteins in Nc Spain 1H at 36 hours pi that could be analyzed for GO enrichment. Four biological process terms were significantly enriched, as shown in appendix Table 6. These were proteolysis, carbohydrate biosynthetic process, protein metabolic process and metabolic process.

Four proteins associated with the proteolysis term were putative cathepsin C2 (TgCPC2) (NCLIV_007010), protein F23B2.12, partially confirmed by transcript evidence, related (NCLIV_008320), Renin, related (NCLIV_022920) and Peptidase, S9A/B/C family, catalytic domain protein, related (NCLIV_063570). The proteolysis GO term is discussed below as this GO term was also found at 56 hr pi of Nc Spain1H

Among GO terms applied to 44 preferentially expressed proteins in Spain 1H at 56 hours pi, 59 biological process terms were enriched (Benjamini < 0.05), as shown in appendix Table 7. However, only proteolysis (GO:0006508) showed statistically significance and non-redundancy with other GO terms. Nine proteins were associated with proteolysis; these wereputative microneme protein MIC4 (NCLIV_002940), cathepsin L, related (NCLIV_004380), putative cathepsin C2 (TgCPC2) (NCLIV_007010), protein F23B2.12, partially confirmed by transcript evidence, related (NCLIV_008320), unspecified product (NCLIV_021050), Renin, related (NCLIV_022920), hypothetical protein (NCLIV_038110), putative peptidase M16 domain containing protein (NCLIV_044230) and eptidase, S9A/B/C family, catalytic domain protein, related (NCLIV_063570).

The ‘proteolysis’ term is enriched in Nc Spain 1H at 36 and 56 hours pi. Proteolysis covers all process involved in breaking peptide bonds. This process plays a vital role in cell and tissue invasion as well as parasite egression from host cells (Miller et al., 2001; Tampaki et al., 2015). Inhibition of proteases in *T. gondii* resulted in blocking parasite invasion and intracellular development (Conseil et al., 1999; Shaw et al., 2002). In this study, preferentially expressed proteins in Nc Spain 1H associated to

proteolysis were composed of a group of micronemes such as MIC4 and TgMIC17C which found in 56 hours pi. Moreover protease enzymes such as cathepsin L, cathepsin C2 (TgCPC2), serine carboxypeptidase s28 protein, renin, S9A/B/C family peptidase, subtilisin SUB1 (TgSUB1), a M16 domain-containing peptidase are also involved in the proteolytic process. The microneme secretes numerous soluble and transmembrane1-anchored proteins during the apical attachment of the parasite to host cells. Microneme proteins have variety of adhesive domains, which can bind to the receptors on mammalian cells (Soldati et al., 2001). These groups of proteins help the parasite to penetrate the host cell. Cathepsins are also proteases that are important for invading cells by apicomplexan parasite such as *Toxoplasma* and *Plasmodium* (Mayer et al., 1991; Que et al., 2002**).** Cell invasion is facilitated by extensive proteolysis of surface protein help for adhesion and penetration **(**Dou & Carruthers, 2011**).** Proteolysis is also associated with micronemes and rhoptry proteins, which are involved in three steps of invasion process. Firstly, proteolytic maturation en route to secretory organelles. Secondly, microneme proteins are trimmed to produce their functional forms after they are secreted on the parasite surface. Finally, microneme is shedding their products, which are released from the parasite surface during, and after they move toward the posterior end. For example, microneme protein 4 (MIC4) is secreted by *T. gondii* tachyzoites during microneme stimulation and exocytosis. Also, the rhoptry protein (ROP1) functions in post-secretory proteolysis within the parasitophorous vacuole after parasite invasion (Bradley et al., 2002). The reason why proteolytic functions are enriched in low-virulence Nc Spain 1H may be due to its essential role in cell invasion by *N. caninum* and all apicomplexan parasites. Greater efficiency in cell invasion by this strain may result in lower virulence since a more rapid invasion may cause less severe pathogenesis in the host.

4.4.2 Preferential protein expression in the low-virulence strain Nc Spain 1H

The high abundance proteins of Nc Spain 1H found in all three time points (12, 36 and 56 pi) were cathepsin C2 protein and two surface antigens known as SRS domain-containing protein (TgSRS35A) and SAG3. Other high abandance proteins involved in host cell attachment and invasion found at 12 hours pi only included

rhoptry protein ROP1 (ROP1), microneme protein (MIC), MIC2, MIC3, MIC4 and MIC9, putative protein phosphatase 2C. The group of microneme proteins such as MIC1, MIC2, MIC3, MIC4, MIC 10, MIC11 were more abundant at 56 hours pi as were three further SRS domain-containing proteins. Cathepsin L also showed high abundance at 56 hours pi. These virulence-related proteins are discussed in the following paragraphs.

4.4.2.1 Surface antigen proteins

The surfaces of *T. gondii* and *N. caninum* tachyzoites are coated with glycosylphosphatidylinositol (GPI) -linked proteins that are homologous to the surface antigen SAG1 (Jung et al., 2004). The surface antigens are recognised as the SRS or SAG1-related superfamily of proteins which mediate attachment to host cells, and are thought to mediate the host immune response and so regulate virulence in *T. gondii* (Lekutis et al., 2001). The two surface antigens proteins shown to be preferentially expression at all three time points are SAG3 and SRS domain-containing protein, the latter an orthologue of SRS35A in *T. gondii* (TgSRS35A).

SAG3 is one out of five major surface antigens which play an important role in host cell attachment and invasion by *T. gondii* (Tomavo, 1996). It is assumed that SAG3 is a *T. gondii* receptor that binds host cell ligands to mediate cell attachment (Dzierszinski et al., 2000). SAG3 can be found both in tachyzoites and bradyzoites stage; it binds sulphated proteoglycans (SPGs) with high affinity (Jacquet et al., 2001). and can interact specifically with cellular heparin sulfate proteoglycan that allows *T. gondii* to attach to host cells (Burg et al., 1988). A tachyzoites mutant that lack SAG3 display a decreased rate of host cell adhesion and reduced virulence in mice (Dzierszinski et al., 2000).

SRS domain-containing protein (i.e. NcSRS35A) is highly abundant in Nc Spain 1H, with the highest fold-change at all three time points. Two other SRS proteins (NcSRS16E and NcSRS67) were most abundant at 56 hours pi. The SRS proteins of *N. caninum* are also localised at the surface of the parasite (Reid et al., 2012). Most of them are members of the surface antigen 1 (SAG1) or SAG2 families (Boothroyd

et al., 1998; Lekutis et al., 2000; Manger et al., 1998). These proteins seem to play important roles in host cell invasion, immune modulation and/or virulence attenuation (Lekutis et al., 2001). As mentioned previously in chapter 2, overexpression of SRS proteins attenuates the virulence of *T. gondii*. (Wasmuth et al. (2012)) found that the transgenic overexpression of SRS29C protein can reduce the virulence of *T. gondii* RH type I. Moreover, outbred mice experimentally infected with transgenic *T. gondii* overexpressing SRS29C survived, while all outbred mice infected with WT strain died. These surface antigen proteins may lower virulence of *N. caninum* also, explaining their particular abundance in Nc Spain 1H. However, studies of SRS proteins in *N. caninum* is very limited and further studies are need, for example, oblating SRS35A function in virulent strains of *N. caninum* and measuring their virulence in *in vitro* or *in vivo.*

4.4.2.2 Putative cathepsin C2

There are five cathepsins that are members of the cysteine proteases family, namely cathepsin L-like, cathepsin B-like and three cathepsin C –like (1,2 and 3) proteases (Dou and Carruthers, 2011). Cathepsins have a role in *T. gondii* microneme and rhoptry protein maturation, replication, host cell invasion and nutrient achievement. In apicomplexan protozoa, cysteine proteases play critical roles in host cell invasion, organellar biogenesis, and intracellular survival (Que et al., 2007). Cathepsin C or dipeptidyl peptidase I (DPPI) has principally exopeptidase activity, which subsequently eliminates dipeptides from unblocked N-terminal substrates of proteins and peptides (Dahl et al., 2001). Cathepsin proteases play the part of lysosomal hydrolases, digesting endogenous and exogeneous polypeptides involved in endocytosis of eukaryotics organisms (Barrett and Kirschke, 1981). The cathepsin C in *Toxoplasma* is important for tachyzoite growth and differentiation (Que et al., 2007).

Moreover, putative cathepsin L was highly abundant in Nc Spain 1H at 56 hours pi**.** The role of cathepsin L-like protease has been described for *Plasmodium***.** (Dahl and Rosenthal (2005)) stated that cathepsin L-like protease is critical for haemoglobin degradation within the parasite food vacuole through erythrocyte infection for

Plasmodium falciparum. The cathepsin C2 showed high abundance in the low virulence strain all three time points**.** Producing large quantities of cathepsin might be an adaptation of the low virulence parasite, allowing it to invade host cells faster, while the high-virulence strain remains outside the host cell for longer**.**

4.4.2.3 Microneme proteins

Microneme proteins are associated with adhesion and invasion of host cells and are released immediately after attachment to the host cells**.** TgMIC4, TgMIC1 and TgMIC6 are formed the complex which has been shown important for host cell penetration in *T. gondii* (Saouros et al., 2005b)**.** Tg MIC1 and TgMIC4 function as adhesins while TgMIC 6 play role as escorter proteins (Reiss et al., 2001b)**.** The deletion of TgMIC1 and TgMIC3 leads to a reduced in virulence of *T. gondii* (Cerede et al., 2005)**.** NcMIC3 protein interact with surface proteoglycan of host cells and initiates attachment and invasion process to host cells (Naguleswaran et al., 2002)**.** MIC3 TgMIC2 and other transmembrane proteins make a link to the parasite actinomysin sytem through their cytoplasmics tail (Jewett and Sibley, 2003) which provide the force for host cell invasion (Soldati and Meissner, 2004)**.** TgMIC10 and TgMIC11 do not relate to the parasite surface during invasion**.** They are soluble microneme proteins**.** TgMIC10 proteins are secreted at the time of penetration and are highly express in *T.gondii* tachyzoites which is responsible for active infection (Hoff et al., 2001)**.** TgMIC11 consists of an alpha-chain and beta-chain tied by disulphide bond due to the internal propeptide is removed by two successive proteolytic events (Harper et al., 2004)**.**

High abundance of microneme proteins at 56 hours pi may be due to the parasite egressing and finding new host cells to invade**.** The putative MIC2-associated protein M2AP was found more abundant at 56 hours pi**.** M2AP plays important role in host cell entry by forming the complex with MIC2**.** The MIC2/M2AP complex is moved over the parasite to the posterior end where the MIC2 ectodomain is secreted through C-terminal proteolysis of MIC2**.** The evident of secreted the content and translocation of MIC2/M2AP complex are associated with the host cell invasion **(**Rabenau et al**.**, 2001**).** Huynh et al. (2003) found that the deletion of M2AP in *T.*

gondii were reduced more than 80**%** of host cell invasion**.** It seemed like deficient of the expression of MIC2 which is important for the parasite to attach the host cells**.** The finding of MICs protein from this study is contrast with the studied by (Regidor-Cerrillo et al., 2012a)**.** They studied proteomics expression change between attenuated **(**Nc Spain 1H**)** and virulent **(**Nc Spain 7 and Nc-Liv**)** of *N. caninum* using DIGE and MALDI-TOF MS technique**.** They found NcMIC1 showing higher abundant in Nc Spain 7 versus Nc Spain 1H at a ratio 1**.**57 of 3**.**5 days tachyzoites culture sample**.** The reason why many invasive microneme proteins were found more abundant in only Nc Spain 1H but not in Nc Spain 7, this may because the microneme proteins are a fundamental protein in all apicomplexan parasites**.** That means a group of microneme are essential for parasite survival and play important role for host cell attachment and invasion and this process does not indicate the virulence as mentioned before Nc low virulent might have higher ability to attach the host cell but the parasite itself may not cause severe harm to the host after invasion**.**

4.4.2.4 Rhoptry protein ROP1

Rhoptry proteins play a vital role in the initial and later stages of host cell invasion**.** Rhoptry kinase are known to be determinants of *T. gondii* virulence (Hunter and Sibley, 2012)**.** ROP proteins are secreted after the invasion process following the microneme protein**.** ROP proteins are a large family proteins with conserved serine/threonine kinases domains, which are associated with the development and maintenance of the parasitophorous vacuole and regulation of host cell processes (Qiu et al., 2009)**.** ROP5, ROP16 and ROP 18 have been determined as a genetically polymorphic virulence factors in *T. gondii* (El Hajj et al., 2007; Saeij et al., 2006; Taylor et al., 2006)**.** ROP5 is essential for pathogenesis in *T. gondii*-infected mice (Behnke et al., 2011)**.** ROP1 is released into the parasitophorous vacuole **(**PV**)** by the time invasion was completed (Carruthers and Sibley, 1997b)**.**

4.4.2.5 Apical membrane antigen1

Apical membrane antigen 1 **(**AMA1**)** localizes to the parasite surface **(**Mital et al., **2005)** and binds to microneme proteins and is highly conserved in apicomplexan

parasites**.** It is one of the major components, along with rhoptry neck proteins RON 2/4/5/8, of the moving junction **(**MJ**)** crucial to the invasion process (Besteiro et al., 2009)**.** Mital et al. (2005) showed that AMA1 depletion in *T. gondii* reduces the ability of parasite to invade host cells**.** The deficiency of TgAMA1 does not influence microneme secretion during the host cell attachment but it does inhibiting secretion from the rhoptries**.**

4.4.2.6 Subtilisin SUB1 (SUB1)

Subtilisin SUB1 is a microneme-secreted protein, released during the infection process. A TgSUB1 protein is determined as a key marker for acute toxoplasmosis (Hruzik et al., 2011). This protein was highly abundant in Nc Spain 1H at 56 hr pi, during which *N. caninum* is likely to exit infected cells and attach the new host cells. The overall proteomic profiles of low-virulence Nc Spain 1H shows a high abundance of proteins involved in host cell invasion, host cell attachment, tyachyzoite growth and differentiation**.** This might mean that Nc Spain 1H has a better ability to attach and invade the host cells than the high-virulence strain Nc Spain 7**.** However, these findings contradict the previous study by Regidor-Cerrillo et al. (2011) who investigated the *in vitro* invasion and proliferation capabilities of the Nc-Liv and ten Spanish *N. caninum* isolates at 2, 4, and 6 hours post infection. Nc Spain 3H and Nc Spain 1H revealed the lowest invation rate by Dunn's test. They found no significant difference in invasion rate between Nc Spain 1H and Nc Spain 7 at any time points.

4.4.3 Preferential proteins expression in high-virulence strains of *N. caninum* (Nc Spain 7)

The proteins preferentially expressed in Nc Spain 7 that were found in all three time courses were a predicted rhoptry kinase, subfamily ROP24 (NCLIV_068850) and SRS39 domain-containing protein (NCLIV_023620).

4.4.3.1 Predicted rhoptry kinase, subfamily ROP24

TgROP 24 is a member of the rhoptry kinase family. Talevich and Kannan (2013) found unusual motifs in non-canonical subfamilies ROP24 and ROP45 from *T. gondii* that could indicate novel kinase mechanisms of activation, ATP positioning and catalysis. A number of rhoptry kinases are implicated in the virulence of *T. gondii*, such as pesudokinase ROP5 (Talevich and Kannan, 2013), rhoptry kinase ROP18 (Reid et al., 2012) and rhoptry kinase ROP16 (Denkers et al., 2012; Saeij et al., 2007b). However, these three rhoptry proteins in *N. caninum* are different from *T. gondii* as they differ in the specific location of loci. The rhoptry kinase ROP18 in *N. caninum* is pseudogenised, which means that *N. caninum* is incapable of phosphorylating host immunity-related GTPases (Reid et al., 2012). This finding is contradicted with the study of (Regidor-Cerrillo et al., 2012b). They found that ROP40 and ROP9 were more abundant in Nc Spain 7 than Nc Spain 1H and Nc Liv. They suggested that these two proteins might play different roles in the two parasite strains. The role of ROP24 is still unknown in *N. caninum*; however, it may be associated with *N. caninum* virulence since it is significantly more abundant in Nc Spain 7 than Nc Spain 1H in all three time points.

4.4.3.2 SRS domain containing protein (SRS39)

As described above, the SRS proteins are thought to be responsible for host cells attachment and host immunity activation to regulate the parasite's virulence (Jung et al., 2004). ROP24 and SRS39 proteins may be involved in *N. caninum* virulence of as both proteins were significantly more abundant in Nc Spain 7 at all time points. These two proteins may work together in order to generate the *Neospora* virulence.

4.4.4 Validation of identified protein

Western blotting is a method to validate protein expression in a proteomics analysis (Nirmalan et al., 2009). It need to do validation for confirming protein expression over a larger sample size. Validation using antibodies and/or quantitative mass spectrometry are really crucial for being able to draw conclusions from identified protein.

4.4.5 Limitation of the study

This study has no negative control to quantify the background level of population variation in protein expression**.** The study should also contain comparisons of strains with the same virulence phenotypes, and comparisons of different phenotypes that are controlled for the population history of the strains**.** There is also no replication; ideally, we should compare multiple, independent instances of the low/high virulence differential**.**

4.4.6 Establish functional role of the identified proteins

The functional role could be established by using the ToxoDB as well as the protein databases. Known genes and proteins do not have any problem to know the functional role directly from the database. However, hypothetical proteins and unspecified products need to do further investigation. The databases of protein families, domains and functional sites such as PFAM (Finn et al., 2014), Interpro (Armean et al., 2018), UniProt (Gabella and Durinx, 2017) can be employed to investigate the unknown protein by comparing or matching the protein sequence in order to know or predict their function and characterise them.

4.4.7 Conclusions and future purspectives

In conclusion, the preferential expression of SRS domain-containing protein (NCLIV_019580) in Nc Spain 1H (low virulent strain) over the whole time course is thought to reflect a mechanistic role of SRS in limiting the virulence of the parasite. Conversely, TgROP24 and SRS39 proteins may be associated with increased virulence in Nc Spain 7.

SRS domain-containing protein (NCLIV_019580) should be studied further to understand how it limits virulence. They might be performed by over expressing this protein in high virulent strains and measuring pathogenicity *in vivo*. Similarly, it would be beneficial to repress ROP24 in a virulent strain to better understand its stimulatory effect on virulence in *N. caninum*.

Chapter 5: General discussion

This thesis comprises a comparative study of gene expression in different *N. caninum* developmental stages and in parasite strains with distinct virulence phenotypes using advanced techniques to investigate global transcript and protein abundance.

The interest to study the virulence of *N. caninum* was raised because *N. caninum* strains shows significant variation in this phenotype within the population, as defined by the pathogenesis associated with their infections (Al-Qassab et al., 2010). Similar to the tachyzoite and bradyzoite stages of *N. caninum*, strains varying in virulence are different in terms of morphology and their abilities to invade host cells. Together, understanding the genetic factors regulating virulence and development in *N. caninum* is necessary as this can promote the identification of vaccine and drug targets for disease prevention and treatment. Previous studies of *N. caninum* virulence mostly focused on the pathogenicity of the disease in clinically affected animals or neonatal infections (Arranz-Solis et al., 2015; Cavalcante et al., 2012; Regidor-Cerrillo et al., 2014; Rojo-Montejo et al., 2009a). The major finding of this thesis is the identification of transcripts and proteins that are preferentially expressed in either a low- and high- virulence condition in *N. caninum*, as well as the protein expression profiles during stage conversion of the parasites.

In chapter two, the protein expression during tachyzoites-bradyzoites interconversion of *N. caninum* was investigated using label-free quantitative proteomics. Larger numbers of proteins were identified using label-free quantitative techniques than in previous studies that used gel-based methods. The result shows most of apical secretory proteins such as miconemes, dense granules and rhoptries proteins as well as the proteins associated to the parasite movement and invasion process have shown decreased expression in bradyzoites stage. The reduction of these groups of proteins suggests that they are not necessary in bradyzoites stage as these proteins role are involved in host cell attachment, invasion and PV formation. However, the general GO enrichment terms were introduced the general board terms which could not provide a biological information of predicted function related to each *N. caninum* stage.

However, the development of bradyzoites from this study is from *in vitro* culture which is different from the bradyzoites from *in vivo*. Therefore bradyzoites from animal model has the influence from host immunity rather than receiving stress from chemical induction *in vitro* and the duration to be converted from tachyzoites to bradyzoites stage *in vivo* has shown longer time than *in vitro* (Skariah et al., 2010). These all affect the protein expression in stage conversion. As a consequence, the different pattern of proteins expressions might not be the same and not totally reflect the biological knowledge in *in vivo*. However, it is difficult to isolate the tissue cyst from animal model especially from natural infection (Hemphill et al., 2004).

A proteomics analysis of gene expression during the reverse stage conversion of bradyzoites to tachyzoites might be studied in future using an alkaline medium to induce reversion (Soete et al., 1993). This might show whether the same patterns of gene expression changes occur with tachyzoites-bradyzoites interconversion or not. Moreover, other developmental changes during the *N. caninum* life cycle might be examined, such as differentiation from oocyst to tachyzoites. Since *N. caninum* oocyst also plays the important role in *N. caninum* transmission. Most of *N. caninum* abortion outbreak are most likely come from the cattle ingest the oocyst contaminated in food or water (McAllister et al., 2000; McAllister et al., 1996). However, it might be difficult to obtain sufficient *N. caninum* oocysts and induce stage conversion *in vitro*. The stage conversion might be studied *in vivo* to achieved the data that reflect the real situation in animal.

Another type of cell culture might improve research on tachyzoites-bradyzoites stage conversion, for example astrocyte or neurons that are cell types related to the nervous system. The CNS is a favoured target of *N. caninum* and the parasites might show a different protein expression pattern against a different host cell background. *N. caninum* can successfully infect astrocyte cell culture (Jesus et al., 2013; Pinheiro et al., 2006). This kind of information might give more understanding of the *N. caninum* stage biology.

In chapter three, the transcriptomics expression change between low and high-virulence *N. caninum* was investigated using RNA-Seq technology applications. The

main finding from these transcriptomic profiles was that several SRS domain-containing transcripts were preferentially expressed in the low virulence strain. Reid et al. (2012) found a group family of surface antigens (SAGs) or SRS (SAG-1 related sequence) proteins which were vastly divergent in *N. caninum* than *T. gondii.* This might explain its narrower host range. These SRS proteins might associate with reduction of the virulence in *N. caninum.* On the other hand, protein phosphatase 2C might be involved in *N. caninum* virulence. This protein is localized to rhoptries and released during the host cell invasion. It transports the parasite protein into the host cell which might increase the possibility to effect the host cells (Gilbert et al., 2007).

The major limitation of the transcriptomics study is the small number of significantly preferentially expressed transcripts, which means that the data cannot provide much insight into the functional consequences of differences in gene expression. This might be because of too few genes to be captured by GO term enrichment, which needs hundreds of genetic differences to be functionally characterised to produce anything compelling. In addition, the most of the transcript found to be differentially expressed encoded hypothetical proteins, which by definition do not have functional ontological terms associated with them, and which cannot help a functional interpretation based on GO term enrichment. However, *N. caninum* strains used in this study came from the same country and population, such that there might not be vast differences in terms of genetic diversity. The genetic diversity seems to be restricted in *N. caninum* and the clonality is a prominent feature in this organism which transmission always occurs within intermediate host (Sibley and Ajioka, 2008). Efficient vertical transmission of this parasite means that the parasite can live without definitive host and sexual reproduction is uncommon leading population diversity presumably low (Perez-Zaballos et al., 2005). There are no clear correlation between the pathogenicity and genetic diversity among the *N. caninum* isolates so far (Al-Qassab et al., 2010; Regidor-Cerrillo et al., 2006). One possibility reason is that most *N. caninum* used in many study were from the cell culture and the long term passage causes lower the pathogenicity and other features of the parasite *in vivo* (Bartley et al., 2006). The strict control and report number of the passage used in each experimental study as well as well maintenance the culture seed stock of *N. caninum* are suggested to be performed (Goodswen et al., 2013).

In future of the transcriptomics research, more *N. caninum* strains displaying both low and high-virulence phenotypes should be identified and examined because only one example of each condition cannot capture the full range of virulence phenotypes, or hope to identify characteristic variations in gene expression that are distinct from strain variations in the population at large. Moreover, the standard reference strains should be included in the experiment as the control to provide a measure of standard strain variation in transcript abundance. The data from this thesis showed that PP2C could be a virulent factor for *N. caninum.* This protein might be studied more *in vivo* such as knockout it from the parasite and infect to animal model to observe the pathogenicity occurring in infected animal. Additionally, a group of SRS proteins which were expected to lowering the virulence of the parasite have to investigate further for their functional role.

In chapter four, differences in expression between low and high-virulence *N. caninum* were recorded during infections using a proteomic analysis at 12, 36 and 56 hours pi. The predicted rhoptry kinase ROP24 is expected to be associated with virulence as this ROP protein was preferentially expressed in high-virulence strains of *N. caninum* at all three time points. This protein is involved in host cell invasion and might be one variable to cause impairment to the host. However this finding contradicts with the study of Regidor-Cerrillo et al. (2012a), who found another ROP protein (TgRop40) was more abundance in Nc Spain 7. This could be that both rhoptry proteins might play role in virulence of *N. caninum.* Additional study related to virulence in these two strains need to be investigated in *N. caninum* such as disruption of these proteins and examine under either *in vitro* or *in vivo*. Few known rhoptry proteins have been implicated in virulence in *T. gondii* such as ROP5 (Reese et al., 2011), ROP16 (Saeij et al., 2007b) and ROP18 (El Hajj et al., 2007). The rhoptries release effector proteins named ROPs which are brought into the host cytoplasm which hijack host cells function. Recently, Kim et al. (2016) had found the rhoptry peudokinase ROP54 in *T. gondii* can modulate its virulence. They demonstrated that the knockout of the ROP54 in type II strain of *T. gondii* did not affect *in vitro* growth but caused 100-fold reduction in virulence *in vivo.* They suggested that ROP54 is involved in invasion of innate immune effectors which

increased of guanylate binding protein 2 (GBP2) loading onto the PV of the ROP54-disrupted *T. gondii* to promote the *Toxoplasma* infection.

Surprisingly, most of host cell attachment-associated proteins are more abundant in low-virulence Nc Spain 1H. For example, MICs were most abundant at 12 and 56 hr pi. as well as putative cathepsin C2 and subtilisin SUB1proteins. It might indicate the better ability for host cell attachment than high virulence Nc Spain7.

The limitation of this analysis is that no control was included to quantify the background level of protein expression variation between strains of the same virulence. Moreover, there is only one strain representing each virulence phenotype and therefore the observations are not replicated.

Future proteomic research on *N. caninum* virulence should include more *N. caninum* strains that have defined virulence phenotypes. The functional consequences of modifying ROP24 expression should be explored, perhaps by observing the effects of enhancing or ablating its function on virulence *in vivo*.

From the study, there are some identified protein involved in the stage conversion which decrease in expression during the differentiation process such as apical secretory proteins as microneme, rhoptries and dense granule. The transcript expression of PP2C might involve in virulence of *N. caninum* and ROP 24 and SRS39 proteins are expected to be associated in virulence of high virulent strains. Further investigation of these proteins and transcripts need to be performed such as gene/protein validation by realtime PCR and western blot which leads to know that they are definitely preferentially expressed. This outcome can be used to develop drug and vaccine against neosporosis. Up to date, no available commercial vaccine to *N. caninum* is on the market. Zhang et al. (2013) performed the cost of management practice to control neosporosis using the mathematical modelling to predict. They found that vaccine could be the efficient intervention strategy (Reichel and Ellis, 2009). Most of the vaccine research using recombinant subunit vaccine candidates as antigents represent surface proteins, heat shock protein assessed to date. The proteins that are secreted from the secretory organelles; rhoptries, micronemes and dense

granule have been applied for producing vaccine. Recently, the chemotherapeutic drug for neosporosis treatment is not regarded as an economically variable factor because of long withdrawal duration period. The drug candidate should be prior tested *in vitro*. There is currently no effective drug on the market. Some genes and proteins from this study might be the candidate for drug and vaccine development in the future.

In conclusion, this thesis has provided a view of *N. caninum* gene expression relating to developmental stage and virulence using transcriptomic and proteomic approaches. The comparative analysis between low- and high-virulence as well as tachyzoites and bradyzoites provide more biological information insight into the key genes and proteins related to these critical features of pathogenesis and parasite transmission. Thus, these findings contribute to an improved understanding of *N. caninum* cell function and move us closer to finding new ways of preventing and controlling neosporosis.

APPENDIX

Appendix Table I: The list of proteins expression in cluster 1 showing downward expression trend throughout three time points

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_025600	putative calmodulin	calmodulin CAM2 (CAM2)	265739.36	263341.31	25032.09	10.61595	0.00139
NCLIV_043870	conserved hypothetical protein	hypothetical protein	632804.95	491318.85	83413.24	7.58639	0.00066
NCLIV_036200	putative cyclophilin	peptidyl-prolyl cis-trans isomerase family 1	440164.70	311944.18	61153.88	7.19766	0.00194
NCLIV_053590	WD-40 repeat protein, related	transducin beta-like protein TBL1 (TBL1)	1078432.05	1120640.79	167584.27	6.68703	0.00339
NCLIV_043320	rna recognition motif (RRM)-containing protein,related	U2 snRNP auxilliary factor, large subunit, splicing factor subfamily protein	999741.20	910921.59	149615.12	6.68209	0.00293
NCLIV_017040	Phosphofructokinase, related	phosphofructokinase PFKII (PFKII)	846932.48	613245.67	156661.14	5.40614	0.00459
NCLIV_047600	hypothetical protein	hypothetical protein	245166.59	265982.91	50582.39	5.25841	0.00485
NCLIV_064340	putative prefoldin subunit	prefoldin subunit 6, putative	252216.14	225237.83	49038.51	5.14323	0.00198
NCLIV_042770	hypothetical protein	hydrolase, NUDIX family protein	191186.70	246180.35	48626.41	5.06269	0.02926
NCLIV_048800	hypothetical protein	ubiquitin family protein	1616946.72	1243199.11	319526.01	5.06045	0.00152
NCLIV_019600	hypothetical protein	clathrin adaptor complex small chain subfamily protein	584126.80	516189.27	119401.46	4.89212	0.00175
NCLIV_053360	conserved hypothetical protein	transcription elongation factor SPT6 (SPT6)	133160.44	107140.26	30173.92	4.41310	0.00156
NCLIV_001770	putative DNAK family domain containing protein	tetratricopeptide repeat-containing protein	1353555.83	938777.78	321447.71	4.21081	0.00267
NCLIV_056590	putative phosphatidylinositol-4-phosphate 5-kinase	MORN repeat-containing protein	1058324.45	896339.46	251841.30	4.20235	0.00751
NCLIV_024440	hypothetical protein	vacuolar protein sorting-associated protein 26, putative	265255.35	220355.64	66567.25	3.98477	0.00371
NCLIV_056360	Eukaryotic initiation factor, related	DEAD (Asp-Glu-Ala-Asp) box polypeptide DDX6 (DDX6)	512884.26	586998.49	151636.41	3.87109	0.02395
NCLIV_054550	Cytochrome c oxidase subunit 2, related	cytochrome C oxidase subunit IIb, putative	3972133.77	2765119.30	1059769.0	3.74811	0.00058
NCLIV_019450	hypothetical protein	signal peptidase	551941.50	669909.50	192573.35	3.47872	0.00073
NCLIV_025710	unspecified product	microneme protein MIC7 (MIC7)	4010693.96	3108615.04	1165075.8	3.44243	0.00528
NCLIV_005190	putative transcription elongation factor FACT 140 kDa	transcriptional elongation factor FACT140 (FACT140)	1102959.20	893566.69	329756.63	3.34477	0.00230
NCLIV_045870	unspecified product	dense granule protein GRA3 (GRA3)	29563772.10	20964950.60	8937345.3	3.30789	0.00146

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_062570	Contig An13c0020, complete genome, related	PCI domain-containing protein	3194546.53	4019722.08	1238916.0	3.24455	0.00382
NCLIV_054110	YHL017Wp-like protein, related	membrane protein, putative	1935637.03	1499365.93	599800.64	3.22713	0.00267
NCLIV_036970	putative calmodulin	calmodulin, putative	1495446.50	1301915.15	468635.32	3.19107	0.01038
NCLIV_036720	CBR-CSN-5 protein, related	proteasome regulatory subunit	887029.48	695815.78	287500.07	3.08532	0.00608
NCLIV_002610	putative isoleucine-tRNAsynthetase	isoleucyl-tRNA synthetase family protein	2307714.55	1989230.48	783400.08	2.94577	0.00293
NCLIV_028240	putative Ras family domain-containing protein	Rab1 protein	349939.55	375387.59	129222.97	2.90496	0.00156
NCLIV_001600	possible RNA-binding protein, related	RNA recognition motif-containing protein	959189.50	840035.66	332709.65	2.88296	0.00369
NCLIV_032620	Cs1 protein, related	Hsp70 interacting protein HIP (HIP)	3553222.50	3090527.14	1286645.9	2.76162	0.00232
NCLIV_017510	putative 40S ribosomal protein S29	N/A	1600757.06	1792729.13	654380.64	2.73958	0.00442
NCLIV_030120	putative importin beta-3 subunit	HEAT repeat-containing protein	2379127.75	1759619.30	887190.09	2.68164	0.00455
NCLIV_059410	hypothetical protein	cytokine induced apoptosis inhibitor 1 family protein	1556791.69	1422897.08	581884.56	2.67543	0.00480
NCLIV_060560	conserved hypothetical protein	hypothetical protein	1466291.69	1202805.36	552761.25	2.65267	0.01931
NCLIV_049740	conserved hypothetical protein	hypothetical protein	656117.74	582779.72	252292.50	2.60062	0.03507
NCLIV_019480	proteasome A-type and B-type domain-containing protein	20S proteasome subunit beta 7, putative	3769139.21	3385413.13	1457230.5	2.58651	0.00264
NCLIV_016370	Galactosyltransferase, related	hypothetical protein	221053.48	215214.31	86240.68	2.56322	0.01653
NCLIV_027100	putative orotidine-monophosphate-decarboxylase	orotate phosphoribosyltransferase	108878.97	96123.69	43474.92	2.50441	0.01336
NCLIV_027660	conserved hypothetical protein	hypothetical protein	940014.22	820373.02	375591.06	2.50276	0.01836
NCLIV_002680	agap001651-PA, related	PCI domain-containing protein	1901213.38	2206691.82	906066.58	2.43546	0.00854
NCLIV_062220	Glutaminyl-tRNA synthetase,related	glutaminyl-tRNA synthetase (GlnRS)	11921452.32	10815435.02	4897269.5	2.43431	0.00085
NCLIV_027930	unspecified product	rhoptry protein ROP17 (ROP17)	1141984.84	869393.11	475372.45	2.40229	0.01784
NCLIV_028320	putative actin-like family protein ARP4a	actin-related protein ARP4A (ARP4A)	824039.22	627785.86	344666.43	2.39083	0.00755
NCLIV_039960	hypothetical protein	heat shock protein 40, putative	3269089.23	2546498.32	1371799.4	2.38307	0.00758
NCLIV_019200	Eukaryotic translation initiation factor 5,related	eukaryotic initiation factor-5, putative	2509886.44	1971241.02	1060635.7	2.36640	0.00332
NCLIV_024220	putative glycerol-3-phosphate	FAD-dependent glycerol-3-phosphate	1067905.61	869111.34	458320.76	2.33004	0.02024

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
	dehydrogenase	dehydrogenase					
NCLIV_064880	conserved hypothetical protein	hypothetical protein	5308282.98	4378444.63	2290038.3	2.31799	0.02948
NCLIV_027290	Ribosomal protein S21-maize (ISS), related	ribosomal protein RPS21 (RPS21)	4131812.01	3303550.25	1785066.1	2.31465	0.04119
NCLIV_005280	conserved hypothetical protein	hypothetical protein	719881.37	693685.18	311969.84	2.30754	0.02944
NCLIV_033830	hypothetical protein	hypothetical protein	945120.67	807206.67	414628.18	2.27944	0.00367
NCLIV_017020	hypothetical protein	beta adaptin protein, putative	321899.29	287376.23	144604.48	2.22607	0.01204
NCLIV_041210	putative Ubiquinol-cytochrome c reductase complex 14 kDa protein	ubiquinol-cytochrome c reductase	5696938.76	5387596.52	2562340.4	2.22333	0.00405
NCLIV_007110	hypothetical protein	ribosomal protein RPS10 (RPS10)	1611162.07	1383210.91	739344.96	2.17918	0.01788
NCLIV_012830	putative MORN repeat-containing protein	phosphatidylinositol-4-phosphate 5-kinase	1841649.52	1794299.75	864260.51	2.13090	0.00655
NCLIV_050300	GH18750, related	eukaryotic initiation factor-2 gamma, putative	8906939.60	8561857.33	4260432.2	2.09062	0.00589
NCLIV_035220	hypothetical protein	hypothetical protein	6119278.87	5235311.74	2949403.1	2.07475	0.00094
NCLIV_065970	conserved hypothetical protein	GAP40 protein (GAP40)	2492858.69	2120046.31	1208655.4	2.06251	0.03775
NCLIV_044440	hypothetical protein	eukaryotic peptide chain release factor, putative	672906.17	690492.46	337929.55	2.04330	0.04579
NCLIV_036400	unspecified product	dense granule protein GRA1 (GRA1)	801663933.5	693263794.7	395255100	2.02822	0.00304

Appendix Table II: The list of proteins expression in cluster 2 showing downward expression trend and then not change in expression

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_021510	conserved hypothetical protein	DUF3228 domain-containing protein	396479.64	66940.64	20525.99	19.31598	0.01330
NCLIV_021410	conserved hypothetical protein	hypothetical protein	1934838.9	217867.71	638107.78	8.88080	0.00297
NCLIV_056730	conserved hypothetical protein	hypothetical protein	588892.91	92014.19	80675.27	7.29955	0.01047
NCLIV_020430	hypothetical protein	cleft lip and palate transmembrane protein 1 (clptm1) protein	714046.52	151961.38	122149.83	5.84566	0.04728
NCLIV_018260	putative U4/U6 small nuclear ribonucleoprotein	WD domain, G-beta repeat-containing protein	155932.92	44142.10	27703.00	5.62874	0.03875
NCLIV_040600	hypothetical protein	signal recognition particle receptor beta subunit protein	5691004.5	1028281.60	1039360.9	5.53448	0.03800
NCLIV_008860	hypothetical protein	UDP-galactose transporter family protein	671184.72	203690.15	142488.38	4.71045	0.00485
NCLIV_065570	putative translation initiation factor SUI1	translation initiation factor SUI1, putative	1995635.1	558339.68	505226.56	3.94998	0.00066
NCLIV_046970	conserved hypothetical protein	hypothetical protein	1103666.8	301124.81	411043.97	3.66515	0.00993
NCLIV_048040	conserved hypothetical protein	hypothetical protein	1109215.0	316639.53	401233.78	3.50308	0.01931
NCLIV_051590	putative GTP binding protein	GTP-binding protein, putative	8497994.9	3170482.10	2547632.1	3.33564	0.00293
NCLIV_036630	putative 14-3-3 protein	14-3-3 superfamily protein	153305.83	46317.57	114537.63	3.30989	0.02400
NCLIV_004510	putative kelch motif domain-containing protein	kelch repeat-containing protein	136045.70	41180.82	88555.08	3.30362	0.04414
NCLIV_005050	putative coatomer delta subunit	adaptor complexes medium subunit family protein	3716585.0	1425511.92	1139790.5	3.26076	0.00156
NCLIV_005040	conserved hypothetical protein	hypothetical protein	2968418.0	943944.08	1068584.0	3.14470	0.00273
NCLIV_032240	Catalase (EC 1.11.1.6), related	catalase	66004502.	21314140.59	23758910	3.09675	0.00282
NCLIV_058300	hypothetical protein	hypothetical protein	2576396.5	840942.93	1649767.2	3.06370	0.03922
NCLIV_035190	conserved hypothetical protein	glideosome-associated protein with multiple-membrane spans GAPM3	3676957.3	1267464.45	2785751.4	2.90103	0.00854
NCLIV_045490	conserved hypothetical protein	hypothetical protein	1081874.9	394781.00	617225.23	2.74044	0.00548
NCLIV_058000	putative alanine dehydrogenase	NAD(P) transhydrogenase subunit beta, putative	12759821	4755062.00	5570964.7	2.68342	0.02397
NCLIV_028500	GJ17676, related	homocysteine s-methyltransferase domain-containing protein	4317622.0	1628409.34	2028678.9	2.65144	0.00822
NCLIV_037760	Rhomboid-6, isoform A, related	rhomboid protease ROM4 (ROM4)	1003097.6	384428.44	444785.57	2.60932	0.04511
NCLIV_063810	conserved hypothetical protein	hypothetical protein	3130555.7	1223202.26	1408204.15	2.55931	0.03104
NCLIV_015290	hypothetical protein	endomembrane protein 70 subfamily protein	1350738.4	574436.39	528568.87	2.55546	0.01453
NCLIV_052900	conserved hypothetical protein	hypothetical protein	5306909.2	2080671.92	3589507.33	2.55057	0.00262

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_042760	dehydrogenases with different specificities,related	carbonyl reductase 1, putative	1560959.3	640367.74	1290273.37	2.43760	0.01064
NCLIV_006000	hypothetical protein	inositol(myo)-1(or 4)-monophosphatase 2, putative	965612.84	401140.43	427750.64	2.40717	0.02944
NCLIV_028050	conserved hypothetical protein	hypothetical protein	3300756.2	1515466.09	1384822.38	2.38352	0.01601
NCLIV_069460	hypothetical protein	N/A	487802.59	206080.78	378059.47	2.36705	0.00577
NCLIV_022430	hypothetical protein	GAPM1a	5163613.3	2196494.43	3621235.01	2.35084	0.04303
NCLIV_010370	ubiquitin carrier protein, related	ubiquitin-conjugating enzyme subfamily protein	10574042	4556562.03	4528909.29	2.33479	0.01390
NCLIV_043180	hypothetical protein	Sec23/Sec24 trunk domain-containing protein	3937203.8	1719769.23	2680803.81	2.28938	0.00462
NCLIV_018020	hypothetical protein	myosin A	434415.96	211605.64	191959.78	2.26306	0.01179
NCLIV_044350	conserved hypothetical protein	glideosome-associated protein with multiple-membrane spans GAPM2B (GAPM2B)	4400460.2	1959765.79	2343838.50	2.24540	0.02756
NCLIV_012610	conserved hypothetical protein	hypothetical protein	412168.46	191188.75	201788.93	2.15582	0.04134
NCLIV_018420	unspecified product	rhoptry protein ROP9 (ROP9)	299806549	154910479.8	139926402.5	2.14260	0.00114
NCLIV_013180	GM04207p, related	hypothetical protein	7570779.8	3548726.29	4326689.82	2.13338	0.00842
NCLIV_040650	conserved hypothetical protein	hypothetical protein	1660775.0	794100.13	803617.55	2.09139	0.01485
NCLIV_024740	Myosin, heavy polypeptide 1, skeletal muscle,adult, related	myosin A	15700034.	7699189.39	7565961.09	2.07509	0.00560
NCLIV_025730	conserved hypothetical protein	rhoptry neck protein RON10 (RON10)	1615724.7	779488.43	899737.63	2.07280	0.02139
NCLIV_026340	hypothetical protein	IMC sub-compartment protein ISP2 (ISP2)	34275112.	16686366.16	26602726.71	2.05408	0.00594
NCLIV_006780	conserved hypothetical protein	dense granule protein DG32	26762962.	13139404.22	15237274.41	2.03685	0.02584
NCLIV_060810	conserved hypothetical protein	hypothetical protein	6636406.5	3554296.22	3296969.09	2.01288	0.01640
NCLIV_027850	unspecified product	rhoptry protein ROP6 (ROP6)	21488945	10728440.37	19145440.15	2.00299	0.04332

Appendix Table III: The list of expressed proteins of cluster 3 showing upward expression trend throughout three time points

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_012950	putative ubiquitin family domain-containing protein	ubiquitin family protein	32183.17	352394.63	1089634.23	33.85726	0.00250
NCLIV_001290	conserved hypothetical protein	HEAT repeat-containing protein	385271.78	820139.66	4696425.55	12.18990	0.00139
NCLIV_035590	conserved hypothetical protein	hypothetical protein	1513363.22	9352111.17	12779680.14	8.44456	0.01333
NCLIV_033910	hypothetical protein	WD domain, G-beta repeat-containing protein	665310.80	1515968.93	5159820.96	7.75550	0.00714
NCLIV_012600	hypothetical protein	UBX domain-containing protein	1216519.76	3110979.71	9384912.85	7.71456	0.01076
NCLIV_031460	conserved hypothetical protein	hypothetical protein	4253666.60	7494357.69	28251210.99	6.64161	0.00156
NCLIV_066970	putative enoyl-acyl carrier reductase	enoyl-acyl carrier reductase ENR (ENR)	7574215.54	22602986.54	41683983.44	5.50341	0.04438
NCLIV_015620	60S ribosomal protein L36, related	ribosomal protein RPL36 (RPL36)	3902566.58	7980362.45	18483879.84	4.73634	0.00329
NCLIV_044410	unspecified product	rhoptry kinase family protein ROP35 (ROP35)	20917373.66	28776282.42	95433729.32	4.56241	0.00762
NCLIV_038590	conserved hypothetical protein	hypothetical protein	150718.32	429752.79	655892.18	4.35177	0.00589
NCLIV_065440	hypothetical protein	calmodulin, putative	8454788.39	15033666.83	34716268.40	4.10611	0.00415
NCLIV_065210	KLLA0F09449p, related	heat shock protein HSP60 (HSP60)	36358177.28	54507306.08	140030918.30	3.85143	0.00569
NCLIV_014050	conserved hypothetical protein	microtubule associated protein SPM2 (SPM2)	522829.03	827953.23	1913522.53	3.65994	0.01407
NCLIV_030650	putative 26S protease regulatory subunit 6b	26S proteasome regulatory subunit 6b, putative	1195117.95	1771237.37	4252425.39	3.55816	0.00085
NCLIV_024860	Proteasome/cyclosome repeat family protein,related	Proteasome/cyclosome repeat-containing protein	2564313.44	3917997.98	9027531.44	3.52045	0.00419
NCLIV_006060	conserved hypothetical protein	hypothetical protein	728248.46	1356145.96	2366549.83	3.24965	0.00655
NCLIV_034090	putative kinesin heavy chain	kinesin heavy chain, putative	2048222.47	4914383.62	6617486.51	3.23084	0.01873
NCLIV_022370	Ribophorin I, related	hypothetical protein	125163.22	218550.97	383389.85	3.06312	0.02062
NCLIV_002190	conserved hypothetical protein	hypothetical protein	1368076.06	2540517.20	4097414.88	2.99502	0.01794
NCLIV_037200	GJ21226, related	MIF4G domain-containing protein	836413.49	1010395.90	2482886.14	2.96849	0.00200
NCLIV_054800	conserved hypothetical protein	emp24/gp25L/p24 family protein	8619496.12	13501701.49	25488973.60	2.95713	0.00678
NCLIV_051960	conserved hypothetical protein	hypothetical protein	3098086.23	5689450.17	8543120.04	2.75755	0.01073
NCLIV_013380	hypothetical protein	small nuclear ribonucleoprotein	332626.02	401441.27	907857.52	2.72936	0.00534

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
		f (snrnp-f), putative					
NCLIV_015920	Histone H4, related	histone H4	206581257.00	373831434.00	555071245.90	2.68694	0.01653
NCLIV_056910	conserved hypothetical protein	hypothetical protein	2935570.96	3948971.66	7735062.70	2.63494	0.02153
NCLIV_067140	Myosin, related	myosin I	1288308.68	2056757.40	3162016.98	2.45439	0.00197
NCLIV_066760	Translationally-controlled tumor protein,related	histamine-releasing factor, putative	1639514.27	2315527.11	3901201.33	2.37949	0.01216
NCLIV_016800	putative TCP-1/cpn60 chaperonin family protein	chaperonin cpn60, putative	19183439.88	22058701.47	45562092.73	2.37507	0.00415
NCLIV_030890	putative high molecular mass nuclear antigen	hypothetical protein	3697571.32	4226171.79	8680097.38	2.34751	0.00326
NCLIV_031440	60S ribosomal protein L38, related	ribosomal protein RPL38 (RPL38)	1905393.17	2677348.09	4466483.19	2.34413	0.00560
NCLIV_004810	conserved hypothetical protein	hypothetical protein	1178240.02	1686524.52	2741350.26	2.32665	0.01176
NCLIV_038540	rab22a, member RAS oncogene family, related	Rab 5	4231851.02	6333408.81	9841077.71	2.32548	0.00536
NCLIV_053190	hypothetical protein	phenylalanine--tRNA ligase, beta subunit protein	3436435.52	4447971.60	7816649.89	2.27464	0.01038
NCLIV_009030	metallophosphoesterase, related	Ser/Thr phosphatase family protein	652694.57	1444930.12	1464228.51	2.24336	0.01023
NCLIV_022270	unspecified product	polo kinase	1438094.89	3194609.96	3054610.10	2.22142	0.01900
NCLIV_020840	hypothetical protein	ATPase synthase subunit alpha, putative	44966319.20	54487643.99	98864290.19	2.19863	0.01163
NCLIV_040880	hsp90, related	heat shock protein 90, putative	125472487.50	151042013.10	271525623.00	2.16403	0.00145
NCLIV_007330	similar to uniprot|P15705 Saccharomyces cerevisiae YOR027w STI1, related	tetratricopeptide repeat domain containing protein	4896260.47	5635558.70	10442762.65	2.13280	0.00536
NCLIV_025670	ATP synthase subunit beta, related	ATP synthase beta subunit ATP-B (ATPB)	128785733.80	151423081.60	272190341.80	2.11351	0.00854
NCLIV_042810	hypothetical protein	EMP/nonaspanin domain family protein	2751840.09	3009979.57	5739587.73	2.08573	0.00720
NCLIV_024610	conserved hypothetical protein	hypothetical protein	3832033.97	4456824.89	7949808.85	2.07457	0.00066
NCLIV_050070	Fructose-1,6-bisphosphatase,related	sedoheptulose-1,7-bisphosphatase	2068309.82	2600648.55	4224774.74	2.04262	0.00420

Appendix Table IV: The list of proteins expression in cluster 4 showing not change or slightly down expression trend at the beginning and then upward expression trtend

ID	*N. caninum* Description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
NCLIV_020970	conserved hypothetical protein	GYF domain-containing protein	451657.79	625660.82	9665977.74	21.40111	0.00032
NCLIV_019580	SRS domain-containing protein	SAG-related sequence SRS35A (SRS35A)	303766.07	373644.29	3658692.90	12.04444	0.00114
NCLIV_004250	putative nuclear RNA binding protein	hypothetical protein	2796716.17	4026163.48	23359224.10	8.35238	0.00066
NCLIV_060500	conserved hypothetical protein	hypothetical protein	37674990.75	53448716.98	297628121.50	7.89989	0.00417
NCLIV_016250	putative cytidine deaminase	cytidine and deoxycytidylate deaminase zinc-binding region domain-containing protein	1710510.18	2416250.81	12384871.19	7.24045	0.00066
NCLIV_007800	unspecified product	hypothetical protein	57927434.60	69343600.66	343683829.60	5.93301	0.00133
NCLIV_031700	DEHA2G03652p, related	LsmAD domain-containing protein	528306.96	101558.13	579518.71	5.70628	0.04211
NCLIV_048330	hypothetical protein	serine/threonine-protein phosphatase 2A catalytic subunit beta isoform, putative	354085.24	232071.42	1157635.38	4.98827	0.00392
NCLIV_007510	conserved hypothetical protein	hypothetical protein	2835411.64	2798676.06	13455993.20	4.80799	0.00473
NCLIV_039280	isocitrate dehydrogenase-like protein, related	isocitrate dehydrogenase	3402418.68	4394418.36	16146865.25	4.74570	0.00196
NCLIV_017840	conserved hypothetical protein	membrane protein	3742394.05	3674093.12	15954504.80	4.34243	0.00170
NCLIV_016330	putative phosphomannomutase 2	phosphomannomutase	78653.89	66253.03	273498.16	4.12809	0.02291
NCLIV_031970	hypothetical protein	pre-mRNA processing splicing factor PRP8 (PRP8)	4616679.00	5443197.93	18867085.16	4.08672	0.00499
NCLIV_041270	hypothetical protein	RuvB family 2 protein	957256.25	658961.86	2433646.21	3.69315	0.00118
NCLIV_055690	hypothetical protein	small GTPase Rab2, putative	2685377.41	2725196.60	9531330.85	3.54934	0.00298
NCLIV_063230	hypothetical protein	CRAL/TRIO domain-containing protein	989286.38	817158.05	2592223.66	3.17224	0.02502
NCLIV_051820	hypothetical protein	small GTP binding protein rab1a, putative	7441716.63	7966699.96	23471753.59	3.15408	0.00367
NCLIV_054750	hypothetical protein	serine/threonine phosophatase PP1	6688114.92	5946345.40	18412003.87	3.09636	0.00332
NCLIV_009170	proteasome (Prosome,	proteasome subunit beta type 2,	1264325.79	808412.90	2493083.11	3.08392	0.01410

ID	*N. caninum* Description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
	macropain) subunit, beta type, 1, related	putative					
NCLIV_052510	hypothetical protein	hypothetical protein	5436994.54	5025834.72	14992101.37	2.98301	0.00937
NCLIV_010850	mgc81714 protein, related	ATP-binding cassette sub-family F member 1	564284.92	499364.18	1487229.03	2.97825	0.00793
NCLIV_026430	DnaJ domain containing protein, related	DnaJ family Sec63 protein	3877682.42	4019747.38	10964191.90	2.82751	0.00066
NCLIV_057550	unspecified product	subtilisin SUB2 (SUB2)	1877603.77	2090967.91	5267163.96	2.80526	0.00697
NCLIV_062940	putative pyruvate dehydrogenase	pyruvate dehydrogenase complex subunit PDH-E1Alpha (PDHE1A)	3896222.95	4070465.63	10646651.38	2.73256	0.00066
NCLIV_056950	hypothetical protein	adaptin n terminal region domain-containing protein	364118.28	235458.95	630987.42	2.67982	0.02427
NCLIV_005150	hypothetical protein	beta-1 tubulin, putative	46973635.73	38482211.49	102357621.80	2.65987	0.00085
NCLIV_042680	conserved hypothetical protein	hypothetical protein	2804776.97	2017620.07	5262147.40	2.60810	0.00294
NCLIV_003050	putative myosin heavy chain	myosin heavy chain, putative	24804590.50	19699025.92	50219407.41	2.54933	0.00066
NCLIV_007000	adp-ribosylation factor 4, related	ADP ribosylation factor ARF1 (ARF1)	13384885.98	12113051.14	30380291.34	2.50806	0.00667
NCLIV_062350	Ribosomal protein S27, related	ribosomal protein RPS27 (RPS27)	8523545.94	8054591.67	20069702.66	2.49171	0.00175
NCLIV_026540	conserved hypothetical protein	COPI associated protein, putative	1830119.98	1236842.70	2974515.40	2.40493	0.00198
NCLIV_017500	hypothetical protein	ribosomal protein RPS5 (RPS5)	21606091.19	22839638.84	51847984.72	2.39969	0.00496
NCLIV_064810	hypothetical protein	V-type H(+)-translocating pyrophosphatase VP1	1047602.74	745766.34	1768839.74	2.37184	0.00751
NCLIV_022560	Thermosome subunit, related	T-complex protein 1, epsilon subunit (TCP-1-epsilon), putative	6494927.85	5457768.29	12872209.09	2.35851	0.00156
NCLIV_003440	actin, related	actin ACT1 (ACT1)	773369556.50	636090488.10	1486659804.00	2.33718	0.00175
NCLIV_034530	putative TCP-1/cpn60 family chaperonin	T-complex protein 1 delta subunit	6550443.57	6401245.06	14641121.07	2.28723	0.00226
NCLIV_061460	Family T1, proteasome beta subunit, threonine peptidase, related	proteasome subunit beta type, putative	2463404.56	2073481.02	4683982.90	2.25899	0.00145
NCLIV_065640	putative Rhoptry kinase family protein, truncated (incomplete catalytic triad)	Rhoptry kinase family protein, truncated (incomplete catalytic triad)	3193954.51	1641152.97	3673151.23	2.23815	0.00788
NCLIV_018620	Os08g0464000 protein, related	activator of hsp90 atpase 1 family	1452022.50	756348.86	1688512.59	2.23245	0.00842

ID	*N. caninum* Description	*T. gondii* orthologues	D0	D3	D6	Fold change	q Value
		protein					
NCLIV_027480	hypothetical protein	vacuolar ATP synthase subunit d, putative	9420742.98	8570579.18	19038517.24	2.22138	0.03355
NCLIV_056300	conserved hypothetical protein	hypothetical protein	18091603.15	11119003.99	24595182.93	2.21200	0.00560
NCLIV_057960	unspecified product	rhoptry protein, putative	4988515.30	3781467.18	8363598.99	2.21173	0.00536
NCLIV_005770	hypothetical protein	importin-beta N-terminal domain-containing protein	4185123.00	2841306.19	6130854.85	2.15776	0.00328
NCLIV_039940	hypothetical protein	cell-cycle-associated protein kinase GSK, putative	10954372.92	5460253.81	11639629.64	2.13170	0.01703
NCLIV_058890	tubulin alpha chain	alpha-tubulin I, putative	266952606.30	240376973.30	510267636.20	2.12278	0.00129
NCLIV_035920	Adenylate/guanylate cyclase with GAF sensor and FHA domain, related	adenylate cyclase, putative	849211.24	569580.71	1162695.62	2.04132	0.01611
NCLIV_063590	Glutathione reductase, related	glutathione reductase	11763886.13	7666990.21	15431498.68	2.01272	0.00786

Appendix Table V: The list of proteins expression in cluster 5 showing upward expression and then downward expression trend

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_042540	putative PBS lyase HEAT-like repeat domain-containing protein	HEAT repeat-containing protein	120685.78	233588.55	25276.21	9.24144	0.00337
NCLIV_001550	putative ribonucleoside-diphosphate reductase, large subunit	ribonucleoside-diphosphate reductase large chain	87340.33	583783.56	418757.21	6.68401	0.00638
NCLIV_051450	putative centromere/microtubule binding protein	rRNA pseudouridine synthase	223142.17	1220448.03	675565.49	5.46937	0.00778
NCLIV_033680	Solute carrier family 25 (Mitochondrial carrier,dicarboxylate transporter), member 10, related	2-oxoglutarate/malate translocase OMT (OMT)	329636.97	1525846.96	793626.42	4.62887	0.00066
NCLIV_013700	putative CHCH domain-containing protein	CHCH domain-containing protein	856631.00	3695217.99	2024348.95	4.31366	0.01140
NCLIV_032940	putative adenosine transporter	nucleoside transporter protein	891615.53	2919848.11	767821.02	3.80277	0.00226
NCLIV_037590	conserved hypothetical protein	hypothetical protein	1270323.18	2177673.77	611453.73	3.56147	0.04333
NCLIV_032950	putative deoxyuridine 5'-triphosphate nucleotidohydrolase	deoxyuridine 5'-triphosphate nucleotidohydrolase, putative	564550.47	1609532.21	768416.03	2.85100	0.00751
NCLIV_010050	srs domain-containing protein	SAG-related sequence SRS16E (SRS6)	1285383.80	2211278.86	835739.05	2.64590	0.00175
NCLIV_032910	hypothetical protein	SPFH domain / Band 7 family protein	1312561.54	3427174.12	2072315.96	2.61106	0.00802
NCLIV_061760	putative microneme protein MIC6	microneme protein MIC6 (MIC6)	99847150.72	145373153.80	57170746.58	2.54279	0.00214
NCLIV_051890	unspecified product	AP2 domain transcription factor AP2X-7 (AP2X7)	512977.59	796420.85	316090.41	2.51960	0.01159
NCLIV_022970	unspecified product	microneme protein MIC2 (MIC2)	19596185.13	34068901.48	13797268.62	2.46925	0.00674
NCLIV_001620	conserved hypothetical protein	hypothetical protein	197451.04	472436.90	366920.74	2.39268	0.00720
NCLIV_010600	putative microneme protein MIC3	microneme protein MIC3 (MIC3)	460862170.3	597623331.90	263553637.7	2.26756	0.00066
NCLIV_012130	eukaryotic translation initiation factor 3 subunit 7-like protein, related	eukaryotic translation initiation factor 3 subunit 7, putative	3058435.49	4282803.69	1901774.7	2.25200	0.00720
NCLIV_060470	hypothetical protein	cysteine-tRNA synthetase (CysRS)	1168807.61	2187167.46	993506.64	2.20146	0.04842
NCLIV_046540	putative oligoendopeptidase F	peptidase family M3 protein	559748.68	1227000.02	598720.08	2.19206	0.00556
NCLIV_013360	putative plectin	surface antigen repeat-containing protein	161159.37	212122.31	97005.97	2.18669	0.00740
NCLIV_016150	Serine--pyruvate transaminase,related	alanine-glyoxylate aminotransferase	604348.13	1232329.87	570792.00	2.15898	0.01293
NCLIV_046830	putative ATP synthase	ATP synthase, putative	4028206.45	5389668.82	2508110.52	2.14890	0.00273
NCLIV_066250	unspecified product	microneme protein MIC10 (MIC10)	15263145.84	31405245.90	21647544.12	2.05759	0.00668
NCLIV_020720	putative microneme protein MIC11	microneme protein MIC11 (MIC11)	37080653.29	69913253.09	34154350.16	2.04698	0.00197

Appendix Table VI: The list of proteins expression in cluster 6 showing similar expression trend throughout three time points

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_004050	formate/nitrite transporter family protein,related	formate/nitrite transporter protein	526974.15	182506.23	8223.52	64.08132	0.00499
NCLIV_0275	peptidyl-prolyl cis-trans isomerase NIMA-interacting 1	peptidylprolyl isomerase	1247517.81	221748.16	49890.12	25.00531	0.00332
NCLIV_045400	hypothetical protein	splicing factor 3A subunit 2, putative	369688.77	177245.27	33584.01	11.00788	0.01333
NCLIV_056130	conserved hypothetical protein	hypothetical protein	1531949.74	740896.97	178443.55	8.58507	0.00078
NCLIV_057250	conserved hypothetical protein	hypothetical protein	3146278.19	1580093.20	376648.63	8.35335	0.00236
NCLIV_046950	UDP-glucose 4-epimerase, related	UDP-glucose 4-epimerase	206692.55	85457.45	26501.99	7.79913	0.00739
NCLIV_039080	adaptin N terminal region family protein,related	beta-COP	2967373.03	1361368.99	393575.40	7.53953	0.00259
NCLIV_050780	conserved hypothetical protein	hypothetical protein	1641546.33	736822.23	226831.19	7.23686	0.00282
NCLIV_048930	hypothetical protein	importin-beta N-terminal domain-containing protein	1152760.57	366793.78	161921.32	7.11926	0.00409
NCLIV_060900	conserved hypothetical protein	hypothetical protein	343669.12	126206.43	49841.19	6.89528	0.00655
NCLIV_007120	hypothetical protein	microneme protein, putative	1298651.02	635904.54	206501.33	6.28883	0.00467
NCLIV_055980	conserved hypothetical protein	hypothetical protein	5287578.25	2117710.36	880652.61	6.00416	0.00066
NCLIV_063240	DnaJ homologue, related	DnaJ domain-containing protein	379609.76	206332.62	66489.75	5.70930	0.00876
NCLIV_044100	conserved hypothetical protein	hypothetical protein	1075443.92	373675.67	193536.67	5.55680	0.01201
NCLIV_057280	Peptidylprolyl isomerase D (Cyclophilin D),related	tetratricopeptide repeat-containing protein	389988.54	192061.67	70593.74	5.52441	0.01568
NCLIV_011690	unspecified product	rhoptry protein ROP15 (ROP15)	3193011.07	1200586.02	608179.47	5.25011	0.04398
NCLIV_058260	hypothetical protein	N/A	6073953.30	3533331.29	1282677.41	4.73537	0.00145
NCLIV_010110	hypothetical protein	ubiquinol-cytochrome c reductase hinge protein, putative	1529043.10	938260.82	333063.12	4.59085	0.00372
NCLIV_006620	trehalose-6-phosphate synthase of likely plant origin, related	trehalose-phosphatase	157243.48	69697.86	34737.52	4.52662	0.04629
NCLIV_019240	conserved hypothetical protein	hypothetical protein	2088544.07	1098747.25	465879.65	4.48301	0.00241
NCLIV_053680	hypothetical protein	oxidoreductase, short chain dehydrogenase/reductase family protein	171812.59	113404.50	40062.08	4.28866	0.00930
NCLIV_007580	hypothetical protein	glycosyl hydrolase, family 31 protein	1118021.53	740504.92	260931.06	4.28474	0.01778

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_055370	conserved hypothetical protein	hypothetical protein	4586380.78	2070334.04	1203494.55	3.81089	0.00502
NCLIV_061030	hypothetical protein	RuvB family 1 protein	1127153.72	564791.71	306764.83	3.67433	0.00304
NCLIV_063980	conserved hypothetical protein	hypothetical protein	1312187.43	578545.94	368569.44	3.56022	0.01488
NCLIV_034900	putative casein kinase II beta chain	ck2 beta subunit	650239.43	382974.10	183041.91	3.55241	0.00379
NCLIV_005860	IMP-specific 5'-nucleotidase, related	IMP-specific 5'-nucleotidase 1, putative	580921.33	291978.48	166599.48	3.48693	0.00180
NCLIV_023610	putative vacuolar ATP synthase subunit d	V-type ATPase, D subunit protein	142938.57	94708.37	41048.32	3.48220	0.02397
NCLIV_004570	hypothetical protein	N/A	10511384.56	6728637.61	3137854.36	3.34986	0.00609
NCLIV_066360	conserved hypothetical protein	hypothetical protein	630736.47	282140.95	190170.78	3.31668	0.02648
NCLIV_063150	Serpin peptidase inhibitor, clade B (Ovalbumin),member 1, like 3, related	protease inhibitor PI1 (PI1)	16691405.78	7220226.92	5090209.34	3.27912	0.01204
NCLIV_056660	conserved hypothetical protein	hypothetical protein	181215.13	73395.65	56346.76	3.21607	0.01038
NCLIV_044600	conserved hypothetical protein	N/A	3233199.15	2043009.78	1006283.70	3.21301	0.00262
NCLIV_023780	conserved hypothetical protein	hypothetical protein	2031715.11	1084917.10	634411.40	3.20252	0.01101
NCLIV_037540	YOR039Wp-like protein, related	ck2 beta subunit	511020.22	344545.21	162456.59	3.14558	0.00877
NCLIV_024030	conserved hypothetical protein	hypothetical protein	5689192.84	3941583.71	1824754.94	3.11778	0.00190
NCLIV_021640	unspecified product	dense granule protein GRA7 (GRA7)	618276610.00	326387253.80	199183314.30	3.10406	0.00337
NCLIV_024320	conserved hypothetical protein	hypothetical protein	3981815.13	2718709.96	1306110.09	3.04861	0.00057
NCLIV_003120	unspecified product	protease inhibitor PI2 (PI2)	8002697.31	5619635.96	2704237.69	2.95932	0.00066
NCLIV_028820	hypothetical protein	kelch repeat-containing protein	2916345.61	1858850.18	986041.21	2.95763	0.00377
NCLIV_010390	novel protein (Zgc:77804), related	splicing factor U2AF family SnRNP auxilary factor large subunit, RRM domain-containing protein	4239376.50	2740188.69	1453589.39	2.91649	0.00114
NCLIV_003160	nicotinate phosphoribosyltransferase, related	nicotinate phosphoribosyltransferase	942002.72	582945.66	325336.24	2.89547	0.00369
NCLIV_020650	putative splicing factor 3B subunit 1	U2 small nuclear ribonucleoprotein family protein, putative	447408.15	315823.69	156321.45	2.86210	0.00316
NCLIV_025910	Histone H2A, related	26s proteasome subunit p55,	8532398.88	4526686.83	3144178.83	2.71371	0.00066

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
		putative					
NCLIV_052880	unspecified product	dense granule protein GRA6 (GRA6)	108609131.00	60518770.01	40076337.39	2.71006	0.01293
NCLIV_044690	conserved hypothetical protein	hypothetical protein	581942.89	271475.56	215358.05	2.70221	0.01176
NCLIV_062730	hypothetical protein	mitotic checkpoint protein, BUB3 family protein	909942.70	543893.41	350182.10	2.59848	0.04129
NCLIV_014360	hypothetical protein	elongation factor 2 family protein	2310371.76	1446449.93	895882.47	2.57888	0.01931
NCLIV_000520	conserved hypothetical protein	prefoldin subunit protein	252553.53	181093.99	99943.97	2.52695	0.032541
NCLIV_046530	putative reticulon domain-containing protein	reticulon protein	334547.43	248164.11	136183.28	2.45660	0.02728
NCLIV_050000	putative deoxyhypusine synthase,related	deoxyhypusine synthase	1934000.51	1189742.03	792001.16	2.44192	0.03219
NCLIV_009790	putative Rab 11b	Rab11b	1687805.37	939852.19	696915.30	2.42182	0.02143
NCLIV_045260	hypothetical protein	acid phosphatase	7027822.22	4247139.50	2908004.04	2.41672	0.00674
NCLIV_020380	conserved hypothetical protein	hypothetical protein	608311.00	306796.52	251817.89	2.41568	0.02674
NCLIV_020340	unspecified product	hypothetical protein	1336222.10	697723.19	557333.42	2.39753	0.01140
NCLIV_036640	conserved hypothetical protein	hypothetical protein	16650725.79	8620559.91	7061373.27	2.35800	0.00175
NCLIV_036830	conserved hypothetical protein	hypothetical protein	8699374.91	5567171.76	3707757.33	2.34626	0.00085
NCLIV_068460	unspecified product	NTPase I	57901902.25	35903346.15	24898096.57	2.32556	0.00110
NCLIV_047110	conserved hypothetical protein	N/A	731201.03	464361.89	315671.65	2.31633	0.00534
NCLIV_070060	RNA binding protein, putative	RNA recognition motif-containing protein	14347608.80	9019075.82	6227731.82	2.30383	0.01336
NCLIV_061200	putative actin	actin like protein ALP1 (ALP1)	1629642.80	1081094.38	708821.73	2.29909	0.02298
NCLIV_001370	putative DEAD/DEAH box helicase	DEAD/DEAH box helicase domain-containing protein	1500299.11	993070.44	654154.77	2.29349	0.01900
NCLIV_067470	hypothetical protein	multiprotein bridging factor type 1 family transcriptional co-activator, putative	1176164.21	840727.53	514204.23	2.28735	0.00284
NCLIV_069630	hypothetical protein, conserved	hypothetical protein	5748488.26	3387961.65	2567834.16	2.23865	0.02048
NCLIV_063760	hypothetical protein	ubiquitin carboxyl-terminal hydrolase, family 1 protein	1999938.21	1392250.99	895133.71	2.23423	0.00232
NCLIV_031750	putative importin	HEAT repeat-containing protein	2596120.33	1581932.50	1168221.72	2.22228	0.02113
NCLIV_019780	putative KH domain containing protein	hypothetical protein	8771405.31	6430697.48	3950319.81	2.22043	0.00187
NCLIV_031530	hypothetical protein	hypothetical protein	13254567.02	7992470.97	5988436.38	2.21336	0.00770

ID	*N. caninum* description	*T. gondii* orthologues	D0	D3	D6	Max fold change	q Value
NCLIV_008170	hypothetical protein	hypothetical protein	16013829.89	9788103.29	7248816.55	2.20916	0.00503
NCLIV_025000	hypothetical protein	trypsin domain-containing protein	4138686.77	2628971.02	1874607.73	2.20776	0.00766
NCLIV_045860	hypothetical protein	histone deacetylase HDAC3 (HDAC3)	703371.86	540945.61	321244.59	2.18952	0.00775
NCLIV_003100	unspecified product	serine proteinase inhibitor PI-2, putative	4546940.29	3356091.04	2093847.01	2.17157	0.02069
NCLIV_068890	unspecified product	hypothetical protein	11501016.82	6773863.20	5345851.67	2.15139	0.00505
NCLIV_053370	putative U2 small nuclear ribonucleoprotein auxiliary factor U2AF	splicing factor U2AF protein	624439.79	386557.49	290806.17	2.14727	0.04941
NCLIV_022000	putative para-aminobenzoate synthase	N/A	1744171.81	1309820.70	834935.81	2.08899	0.01692
NCLIV_012090	putative ATP-dependent protease ATP-binding subunit	N/A	743237.21	411773.61	356410.79	2.08534	0.03659
NCLIV_013150	conserved hypothetical protein	hypothetical protein	10411671.65	7444974.26	4997737.59	2.08328	0.01653
NCLIV_013460	hypothetical protein	lanthionine synthetase C family protein	3473036.78	2412415.31	1673632.76	2.07515	0.03516
NCLIV_068850	unspecified product	rhoptry kinase family protein ROP24 (incomplete catalytic triad) (ROP24)	3384889.18	1981671.93	1644380.20	2.05846	0.00534
NCLIV_041740	conserved hypothetical protein	hypothetical protein	18042761.99	11078187.63	8788788.17	2.05293	0.00675
NCLIV_062890	hypothetical protein	ribosomal-ubiquitin protein RPS27A (RPS27A)	1592854.49	967630.17	789082.18	2.01862	0.03616
NCLIV_041790	conserved hypothetical protein	hypothetical protein	4117481.52	2802880.94	2049093.08	2.00942	0.00298
NCLIV_011960	conserved hypothetical protein	hypothetical protein	2550948.21	1955209.25	1269521.48	2.00938	0.00993
NCLIV_054810	hypothetical protein	protein phosphatase 2C domain-containing protein	661665.95	463870.77	329676.27	2.00702	0.02949
NCLIV_061750	conserved hypothetical protein	proteasome-interacting thioredoxin domain-containing protein	1222449.17	859812.02	609298.88	2.00632	0.03468

Appendix Table VII: The preferentially expressed genes in Nc Spain 1H by FPKM (Fragments per kilobase of exon per million fragments mapped)

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold change (1H/7)	q value
NCLIV_005760	putative SRS13	10.88	1.95	5.5784	0.0054
NCLIV_060015	unspecified product	44.03	9.91	4.4425	0.0054
NCLIV_027990	putative WD domain, G-beta repeat-containing protein	4.49	1.04	4.3317	0.0498
NCLIV_046230	putative cGMP-inhibited 3',5'-cyclic phosphodiesterase	2.84	0.66	4.2684	0.0054
NCLIV_067430	conserved hypothetical protein	5.39	1.28	4.2081	0.0054
NCLIV_016130	conserved hypothetical protein	10.84	2.60	4.1759	0.0262
NCLIV_042840	hypothetical protein	23.65	5.82	4.0629	0.0054
NCLIV_069900	zinc finger (CCCH type) protein, putative	2.87	0.73	3.9323	0.0349
NCLIV_042910	malate dehydrogenase, related	118.62	30.31	3.9142	0.0054
NCLIV_003260	putative NcMCP3	95.71	24.78	3.8627	0.0054
NCLIV_038930	srs domain-containing protein	19.96	5.29	3.7700	0.0174
NCLIV_000750	conserved hypothetical protein	75.53	20.05	3.7666	0.0054
NCLIV_017300	putative 3', 5'-cyclic nucleotide phosphodiesterase	49.95	13.48	3.7067	0.0054
NCLIV_031960	conserved hypothetical protein	13.80	3.79	3.6409	0.0054
NCLIV_015120	hypothetical protein	16.51	4.63	3.5659	0.0054
NCLIV_013310	hypothetical protein	6.41	1.81	3.5423	0.0114
NCLIV_068830	PAN domain-containing protein (EC 3.4.21.27),related	13.91	4.04	3.4439	0.0141
NCLIV_019580	SRS domain-containing protein	1086.73	319.31	3.4034	0.0054
NCLIV_009520	conserved hypothetical protein	11.34	3.35	3.3876	0.0162
NCLIV_021430	hypothetical protein	8.56	2.60	3.2903	0.0054
NCLIV_010810	putative deoxyribose-phosphate aldolase	36.31	11.36	3.1967	0.0141
NCLIV_069520	hypothetical protein	36.03	11.44	3.1497	0.0054

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold change (1H/7)	q value
NCLIV_022840	conserved hypothetical protein	14.04	4.49	3.1253	0.0141
NCLIV_025840	conserved hypothetical protein	18.66	6.01	3.1035	0.0054
NCLIV_003250	putative NcMCP4	126.36	40.85	3.0930	0.0054
NCLIV_037560	conserved hypothetical protein	26.60	8.66	3.0704	0.0136
NCLIV_043730	hypothetical protein	10.34	3.37	3.0642	0.0054
NCLIV_037490	unspecified product	127.20	41.62	3.0565	0.0054
NCLIV_006950	conserved hypothetical protein	7.08	2.34	3.0279	0.0054
NCLIV_067410	conserved hypothetical protein	3.33	1.11	2.9940	0.0349
NCLIV_058130	conserved hypothetical protein	10.67	3.57	2.9884	0.0141
NCLIV_066170	hypothetical protein	88.39	29.80	2.9660	0.0054
NCLIV_034870	MAC/Perforin domain containing protein, related	3.03	1.03	2.9407	0.0392
NCLIV_069935	unspecified product	9.57	3.31	2.8922	0.0114
NCLIV_026990	conserved hypothetical protein	18.43	6.40	2.8775	0.0054
NCLIV_037830	putative RNA recognition motif-containing protein	6.17	2.17	2.8483	0.0054
NCLIV_027630	conserved hypothetical protein	38.62	13.62	2.8354	0.0054
NCLIV_027470	putative bradyzoite antigen	201.69	71.24	2.8310	0.0141
NCLIV_020100	SRS domain-containing protein	13.86	4.95	2.7995	0.0141
NCLIV_000760	conserved hypothetical protein	196.74	70.63	2.7855	0.0092
NCLIV_038440	conserved hypothetical protein	559.18	201.23	2.7788	0.0054
NCLIV_070090	hypothetical protein, conserved	36.99	13.33	2.7759	0.0054
NCLIV_038100	hypothetical protein	11.72	4.23	2.7727	0.0402
NCLIV_002790	conserved hypothetical protein	1.77	0.65	2.7404	0.0162
NCLIV_043600	srs domain-containing protein	27.05	10.16	2.6621	0.0114
NCLIV_025520	conserved hypothetical protein	7.71	2.96	2.6010	0.0272
NCLIV_022240	ATPase, related	45.58	17.55	2.5972	0.0136

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold change (1H/7)	q value
NCLIV_054090	hypothetical protein	4.73	1.85	2.5555	0.0054
NCLIV_034731	SRS domain-containing protein	31.25	12.23	2.5545	0.0114
NCLIV_005420	phosphoglycerate kinase, related	10.49	4.11	2.5497	0.0114
NCLIV_064160	conserved hypothetical protein	15.87	6.27	2.5302	0.0174
NCLIV_022830	conserved hypothetical protein	10.09	4.00	2.5253	0.0287
NCLIV_025120	unspecified product	6.63	2.65	2.4970	0.0315
NCLIV_0269	GDA1/CD39 (nucleoside phospha tase) family domain containing protein	25.03	10.07	2.4858	0.0054
NCLIV_048010	putative regulator of chromosome condensation domain-containing protein	12.04	4.86	2.4756	0.0141
NCLIV_062510	hypothetical protein	13.59	5.52	2.4638	0.0449
NCLIV_007660	hypothetical protein	16.50	6.73	2.4522	0.0114
NCLIV_044990	putative high-affinity cGMP-specific 3',5'-cyclic phosphodies terase 9A	48.16	19.68	2.4477	0.0054
NCLIV_034170	hypothetical protein	11.37	4.65	2.4441	0.0054
NCLIV_062990	hypothetical protein	11.39	4.69	2.4268	0.0209
NCLIV_033540	conserved hypothetical protein	18.40	7.60	2.4204	0.0054
NCLIV_051000	Ethylene-inducible protein hever, related	16.81	6.95	2.4185	0.0299
NCLIV_043550	glycerol-3-phosphate dehydrogenase, related	90.35	37.37	2.4178	0.0209
NCLIV_056990	Proteophosphoglycan 5, related	46.85	19.40	2.4155	0.0092
NCLIV_065960	Dynein heavy chain 1, axonemal, related	7.73	3.21	2.4046	0.0054
NCLIV_0380	transporter, major facilitator family domain containing protein	3.10	1.29	2.4046	0.0349
NCLIV_006120	hypothetical protein	172.10	72.61	2.3703	0.0054
NCLIV_031040	Peptidyl-prolyl cis-trans isomerase,related	187.32	79.14	2.3669	0.0054
NCLIV_009880	conserved hypothetical protein	23.08	9.81	2.3527	0.0054
NCLIV_069530	hypothetical protein	7.25	3.09	2.3450	0.0141

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold change (1H/7)	q value
NCLIV_009830	conserved hypothetical protein	4.44	1.90	2.3376	0.0162
NCLIV_023990	Acyl carrier protein, related	184.02	79.39	2.3178	0.0054
NCLIV_050200	GF18580, related	3.15	1.36	2.3136	0.0162
NCLIV_061910	hypothetical protein	56.44	24.54	2.2999	0.0299
NCLIV_015830	hypothetical protein	8.92	3.89	2.2910	0.0136
NCLIV_062520	3-ketoacyl-(Acyl-carrier-protein) reductase, related	28.86	12.62	2.2869	0.0193
NCLIV_028835	Glucose-6-phosphate isomerase,related	7.95	3.49	2.2759	0.0456
NCLIV_038280	conserved hypothetical protein	92.32	40.75	2.2657	0.0272
NCLIV_061740	Hydrolase, alpha/beta fold family, related	20.57	9.09	2.2630	0.0262
NCLIV_024130	putative ATP-dependent RNA helicase	20.91	9.25	2.2602	0.0272
NCLIV_049510	SRS domain-containing protein	78.76	35.16	2.2399	0.0136
NCLIV_066970	putative enoyl-acyl carrier reductase	16.35	7.33	2.2314	0.0456
NCLIV_041630	conserved hypothetical protein	80.84	36.23	2.2314	0.0054
NCLIV_032950	putative deoxyuridine 5'-triphosphate nucleotidohydrolase	247.83	112.74	2.1982	0.0054
NCLIV_067820	conserved hypothetical protein	11.93	5.44	2.1930	0.0349
NCLIV_060330	putative molybdopterin cofactor sulfurase	6.87	3.16	2.1712	0.0450
NCLIV_034180	conserved hypothetical protein	35.67	16.47	2.1655	0.0449
NCLIV_048760	conserved hypothetical protein	14.85	6.86	2.1641	0.0092
NCLIV_048340	putative pseudouridylate synthase 1	6.57	3.05	2.1584	0.0174
NCLIV_063000	Coiled-coil protein, related	11.02	5.11	2.1567	0.0114
NCLIV_055730	hypothetical protein	28.10	13.07	2.1505	0.0114
NCLIV_068890	unspecified product	92.09	43.02	2.1406	0.0054
NCLIV_014020	Peroxiredoxin-2E-1, related	76.10	35.60	2.1377	0.0054
NCLIV_008230	delta-aminolevulinic acid dehydratase, related	27.99	13.10	2.1364	0.0054

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold change (1H/7)	q value
NCLIV_044100	conserved hypothetical protein	57.62	26.99	2.1348	0.0141
NCLIV_056970	conserved hypothetical protein	13.92	6.53	2.1303	0.0092
NCLIV_030730	putative acetyltransferase domain-containing protein	30.30	14.23	2.1291	0.0262
NCLIV_042760	dehydrogenases with different specificities,related	74.90	35.36	2.1184	0.0054
NCLIV_008220	putative M16 family peptidase	8.71	4.14	2.1053	0.0114
NCLIV_014420	conserved hypothetical protein	17.89	8.53	2.0975	0.0378
NCLIV_026260	GA26239, related	24.22	11.55	2.0969	0.0054
NCLIV_0155	alpha/beta hydrolase fold domain containing protein	23.55	11.28	2.0885	0.0141
NCLIV_001280	putative ribokinase	59.24	28.90	2.0499	0.0092
NCLIV_059130	hypothetical protein	1576.34	770.60	2.0456	0.0054
NCLIV_023380	putative protein kinase, PfPK7 homolog	19.81	9.70	2.0421	0.0449
NCLIV_062650	putative DNA polymerase alpha catalytic subunit	10.10	5.02	2.0094	0.0054

Appendix Table VIII: The preferentially expressed genes in Nc Spain 7 by FPKM (Fragments per kilobase of exon per million fragments mapped)

Gene ID	Description	FPKM Nc Spain 1H	FPKM Nc Spain 7	Fold Change (7/1H)	q value
NCLIV_047150	conserved hypothetical protein	64.28	513.26	7.9853	0.0054
NCLIV_015810	putative retinitis pigmentosa GTPase regulator	0.38	1.69	4.4847	0.0092
NCLIV_038025	unspecified product	0.72	2.62	3.6505	0.0092
NCLIV_047420	conserved hypothetical protein	0.53	1.84	3.4888	0.0092
NCLIV_019010	putative protein phosphatase 2C	1.48	4.64	3.1355	0.0054
NCLIV_006200	hypothetical protein	5.49	16.87	3.0747	0.0054
NCLIV_069070	hypothetical protein	110.04	310.79	2.8243	0.0141
NCLIV_011060	conserved hypothetical protein	10.85	29.42	2.7107	0.0114
NCLIV_058960	conserved hypothetical protein	2.43	6.51	2.6747	0.0054
NCLIV_038021	unspecified product	3.41	8.73	2.5585	0.0054
NCLIV_050550	hypothetical protein	6.38	15.86	2.4874	0.0114
NCLIV_006110	conserved hypothetical protein	8.67	21.15	2.4407	0.0054
NCLIV_004950	putative actin	17.14	41.13	2.4001	0.0054
NCLIV_046220	hypothetical protein	2.65	6.27	2.3639	0.0114
NCLIV_019000	putative adenosine transporter	273.57	626.16	2.2889	0.0136
NCLIV_027970	conserved hypothetical protein	243.12	555.45	2.2847	0.0249
NCLIV_053760	unspecified product	19.39	44.28	2.2839	0.0054
NCLIV_012010	hypothetical protein	64.07	144.57	2.2564	0.0054
NCLIV_064250	putative zinc finger (CCCH type) protein	9.51	21.20	2.2306	0.0054
NCLIV_066300	putative PPR repeat-containing protein	5.47	11.68	2.1356	0.0092
NCLIV_027980	hypothetical protein	76.14	160.91	2.1133	0.0287
NCLIV_030830	conserved hypothetical protein	29.53	61.02	2.0662	0.0141
NCLIV_022340	putative protein phosphatase 2C	7.99	16.38	2.0498	0.0114
NCLIV_007720	conserved hypothetical protein	978.93	1991.08	2.0339	0.0341
NCLIV_006870	conserved hypothetical protein	119.30	241.70	2.0260	0.0114
NCLIV_051850	conserved hypothetical protein	7.32	14.64	2.0003	0.0174

Appendix Table IX**:** Gene Ontology enrichment: Biological Process, Cellular Component and Molecular Function of 106 preferentially expressed genes in Nc Spain 1H obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0007018	microtubule-based movement	29	5	17.2	10.04	10.89	0.000220797	0.005393107
GO:0006928	cellular component movement	31	5	16.1	9.39	10.18	0.000291519	0.005393107
GO:0007017	microtubule-based process	41	5	12.2	7.1	7.67	0.000931729	0.011491325
GO:0008150	biological_process	1771	40	2.3	1.31	2.01	0.008398139	0.04983305
GO:0005975	carbohydrate metabolic process	86	5	5.8	3.38	3.61	0.017750058	0.04983305
GO:0006633	fatty acid biosynthetic process	12	2	16.7	9.7	10.01	0.022399471	0.04983305
GO:0071704	organic substance metabolic process	65	4	6.2	3.58	3.77	0.028068384	0.04983305
GO:1901135	carbohydrate derivative metabolic process	65	4	6.2	3.58	3.77	0.028068384	0.04983305
GO:0006096	glycolysis	14	2	14.3	8.32	8.58	0.028908028	0.04983305
GO:0006631	fatty acid metabolic process	14	2	14.3	8.32	8.58	0.028908028	0.04983305
GO:0046080	dUTP metabolic process	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0009211	pyrimidine deoxyribonucleoside triphosphate metabolic process	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0009211	pyrimidine deoxyribonucleoside triphosphate metabolic process	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0006200	ATP catabolic process	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0009200	deoxyribonucleoside triphosphate metabolic process	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0042816	vitamin B6 metabolic process	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0042822	pyridoxal phosphate metabolic process	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0046836	glycolipid transport	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0042823	pyridoxal phosphate biosynthetic process	1	1	100	58.21	59.21	0.033499416	0.04983305
GO:0042819	vitamin B6 biosynthetic process	1	1	100	58.21	59.21	0.033499416	0.04983305

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
GO:0032787	monocarboxylic acid metabolic process	16	2	12.5	7.28	7.5	0.036073864	0.04983305
GO:0007596	blood coagulation	18	2	11.1	6.47	6.66	0.043847232	0.04983305
GO:0050878	regulation of body fluid levels	18	2	11.1	6.47	6.66	0.043847232	0.04983305
GO:0007599	hemostasis	18	2	11.1	6.47	6.66	0.043847232	0.04983305
GO:0042060	wound healing	18	2	11.1	6.47	6.66	0.043847232	0.04983305
GO:0050817	coagulation	18	2	11.1	6.47	6.66	0.043847232	0.04983305
GO:0009611	response to wounding	18	2	11.1	6.47	6.66	0.043847232	0.04983305
GO:0044283	small molecule biosynthetic process	44	3	6.8	3.97	4.13	0.043889576	0.04983305
GO:0006007	glucose catabolic process	19	2	10.5	6.13	6.31	0.04794688	0.04983305
GO:0032501	multicellular organismal process	19	2	10.5	6.13	6.31	0.04794688	0.04983305
GO:0019320	hexose catabolic process	19	2	10.5	6.13	6.31	0.04794688	0.04983305
GO:0046365	monosaccharide catabolic process	19	2	10.5	6.13	6.31	0.04794688	0.04983305
GO:0006108	malate metabolic process	2	1	50	29.1	29.6	0.04983305	0.04983305
GO:0009394	2'-deoxyribonucleotide metabolic process	2	1	50	29.1	29.6	0.04983305	0.04983305
GO:0046168	glycerol-3-phosphate catabolic process	2	1	50	29.1	29.6	0.04983305	0.04983305
GO:0046434	organophosphate catabolic process	2	1	50	29.1	29.6	0.04983305	0.04983305
GO:0009219	pyrimidine deoxyribonucleotide metabolic process	2	1	50	29.1	29.6	0.04983305	0.04983305
GO:0006072	glycerol-3-phosphate metabolic process	2	1	50	29.1	29.6	0.04983305	0.04983305
Cellular Component:								
GO:0030286	dynein complex	10	5	50	29.1	31.75	3.05E-06	2.13E-05
GO:0005875	microtubule associated complex	17	5	29.4	17.12	18.64	2.44E-05	8.54E-05
GO:0015630	microtubule cytoskeleton	22	5	22.7	13.23	14.38	7.01E-05	0.000164
GO:0044430	cytoskeletal part	35	5	14.3	8.32	9.01	0.000483	0.000846
GO:0005856	cytoskeleton	40	5	12.5	7.28	7.87	0.000841	0.001178

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
GO:0005576	extracellular region	25	3	12	6.98	7.31	0.011103	0.012954
GO:0000015	phosphopyruvate hydratase complex	2	1	50	29.1	29.6	0.049833	0.049833
Molecular Function:								
GO:0003777	microtubule motor activity	29	5	17.2	10.04	10.89	0.000221	0.005299
GO:0003774	motor activity	40	5	12.5	7.28	7.87	0.000841	0.010095
GO:0016887	ATPase activity	111	7	6.3	3.67	4.04	0.003326	0.019487
GO:0003824	catalytic activity	1365	34	2.5	1.45	2.09	0.004211	0.019487
GO:0016616	oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor	18	3	16.7	9.7	10.18	0.004901	0.019487
GO:0016614	oxidoreductase activity, acting on CH-OH group of donors	18	3	16.7	9.7	10.18	0.004901	0.019487
GO:0048037	cofactor binding	64	5	7.8	4.55	4.88	0.005684	0.019487
GO:0050662	coenzyme binding	50	4	8	4.66	4.93	0.012371	0.037113
GO:0051287	NAD binding	9	2	22.2	12.93	13.36	0.013984	0.037291
GO:0016836	hydro-lyase activity	10	2	20	11.64	12.02	0.016601	0.039841
GO:0016835	carbon-oxygen lyase activity	11	2	18.2	10.58	10.93	0.019408	0.042345
GO:0016829	lyase activity	37	3	8.1	4.72	4.92	0.029042	0.049833
GO:0004655	porphobilinogen synthase activity	1	1	100	58.21	59.21	0.033499	0.049833
GO:0051861	glycolipid binding	1	1	100	58.21	59.21	0.033499	0.049833
GO:0004316	3-oxoacyl-[acyl-carrier-protein] reductase (NADPH) activity	1	1	100	58.21	59.21	0.033499	0.049833
GO:0017089	glycolipid transporter activity	1	1	100	58.21	59.21	0.033499	0.049833
GO:0036094	small molecule binding	641	17	2.7	1.54	1.77	0.039846	0.049833
GO:0019842	vitamin binding	18	2	11.1	6.47	6.66	0.043847	0.049833
GO:0030060	L-malate dehydrogenase activity	2	1	50	29.1	29.6	0.049833	0.049833
GO:0016615	malate dehydrogenase activity	2	1	50	29.1	29.6	0.049833	0.049833
GO:0004312	fatty acid synthase activity	2	1	50	29.1	29.6	0.049833	0.049833
GO:0004618	phosphoglycerate kinase activity	2	1	50	29.1	29.6	0.049833	0.049833

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
GO:0004634	phosphopyruvate hydratase activity	2	1	50	29.1	29.6	0.049833	0.049833
GO:0004367	glycerol-3-phosphate dehydrogenase [NAD+] activity	2	1	50	29.1	29.6	0.049833	0.049833

Appendix Table X**:** Gene Ontology enrichment: Molecular Function of 26 preferentially genes in Nc Spain 7 obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Molecular Function:								
GO:0015932	nucleobase-containing compound transmembrane transporter activity	3	1	33.3	225.07	281.08	0.005905	0.008857
GO:0005337	nucleoside transmembrane transporter activity	3	1	33.3	225.07	281.08	0.005905	0.008857
GO:0008270	zinc ion binding	232	2	0.9	5.82	9.03	0.041488	0.041488

Appendix Table XI**:** List of the preferentially expressed proteins between Nc Spain 1H and Nc Spain 7 at 12 hours post infection

Accession	Description	Nc 1H-12hr	Nc 7-12hr	fold change	q Value
NCLIV_019580	SRS domain-containing protein	231858.78	35504.03	6.53	0.010746
NCLIV_036640	conserved hypothetical protein	108409.67	20405.09	5.31	0.008713
NCLIV_022920	Renin, related	34023.50	6502.10	5.23	0.030770
NCLIV_049520	conserved hypothetical protein	4166595.93	819051.67	5.09	0.008507
NCLIV_053400	putative histidine acid phosphatase domain containing protein	42523.80	11744.66	3.62	0.022182
NCLIV_027130	putative orotate phosphoribosyltransferase	94128.62	26321.87	3.58	0.028098
NCLIV_019740	putative 3',5'-cyclic phosphodiesterase	1855.22	537.45	3.45	0.037773
NCLIV_012090	putative ATP-dependent protease ATP-binding subunit	23024.58	6765.14	3.40	0.008507
NCLIV_018300	HIT family protein, related	1402055.32	421821.52	3.32	0.015779
NCLIV_012460	Lysyl-tRNA synthetase, related	2738.42	826.68	3.31	0.033022
NCLIV_020990	perforin-like protein PLP1 (PLP1)	202166.73	65566.06	3.08	0.008713
NCLIV_037460	citrate synthase I	404581.56	152092.37	2.66	0.024397
NCLIV_065500	conserved hypothetical protein	445504.88	169497.34	2.63	0.019484
NCLIV_007010	putative cathepsin C2 (TgCPC2)	274324.17	106677.59	2.57	0.024397
NCLIV_011110	putative DEAH-box RNA/DNA helicase	20806.13	8151.25	2.55	0.017953
NCLIV_041850	putative Hsp20/alpha crystallin domain-containing protein	2691566.63	1056490.91	2.55	0.010746
NCLIV_069110	Hypothetical protein	27187.72	11024.39	2.47	0.049761
NCLIV_022970	microneme protein MIC2 (TgMIC2)	1968924.78	807753.17	2.44	0.008713
NCLIV_062770	microneme protein MIC9 (TgMIC9)	1338525.62	552301.84	2.42	0.037391
NCLIV_010600	putative microneme protein MIC3	34648535.79	14310252.42	2.42	0.008507
NCLIV_021050	subtilisin SUB1 (SUB1)	31262506.44	13006985.74	2.40	0.008507
NCLIV_065650	putative TPR domain-containing protein	13131.28	5468.16	2.40	0.048395
NCLIV_034740	SAG3 protein, related	1113514.51	465065.22	2.39	0.012491
NCLIV_040860	tryptophanyl-tRNAsynthetase,related	6121008.97	2562229.27	2.39	0.012212
NCLIV_035300	putative lipoate-protein ligase a	51071.75	21468.28	2.38	0.030770
NCLIV_063810	conserved hypothetical protein	87183.75	36819.63	2.37	0.012212
NCLIV_040360	conserved hypothetical	7030.48	2991.37	2.35	0.040856

Accession	Description	Nc 1H-12hr	Nc 7-12hr	fold change	q Value
	protein				
NCLIV_054180	Oligoribonuclease, related	29805.83	13001.33	2.29	0.034743
NCLIV_004200	glycosyltransferase	16610.25	7300.75	2.28	0.024441
NCLIV_028680	putative apical membrane antigen 1	4014612.86	1813866.83	2.21	0.008507
NCLIV_050070	Fructose-1,6-bisphosphatase,related	322702.68	146592.73	2.20	0.013771
NCLIV_041900	phosphoenolpyruvate carboxykinase (ATP), related	2037292.77	928541.53	2.19	0.008713
NCLIV_068230	putative nucleoside-triphosphatase	90478.38	41289.12	2.19	0.030846
NCLIV_041100	novel protein (Zgc:66430), related	1796958.14	835818.69	2.15	0.041156
NCLIV_010960	putative phosphoglucomutase	440019.73	205599.99	2.14	0.008507
NCLIV_033690	microneme protein MIC2 (MIC2)	3492514.19	1641563.85	2.13	0.008507
NCLIV_031610	Glucosamine--fructose-6-phosphate aminotransferase (Isomerizing), related	19838.00	9328.99	2.13	0.032766
NCLIV_007105	unspecified product	1031577.09	487742.66	2.12	0.040856
NCLIV_002940	putative microneme protein MIC4	14034172.45	6669575.98	2.10	0.012491
NCLIV_006000	inositol(myo)-1(or 4)-monophosphatase 2, putative	14991.28	7142.59	2.10	0.040900
NCLIV_054840	conserved hypothetical protein	336594.75	161187.10	2.09	0.033319
NCLIV_067490	putative protein phosphatase 2C	579671.36	278164.64	2.08	0.014900
NCLIV_030630	p25-alpha family protein, related	855722.54	411525.19	2.08	0.019427
NCLIV_029160	Pyridine nucleotide-disulphide oxidoreductase family protein, related	13199.86	6348.73	2.08	0.046223
NCLIV_060910	putative DNA replication licensing factor	5165.89	2489.96	2.07	0.014900
NCLIV_054140	putative adenylyl cyclase associated protein	6297412.92	3047581.05	2.07	0.010728
NCLIV_001180	cbr-GPDH-1 protein, related	30393.72	14727.40	2.06	0.018399
NCLIV_058830	Superoxide dismutase, related	949048.26	466515.17	2.03	0.008507
NCLIV_050080	Fructose-1,6-bisphosphatase class 1, related	380865.11	187603.60	2.03	0.037924
NCLIV_060860	Cytochrome c, related	1576594.78	779486.43	2.02	0.009868
NCLIV_019540	putative signal recognition particle protein	57677.51	28535.99	2.02	0.022182
NCLIV_064760	Contig An16c0190, complete genome. (Precursor), related	110329.28	54614.20	2.02	0.041113
NCLIV_010370	ubiquitin carrier protein, related	4119127.04	2043757.57	2.02	0.038979
NCLIV_060920	conserved hypothetical	521150.48	258834.59	2.01	0.045450

Accession	Description	Nc 1H-12hr	Nc 7-12hr	fold change	q Value
	protein				
NCLIV_011830	conserved hypothetical protein	1314349.61	653788.87	2.01	0.012491
NCLIV_027230	Dihydropteroate synthase, related	1431083.85	713607.59	2.01	0.013090
NCLIV_001070	histone H2B, related	7791806.08	15605528.63	-2.00	0.017404
NCLIV_043670	high mobility group protein, related	1151938.37	2390237.21	-2.07	0.017404
NCLIV_003700	conserved hypothetical protein	14446.79	32043.80	-2.22	0.014900
NCLIV_004680	conserved hypothetical protein	20769.13	46693.30	-2.25	0.019397
NCLIV_013660	putative regulator of chromosome condensation	6377.03	14349.80	-2.25	0.020239
NCLIV_005800	unspecified product	7195.45	16205.26	-2.25	0.022798
NCLIV_014020	Peroxiredoxin-2E-1, related	282524.95	673761.93	-2.38	0.013090
NCLIV_057260	conserved hypothetical protein	82029.28	199503.64	-2.43	0.039167
NCLIV_066900	putative serine/threonine protein phosphatase 5	9279.11	22678.42	-2.44	0.017355
NCLIV_004340	beta-hydroxyacyl-acyl carrier protein dehydratase (FABZ)	56748.15	139170.82	-2.45	0.039981
NCLIV_032350	putative CW-type zinc finger domain-containing protein	2525.99	6259.59	-2.48	0.015322
NCLIV_056130	conserved hypothetical protein	13192.10	33564.92	-2.54	0.040900
NCLIV_030730	putative acetyltransferase domain-containing protein	14096.68	36815.12	-2.61	0.017404
NCLIV_058260	hypothetical protein	1164911.44	3279434.40	-2.82	0.047104
NCLIV_026210	apicoplast triosephosphate translocator APT1 (APT1)	184546.29	520631.34	-2.82	0.012491
NCLIV_068850	Predicted rhoptry kinase, subfamily ROP24	307278.32	902109.59	-2.94	0.013006
NCLIV_023610	putative vacuolar ATP synthase subunit d	15790.05	46768.32	-2.96	0.017404
NCLIV_054690	conserved hypothetical protein	1986.18	6184.76	-3.11	0.036273
NCLIV_039290	transporter/permease protein	27564.48	91771.33	-3.33	0.013090
NCLIV_010650	conserved hypothetical protein	8883.40	41789.64	-4.70	0.034678
NCLIV_023620	SRS domain-containing protein	11502.31	86069.13	-7.48	0.010746

Appendix Table XII**:** List of the preferentially expressed proteins between Nc Spain 1H and Nc Spain 7 at 36 hours post infection

Accession	Description	Nc 1H-12hr	Nc 7-12hr	Fold change	q Value
NCLIV_019580	SRS domain-containing protein	310976.44	24221.35	12.84	0.010883
NCLIV_049520	conserved hypothetical protein	2507869.20	373570.09	6.71	0.010883
NCLIV_036640	conserved hypothetical protein	127257.07	24347.76	5.23	0.010883
NCLIV_022920	Renin, related	32867.31	7075.07	4.65	0.038356
NCLIV_053400	putative histidine acid phosphatase domain containing protein	40155.73	9350.73	4.29	0.028609
NCLIV_007010	putative cathepsin C2 (TgCPC2)	233283.80	85862.13	2.72	0.014564
NCLIV_008320	protein F23B2.12, partially confirmed by transcript evidence, related	12493.30	4700.62	2.66	0.044566
NCLIV_034740	SAG3 protein, related	1041019.33	396426.25	2.63	0.031809
NCLIV_054180	Oligoribonuclease, related	41807.20	17333.89	2.41	0.028769
NCLIV_063570	Peptidase, S9A/B/C family, catalytic domain protein, related	21390.60	9018.98	2.37	0.011348
NCLIV_031190	hypothetical protein	18011.58	7713.01	2.34	0.028609
NCLIV_029980	putative glycerol-3-phosphate acyltransferase	3380.46	1537.80	2.20	0.031926
NCLIV_031610	Glucosamine--fructose-6-phosphate aminotransferase (Isomerizing), related	19100.03	9354.26	2.04	0.045253
NCLIV_014020	Peroxiredoxin-2E-1, related	374177.49	877317.10	-2.34	0.015602
NCLIV_041790	conserved hypothetical protein	25897.95	61260.14	-2.37	0.015602
NCLIV_021440	putative RNA binding protein	2927.82	6963.26	-2.38	0.028897
NCLIV_050060	conserved hypothetical protein	168618.80	416210.17	-2.47	0.010883
NCLIV_068850	Predicted rhoptry kinase, subfamily ROP24	258639.81	718755.68	-2.78	0.019337
NCLIV_023620	SRS domain-containing protein	25932.55	92463.37	-3.57	0.015602

Appendix Table XIII**:** List of the preferentially expressed proteins between Nc Spain 1H and Nc Spain 7 at 56 hours post infection

Accession	Description	Nc 1H-12hr	Nc 7-12hr	Fold change	q Value
NCLIV_019580	SRS domain-containing protein	405070.47	27750.65	14.60	0.029754
NCLIV_049520	conserved hypothetical protein	4695254.48	398502.96	11.78	0.027763
NCLIV_053400	putative histidine acid phosphatase domain containing protein	50228.98	6897.47	7.28	0.027695
NCLIV_022920	Renin, related	48104.27	6882.95	6.99	0.030872
NCLIV_038110	microneme protein MIC17C (MIC17C)	53036.03	9072.69	5.85	0.036448
NCLIV_036640	conserved hypothetical protein	136244.42	27137.77	5.02	0.038383
NCLIV_007010	putative cathepsin C2 (TgCPC2)	436309.74	114586.24	3.81	0.041009
NCLIV_034740	SAG3 protein, related	1452783.02	400652.78	3.63	0.016437
NCLIV_008320	protein F23B2.12, partially confirmed by transcript evidence, related	18642.01	5178.58	3.60	0.026346
NCLIV_057890	Protein-L-isoaspartate O-methyltransferase, related	43970.20	13736.82	3.20	0.024390
NCLIV_010050	srs domain-containing protein	220825.35	70365.37	3.14	0.028942
NCLIV_033690	microneme protein MIC2 (MIC2)	5164280.08	1867479.75	2.77	0.033339
NCLIV_002940	putative microneme protein MIC4	24897767.32	9017437.25	2.76	0.016437
NCLIV_019740	putative 3',5'-cyclic phosphodiesterase	3620.19	1317.63	2.75	0.029575
NCLIV_028680	putative apical membrane antigen 1	6386731.22	2354879.39	2.71	0.049602
NCLIV_020990	perforin-like protein PLP1 (PLP1)	311947.73	115987.59	2.69	0.027695
NCLIV_018530	conserved hypothetical protein	301938.58	113617.22	2.66	0.018488
NCLIV_068040	dolichol-phosphate-mannose synthase family protein	29312.09	11032.95	2.66	0.025724
NCLIV_043270	putative microneme protein MIC1	35922016.21	14004447.55	2.57	0.025376
NCLIV_012460	Lysyl-tRNA synthetase, related	2906.03	1148.66	2.53	0.039886
NCLIV_063570	Peptidase, S9A/B/C family, catalytic domain protein, related	18390.84	7307.77	2.52	0.014812
NCLIV_019540	putative signal recognition particle protein	57604.59	23196.15	2.48	0.024390
NCLIV_000740	class I chitinase, related	44813.57	18104.39	2.48	0.007757
NCLIV_065500	conserved hypothetical protein	483184.45	203994.24	2.37	0.000810
NCLIV_021050	subtilisin SUB1 (SUB1)	42177088.26	18189478.66	2.32	0.045328
NCLIV_010600	putative microneme	47685525.48	20568374.02	2.32	0.007757

Accession	Description	Nc 1H-12hr	Nc 7-12hr	Fold change	q Value
	protein MIC3				
NCLIV_061620	hypothetical protein	94758.35	41032.99	2.31	0.025253
NCLIV_060220	conserved hypothetical protein	235590.26	103967.58	2.27	0.001965
NCLIV_044230	putative peptidase M16 domain containing protein	193502.92	86978.56	2.22	0.017098
NCLIV_031300	conserved hypothetical protein	19532.98	8858.65	2.20	0.033525
NCLIV_025760	GD24343, related	11936.57	5472.93	2.18	0.030673
NCLIV_004380	cathepsin L, related	209872.33	96866.38	2.17	0.016437
NCLIV_051970	putative MIC2-associated protein M2AP	13344397.82	6185665.56	2.16	0.016167
NCLIV_026500	ALVEOLIN1, related	253787.20	118003.62	2.15	0.037862
NCLIV_009810	ABC transporter transmembrane region domain-containing protein	6179.38	2888.94	2.14	0.025253
NCLIV_022970	microneme protein MIC2 (MIC2)	2926575.86	1385297.87	2.11	0.014471
NCLIV_054250	Acyl-CoA synthetase, related	237789.24	114258.22	2.08	0.036672
NCLIV_046140	SAG-related sequence SRS67 (SRS67)	450194.60	217267.62	2.07	0.033379
NCLIV_066250	microneme protein MIC10 (MIC10)	7031677.34	3398518.59	2.07	0.017098
NCLIV_055690	small GTPase Rab2, putative	430632.31	209274.21	2.06	0.008960
NCLIV_011110	putative DEAH-box RNA/DNA helicase	19478.67	9483.93	2.05	0.049702
NCLIV_029980	putative glycerol-3-phosphate acyltransferase	4775.41	2325.68	2.05	0.030888
NCLIV_047570	putative rRNA processing protein	395877.67	195957.40	2.02	0.024390
NCLIV_020720	putative microneme protein MIC11	4785446.63	2376646.56	2.01	0.025376
NCLIV_011740	conserved hypothetical protein	180637.34	361764.45	-2.00	0.007514
NCLIV_025990	hydrolase	53559.04	109114.64	-2.04	0.044545
NCLIV_014040	conserved hypothetical protein	215092.43	443753.72	-2.06	0.035656
NCLIV_012950	putative ubiquitin family domain-containing protein	5902.62	12357.41	-2.09	0.021485
NCLIV_005150	beta-1 tubulin, putative	1982208.85	4173581.92	-2.11	0.013384
NCLIV_030730	putative acetyltransferase domain-containing protein	14445.32	32315.82	-2.24	0.025376
NCLIV_007770	putative Rhoptry kinase family protein, truncated (incomplete catalytic triad)	433281.75	981688.73	-2.27	0.007757
NCLIV_0292	YbaK / prolyl-tRNA synthetases associated domain containing protein	11133.55	25340.18	-2.28	0.024390
NCLIV_026210	apicoplast triosephosphate translocator APT1 (APT1)	150238.98	343906.78	-2.29	0.010723
NCLIV_021440	putative RNA binding protein	3683.31	8433.30	-2.29	0.018227
NCLIV_050060	conserved hypothetical	140363.12	360203.43	-2.57	0.014471

Accession	Description	Nc 1H-12hr	Nc 7-12hr	Fold change	q Value
	protein				
NCLIV_037480	putative TPR domain-containing protein	7415.09	20097.99	-2.71	0.027695
NCLIV_042450	putative opine dehydrogenase	261448.45	725179.73	-2.77	0.024390
NCLIV_023620	SRS domain-containing protein	27474.34	80197.10	-2.92	0.000768
NCLIV_063510	GH23102, related	3411.84	10578.65	-3.10	0.000387
NCLIV_068850	Predicted rhoptry kinase, subfamily ROP24	335590.44	1111069.09	-3.31	0.002306
NCLIV_046320	conserved hypothetical protein	361.17	1890.59	-5.23	0.000387
NCLIV_019660	putative ubiquitin conjugation factor	137.21	3690.97	-26.90	0.030673

Appendix Table XIV: Gene Ontology enrichment: Biological Process, Cellular Component and Molecular Function of 56 preferentially expressed proteins in Nc Spain 1H at 12 hours post infection obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0005975	carbohydrate metabolic process	86	8	9.3	7.48	9	1.33E-05	0.000185783
GO:0016051	carbohydrate biosynthetic process	10	2	20	16.08	16.83	0.009000007	0.049090616
GO:0008152	metabolic process	1317	24	1.8	1.46	2.08	0.01367035	0.049090616
GO:0009249	protein lipoylation	1	1	100	80.38	82.32	0.024428335	0.049090616
GO:0006436	tryptophanyl-tRNA aminoacylation	1	1	100	80.38	82.32	0.024428335	0.049090616
GO:0018065	protein-cofactor linkage	1	1	100	80.38	82.32	0.024428335	0.049090616
GO:0044238	primary metabolic process	1075	20	1.9	1.5	1.95	0.024545308	0.049090616
GO:0046434	organophosphate catabolic process	2	1	50	40.19	41.15	0.036423069	0.049803173
GO:0046168	glycerol-3-phosphate catabolic process	2	1	50	40.19	41.15	0.036423069	0.049803173
GO:0006072	glycerol-3-phosphate metabolic process	2	1	50	40.19	41.15	0.036423069	0.049803173
GO:0072593	reactive oxygen species metabolic process	3	1	33.3	26.79	27.42	0.048273795	0.049803173
GO:0006801	superoxide metabolic process	3	1	33.3	26.79	27.42	0.048273795	0.049803173
GO:1901135	carbohydrate derivative metabolic process	65	3	4.6	3.71	3.92	0.049803173	0.049803173
GO:0071704	organic substance metabolic process	65	3	4.6	3.71	3.92	0.049803173	0.049803173
Cellular component:								
GO:0005785	signal recognition particle receptor complex	1	1	100	80.38	82.32	0.02443	0.02443
GO:0005791	rough endoplasmic reticulum	1	1	100	80.38	82.32	0.02443	0.02443
GO:0030867	rough endoplasmic reticulum membrane	1	1	100	80.38	82.32	0.02443	0.02443
Molecular function:								
GO:0004360	glutamine-fructose-6-phosphate transaminase (isomerizing) activity	1	1	100	80.38	82.32	0.024428	0.048273795

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
GO:0005047	signal recognition particle binding	1	1	100	80.38	82.32	0.024428	0.048273795
GO:0004156	dihydropteroate synthase activity	1	1	100	80.38	82.32	0.024428	0.048273795
GO:0070548	L-glutamine aminotransferase activity	1	1	100	80.38	82.32	0.024428	0.048273795
GO:0004830	tryptophan-tRNA ligase activity	1	1	100	80.38	82.32	0.024428	0.048273795
GO:0003848	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase activity	1	1	100	80.38	82.32	0.024428	0.048273795
GO:0004612	phosphoenolpyruvate carboxykinase (ATP) activity	2	1	50	40.19	41.15	0.036423	0.048273795
GO:0008483	transaminase activity	2	1	50	40.19	41.15	0.036423	0.048273795
GO:0004611	phosphoenolpyruvate carboxykinase activity	2	1	50	40.19	41.15	0.036423	0.048273795
GO:0004367	glycerol-3-phosphate dehydrogenase [NAD+] activity	2	1	50	40.19	41.15	0.036423	0.048273795
GO:0016769	transferase activity, transferring nitrogenous groups	2	1	50	40.19	41.15	0.036423	0.048273795
GO:0003824	catalytic activity	1365	23	1.7	1.35	1.78	0.043708	0.048273795
GO:0048037	cofactor binding	64	3	4.7	3.77	3.98	0.048007	0.048273795
GO:0004784	superoxide dismutase activity	3	1	33.3	26.79	27.42	0.048274	0.048273795
GO:0042802	identical protein binding	3	1	33.3	26.79	27.42	0.048274	0.048273795
GO:0016721	oxidoreductase activity, acting on superoxide radicals as acceptor	3	1	33.3	26.79	27.42	0.048274	0.048273795
GO:0042803	protein homodimerization activity	3	1	33.3	26.79	27.42	0.048274	0.048273795
GO:0043021	ribonucleoprotein complex binding	3	1	33.3	26.79	27.42	0.048274	0.048273795

Appendix Table XV: Gene Ontology enrichment: Biological Process, Cellular Component and Molecular Function of 21 decreased abundance proteins in Nc Spain 1H at 12-hours post infection obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0008643	carbohydrate transport	5	1	20	56.27	61.29	0.02108	0.04855
GO:0015693	magnesium ion transport	7	1	14.3	40.19	43.75	0.02802	0.04855
GO:0070838	divalent metal ion transport	8	1	12.5	35.17	38.27	0.03147	0.04855
GO:0072511	divalent inorganic cation transport	8	1	12.5	35.17	38.27	0.03147	0.04855
GO:0071702	organic substance transport	10	1	10	28.13	30.6	0.03833	0.04855
GO:0006633	fatty acid biosynthetic process	12	1	8.3	23.44	25.48	0.04516	0.04855
GO:0071824	protein-DNA complex subunit organization	13	1	7.7	21.64	23.52	0.04855	0.04855
GO:0031497	chromatin assembly	13	1	7.7	21.64	23.52	0.04855	0.04855
GO:0034728	nucleosome organization	13	1	7.7	21.64	23.52	0.04855	0.04855
GO:0065004	protein-DNA complex assembly	13	1	7.7	21.64	23.52	0.04855	0.04855
GO:0006334	nucleosome assembly	13	1	7.7	21.64	23.52	0.04855	0.04855
GO:0006323	DNA packaging	13	1	7.7	21.64	23.52	0.04855	0.04855
Cellular Component:								
GO:0000785	chromatin	10	1	10	28.13	30.6	0.03833	0.04516
GO:0000786	nucleosome	10	1	10	28.13	30.6	0.03833	0.04516
GO:0032993	protein-DNA complex	10	1	10	28.13	30.6	0.03833	0.04516
GO:0000139	Golgi membrane	11	1	9.1	25.58	27.81	0.04175	0.04516
GO:0044427	chromosomal part	12	1	8.3	23.44	25.48	0.04516	0.04516
Molecular Function:								
GO:0008324	cation transmembrane transporter activity	50	2	4	11.25	13.3	0.01382	0.041953676
GO:0005351	sugar:hydrogen symporter activity	4	1	25	70.33	76.64	0.01759	0.041953676
GO:0051119	sugar transmembrane transporter activity	4	1	25	70.33	76.64	0.01759	0.041953676

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
GO:0015295	solute:hydrogen symporter activity	4	1	25	70.33	76.64	0.01759	0.041953676
GO:0005402	cation:sugar symporter activity	4	1	25	70.33	76.64	0.01759	0.041953676
GO:0015144	carbohydrate transmembrane transporter activity	5	1	20	56.27	61.29	0.02108	0.041953676
GO:0015075	ion transmembrane transporter activity	64	2	3.1	8.79	10.35	0.02175	0.041953676
GO:0022891	substrate-specific transmembrane transporter activity	71	2	2.8	7.92	9.31	0.02629	0.041953676
GO:0015095	magnesium ion transmembrane transporter activity	7	1	14.3	40.19	43.75	0.02802	0.041953676
GO:0022857	transmembrane transporter activity	76	2	2.6	7.4	8.68	0.02974	0.041953676
GO:0072509	divalent inorganic cation transmembrane transporter activity	8	1	12.5	35.17	38.27	0.03147	0.041953676
GO:0015293	symporter activity	9	1	11.1	31.26	34.01	0.03491	0.041953676
GO:0015294	solute:cation symporter activity	9	1	11.1	31.26	34.01	0.03491	0.041953676
GO:0046982	protein heterodimerization activity	10	1	10	28.13	30.6	0.03833	0.041953676
GO:0016836	hydro-lyase activity	10	1	10	28.13	30.6	0.03833	0.041953676
GO:0016835	carbon-oxygen lyase activity	11	1	9.1	25.58	27.81	0.04175	0.041953676
GO:0022892	substrate-specific transporter activity	92	2	2.2	6.12	7.14	0.04195	0.041953676

Appendix Table XVI: Gene Ontology enrichment: Biological Process, Cellular Component and Molecular Function of 13 preferentially expressed proteins in Nc Spain 1H at 36 hours post infection obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0006508	proteolysis	133	4	3	12.69	24.38	0.00016	0.00063
GO:0016051	carbohydrate biosynthetic process	10	1	10	42.2	48.09	0.02574	0.04425
GO:0019538	protein metabolic process	607	4	0.7	2.78	4.56	0.03996	0.04425
GO:0008152	metabolic process	1317	6	0.5	1.92	4.69	0.04425	0.04425
Molecular Function:								
GO:0070011	peptidase activity, acting on L-amino acid peptides	99	4	4	17.05	33.1	5.16E-05	0.00047
GO:0008233	peptidase activity	106	4	3.8	15.92	30.85	6.69E-05	0.00047
GO:0017171	serine hydrolase activity	24	2	8.3	35.17	46.56	0.00154523	0.00541
GO:0008236	serine-type peptidase activity	24	2	8.3	35.17	46.56	0.00154523	0.00541
GO:0016787	hydrolase activity	537	5	0.9	3.93	8.81	0.00380201	0.00945
GO:0004360	glutamine-fructose-6-phosphate transaminase (isomerizing) activity	1	1	100	422	482.14	0.00472324	0.00945
GO:0070548	L-glutamine aminotransferase activity	1	1	100	422	482.14	0.00472324	0.00945
GO:0016769	transferase activity, transferring nitrogenous groups	2	1	50	211	241	0.00707753	0.01101
GO:0008483	transaminase activity	2	1	50	211	241	0.00707753	0.01101
GO:0004190	aspartic-type endopeptidase activity	9	1	11.1	46.89	53.44	0.02342158	0.02981
GO:0008234	cysteine-type peptidase activity	15	1	6.7	28.13	32.01	0.0372429	0.03953
GO:0004527	exonuclease activity	16	1	6.3	26.37	30	0.03952973	0.03953
GO:0030246	carbohydrate binding	16	1	6.3	26.37	30	0.03952973	0.03953

Appendix Table XVII: Gene Ontology enrichment: Biological Process, Cellular Component and Molecular Function of 44 preferentially expressed proteins in Nc Spain 1H at 56 hours post infection obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0006508	proteolysis	133	9	6.8	8.79	12.91	5.09E-07	3.01E-05
GO:1901136	carbohydrate derivative catabolic process	30	3	10	12.98	14.55	0.001857789	0.028737471
GO:0019538	protein metabolic process	607	11	1.8	2.35	3.35	0.003562624	0.028737471
GO:0008150	biological_process	1771	20	1.1	1.47	3.02	0.009604511	0.028737471
GO:0007599	hemostasis	18	2	11.1	14.43	15.55	0.009807769	0.028737471
GO:0007596	blood coagulation	18	2	11.1	14.43	15.55	0.009807769	0.028737471
GO:0042060	wound healing	18	2	11.1	14.43	15.55	0.009807769	0.028737471
GO:0050878	regulation of body fluid levels	18	2	11.1	14.43	15.55	0.009807769	0.028737471
GO:0009611	response to wounding	18	2	11.1	14.43	15.55	0.009807769	0.028737471
GO:0050817	coagulation	18	2	11.1	14.43	15.55	0.009807769	0.028737471
GO:0032501	multicellular organismal process	19	2	10.5	13.67	14.72	0.010789502	0.028737471
GO:0043170	macromolecule metabolic process	823	12	1.5	1.89	2.66	0.01303924	0.028737471
GO:1901135	carbohydrate derivative metabolic process	65	3	4.6	5.99	6.64	0.014295634	0.028737471
GO:0071704	organic substance metabolic process	65	3	4.6	5.99	6.64	0.014295634	0.028737471
GO:0006032	chitin catabolic process	1	1	100	129.85	135	0.015228948	0.028737471
GO:0006026	aminoglycan catabolic process	1	1	100	129.85	135	0.015228948	0.028737471
GO:0046348	amino sugar catabolic process	1	1	100	129.85	135	0.015228948	0.028737471
GO:0000272	polysaccharide catabolic process	1	1	100	129.85	135	0.015228948	0.028737471
GO:1901072	glucosamine-containing compound catabolic process	1	1	100	129.85	135	0.015228948	0.028737471
GO:0046130	purine ribonucleoside catabolic process	24	2	8.3	10.82	11.64	0.016311915	0.028737471
GO:0006152	purine nucleoside catabolic process	24	2	8.3	10.82	11.64	0.016311915	0.028737471

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
GO:1901069	guanosine-containing compound catabolic process	24	2	8.3	10.82	11.64	0.016311915	0.028737471
GO:0006184	GTP catabolic process	24	2	8.3	10.82	11.64	0.016311915	0.028737471
GO:0044238	primary metabolic process	1075	14	1.3	1.69	2.5	0.016876802	0.028737471
GO:0008152	metabolic process	1317	16	1.2	1.58	2.5	0.017163759	0.028737471
GO:0050896	response to stimulus	130	4	3.1	4	4.54	0.017534593	0.028737471
GO:0009154	purine ribonucleotide catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0006195	purine nucleotide catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0072523	purine-containing compound catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0009164	nucleoside catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0009143	nucleoside triphosphate catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0042454	ribonucleoside catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0009207	purine ribonucleoside triphosphate catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0009261	ribonucleotide catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0009203	ribonucleoside triphosphate catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0009146	purine nucleoside triphosphate catabolic process	25	2	8	10.39	11.17	0.017534728	0.028737471
GO:0046700	heterocycle catabolic process	27	2	7.4	9.62	10.34	0.020093753	0.031198196
GO:0046039	GTP metabolic process	27	2	7.4	9.62	10.34	0.020093753	0.031198196
GO:0009166	nucleotide catabolic process	28	2	7.1	9.27	9.96	0.021428761	0.032417869
GO:0044270	cellular nitrogen compound catabolic process	29	2	6.9	8.95	9.62	0.022799974	0.032809719
GO:0034655	nucleobase-containing compound catabolic process	29	2	6.9	8.95	9.62	0.022799974	0.032809719
GO:1901068	guanosine-containing compound metabolic process	31	2	6.5	8.38	8.99	0.025648679	0.036030287
GO:0044036	cell wall macromolecule metabolic	3	1	33.3	43.28	44.97	0.030234764	0.037471116

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
	process							
GO:0071554	cell wall organization or biogenesis	3	1	33.3	43.28	44.97	0.030234764	0.037471116
GO:0016998	cell wall macromolecule catabolic process	3	1	33.3	43.28	44.97	0.030234764	0.037471116
GO:0046128	purine ribonucleoside metabolic process	35	2	5.7	7.42	7.95	0.031755183	0.037471116
GO:0009144	purine nucleoside triphosphate metabolic process	35	2	5.7	7.42	7.95	0.031755183	0.037471116
GO:0009199	ribonucleoside triphosphate metabolic process	35	2	5.7	7.42	7.95	0.031755183	0.037471116
GO:0042278	purine nucleoside metabolic process	35	2	5.7	7.42	7.95	0.031755183	0.037471116
GO:0009205	purine ribonucleoside triphosphate metabolic process	35	2	5.7	7.42	7.95	0.031755183	0.037471116
GO:0009141	nucleoside triphosphate metabolic process	36	2	5.6	7.21	7.73	0.033363771	0.038597303
GO:0009119	ribonucleoside metabolic process	37	2	5.4	7.02	7.52	0.035004048	0.039716131
GO:0065007	biological regulation	165	4	2.4	3.15	3.54	0.037344695	0.040393524
GO:0051704	multi-organism process	4	1	25	32.46	33.72	0.03765498	0.040393524
GO:0009405	pathogenesis	4	1	25	32.46	33.72	0.03765498	0.040393524
GO:0009150	purine ribonucleotide metabolic process	40	2	5	6.49	6.95	0.040109664	0.042258396
GO:0009259	ribonucleotide metabolic process	41	2	4.9	6.33	6.78	0.041871379	0.04334055
GO:0009116	nucleoside metabolic process	42	2	4.8	6.18	6.62	0.043662151	0.044414947
GO:0006163	purine nucleotide metabolic process	43	2	4.7	6.04	6.46	0.045481473	0.045481473
Cellular Component:								
GO:0005785	signal recognition particle receptor complex	1	1	100	129.85	135	0.01523	0.01753
GO:0005791	rough endoplasmic reticulum	1	1	100	129.85	135	0.01523	0.01753
GO:0030867	rough endoplasmic reticulum membrane	1	1	100	129.85	135	0.01523	0.01753
GO:0005576	extracellular region	25	2	8	10.39	11.17	0.01753	0.01753

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Molecular Function:								
GO:0070011	peptidase activity, acting on L-amino acid peptides	99	7	7.1	9.18	12.2	9.44E-06	0.00012
GO:0008233	peptidase activity	106	7	6.6	8.57	11.37	1.44E-05	0.00012
GO:0016787	hydrolase activity	537	12	2.2	2.9	4.53	0.000301538	0.00171
GO:0017171	serine hydrolase activity	24	3	12.5	16.23	18.22	0.001026753	0.00349
GO:0008236	serine-type peptidase activity	24	3	12.5	16.23	18.22	0.001026753	0.00349
GO:0008234	cysteine-type peptidase activity	15	2	13.3	17.31	18.67	0.007119735	0.02017
GO:0004568	chitinase activity	1	1	100	129.85	135	0.015228948	0.02834
GO:0004719	protein-L-isoaspartate (D-aspartate) O-methyltransferase activity	1	1	100	129.85	135	0.015228948	0.02834
GO:0005047	signal recognition particle binding	1	1	100	129.85	135	0.015228948	0.02834
GO:0004175	endopeptidase activity	69	3	4.3	5.65	6.25	0.016672522	0.02834
GO:0051998	protein carboxyl O-methyltransferase activity	2	1	50	64.92	67.48	0.02275955	0.03224
GO:0010340	carboxyl-O-methyltransferase activity	2	1	50	64.92	67.48	0.02275955	0.03224
GO:0008171	O-methyltransferase activity	3	1	33.3	43.28	44.97	0.030234764	0.03671
GO:0043021	ribonucleoprotein complex binding	3	1	33.3	43.28	44.97	0.030234764	0.03671
GO:0008276	protein methyltransferase activity	5	1	20	25.97	26.97	0.045020588	0.04502
GO:0016798	hydrolase activity, acting on glycosyl bonds	5	1	20	25.97	26.97	0.045020588	0.04502
GO:0004553	hydrolase activity, hydrolyzing O-glycosyl compounds	5	1	20	25.97	26.97	0.045020588	0.04502

Appendix Table XVIII: Gene Ontology enrichment: Biological Process, Cellular Component and Molecular Function of 18 decreased abundant proteins in Nc Spain 1H at 56 hours post infection obtained from ToxoDB website

Term ID	Description	Bgd count	Result count	Pct of bgd	Fold enrichment	Odds ratio	P-value	Benjamini
Biological Process:								
GO:0016567	protein ubiquitination	3	1	33.3	102.3	112.43	0.01293	0.0396
GO:0032446	protein modification by small protein conjugation	4	1	25	76.73	84.3	0.01614	0.0396
GO:0070647	protein modification by small protein conjugation or removal	4	1	25	76.73	84.3	0.01614	0.0396
GO:0036211	protein modification process	220	3	1.4	4.19	5.38	0.03132	0.0396
GO:0006464	cellular protein modification process	220	3	1.4	4.19	5.38	0.03132	0.0396
GO:0051258	protein polymerization	9	1	11.1	34.1	37.41	0.03205	0.0396
GO:0044248	cellular catabolic process	89	2	2.2	6.9	8.21	0.03354	0.0396
GO:0043623	cellular protein complex assembly	10	1	10	30.69	33.66	0.0352	0.0396
GO:0043412	macromolecule modification	256	3	1.2	3.6	4.57	0.04604	0.04604
Cellular Components:								
GO:0005869	dynactin complex	1	1	100	306.91	337.5	0.00649	0.01902
GO:0043234	protein complex	156	3	1.9	5.9	7.74	0.01268	0.01902
GO:0000151	ubiquitin ligase complex	9	1	11.1	34.1	37.41	0.03205	0.03205
Molecular Functions:								
GO:0034450	ubiquitin-ubiquitin ligase activity	1	1	100	306.91	337.5	0.00649	0.019457496
GO:0004842	ubiquitin-protein ligase activity	6	1	16.7	51.15	56.17	0.02253	0.02253345
GO:0019787	small conjugating protein ligase activity	6	1	16.7	51.15	56.17	0.02253	0.02253345

Reference

. WHO. World Malaria Report 2015. Geneva: World Health Organization; 2015. http://www.who.int/malaria/publications/world-malaria-report-2015/report/en.

Achbarou, A., Mercereau-Puijalon, O., Sadak, A., Fortier, B., Leriche, M.A., Camus, D., Dubremetz, J.F., 1991. Differential targeting of dense granule proteins in the parasitophorous vacuole of *Toxoplasma gondii*. Parasitology 103 Pt 3, 321-329.

Al-Qassab, S.E., Reichel, M.P., Ellis, J.T., 2010. On the Biological and Genetic Diversity in *Neospora caninum*. Diversity 2, 411.

Alexa, A., Rahnenfuhrer, J., 2016. topGO: Enrichment Analysis for Gene Ontology. R package version 2.24.0.

Alexander, D.L., Mital, J., Ward, G.E., Bradley, P., Boothroyd, J.C., 2005. Identification of the moving junction complex of *Toxoplasma gondii*: a collaboration between distinct secretory organelles. PLoS Pathog 1, e17.

Almelli, T., Nuel, G., Bischoff, E., Aubouy, A., Elati, M., Wang, C.W., Dillies, M.A., Coppee, J.Y., Ayissi, G.N., Basco, L.K., Rogier, C., Ndam, N.T., Deloron, P., Tahar, R., 2014. Differences in gene transcriptomic pattern of Plasmodium falciparum in children with cerebral malaria and asymptomatic carriers. PloS one 9, e114401.

America, A.H., Cordewener, J.H., 2008. Comparative LC-MS: a landscape of peaks and valleys. Proteomics 8, 731-749.

Anderson, N.L., Anderson, N.G., 1998. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis 19, 1853-1861.

Armean, I.M., Lilley, K.S., Trotter, M.W.B., Pilkington, N.C.V., Holden, S.B., 2018. Co-complex protein membership evaluation using Maximum Entropy on GO ontology and InterPro annotation. Bioinformatics (Oxford, England).

Arranz-Solis, D., Aguado-Martinez, A., Muller, J., Regidor-Cerrillo, J., Ortega-Mora, L.M., Hemphill, A., 2015. Dose-dependent effects of experimental infection with the virulent *Neospora caninum* Nc-Spain7 isolate in a pregnant mouse model. Veterinary parasitology 211, 133-140.

Atkinson, R., Harper, P.A., Ryce, C., Morrison, D.A., Ellis, J.T., 1999. Comparison of the biological characteristics of two isolates of *Neospora caninum*. Parasitology 118 (Pt 4), 363-370.

Atkinson, R.A., Ryce, C., Miller, C.M., Balu, S., Harper, P.A., Ellis, J.T., 2001. Isolation of *Neospora caninum* genes detected during a chronic murine infection. International journal for parasitology 31, 67-71.

Bannister, L.H., Hopkins, J.M., Dluzewski, A.R., Margos, G., Williams, I.T., Blackman, M.J., Kocken, C.H., Thomas, A.W., Mitchell, G.H., 2003. Plasmodium falciparum apical membrane antigen 1 (PfAMA-1) is translocated within micronemes along subpellicular microtubules during merozoite development. Journal of Cell Science 116, 3825.

Bantscheff, M., Schirle, M., Sweetman, G., Rick, J., Kuster, B., 2007. Quantitative mass spectrometry in proteomics: a critical review. Anal Bioanal Chem 389, 1017-1031.

Barbazuk, W.B., Emrich, S.J., Chen, H.D., Li, L., Schnable, P.S., 2007. SNP discovery via 454 transcriptome sequencing. Plant J 51, 910-918.
Barber, J., Trees, A.J., Owen, M., Tennant, B., 1993. Isolation of *Neospora caninum* from a British dog. Vet Rec 133, 531-532.
Barber, J.S., Holmdahl, O.J., Owen, M.R., Guy, F., Uggla, A., Trees, A.J., 1995. Characterization of the first European isolate of *Neospora caninum* (Dubey, Carpenter, Speer, Topper and Uggla). Parasitology 111 (Pt 5), 563-568.
Barber, J.S., Trees, A.J., 1998. Naturally occurring vertical transmission of *Neospora caninum* in dogs. International journal for parasitology 28, 57-64.
Barling, K.S., Lunt, D.K., Snowden, K.F., Thompson, J.A., 2001. Association of serologic status for *Neospora caninum* and postweaning feed efficiency in beef steers. J Am Vet Med Assoc 219, 1259-1262.
Barr, B.C., Conrad, P.A., Breitmeyer, R., Sverlow, K., Anderson, M.L., Reynolds, J., Chauvet, A.E., Dubey, J.P., Ardans, A.A., 1993. Congenital *Neospora* infection in calves born from cows that had previously aborted *Neospora* -infected fetuses: four cases (1990-1992). J Am Vet Med Assoc 202, 113-117.
Barr, B.C., Conrad, P.A., Dubey, J.P., Anderson, M.L., 1991. *Neospora* -like encephalomyelitis in a calf: pathology, ultrastructure, and immunoreactivity. J Vet Diagn Invest 3, 39-46.
Barragan, A., Sibley, L.D., 2002. Transepithelial migration of *Toxoplasma gondii* is linked to parasite motility and virulence. The Journal of experimental medicine 195, 1625-1633.
Barrett, A.J., Kirschke, H., 1981. Cathepsin B, Cathepsin H, and cathepsin L. Methods Enzymol 80 Pt C, 535-561.
Bartley, P.M., Wright, S., Sales, J., Chianini, F., Buxton, D., Innes, E.A., 2006. Long-term passage of tachyzoites in tissue culture can attenuate virulence of *Neospora caninum in vivo*. Parasitology 133, 421-432.
Beckers, C.J., Dubremetz, J.F., Mercereau-Puijalon, O., Joiner, K.A., 1994. The *Toxoplasma gondii* rhoptry protein ROP 2 is inserted into the parasitophorous vacuole membrane, surrounding the intracellular parasite, and is exposed to the host cell cytoplasm. J Cell Biol 127, 947-961.
Behnke, M.S., Khan, A., Wootton, J.C., Dubey, J.P., Tang, K., Sibley, L.D., 2011. Virulence differences in *Toxoplasma* mediated by amplification of a family of polymorphic pseudokinases. Proc Natl Acad Sci U S A 108, 9631-9636.
Besteiro, S., Michelin, A., Poncet, J., Dubremetz, J.F., Lebrun, M., 2009. Export of a *Toxoplasma gondii* rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. PLoS Pathog 5, e1000309.
Beyer, M., Mallmann, M.R., Xue, J., Staratschek-Jox, A., Vorholt, D., Krebs, W., Sommer, D., Sander, J., Mertens, C., Nino-Castro, A., Schmidt, S.V., Schultze, J.L., 2012. High-resolution transcriptome of human macrophages. PloS one 7, e45466.
Bjerkas, I., Dubey, J.P., 1991. Evidence that *Neospora caninum* is identical to the *Toxoplasma*-like parasite of Norwegian dogs. Acta Vet Scand 32, 407-410.
Bjerkas, I., Mohn, S.F., Presthus, J., 1984. Unidentified cyst-forming sporozoon causing encephalomyelitis and myositis in dogs. Z Parasitenkd 70, 271-274.
Bjerkas, I., Presthus, J., 1989. The neuropathology in toxoplasmosis-like infection caused by a newly recognized cyst-forming sporozoon in dogs. APMIS 97, 459-468.

Bjorkman, C., Hemphill, A., 1998. Characterization of *Neospora caninum* iscom antigens using monoclonal antibodies. Parasite immunology 20, 73-80.
Bohne, W., Heesemann, J., Gross, U., 1994. Reduced replication of *Toxoplasma gondii* is necessary for induction of bradyzoite-specific antigens: a possible role for nitric oxide in triggering stage conversion. Infection and Immunity 62, 1761-1767.
Bohne, W., Holpert, M., Gross, U., 1999. Stage differentiation of the protozoan parasite *Toxoplasma gondii*. Immunobiology 201, 248-254.
Bondarenko, P.V., Chelius, D., Shaler, T.A., 2002. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. Anal Chem 74, 4741-4749.
Boothroyd, J.C., Hehl, A., Knoll, L.J., Manger, I.D., 1998. The surface of *Toxoplasma*: more and less. International journal for parasitology 28, 3-9.
Boucher, L.E., Bosch, J., 2014. Structure of *Toxoplasma gondii* fructose-1,6-bisphosphate aldolase. Acta Crystallogr F Struct Biol Commun 70, 1186-1192.
Boulton, J.G., Gill, P.A., Cook, R.W., Fraser, G.C., Harper, P.A., Dubey, J.P., 1995. Bovine *Neospora* abortion in north-eastern New South Wales. Aust Vet J 72, 119-120.
Bradley, P.J., Hsieh, C.L., Boothroyd, J.C., 2002. Unprocessed *Toxoplasma* ROP1 is effectively targeted and secreted into the nascent parasitophorous vacuole. Mol Biochem Parasitol 125, 189-193.
Bradley, P.J., Sibley, L.D., 2007. Rhoptries: an arsenal of secreted virulence factors. Current opinion in microbiology 10, 582-587.
Bradley, P.J., Ward, C., Cheng, S.J., Alexander, D.L., Coller, S., Coombs, G.H., Dunn, J.D., Ferguson, D.J., Sanderson, S.J., Wastling, J.M., Boothroyd, J.C., 2005. Proteomic analysis of rhoptry organelles reveals many novel constituents for host-parasite interactions in *Toxoplasma gondii*. J Biol Chem 280, 34245-34258.
Braun, L., Brenier-Pinchart, M.P., Yogavel, M., Curt-Varesano, A., Curt-Bertini, R.L., Hussain, T., Kieffer-Jaquinod, S., Coute, Y., Pelloux, H., Tardieux, I., Sharma, A., Belrhali, H., Bougdour, A., Hakimi, M.A., 2013. A *Toxoplasma* dense granule protein, GRA24, modulates the early immune response to infection by promoting a direct and sustained host p38 MAPK activation. The Journal of experimental medicine 210, 2071-2086.
Brossier, F., Jewett, T.J., Sibley, L.D., Urban, S., 2005. A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by *Toxoplasma*. Proc Natl Acad Sci U S A 102, 4146-4151.
Buguliskis, J.S., Brossier, F., Shuman, J., Sibley, L.D., 2010. Rhomboid 4 (ROM4) affects the processing of surface adhesins and facilitates host cell invasion by *Toxoplasma gondii*. PLoS Pathog 6, e1000858.
Burg, J.L., Perelman, D., Kasper, L.H., Ware, P.L., Boothroyd, J.C., 1988. Molecular analysis of the gene encoding the major surface antigen of *Toxoplasma gondii*. J Immunol 141, 3584-3591.
Buxton, D., McAllister, M.M., Dubey, J.P., 2002. The comparative pathogenesis of neosporosis. Trends in parasitology 18, 546-552.

Carruthers, V., Sibley, L., 1997a. Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. Eur J Cell Biol 73, 114 - 123.

Carruthers, V.B., Sibley, L.D., 1997b. Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. Eur J Cell Biol 73, 114-123.

Carruthers, V.B., Sibley, L.D., 1999. Mobilization of intracellular calcium stimulates microneme discharge in *Toxoplasma gondii*. Mol Microbiol 31, 421-428.

Caspe, S.G., Moore, D.P., Leunda, M.R., Cano, D.B., Lischinsky, L., Regidor-Cerrillo, J., Alvarez-Garcia, G., Echaide, I.G., Bacigalupe, D., Ortega Mora, L.M., Odeon, A.C., Campero, C.M., 2012. The *Neospora caninum*-Spain 7 isolate induces placental damage, fetal death and abortion in cattle when inoculated in early gestation. Veterinary parasitology 189, 171-181.

Cavalcante, G.T., Soares, R.M., Nishi, S.M., Hagen, S.C.F., Vannucchi, C.I., Maiorka, P.C., Paixão, A.S., Gennari, S.M., 2012. Experimental infection with *Neospora caninum* in pregnant bitches. Revista Brasileira de Parasitologia Veterinária 21, 232-236.

Cerede, O., Dubremetz, J., Soete, M., Deslee, D., Vial, H., Bout, D., 2005. Synergistic role of micronemal proteins in *Toxoplasma gondii* virulence. The Journal of experimental medicine 201, 453 - 463.

Cesbron-Delauw, M.F., 1994. Dense-granule organelles of *Toxoplasma gondii*: their role in the host-parasite relationship. Parasitol Today 10, 293-296.

Cesbron-Delauw, M.F., Guy, B., Torpier, G., Pierce, R.J., Lenzen, G., Cesbron, J.Y., Charif, H., Lepage, P., Darcy, F., Lecocq, J.P., et al., 1989. Molecular characterization of a 23-kilodalton major antigen secreted by *Toxoplasma gondii*. Proc Natl Acad Sci U S A 86, 7537-7541.

Chang, S., Shan, X., Li, X., Fan, W., Zhang, S.Q., Zhang, J., Jiang, N., Ma, D., Mao, Z., 2015. *Toxoplasma gondii* Rhoptry Protein ROP16 Mediates Partially SH-SY5Y Cells Apoptosis and Cell Cycle Arrest by Directing Ser15/37 Phosphorylation of p53. International Journal of Biological Sciences 11, 1215-1225.

Chen, E.I., Yates Iii, J.R., 2007. Cancer proteomics by quantitative shotgun proteomics. Molecular Oncology 1, 144-159.

Cheung, W.Y., 1980. Calmodulin plays a pivotal role in cellular regulation. Science (New York, N.Y.) 207, 19-27.

Chryssafidis, A.L., Canton, G., Chianini, F., Innes, E.A., Madureira, E.H., Gennari, S.M., 2014. Pathogenicity of Nc-Bahia and Nc-1 strains of *Neospora caninum* in experimentally infected cows and buffaloes in early pregnancy. Parasitology research 113, 1521-1528.

Chu, Y., Corey, D.R., 2012. RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation. Nucleic Acid Therapeutics 22, 271-274.

Cloonan, N., Forrest, A.R.R., Kolle, G., Gardiner, B.B.A., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., Robertson, A.J., Perkins, A.C., Bruce, S.J., Lee, C.C., Ranade, S.S., Peckham, H.E., Manning, J.M., McKernan, K.J., Grimmond, S.M., 2008. Stem cell transcriptome profiling via massive-scale mRNA sequencing. Nat Meth 5, 613-619.

Collantes-Fernandez, E., Arnaiz-Seco, I., Burgos, B.M., Rodriguez-Bertos, A., Aduriz, G., Fernandez-Garcia, A., Ortega-Mora, L.M., 2006. Comparison of

Neospora caninum distribution, parasite loads and lesions between epidemic and endemic bovine abortion cases. Veterinary parasitology 142, 187-191.
Conrad, P.A., Barr, B.C., Sverlow, K.W., Anderson, M., Daft, B., Kinde, H., Dubey, J.P., Munson, L., Ardans, A., 1993. *In vitro* isolation and characterization of a *Neospora* sp. from aborted bovine foetuses. Parasitology 106 (Pt 3), 239-249.
Conseil, V., Soête, M., Dubremetz, J.F., 1999. Serine Protease Inhibitors Block Invasion of Host Cells by *Toxoplasma gondii*. Antimicrobial Agents and Chemotherapy 43, 1358-1361.
Coppens, I., Dunn, J.D., Romano, J.D., Pypaert, M., Zhang, H., Boothroyd, J.C., Joiner, K.A., 2006. *Toxoplasma gondii* sequesters lysosomes from mammalian hosts in the vacuolar space. Cell 125, 261-274.
Craig, R., Beavis, R.C., 2004. TANDEM: matching proteins with tandem mass spectra. Bioinformatics (Oxford, England) 20, 1466-1467.
Cravatt, B.F., Simon, G.M., Yates Iii, J.R., 2007. The biological impact of mass-spectrometry-based proteomics. Nature 450, 991-1000.
Cutillas, P.R., Vanhaesebroeck, B., 2007. Quantitative profile of five murine core proteomes using label-free functional proteomics. Mol Cell Proteomics 6, 1560-1573.
Dahl, E.L., Rosenthal, P.J., 2005. Biosynthesis, localization, and processing of falcipain cysteine proteases of Plasmodium falciparum. Molecular and biochemical parasitology 139, 205-212.
Dahl, S.W., Halkier, T., Lauritzen, C., Dolenc, I., Pedersen, J., Turk, V., Turk, B., 2001. Human recombinant pro-dipeptidyl peptidase I (cathepsin C) can be activated by cathepsins L and S but not by autocatalytic processing. Biochemistry 40, 1671-1678.
Dardé, M.L., Villena, I., Pinon, J.M., Beguinot, I., 1998. Severe Toxoplasmosis Caused by a *Toxoplasma gondii* Strain with a New Isoenzyme Type Acquired in French Guyana. Journal of Clinical Microbiology 36, 324-324.
Davison, H.C., Otter, A., Trees, A.J., 1999. Estimation of vertical and horizontal transmission parameters of *Neospora caninum* infections in dairy cattle. International journal for parasitology 29, 1683-1689.
Delorme, V., Cayla, X., Faure, G., Garcia, A., Tardieux, I., 2003. Actin Dynamics Is Controlled by a Casein Kinase II and Phosphatase 2C Interplay on *Toxoplasma gondii* Toxofilin. Molecular Biology of the Cell 14, 1900-1912.
Denkers, E.Y., Bzik, D.J., Fox, B.A., Butcher, B.A., 2012. An inside job: hacking into Janus kinase/signal transducer and activator of transcription signaling cascades by the intracellular protozoan *Toxoplasma gondii*. Infection and Immunity 80, 476-482.
Dobrowolski, J.M., Sibley, L.D., 1996. *Toxoplasma* invasion of mammalian cells is powered by the actin cytoskeleton of the parasite. Cell 84, 933-939.
Domon, B., Aebersold, R., 2006. Mass Spectrometry and Protein Analysis. Science (New York, N.Y.) 312, 212.
Donahoe, S.L., Lindsay, S.A., Krockenberger, M., Phalen, D., Slapeta, J., 2015. A review of neosporosis and pathologic findings of *Neospora caninum* infection in wildlife. Int J Parasitol Parasites Wildl 4, 216-238.
Dou, Z., Carruthers, V.B., 2011. Cathepsin proteases in *Toxoplasma gondii*. Adv Exp Med Biol 712, 49-61.

Dubey, J.P., 1999a. Neosporosis in cattle: biology and economic impact. J Am Vet Med Assoc 214, 1160-1163.
Dubey, J.P., 1999b. Recent advances in *Neospora* and neosporosis. Veterinary parasitology 84, 349-367.
Dubey, J.P., 2003. Review of *Neospora caninum* and neosporosis in animals. Korean J Parasitol 41, 1-16.
Dubey, J.P., Barr, B.C., Barta, J.R., Bjerkas, I., Bjorkman, C., Blagburn, B.L., Bowman, D.D., Buxton, D., Ellis, J.T., Gottstein, B., Hemphill, A., Hill, D.E., Howe, D.K., Jenkins, M.C., Kobayashi, Y., Koudela, B., Marsh, A.E., Mattsson, J.G., McAllister, M.M., Modry, D., Omata, Y., Sibley, L.D., Speer, C.A., Trees, A.J., Uggla, A., Upton, S.J., Williams, D.J., Lindsay, D.S., 2002. Redescription of *Neospora caninum* and its differentiation from related coccidia. International journal for parasitology 32, 929-946.
Dubey, J.P., Buxton, D., Wouda, W., 2006. Pathogenesis of bovine neosporosis. J Comp Pathol 134, 267-289.
Dubey, J.P., Carpenter, J.L., Speer, C.A., Topper, M.J., Uggla, A., 1988a. Newly recognized fatal protozoan disease of dogs. J Am Vet Med Assoc 192, 1269-1285.
Dubey, J.P., Hattel, A.L., Lindsay, D.S., Topper, M.J., 1988b. Neonatal *Neospora caninum* infection in dogs: isolation of the causative agent and experimental transmission. J Am Vet Med Assoc 193, 1259-1263.
Dubey, J.P., Jenkins, M.C., Rajendran, C., Miska, K., Ferreira, L.R., Martins, J., Kwok, O.C., Choudhary, S., 2011. Gray wolf (Canis lupus) is a natural definitive host for *Neospora caninum*. Veterinary parasitology 181, 382-387.
Dubey, J.P., Koestner, A., Piper, R.C., 1990. Repeated transplacental transmission of *Neospora caninum* in dogs. J Am Vet Med Assoc 197, 857-860.
Dubey, J.P., Leathers, C.W., Lindsay, D.S., 1989. *Neospora caninum*-like protozoon associated with fatal myelitis in newborn calves. The Journal of parasitology 75, 146-148.
Dubey, J.P., Lindsay, D.S., 1996. A review of *Neospora caninum* and neosporosis. Veterinary parasitology 67, 1-59.
Dubey, J.P., Lindsay, D.S., Speer, C.A., 1998. Structures of *Toxoplasma gondii* Tachyzoites, Bradyzoites, and Sporozoites and Biology and Development of Tissue Cysts. Clinical microbiology reviews 11, 267-299.
Dubey, J.P., Schares, G., 2011. Neosporosis in animals--the last five years. Vet Parasitol 180, 90-108.
Dubey, J.P., Schares, G., Ortega-Mora, L.M., 2007. Epidemiology and control of neosporosis and *Neospora caninum*. Clinical microbiology reviews 20.
Dubremetz, J.F., Achbarou, A., Bermudes, D., Joiner, K.A., 1993. Kinetics and pattern of organelle exocytosis during *Toxoplasma gondii*/host-cell interaction. Parasitology research 79, 402-408.
Dubremetz, J.F., Lebrun, M., 2012. Virulence factors of *Toxoplasma gondii*. Microbes Infect 14, 1403-1410.
Dudekula, K., Le Bihan, T., 2016. Data from quantitative label free proteomics analysis of rat spleen. Data in Brief 8, 494-500.
Durand, P., Golinelli-Pimpaneau, B., Mouilleron, S., Badet, B., Badet-Denisot, M.-A., 2008. Highlights of glucosamine-6P synthase catalysis. Archives of Biochemistry and Biophysics 474, 302-317.

Dzierszinski, F., Mortuaire, M., Cesbron-Delauw, M.F., Tomavo, S., 2000. Targeted disruption of the glycosylphosphatidylinositol-anchored surface antigen SAG3 gene in *Toxoplasma gondii* decreases host cell adhesion and drastically reduces virulence in mice. Mol Microbiol 37, 574-582.

Eastick, F.A., Elsheikha, H.M., 2010. Stress-driven stage transformation of *Neospora caninum*. Parasitology research 106, 1009-1014.

Echeverria, P.C., Matrajt, M., Harb, O.S., Zappia, M.P., Costas, M.A., Roos, D.S., Dubremetz, J.F., Angel, S.O., 2005. *Toxoplasma gondii* Hsp90 is a potential drug target whose expression and subcellular localization are developmentally regulated. J Mol Biol 350, 723-734.

El Hajj, H., Lebrun, M., Arold, S.T., Vial, H., Labesse, G., Dubremetz, J.F., 2007. ROP18 is a rhoptry kinase controlling the intracellular proliferation of *Toxoplasma gondii*. PLoS Pathog 3, e14.

Ellis, J.T., Ryce, C., Atkinson, R., Balu, S., Jones, P., Harper, P.A., 2000. Isolation, characterization and expression of a GRA2 homologue from *Neospora caninum*. Parasitology 120 (Pt 4), 383-390.

Emrich, S.J., Barbazuk, W.B., Li, L., Schnable, P.S., 2007. Gene discovery and annotation using LCM-454 transcriptome sequencing. Genome Res 17, 69-73.

Eng, J.K., McCormack, A.L., Yates, J.R., 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry 5, 976-989.

Etheridge, R.D., Alaganan, A., Tang, K., Lou, H.J., Turk, B.E., Sibley, L.D., 2014. The *Toxoplasma* pseudokinase ROP5 forms complexes with ROP18 and ROP17 kinases that synergize to control acute virulence in mice. Cell host & microbe 15, 537-550.

Fernandez-Garcia, A., Risco-Castillo, V., Zaballos, A., Alvarez-Garcia, G., Ortega-Mora, L.M., 2006. Identification and molecular cloning of the *Neospora caninum* SAG4 gene specifically expressed at bradyzoite stage. Molecular and biochemical parasitology 146, 89-97.

Ferreira da Silva Mda, F., Barbosa, H.S., Gross, U., Luder, C.G., 2008. Stress-related and spontaneous stage differentiation of *Toxoplasma gondii*. Mol Biosyst 4, 824-834.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic acids research 42, D222-230.

Fox, B.A., Rommereim, L.M., Guevara, R.B., Falla, A., Hortua Triana, M.A., Sun, Y., Bzik, D.J., 2016. The *Toxoplasma gondii* Rhoptry Kinome Is Essential for Chronic Infection. MBio 7.

French, N.P., Clancy, D., Davison, H.C., Trees, A.J., 1999. Mathematical models of *Neospora caninum* infection in dairy cattle: transmission and options for control. International journal for parasitology 29, 1691-1704.

Friedrich, N., Santos, J.M., Liu, Y., Palma, A.S., Leon, E., Saouros, S., Kiso, M., Blackman, M.J., Matthews, S., Feizi, T., Soldati-Favre, D., 2010. Members of a novel protein family containing microneme adhesive repeat domains act as sialic acid-binding lectins during host cell invasion by apicomplexan parasites. J Biol Chem 285, 2064-2076.

Fuchs, N., Sonda, S., Gottstein, B., Hemphill, A., 1998. Differential expression of cell surface- and dense granule-associated *Neospora caninum* proteins in tachyzoites and bradyzoites. J Parasitol 84, 753-758.
Futschik, M.E., Carlisle, B., 2005. Noise-robust soft clustering of gene expression time-course data. J Bioinform Comput Biol 3, 965-988.
Gabella, C., Durinx, C., 2017. Funding knowledgebases: Towards a sustainable funding model for the UniProt use case. 6.
García-Melo, D.P., Regidor-Cerrillo, J., Ortega-Mora, L.M., Collantes-Fernández, E., de Oliveira, V.S.F., de Oliveira, M.A.P., da Silva, A.C., 2009. Isolation and biological characterisation of a new isolate of *Neospora caninum* from an asymptomatic calf in Brazil. Acta Parasit. 54, 180-185.
Garnett, J.A., Liu, Y., Leon, E., Allman, S.A., Friedrich, N., Saouros, S., Curry, S., Soldati-Favre, D., Davis, B.G., Feizi, T., Matthews, S., 2009. Detailed insights from microarray and crystallographic studies into carbohydrate recognition by microneme protein 1 (MIC1) of *Toxoplasma gondii*. Protein Sci 18, 1935-1947.
Garnick, E., 1992. Parasite virulence and parasite-host coevolution: a reappraisal. J Parasitol 78, 381-386.
Gibbons, I.R., 1995. Dynein family of motor proteins: present status and future questions. Cell Motil Cytoskeleton 32, 136-144.
Gibbons, I.R., 1996. The role of dynein in microtubule-based motility. Cell Struct Funct 21, 331-342.
Gilbert, L.A., Ravindran, S., Turetzky, J.M., Boothroyd, J.C., Bradley, P.J., 2007. *Toxoplasma gondii* targets a protein phosphatase 2C to the nuclei of infected host cells. Eukaryot Cell 6, 73-83.
Gomez, C., Esther Ramirez, M., Calixto-Galvez, M., Medel, O., Rodriguez, M.A., 2010. Regulation of gene expression in protozoa parasites. Journal of biomedicine & biotechnology 2010, 726045.
Gondim, L.F., Gao, L., McAllister, M.M., 2002. Improved production of *Neospora caninum* oocysts, cyclical oral transmission between dogs and cattle, and *in vitro* isolation from oocysts. J Parasitol 88, 1159-1163.
Gondim, L.F., McAllister, M.M., Pitt, W.C., Zemlicka, D.E., 2004. Coyotes (Canis latrans) are definitive hosts of *Neospora caninum*. International journal for parasitology 34, 159-161.
Gondim, L.F., Pinheiro, A.M., Santos, P.O., Jesus, E.E., Ribeiro, M.B., Fernandes, H.S., Almeida, M.A., Freire, S.M., Meyer, R., McAllister, M.M., 2001. Isolation of *Neospora caninum* from the brain of a naturally infected dog, and production of encysted bradyzoites in gerbils. Veterinary parasitology 101, 1-7.
Gong, H., Kobayashi, K., Sugi, T., Takemae, H., Ishiwa, A., Recuenco, F.C., Murakoshi, F., Xuan, X., Horimoto, T., Akashi, H., Kato, K., 2014. Characterization and binding analysis of a microneme adhesive repeat domain-containing protein from *Toxoplasma gondii*. Parasitol Int 63, 381-388.
Goodswen, S.J., Kennedy, P.J., Ellis, J.T., 2013. A review of the infection, genetics, and evolution of *Neospora caninum*: from the past to the present. Infect Genet Evol 13, 133-150.

Haddad, J.P.A., Dohoo, I.R., VanLeewen, J.A., 2005. A review of *Neospora caninum* in dairy and beef cattle — a Canadian perspective. The Canadian Veterinary Journal 46, 230-243.

Haider, S., Pal, R., 2013. Integrated Analysis of Transcriptomic and Proteomic Data. Current Genomics 14, 91-110.

Hall, C.A., Reichel, M.P., Ellis, J.T., 2005. *Neospora* abortions in dairy cattle: diagnosis, mode of transmission and control. Veterinary parasitology 128, 231-241.

Harding, C.R., Egarter, S., Gow, M., Jimenez-Ruiz, E., Ferguson, D.J., Meissner, M., 2016. Gliding Associated Proteins Play Essential Roles during the Formation of the Inner Membrane Complex of *Toxoplasma gondii*. PLoS Pathog 12, e1005403.

Harper, J., Zhou, X., Pszenny, V., Kafsack, B., Carruthers, V., 2004. The novel coccidian micronemal protein MIC11 undergoes proteolytic maturation by sequential cleavage to remove an internal propeptide. International journal for parasitology 34, 1047 - 1058.

Hassan, I.A., Wang, S., Xu, L., Yan, R., Song, X., XiangRui, L., 2014. Immunological response and protection of mice immunized with plasmid encoding *Toxoplasma gondii* glycolytic enzyme malate dehydrogenase. Parasite immunology 36, 674-683.

Hemphill, A., 1999. The host-parasite relationship in neosporosis. Adv Parasitol 43, 47-104.

Hemphill, A., Felleisen, R., Connolly, B., Gottstein, B., Hentrich, B., Muller, N., 1997a. Characterization of a cDNA-clone encoding Nc-p43, a major *Neospora caninum* tachyzoite surface protein. Parasitology 115 (Pt 6), 581-590.

Hemphill, A., Fuchs, N., Sonda, S., Gottstein, B., Hentrich, B., 1997b. Identification and partial characterization of a 36 kDa surface protein on *Neospora caninum* tachyzoites. Parasitology 115 (Pt 4), 371-380.

Hemphill, A., Fuchs, N., Sonda, S., Hehl, A., 1999. The antigenic composition of *Neospora caninum*. International journal for parasitology 29, 1175-1188.

Hemphill, A., Gottstein, B., 2006. *Neospora caninum* and neosporosis — recent achievements in host and parasite cell biology and treatment. Acta Parasitologica 51, 15-25.

Hemphill, A., Gottstein, B., Kaufmann, H., 1996. Adhesion and invasion of bovine endothelial cells by *Neospora caninum*. Parasitology 112 (Pt 2), 183-197.

Hemphill, A., Vonlaufen, N., Naguleswaran, A., Keller, N., Riesen, M., Guetg, N., Srinivasan, S., Alaeddine, F., 2004. Tissue culture and explant approaches to studying and visualizing *Neospora caninum* and its interactions with the host cell. Microsc Microanal 10, 602-620.

Hietala, S.K., Thurmond, M.C., 1999. Postnatal *Neospora caninum* transmission and transient serologic responses in two dairies. International journal for parasitology 29, 1669-1676.

Hobson, J.C., Duffield, T.F., Kelton, D., Lissemore, K., Hietala, S.K., Leslie, K.E., McEwen, B., Cramer, G., Peregrine, A.S., 2002. *Neospora caninum* serostatus and milk production of Holstein cattle. J Am Vet Med Assoc 221, 1160-1164.

Hoff, E., Cook, S., Sherman, G., Harper, J., Ferguson, D., Dubremetz, J., 2001. *Toxoplasma gondii*: molecular cloning and characterization of a novel 18-kDa secretory antigen, TgMIC10. Experimental parasitology 97, 77 - 88.

Howe, D.K., Crawford, A.C., Lindsay, D., Sibley, L.D., 1998. The p29 and p35 Immunodominant Antigens of *Neospora caninum* Tachyzoites Are Homologous to the Family of Surface Antigens of *Toxoplasma gondii*. Infection and Immunity 66, 5322-5328.

Howe, D.K., Sibley, L.D., 1999. Comparison of the major antigens of *Neospora caninum* and *Toxoplasma gondii*. International journal for parasitology 29, 1489-1496.

Hruzik, A., Asif, A.R., Groß, U., 2011. Identification of *Toxoplasma gondii* SUB1 Antigen as a Marker for Acute Infection by Use of an Innovative Evaluation Method. Journal of Clinical Microbiology 49, 2419-2425.

Hunter, C.A., Sibley, L.D., 2012. Modulation of innate immunity by *Toxoplasma gondii* virulence effectors. Nat Rev Microbiol 10, 766-778.

Huynh, M.-H., Carruthers, V.B., 2006. *Toxoplasma* MIC2 Is a Major Determinant of Invasion and Virulence. PLoS Pathogens 2, e84.

Huynh, M.-H., Rabenau, K.E., Harper, J.M., Beatty, W.L., Sibley, L., Carruthers, V.B., 2003. Rapid invasion of host cells by *Toxoplasma* requires secretion of the MIC2–M2AP adhesive protein complex. The EMBO Journal 22, 2082-2090.

Hyun, C., Gupta, G.D., Marsh, A.E., 2003. Sequence comparison of Sarcocystis neurona surface antigen from multiple isolates. Veterinary parasitology 112, 11-20.

Jacquet, A., Coulon, L., De Neve, J., Daminet, V., Haumont, M., Garcia, L., Bollen, A., Jurado, M., Biemans, R., 2001. The surface antigen SAG3 mediates the attachment of *Toxoplasma gondii* to cell-surface proteoglycans. Molecular and biochemical parasitology 116, 35-44.

Jardine, J.E., 1996. The ultrastructure of bradyzoites and tissue cysts of *Neospora caninum* in dogs: absence of distinguishing morphological features between parasites of canine and bovine origin. Veterinary parasitology 62, 231-240.

Jensen, K.D.C., Hu, K., Whitmarsh, R.J., Hassan, M.A., Julien, L., Lu, D., Chen, L., Hunter, C.A., Saeij, J.P.J., 2013. *Toxoplasma gondii* Rhoptry 16 Kinase Promotes Host Resistance to Oral Infection and Intestinal Inflammation Only in the Context of the Dense Granule Protein GRA15. Infection and Immunity 81, 2156-2167.

Jensen, Kirk D.C., Wang, Y., Wojno, Elia D.T., Shastri, Anjali J., Hu, K., Cornel, L., Boedec, E., Ong, Y.-C., Chien, Y.-h., Hunter, Christopher A., Boothroyd, John C., Saeij, Jeroen P.J., 2011. *Toxoplasma* Polymorphic Effectors Determine Macrophage Polarization and Intestinal Inflammation. Cell host & microbe 9, 472-483.

Jesus, E.E., Pinheiro, A.M., Santos, A.B., Freire, S.M., Tardy, M.B., El-Bacha, R.S., Costa, S.L., Costa, M.F., 2013. Effects of IFN-gamma, TNF-alpha, IL-10 and TGF-beta on *Neospora caninum* infection in rat glial cells. Experimental parasitology 133, 269-274.

Jewett, T., Sibley, L., 2004. The *Toxoplasma* proteins MIC2 and M2AP form a hexameric complex necessary for intracellular survival. J Biol Chem 279, 9362 - 9369.

Jewett, T.J., Sibley, L.D., 2003. Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. Mol Cell 11, 885-894.

Jordan, A., Caldwell, D.J., Klein, J., Coppedge, J., Pohl, S., Fitz-Coy, S., Lee, J.T., 2011. Eimeria tenella oocyst shedding and output in cecal or fecal contents following experimental challenge in broilers. Poult Sci 90, 990-995.

Jovanovic, M., Rooney, M.S., Mertins, P., Przybylski, D., Chevrier, N., Satija, R., Rodriguez, E.H., Fields, A.P., Schwartz, S., Raychowdhury, R., Mumbach, M.R., Eisenhaure, T., Rabani, M., Gennert, D., Lu, D., Delorey, T., Weissman, J.S., Carr, S.A., Hacohen, N., Regev, A., 2015. Immunogenetics. Dynamic profiling of the protein life cycle in response to pathogens. Science (New York, N.Y.) 347, 1259038.

Jung, C., Lee, C.Y., Grigg, M.E., 2004. The SRS superfamily of *Toxoplasma* surface proteins. International journal for parasitology 34, 285-296.

Juschke, C., Dohnal, I., Pichler, P., Harzer, H., Swart, R., Ammerer, G., Mechtler, K., Knoblich, J.A., 2013. Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. Genome Biol 14, r133.

Kang, S.W., Lee, E.H., Jean, Y.H., Choe, S.E., Van Quyen, D., Lee, M.S., 2008. The differential protein expression profiles and immunogenicity of tachyzoites and bradyzoites of *in vitro* cultured *Neospora caninum*. Parasitology research 103, 905-913.

Keeley, A., Soldati, D., 2004. The glideosome: a molecular machine powering motility and host-cell invasion by Apicomplexa. Trends Cell Biol 14, 528-532.

Keller, N., Naguleswaran, A., Cannas, A., Vonlaufen, N., Bienz, M., Bjorkman, C., Bohne, W., Hemphill, A., 2002. Identification of a *Neospora caninum* microneme protein (NcMIC1) which interacts with sulfated host cell surface glycosaminoglycans. Infection and Immunity 70, 3187-3198.

Kerr, G., Ruskin, H.J., Crane, M., Doolan, P., 2008. Techniques for clustering gene expression data. Comput Biol Med 38, 283-293.

Khan, A., Taylor, S., Ajioka, J.W., Rosenthal, B.M., Sibley, L.D., 2009. Selection at a single locus leads to widespread expansion of *Toxoplasma gondii* lineages that are virulent in mice. PLoS genetics 5, e1000404.

Kieschnick, H., Wakefield, T., Narducci, C.A., Beckers, C., 2001. *Toxoplasma gondii* attachment to host cells is regulated by a calmodulin-like domain protein kinase. J Biol Chem 276, 12369-12377.

Kim, E.W., Nadipuram, S.M., Tetlow, A.L., Barshop, W.D., Liu, P.T., Wohlschlegel, J.A., Bradley, P.J., 2016. <div xmlns="http://www.w3.org/1999/xhtml">The Rhoptry Pseudokinase ROP54 Modulates *Toxoplasma gondii* Virulence and Host GBP2 Loading</div&gt. mSphere 1.

Kim, J.H., Sohn, H.J., Hwang, W.S., Hwang, E.K., Jean, Y.H., Yamane, I., Kim, D.Y., 2000. *In vitro* isolation and characterization of bovine *Neospora caninum* in Korea. Veterinary parasitology 90, 147-154.

King, J.S., Slapeta, J., Jenkins, D.J., Al-Qassab, S.E., Ellis, J.T., Windsor, P.A., 2010. Australian dingoes are definitive hosts of *Neospora caninum*. International journal for parasitology 40, 945-950.

Kondo, T., 2008. Tissue proteomics for cancer biomarker development: laser microdissection and 2D-DIGE. BMB Rep 41, 626-634.
Koyama, T., Kobayashi, Y., Omata, Y., Yamada, M., Furuoka, H., Maeda, R., Matsui, T., Saito, A., Mikami, T., 2001. Isolation of *Neospora caninum* from the brain of a pregnant sheep. J Parasitol 87, 1486-1488.
Krishna, R., Xia, D., Sanderson, S., Shanmugasundram, A., Vermont, S., Bernal, A., Daniel-Naguib, G., Ghali, F., Brunk, B.P., Roos, D.S., Wastling, J.M., Jones, A.R., 2015. A large-scale proteogenomics study of apicomplexan pathogens-*Toxoplasma gondii* and *Neospora caninum*. Proteomics 15, 2618-2628.
Labruyere, E., Lingnau, M., Mercier, C., Sibley, L.D., 1999. Differential membrane targeting of the secretory proteins GRA4 and GRA6 within the parasitophorous vacuole formed by *Toxoplasma gondii*. Molecular and biochemical parasitology 102, 311-324.
Lagal, V., Binder, E.M., Huynh, M.H., Kafsack, B.F., Harris, P.K., Diez, R., Chen, D., Cole, R.N., Carruthers, V.B., Kim, K., 2010. *Toxoplasma gondii* protease TgSUB1 is required for cell surface processing of micronemal adhesive complexes and efficient adhesion of tachyzoites. Cell Microbiol 12, 1792-1808.
Larson, R.L., Hardin, D.K., Pierce, V.L., 2004. Economic considerations for diagnostic and control options for *Neospora caninum*-induced abortions in endemically infected herds of beef cattle. J Am Vet Med Assoc 224, 1597-1604.
Lauron, E.J., Oakgrove, K.S., Tell, L.A., Biskar, K., Roy, S.W., Sehgal, R.N., 2014. Transcriptome sequencing and analysis of Plasmodium gallinaceum reveals polymorphisms and selection on the apical membrane antigen-1. Malaria Journal 13, 1-14.
Lee, E.G., Kim, J.H., Shin, Y.S., Shin, G.W., Kim, Y.R., Palaksha, K.J., Kim, D.Y., Yamane, I., Kim, Y.H., Kim, G.S., Suh, M.D., Jung, T.S., 2005. Application of proteomics for comparison of proteome of *Neospora caninum* and *Toxoplasma gondii* tachyzoites. J Chromatogr B Analyt Technol Biomed Life Sci 815, 305-314.
Lee, W.K., Ahn, H.J., Yu, Y.G., Nam, H.W., 2014. Rhoptry protein 6 from *Toxoplasma gondii* is an intrinsically disordered protein. Protein Expr Purif 101, 146-151.
Lekutis, C., Ferguson, D.J., Boothroyd, J.C., 2000. *Toxoplasma gondii*: identification of a developmentally regulated family of genes related to SAG2. Experimental parasitology 96, 89-96.
Lekutis, C., Ferguson, D.J., Grigg, M.E., Camps, M., Boothroyd, J.C., 2001. Surface antigens of *Toxoplasma gondii*: variations on a theme. International journal for parasitology 31, 1285-1292.
Li, J., Witten, D.M., Johnstone, I.M., Tibshirani, R., 2012. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. Biostatistics (Oxford, England) 13, 523-538.
Liesenfeld, O., 2002. Oral infection of C57BL/6 mice with *Toxoplasma gondii*: a new model of inflammatory bowel disease? The Journal of infectious diseases 185 Suppl 1, S96-101.
Lindsay, D.S., Dubey, J.P., Duncan, R.B., 1999a. Confirmation that the dog is a definitive host for *Neospora caninum*. Veterinary parasitology 82, 327-333.

Lindsay, D.S., Little, S.E., Davidson, W.R., 2002. Prevalence of antibodies to *Neospora caninum* in white-tailed deer, Odocoileus virginianus, from the southeastern United States. J Parasitol 88, 415-417.
Lindsay, D.S., Speer, C.A., Toivio-Kinnucan, M.A., Dubey, J.P., Blagburn, B.L., 1993. Use of infected cultured cells to compare ultrastructural features of *Neospora caninum* from dogs and *Toxoplasma gondii*. Am J Vet Res 54, 103-106.
Lindsay, D.S., Upton, S.J., Dubey, J.P., 1999b. A structural study of the *Neospora caninum* oocyst. International journal for parasitology 29, 1521-1523.
Liu, H., Sadygov, R.G., Yates, J.R., 3rd, 2004. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 76, 4193-4201.
Liu, Z., Yuan, F., Yang, Y., Yin, L., Liu, Y., Wang, Y., Zheng, K., Cao, J., 2016. Partial protective immunity against toxoplasmosis in mice elicited by recombinant *Toxoplasma gondii* malate dehydrogenase. Vaccine 34, 989-994.
Luan, S., 2003. Protein phosphatases in plants. Annu Rev Plant Biol 54, 63-92.
Lyons, R.E., McLeod, R., Roberts, C.W., 2002. *Toxoplasma gondii* tachyzoite-bradyzoite interconversion. Trends in parasitology 18, 198-201.
Macaldowie, C., Maley, S.W., Wright, S., Bartley, P., Esteban-Redondo, I., Buxton, D., Innes, E.A., 2004. Placental pathology associated with fetal death in cattle inoculated with *Neospora caninum* by two different routes in early pregnancy. J Comp Pathol 131, 142-156.
Maier, T., Guell, M., Serrano, L., 2009. Correlation of mRNA and protein in complex biological samples. FEBS Lett 583, 3966-3973.
Manger, I.D., Hehl, A.B., Boothroyd, J.C., 1998. The surface of *Toxoplasma* tachyzoites is dominated by a family of glycosylphosphatidylinositol-anchored antigens related to SAG1. Infection and Immunity 66, 2237-2244.
Marugan-Hernandez, V., Alvarez-Garcia, G., Risco-Castillo, V., Regidor-Cerrillo, J., Ortega-Mora, L.M., 2010. Identification of *Neospora caninum* proteins regulated during the differentiation process from tachyzoite to bradyzoite stage by DIGE. Proteomics 10, 1740-1750.
Mayer, R., Picard, I., Lawton, P., Grellier, P., Barrault, C., Monsigny, M., Schrevel, J., 1991. Peptide derivatives specific for a Plasmodium falciparum proteinase inhibit the human erythrocyte invasion by merozoites. Journal of medicinal chemistry 34, 3029-3035.
McAllister, M.M., Bjorkman, C., Anderson-Sprecher, R., Rogers, D.G., 2000. Evidence of point-source exposure to *Neospora caninum* and protective immunity in a herd of beef cows. J Am Vet Med Assoc 217, 881-887.
McAllister, M.M., Dubey, J.P., Lindsay, D.S., Jolley, W.R., Wills, R.A., McGuire, A.M., 1998. Dogs are definitive hosts of *Neospora caninum*. International journal for parasitology 28, 1473-1478.
McAllister, M.M., Parmley, S.F., Weiss, L.M., Welch, V.J., McGuire, A.M., 1996. An immunohistochemical method for detecting bradyzoite antigen (BAG5) in *Toxoplasma gondii*-infected tissues cross-reacts with a *Neospora caninum* bradyzoite antigen. J Parasitol 82, 354-355.
McCann, C.M., McAllister, M.M., Gondim, L.F., Smith, R.F., Cripps, P.J., Kipar, A., Williams, D.J., Trees, A.J., 2007. *Neospora caninum* in cattle: experimental infection with oocysts can result in exogenous transplacental

infection, but not endogenous transplacental infection in the subsequent pregnancy. International journal for parasitology 37, 1631-1639.
Megger, D.A., Bracht, T., Meyer, H.E., Sitek, B., 2013. Label-free quantification in clinical proteomics. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1834, 1581-1590.
Melo, M.B., Jensen, K.D., Saeij, J.P., 2011. *Toxoplasma gondii* effectors are master regulators of the inflammatory response. Trends in parasitology 27, 487-495.
Mercier, C., Dubremetz, J.F., Rauscher, B., Lecordier, L., Sibley, L.D., Cesbron-Delauw, M.F., 2002. Biogenesis of nanotubular network in *Toxoplasma* parasitophorous vacuole induced by parasite proteins. Mol Biol Cell 13, 2397-2409.
Milisav, I., 1998. Dynein and dynein-related genes. Cell Motil Cytoskeleton 39, 261-272.
Miller, C.M., Quinn, H.E., Windsor, P.A., Ellis, J.T., 2002. Characterisation of the first Australian isolate of *Neospora caninum* from cattle. Aust Vet J 80, 620-625.
Miller, S., Binder, E., Blackman, M., Carruthers, V., Kim, K., 2001. A conserved subtilisin-like protein TgSUB1 in microneme organelles of *Toxoplasma gondii*. J Biol Chem 276, 45341 - 45348.
Mital, J., Meissner, M., Soldati, D., Ward, G.E., 2005. Conditional expression of *Toxoplasma gondii* apical membrane antigen-1 (TgAMA1) demonstrates that TgAMA1 plays a critical role in host cell invasion. Mol Biol Cell 16, 4341-4349.
Molina-López, R., Cabezón, O., Pabón, M., Darwich, L., Obón, E., Lopez-Gatius, F., Dubey, J.P., Almería, S., 2012. High seroprevalence of *Toxoplasma gondii* and *Neospora caninum* in the Common raven (Corvus corax) in the Northeast of Spain. Research in Veterinary Science 93, 300-302.
Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R., Dermitzakis, E.T., 2010. Transcriptome genetics using second generation sequencing in a Caucasian population. Nature 464, 773-777.
Morisaki, J.H., Heuser, J.E., Sibley, L.D., 1995. Invasion of *Toxoplasma gondii* occurs by active penetration of the host cell. J Cell Sci 108 (Pt 6), 2457-2464.
Morrissette, N., 2015. Targeting *Toxoplasma* Tubules: Tubulin, Microtubules, and Associated Proteins in a Human Pathogen. Eukaryotic Cell 14, 2-12.
Morrissette, N., Gubbels, M.-J., 2014. Chapter 13 - The *Toxoplasma* Cytoskeleton: Structures, Proteins and Processes, In: *Toxoplasma Gondii* (Second Edition). Academic Press, Boston, pp. 455-503.
Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5, 621-628.
Motoyama, A., Yates, J.R., 3rd, 2008. Multidimensional LC separations in shotgun proteomics. Anal Chem 80, 7187-7193.
Munoz-Zanzi, C.A., Fry, P., Lesina, B., Hill, D., 2010. *Toxoplasma gondii* oocyst-specific antibodies and source of infection. Emerging infectious diseases 16, 1591-1593.
Nabet, C., Doumbo, S., Jeddi, F., Konaté, S., Manciulli, T., Fofana, B., L'Ollivier, C., Camara, A., Moore, S., Ranque, S., Théra, M.A., Doumbo, O.K.,

Piarroux, R., 2016. Genetic diversity of Plasmodium falciparum in human malaria cases in Mali. Malaria Journal 15, 1-10.
Nagalakshmi, U., Waern, K., Snyder, M., 2010. RNA-Seq: a method for comprehensive transcriptome analysis. Curr Protoc Mol Biol Chapter 4, Unit 4.11.11-13.
Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. Science (New York, N.Y.) 320, 1344-1349.
Naguleswaran, A., Cannas, A., Keller, N., Vonlaufen, N., Björkman, C., Hemphill, A., 2002. Vero cell surface proteoglycan interaction with the microneme protein NcMIC3 mediates adhesion of *Neospora caninum* tachyzoites to host cells unlike that in *Toxoplasma gondii*. International journal for parasitology 32, 695-704.
Naguleswaran, A., Muller, N., Hemphill, A., 2003. *Neospora caninum* and *Toxoplasma gondii*: a novel adhesion/invasion assay reveals distinct differences in tachyzoite-host cell interactions. Experimental parasitology 104, 149-158.
Naitza, S., Spano, F., Robson, K.J., Crisanti, A., 1998. The Thrombospondin-related Protein Family of Apicomplexan Parasites: The Gears of the Cell Invasion Machinery. Parasitol Today 14, 479-484.
Nam, H.-W., 2009. GRA Proteins of *Toxoplasma gondii*: Maintenance of Host-Parasite Interactions across the Parasitophorous Vacuolar Membrane. The Korean Journal of Parasitology 47, S29-S37.
Niedelman, W., Gold, D.A., Rosowski, E.E., Sprokholt, J.K., Lim, D., Farid Arenas, A., Melo, M.B., Spooner, E., Yaffe, M.B., Saeij, J.P., 2012. The rhoptry proteins ROP18 and ROP5 mediate *Toxoplasma gondii* evasion of the murine, but not the human, interferon-gamma response. PLoS Pathog 8, e1002784.
Nirmalan, N.J., Harnden, P., Selby, P.J., Banks, R.E., 2009. Development and validation of a novel protein extraction methodology for quantitation of protein expression in formalin-fixed paraffin-embedded tissues using western blotting. The Journal of pathology 217, 497-506.
Nischik, N., Schade, B., Dytnerska, K., Dlugonska, H., Reichmann, G., Fischer, H.G., 2001. Attenuation of mouse-virulent *Toxoplasma gondii* parasites is associated with a decrease in interleukin-12-inducing tachyzoite activity and reduced expression of actin, catalase and excretory proteins. Microbes Infect 3, 689-699.
Nishimura, M., Tanaka, S., Ihara, F., Muroi, Y., Yamagishi, J., Furuoka, H., Suzuki, Y., Nishikawa, Y., 2015. Transcriptome and histopathological changes in mouse brain infected with *Neospora caninum*. Sci Rep 5, 7936.
Nwagwu, M., Opperdoes, F.R., 1982. Regulation of glycolysis in Trypanosoma brucei: hexokinase and phosphofructokinase activity. Acta Trop 39, 61-72.
Ogawa, K., Mohri, H., 1996. A dynein motor superfamily. Cell Struct Funct 21, 343-349.
Okada, T., Marmansari, D., Li, Z.M., Adilbish, A., Canko, S., Ueno, A., Shono, H., Furuoka, H., Igarashi, M., 2013. A novel dense granule protein, GRA22, is involved in regulating parasite egress in *Toxoplasma gondii*. Molecular and biochemical parasitology 189, 5-13.

Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., Ahn, N.G., 2005. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Mol Cell Proteomics 4, 1487-1502.

Passador-Gurgel, G., Hsieh, W.P., Hunt, P., Deighton, N., Gibson, G., 2007. Quantitative trait transcripts for nicotine resistance in Drosophila melanogaster. Nat Genet 39, 264-268.

Patel, V.J., Thalassinos, K., Slade, S.E., Connolly, J.B., Crombie, A., Murrell, J.C., Scrivens, J.H., 2009. A comparison of labeling and label-free mass spectrometry-based proteomics approaches. J Proteome Res 8, 3752-3759.

Pedroni, M.J., Sondgeroth, K.S., Gallego-Lopez, G.M., Echaide, I., Lau, A.O., 2013. Comparative transcriptome analysis of geographically distinct virulent and attenuated Babesia bovis strains reveals similar gene expression changes through attenuation. BMC Genomics 14, 763.

Peixoto, L., Chen, F., Harb, O.S., Davis, P.H., Beiting, D.P., Brownback, C.S., Ouloguem, D., Roos, D.S., 2010. Integrative genomic approaches highlight a family of parasite-specific kinases that regulate host responses. Cell host & microbe 8, 208-218.

Perco, P., Muhlberger, I., Mayer, G., Oberbauer, R., Lukas, A., Mayer, B., 2010. Linking transcriptomic and proteomic data on the level of protein interaction networks. Electrophoresis 31, 1780-1789.

Pereira Garcia-Melo, D., Regidor-Cerrillo, J., Collantes-Fernandez, E., Aguado-Martinez, A., Del Pozo, I., Minguijon, E., Gomez-Bautista, M., Aduriz, G., Ortega-Mora, L.M., 2010. Pathogenic characterization in mice of *Neospora caninum* isolates obtained from asymptomatic calves. Parasitology 137, 1057-1068.

Perez-Zaballos, F.J., Ortega-Mora, L.M., Alvarez-Garcia, G., Collantes-Fernandez, E., Navarro-Lozano, V., Garcia-Villada, L., Costas, E., 2005. Adaptation of *Neospora caninum* isolates to cell-culture changes: an argument in favor of its clonal population structure. J Parasitol 91, 507-510.

Perkins, D.N., Pappin, D.J., Creasy, D.M., Cottrell, J.S., 1999. Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20, 3551-3567.

Petersen, E., Dubey, J.P., 2001. Biology of toxoplasmosis. , In: Joynson, D.H.M., Wreghitt, T.G. (Eds.) Toxoplasmosis: A Comprehensive Clinical Guide. Cambridge University Press, Cambridge, United Kingdom, pp. 1-42.

Piergili Fioretti, D., Pasquali, P., Diaferia, M., Mangili, V., Rosignoli, L., 2003. *Neospora caninum* infection and congenital transmission: serological and parasitological study of cows up to the fourth gestation. J Vet Med B Infect Dis Vet Public Health 50, 399-404.

Pinheiro, A.M., Costa, S.L., Freire, S.M., Almeida, M.A., Tardy, M., El Bacha, R., Costa, M.F., 2006. Astroglial cells in primary culture: a valid model to study *Neospora caninum* infection in the CNS. Vet Immunol Immunopathol 113, 243-247.

Pollo-Oliveira, L., Post, H., Acencio, M.L., Lemke, N., van den Toorn, H., Tragante, V., Heck, A.J., Altelaar, A.M., Yatsuda, A.P., 2013. Unravelling the *Neospora caninum* secretome through the secreted fraction (ESA) and quantification of the discharged tachyzoite using high-resolution mass spectrometry-based proteomics. Parasites & Vectors 6, 1-14.

Qiu, W., Wernimont, A., Tang, K., Taylor, S., Lunin, V., Schapira, M., Fentress, S., Hui, R., Sibley, L.D., 2009. Novel structural and regulatory features of rhoptry secretory kinases in *Toxoplasma gondii*. The EMBO Journal 28, 969-979.

Que, X., Engel, J.C., Ferguson, D., Wunderlich, A., Tomavo, S., Reed, S.L., 2007. Cathepsin Cs are key for the intracellular survival of the protozoan parasite, *Toxoplasma gondii*. J Biol Chem 282, 4994-5003.

Que, X., Ngo, H., Lawton, J., Gray, M., Liu, Q., Engel, J., Brinen, L., Ghosh, P., Joiner, K.A., Reed, S.L., 2002. The cathepsin B of *Toxoplasma gondii*, toxopain-1, is critical for parasite invasion and rhoptry protein processing. J Biol Chem 277, 25791-25797.

Quinn, H.E., Ellis, J.T., Smith, N.C., 2002. *Neospora caninum*: a cause of immune-mediated failure of pregnancy? Trends in parasitology 18, 391-394.

Radke, J.R., Striepen, B., Guerini, M.N., Jerome, M.E., Roos, D.S., White, M.W., 2001. Defining the cell cycle for the tachyzoite stage of *Toxoplasma gondii*. Molecular and biochemical parasitology 115, 165-175.

Reese, M.L., Boothroyd, J.C., 2011. A conserved non-canonical motif in the pseudoactive site of the ROP5 pseudokinase domain mediates its effect on *Toxoplasma* virulence. J Biol Chem 286, 29366-29375.

Reese, M.L., Zeiner, G.M., Saeij, J.P., Boothroyd, J.C., Boyle, J.P., 2011. Polymorphic family of injected pseudokinases is paramount in *Toxoplasma* virulence. Proc Natl Acad Sci U S A 108, 9625-9630.

Regidor-Cerrillo, J., Alvarez-Garcia, G., Pastor-Fernandez, I., Marugan-Hernandez, V., Gomez-Bautista, M., Ortega-Mora, L.M., 2012a. Proteome expression changes among virulent and attenuated *Neospora caninum* isolates. J Proteomics 75.

Regidor-Cerrillo, J., Alvarez-Garcia, G., Pastor-Fernandez, I., Marugan-Hernandez, V., Gomez-Bautista, M., Ortega-Mora, L.M., 2012b. Proteome expression changes among virulent and attenuated *Neospora caninum* isolates. J Proteomics 75, 2306-2318.

Regidor-Cerrillo, J., Arranz-Solís, D., Benavides, J., Gómez-Bautista, M., Castro-Hermida, J.A., Mezo, M., Pérez, V., Ortega-Mora, L.M., González-Warleta, M., 2014. *Neospora caninum* infection during early pregnancy in cattle: how the isolate influences infection dynamics, clinical outcome and peripheral and local immune responses. Veterinary Research 45, 10-10.

Regidor-Cerrillo, J., Díez-Fuertes, F., García-Culebras, A., Moore, D.P., González-Warleta, M., Cuevas, C., Schares, G., Katzer, F., Pedraza-Díaz, S., Mezo, M., Ortega-Mora, L.M., 2013. Genetic Diversity and Geographic Population Structure of Bovine *Neospora caninum* Determined by Microsatellite Genotyping Analysis. PloS one 8, e72678.

Regidor-Cerrillo, J., Garcia-Lunar, P., Pastor-Fernandez, I., Alvarez-Garcia, G., Collantes-Fernandez, E., Gomez-Bautista, M., Ortega-Mora, L.M., 2015. *Neospora caninum* tachyzoite immunome study reveals differences among three biologically different isolates. Veterinary parasitology 212, 92-99.

Regidor-Cerrillo, J., Gomez-Bautista, M., Pereira-Bueno, J., Aduriz, G., Navarro-Lozano, V., Risco-Castillo, V., Fernandez-Garcia, A., Pedraza-Diaz, S., Ortega-Mora, L.M., 2008. Isolation and genetic characterization of *Neospora caninum* from asymptomatic calves in Spain. Parasitology 135, 1651-1659.

Regidor-Cerrillo, J., Gomez-Bautista, M., Sodupe, I., Aduriz, G., Alvarez-Garcia, G., Del Pozo, I., Ortega-Mora, L.M., 2011. *In vitro* invasion efficiency and intracellular proliferation rate comprise virulence-related phenotypic traits of *Neospora caninum*. Vet Res 42, 41.

Regidor-Cerrillo, J., Pedraza-Diaz, S., Gomez-Bautista, M., Ortega-Mora, L.M., 2006. Multilocus microsatellite analysis reveals extensive genetic diversity in *Neospora caninum*. J Parasitol 92, 517-524.

Reichel, M.P., Alejandra Ayanegui-Alcerreca, M., Gondim, L.F., Ellis, J.T., 2013. What is the global economic impact of *Neospora caninum* in cattle - the billion dollar question. International journal for parasitology 43, 133-142.

Reichel, M.P., Ellis, J.T., 2009. *Neospora caninum*--how close are we to development of an efficacious vaccine that prevents abortion in cattle? International journal for parasitology 39, 1173-1187.

Reichel, M.P., Ellis, J.T., Dubey, J.P., 2007. Neosporosis and hammondiosis in dogs. J Small Anim Pract 48, 308-312.

Reichel, M.P., McAllister, M.M., Pomroy, W.E., Campero, C., Ortega-Mora, L.M., Ellis, J.T., 2014. Control options for *Neospora caninum*--is there anything new or are we going backwards? Parasitology 141, 1455-1470.

Reichmann, G., Długońska, H., Fischer, H.-G., 2002. Characterization of TgROP9 (p36), a novel rhoptry protein of *Toxoplasma gondii* tachyzoites identified by T cell clone. Molecular and biochemical parasitology 119, 43-54.

Reid, A.J., Vermont, S.J., Cotton, J.A., Harris, D., Hill-Cawthorne, G.A., Koenen-Waisman, S., Latham, S.M., Mourier, T., Norton, R., Quail, M.A., 2012. Comparative genomics of the apicomplexan parasites *Toxoplasma gondii* and *Neospora caninum*: coccidia differing in host range and transmission strategy. Plos Pathog 8.

Reiss, M., Viebig, N., Brecht, S., Fourmaux, M., Soete, M., Di Cristina, M., 2001a. Identification and characterization of an escorter for two secretory adhesions in *Toxoplasma gondii*. J cell boil 152, 563 - 578.

Reiss, M., Viebig, N., Brecht, S., Fourmaux, M.N., Soete, M., Di Cristina, M., Dubremetz, J.F., Soldati, D., 2001b. Identification and characterization of an escorter for two secretory adhesins in *Toxoplasma gondii*. J Cell Biol 152, 563-578.

Rigbolt, K.T., Vanselow, J.T., Blagoev, B., 2011. GProX, a user-friendly platform for bioinformatics analysis and visualization of quantitative proteomics data. Mol Cell Proteomics 10, O110 007450.

Risco-Castillo, V., Fernandez-Garcia, A., Ortega-Mora, L.M., 2004. Comparative analysis of stress agents in a simplified *in vitro* system of *Neospora caninum* bradyzoite production. J Parasitol 90, 466-470.

Risco-Castillo, V., Fernandez-Garcia, A., Zaballos, A., Aguado-Martinez, A., Hemphill, A., Rodriguez-Bertos, A., Alvarez-Garcia, G., Ortega-Mora, L.M., 2007. Molecular characterisation of BSR4, a novel bradyzoite-specific gene from *Neospora caninum*. International journal for parasitology 37, 887-896.

Robert-Gangneux, F., Darde, M.L., 2012. Epidemiology of and diagnostic strategies for toxoplasmosis. Clinical microbiology reviews 25, 264-296.

Rodrigues, A.A., Gennari, S.M., Aguiar, D.M., Sreekumar, C., Hill, D.E., Miska, K.B., Vianna, M.C., Dubey, J.P., 2004. Shedding of *Neospora caninum* oocysts by dogs fed tissues from naturally infected water buffaloes (Bubalus bubalis) from Brazil. Veterinary parasitology 124, 139-150.

Rojo-Montejo, S., Collantes-Fernandez, E., Blanco-Murcia, J., Rodriguez-Bertos, A., Risco-Castillo, V., Ortega-Mora, L.M., 2009a. Experimental infection with a low virulence isolate of *Neospora caninum* at 70 days gestation in cattle did not result in foetopathy. Vet Res 40, 49.

Rojo-Montejo, S., Collantes-Fernández, E., Regidor-Cerrillo, J., Alvarez-García, G., Marugan-Hernández, V., Pedraza-Díaz, S., Blanco-Murcia, J., Prenafeta, A., Ortega-Mora, L.M., 2009b. Isolation and characterization of a bovine isolate of *Neospora caninum* with low virulence. Veterinary parasitology 159, 7-16.

Rosowski, E.E., Saeij, J.P., 2012. *Toxoplasma gondii* clonal strains all inhibit STAT1 transcriptional activity but polymorphic effectors differentially modulate IFNgamma induced gene expression and STAT1 phosphorylation. PloS one 7, e51448.

Rovira-Graells, N., Gupta, A.P., Planet, E., Crowley, V.M., Mok, S., Ribas de Pouplana, L., Preiser, P.R., Bozdech, Z., Cortes, A., 2012. Transcriptional variation in the malaria parasite Plasmodium falciparum. Genome Res 22, 925-938.

Saeij, J.P., Boyle, J.P., Boothroyd, J.C., 2005. Differences among the three major strains of *Toxoplasma gondii* and their specific interactions with the infected host. Trends in parasitology 21, 476-481.

Saeij, J.P., Boyle, J.P., Coller, S., Taylor, S., Sibley, L.D., Brooke-Powell, E.T., Ajioka, J.W., Boothroyd, J.C., 2006. Polymorphic secreted kinases are key virulence factors in toxoplasmosis. Science (New York, N.Y.) 314, 1780-1783.

Saeij, J.P., Coller, S., Boyle, J.P., Jerome, M.E., White, M.W., Boothroyd, J.C., 2007a. *Toxoplasma* co-opts host gene expression by injection of a polymorphic kinase homologue. Nature 445, 324-327.

Saeij, J.P.J., Coller, S., Boyle, J.P., Jerome, M.E., White, M.W., Boothroyd, J.C., 2007b. *Toxoplasma* co-opts host gene expression by injection of a polymorphic kinase homologue. Nature 445, 324-327.

Saffer, L.D., Mercereau-Puijalon, O., Dubremetz, J.F., Schwartzman, J.D., 1992. Localization of a *Toxoplasma gondii* rhoptry protein by immunoelectron microscopy during and after host cell penetration. J Protozool 39, 526-530.

Saouros, S., Edwards-Jones, B., Reiss, M., Sawmynaden, K., Cota, E., Simpson, P., 2005a. A novel galectin-like domain from *Toxoplasma gondii* micronemal protein 1 assists the folding, assemblymand transport of a cell-adhesion complex. J Biol Chem 280, 38583 - 38591.

Saouros, S., Edwards-Jones, B., Reiss, M., Sawmynaden, K., Cota, E., Simpson, P., Dowse, T.J., Jakle, U., Ramboarina, S., Shivarattan, T., Matthews, S., Soldati-Favre, D., 2005b. A novel galectin-like domain from *Toxoplasma gondii* micronemal protein 1 assists the folding, assembly, and transport of a cell adhesion complex. J Biol Chem 280, 38583-38591.

Sarah, V., 2012. Proteomic and transcriptomic analysis of the protozoan parasite *Neospora caninum* (PhD's thesis). University of Liverpool.

Sawada, M., Kondo, H., Tomioka, Y., Park, C., Morita, T., Shimada, A., Umemura, T., 2000. Isolation of *Neospora caninum* from the brain of a naturally infected adult dairy cow. Veterinary parasitology 90, 247-252.

Schares, G., Pantchev, N., Barutzki, D., Heydorn, A.O., Bauer, C., Conraths, F.J., 2005. Oocysts of *Neospora caninum*, Hammondia heydorni, *Toxoplasma*

gondii and Hammondia hammondi in faeces collected from dogs in Germany. International journal for parasitology 35, 1525-1537.
Schock, A., Innes, E.A., Yamane, I., Latham, S.M., Wastling, J.M., 2001. Genetic and biological diversity among isolates of *Neospora caninum*. Parasitology 123, 13-23.
Schwan, T.G., Hinnebusch, B.J., 1998. Bloodstream- versus tick-associated variants of a relapsing fever bacterium. Science (New York, N.Y.) 280, 1938-1940.
Shaw, M.K., Roos, D.S., Tilney, L.G., 2002. Cysteine and serine protease inhibitors block intracellular development and disrupt the secretory pathway of *Toxoplasma gondii*. Microbes Infect 4, 119-132.
Shin, Y.S., Shin, G.W., Kim, Y.R., Lee, E.Y., Yang, H.H., Palaksha, K.J., Youn, H.J., Kim, J.H., Kim, D.Y., Marsh, A.E., 2005. Comparison of proteome and antigenic proteome between two *Neospora caninum* isolates. Veterinary parasitology 134.
Sibley, L.D., Ajioka, J.W., 2008. Population structure of *Toxoplasma gondii*: clonal expansion driven by infrequent recombination and selective sweeps. Annu Rev Microbiol 62, 329-351.
Sibley, L.D., Niesman, I.R., Parmley, S.F., Cesbron-Delauw, M.F., 1995. Regulated secretion of multi-lamellar vesicles leads to formation of a tubulo-vesicular network in host-cell vacuoles occupied by *Toxoplasma gondii*. J Cell Sci 108 (Pt 4), 1669-1677.
Sinai, A.P., Joiner, K.A., 2001. The *Toxoplasma gondii* protein ROP2 mediates host organelle association with the parasitophorous vacuole membrane. J Cell Biol 154, 95-108.
Skariah, S., McIntyre, M.K., Mordue, D.G., 2010. *Toxoplasma gondii*: determinants of tachyzoite to bradyzoite conversion. Parasitology research 107, 253-260.
Soete, M., Camus, D., Dubremetz, J.F., 1994. Experimental induction of bradyzoite-specific antigen expression and cyst formation by the RH strain of *Toxoplasma gondii in vitro*. Experimental parasitology 78, 361-370.
Soete, M., Fortier, B., Camus, D., Dubremetz, J.F., 1993. *Toxoplasma gondii*: kinetics of bradyzoite-tachyzoite interconversion *in vitro*. Experimental parasitology 76, 259-264.
Soldati, D., Dubremetz, J., Lebrun, M., 2001. Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite *Toxoplasma gondii*. International journal for parasitology 31, 1293 - 1302.
Soldati, D., Meissner, M., 2004. *Toxoplasma* as a novel system for motility. Curt opion cell boil 16, 32 - 40.
Sonda, S., Fuchs, N., Gottstein, B., Hemphill, A., 2000. Molecular characterization of a novel microneme antigen in *Neospora caninum*. Molecular and biochemical parasitology 108, 39-51.
Song, H.O., Ahn, M.H., Ryu, J.S., Min, D.Y., Joo, K.H., Lee, Y.H., 2004. Influence of calcium ion on host cell invasion and intracellular replication by *Toxoplasma gondii*. Korean J Parasitol 42, 185-193.
Speer, C.A., Dubey, J.P., 1989. Ultrastructure of tachyzoites, bradyzoites and tissue cysts of *Neospora caninum*. J Protozool 36, 458-463.
Speer, C.A., Dubey, J.P., McAllister, M.M., Blixt, J.A., 1999. Comparative ultrastructure of tachyzoites, bradyzoites, and tissue cysts of *Neospora*

caninum and *Toxoplasma gondii*. International journal for parasitology 29, 1509-1519.
Storey, J.D., 2003. The Positive False Discovery Rate: A Bayesian Interpretation and the q-Value. The Annals of Statistics 31, 2013-2035.
Suzuki, Y., Joh, K., 1994. Effect of the strain of *Toxoplasma gondii* on the development of toxoplasmic encephalitis in mice treated with antibody to interferon-gamma. Parasitology research 80, 125-130.
Talevich, E., Kannan, N., 2013. Structural and evolutionary adaptation of rhoptry kinases and pseudokinases, a family of coccidian virulence factors. BMC Evolutionary Biology 13, 117.
Tampaki, Z., Mwakubambanya, R.S., Goulielmaki, E., Kaforou, S., Kim, K., Waters, A.P., Carruthers, V.B., Siden-Kiamos, I., Loukeris, T.G., Koussis, K., 2015. Ectopic expression of a *Neospora caninum* Kazal type inhibitor triggers developmental defects in *Toxoplasma* and Plasmodium. PloS one 10, e0121379.
Taylor, S., Barragan, A., Su, C., Fux, B., Fentress, S.J., Tang, K., Beatty, W.L., Hajj, H.E., Jerome, M., Behnke, M.S., White, M., Wootton, J.C., Sibley, L.D., 2006. A secreted serine-threonine kinase determines virulence in the eukaryotic pathogen *Toxoplasma gondii*. Science (New York, N.Y.) 314, 1776-1780.
Tilley, M., Fichera, M.E., Jerome, M.E., Roos, D.S., White, M.W., 1997. *Toxoplasma gondii* sporozoites form a transient parasitophorous vacuole that is impermeable and contains only a subset of dense-granule proteins. Infection and Immunity 65, 4598-4605.
Tomavo, S., 1996. The Major Surface Proteins of *Toxoplasma gondii*: Structures and Functions, In: Gross, U. (Ed.) *Toxoplasma gondii*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 45-54.
Tomavo, S., 2001. The differential expression of multiple isoenzyme forms during stage conversion of *Toxoplasma gondii*: an adaptive developmental strategy. International journal for parasitology 31, 1023-1031.
Tomita, T., Bzik, D.J., Ma, Y.F., Fox, B.A., Markillie, L.M., Taylor, R.C., Kim, K., Weiss, L.M., 2013. The *Toxoplasma gondii* Cyst Wall Protein CST1 Is Critical for Cyst Wall Integrity and Promotes Bradyzoite Persistence. PLoS Pathogens 9, e1003823.
Tomley, F.M., Soldati, D.S., 2001. Mix and match modules: structure and function of microneme proteins in apicomplexan parasites. Trends in parasitology 17, 81-88.
Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., Pachter, L., 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7, 562-578.
Travier, L., Mondragon, R., Dubremetz, J.F., Musset, K., Mondragon, M., Gonzalez, S., Cesbron-Delauw, M.F., Mercier, C., 2008. Functional domains of the *Toxoplasma* GRA2 protein in the formation of the membranous nanotubular network of the parasitophorous vacuole. International journal for parasitology 38, 757-773.
Trees, A.J., Davison, H.C., Innes, E.A., Wastling, J.M., 1999. Towards evaluating the economic impact of bovine neosporosis. International journal for parasitology 29, 1195-1200.

Trees, A.J., Williams, D.J., 2005. Endogenous and exogenous transplacental infection in *Neospora caninum* and *Toxoplasma gondii*. Trends in parasitology 21, 558-561.
Tyler, J.S., Boothroyd, J.C., 2011. The C-terminus of *Toxoplasma* RON2 provides the crucial link between AMA1 and the host-associated invasion complex. PloS Pathog 7.
Vera, J.C., Wheat, C.W., Fescemyer, H.W., Frilander, M.J., Crawford, D.L., Hanski, I., Marden, J.H., 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol 17, 1636-1647.
Vogel, C., Marcotte, E.M., 2012. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13, 227-232.
Vonlaufen, N., Guetg, N., Naguleswaran, A., Muller, N., Bjorkman, C., Schares, G., von Blumroeder, D., Ellis, J., Hemphill, A., 2004. *In vitro* induction of *Neospora caninum* bradyzoites in vero cells reveals differential antigen expression, localization, and host-cell recognition of tachyzoites and bradyzoites. Infection and Immunity 72, 576-583.
Vonlaufen, N., Muller, N., Keller, N., Naguleswaran, A., Bohne, W., McAllister, M.M., Bjorkman, C., Muller, E., Caldelari, R., Hemphill, A., 2002. Exogenous nitric oxide triggers *Neospora caninum* tachyzoite-to-bradyzoite stage conversion in murine epidermal keratinocyte cell cultures. International journal for parasitology 32, 1253-1265.
Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10, 57-63.
Wasmuth, J.D., Pszenny, V., Haile, S., Jansen, E.M., Gast, A.T., Sher, A., Boyle, J.P., Boulanger, M.J., Parkinson, J., Grigg, M.E., 2012. Integrated bioinformatic and targeted deletion analyses of the SRS gene superfamily identify SRS29C as a negative regulator of *Toxoplasma* virulence. MBio 3.
Wastling, J.M., Armstrong, S.D., Krishna, R., Xia, D., 2012. Parasites, proteomes and systems: has Descartes' clock run out of time? Parasitology 139, 1103-1118.
Wastling, J.M., Xia, D., 2016. Proteomic, In: Walochnik, J., Duchêne, M. (Eds.) Molecular Parasitology: Protozoan Parasites and their Molecules. Springer Vienna, pp. 49-74.
Wastling, J.M., Xia, D., Sohal, A., Chaussepied, M., Pain, A., Langsley, G., 2009. Proteomes and transcriptomes of the Apicomplexa--where's the message? International journal for parasitology 39, 135-143.
Weiss, L.M., Ma, Y.F., Halonen, S., McAllister, M.M., Zhang, Y.W., 1999. The *in vitro* development of *Neospora caninum* bradyzoites. International journal for parasitology 29, 1713-1723.
Weiss, L.M., Ma, Y.F., Takvorian, P.M., Tanowitz, H.B., Wittner, M., 1998a. Bradyzoite development in *Toxoplasma gondii* and the hsp70 stress response. Infection and Immunity 66, 3295-3302.
Weiss, L.M., Ma, Y.F., Takvorian, P.M., Tanowitz, H.B., Wittner, M., 1998b. Bradyzoite Development in *Toxoplasma gondii* and the hsp70 Stress Response. Infection and Immunity 66, 3295-3302.
Westermeier, R., Naven, T., Höpker, H.R., 2008. Introduction, In: Proteomics in Practice: A Guide to Successful Experimental Design. Wiley, pp. 1-16.
Wilhelm, B.T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C.J., Rogers, J., Bahler, J., 2008. Dynamic repertoire of a eukaryotic

transcriptome surveyed at single-nucleotide resolution. Nature 453, 1239-1243.
Williams, D.J., Guy, C.S., McGarry, J.W., Guy, F., Tasker, L., Smith, R.F., MacEachern, K., Cripps, P.J., Kelly, D.F., Trees, A.J., 2000. *Neospora caninum*-associated abortion in cattle: the time of experimentally-induced parasitaemia during gestation determines foetal survival. Parasitology 121 (Pt 4), 347-358.
Woodbine, K.A., Medley, G.F., Moore, S.J., Ramirez-Villaescusa, A., Mason, S., Green, L.E., 2008. A four year longitudinal sero-epidemiology study of *Neospora caninum* in adult cattle from 114 cattle herds in south west England: associations with age, herd and dam-offspring pairs. BMC Vet Res 4, 35.
Wouda, W., Moen, A.R., Schukken, Y.H., 1998. Abortion risk in progeny of cows after a *Neospora caninum* epidemic. Theriogenology 49, 1311-1316.
Xia, D., Sanderson, S.J., Jones, A.R., Prieto, J.H., Yates, J.R., Bromley, E., Tomley, F.M., Lal, K., Sinden, R.E., Brunk, B.P., Roos, D.S., Wastling, J.M., 2008. The proteome of *Toxoplasma gondii*: integration with the genome provides novel insights into gene expression and annotation. Genome Biol 9, R116.
Xiao, J., Yolken, R.H., 2015. Strain hypothesis of *Toxoplasma gondii* infection on the outcome of human diseases. Acta physiologica (Oxford, England) 213, 828-845.
Yamane, I., Kokuho, T., Shimura, K., Eto, M., Haritani, M., Ouchi, Y., Sverlow, K.W., Conrad, P.A., 1996. *In vitro* isolation of a bovine *Neospora* in Japan. Vet Rec 138, 652.
Yamane, I., Kokuho, T., Shimura, K., Eto, M., Shibahara, T., Haritani, M., Ouchi, Y., Sverlow, K., Conrad, P.A., 1997. *In vitro* isolation and characterisation of a bovine *Neospora* species in Japan. Res Vet Sci 63, 77-80.
Yang, N., Farrell, A., Niedelman, W., Melo, M., Lu, D., Julien, L., Marth, G.T., Gubbels, M.J., Saeij, J.P., 2013. Genetic basis for phenotypic differences between different *Toxoplasma gondii* type I strains. BMC Genomics 14, 467.
Yang, Y., Hu, M., Yu, K., Zeng, X., Liu, X., 2015. Mass spectrometry-based proteomic approaches to study pathogenic bacteria-host interactions. Protein & Cell 6, 265-274.
Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma, B., 2012. PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification. Molecular & Cellular Proteomics : MCP 11, M111.010587.
Zhang, L., Xiao, H., Karlan, S., Zhou, H., Gross, J., Elashoff, D., Akin, D., Yan, X., Chia, D., Karlan, B., Wong, D.T., 2010. Discovery and Preclinical Validation of Salivary Transcriptomic and Proteomic Biomarkers for the Non-Invasive Detection of Breast Cancer. PloS one 5.
Zhang, N.Z., Chen, J., Wang, M., Petersen, E., Zhu, X.Q., 2013. Vaccines against *Toxoplasma gondii*: new developments and perspectives. Expert review of vaccines 12, 1287-1299.
Zhang, Y., Xu, B., Kinoshita, N., Yoshida, Y., Tasaki, M., Fujinaka, H., Magdeldin, S., Yaoita, E., Yamamoto, T., 2015. Datasets from label-free quantitative proteomic analysis of human glomeruli with sclerotic lesions. Data in Brief 4, 180-185.

Zhao, S., Fung-Leung, W.-P., Bittner, A., Ngo, K., Liu, X., 2014. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. PLoS ONE 9, e78644.
Zhu, W., Smith, J.W., Huang, C.-M., 2010. Mass Spectrometry-Based Label-Free Quantitative Proteomics. Journal of Biomedicine and Biotechnology 2010, 6.