Transcriptomic analysis of *Ehrlichia ruminantium* during the developmental stages in bovine and tick cell culture

Mabotse A Tjale[a, b*], Alri Pretorius[a, b] Antoinette Josemans[a], Mirinda Van Kleef[a, b], and Junita Liebenberg[a],

[a]Agricultural Research Council - Onderstepoort Veterinary Research, Private Bag X5, Onderstepoort 0110, South Africa, [b]Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria, Private Bag X04, Onderstepoort 0110, South Africa.

***Corresponding Author**

Mailing address: ARC - Onderstepoort Veterinary Research, Private Bag X5, Onderstepoort 0110, South Africa

Phone: +27 12 529 9474

Fax: +27 12 529 9417

E-mail: mabotsetjale@gmail.com

Email: LiebenbergJ@arc.agric.za

Abstract

The use of bioinformatics tools to search for possible vaccine candidates has been successful in recent years. In an attempt to search for additional vaccine candidates or improve the current heartwater vaccine design, a genome-wide transcriptional profile of *E. ruminantium* (Welgevonden strain) replicating in bovine endothelial cells (BA886) and *Ixodes scapularis* embryonic tick cells (IDE8) was performed. The RNA was collected from the infective extracellular form, the elementary bodies (EBs) and vegetative intracellular form, reticulate bodies (RBs) and was used for transcriptome sequencing. Several genes previously implicated with adhesion, attachment and pathogenicity were exclusively up-regulated in the EBs from bovine and tick cells. Similarly, genes involved in adaptation or survival of *E. ruminantium* in the host cells were up-regulated in the RBs from bovine cells. Thus, it was concluded that those genes expressed in the EBs might be important for infection of mammalian and tick host cells and these may be targets for both cell and humoral mediated immune responses. Alternatively, those exclusively expressed in the RBs may be important for survival in the host cells. Exported or secreted proteins exclusively expressed at this stage are ideal targets for the stimulation of cytotoxic T-lymphocyte (CTL) immune responses in the host.

Keywords: *Ehrlichia ruminantium*, transcriptome

Abbreviations: Supplementary material

1. Introduction

Heartwater is a non-contagious tick-borne disease of domestic ruminants caused by an infection with a bacterium *Ehrlichia ruminantium,* which is an α-proteobacterium in the order Rickettsiales that is transmitted by *Amblyomma* ticks (Allsopp, 2015). This disease poses a major threat to livestock productivity particularly in areas where it is endemic such as sub-Saharan Africa, and has also spread to the eastern Caribbean (Walker and Olwage, 1987). The occurrence of the disease prevents more productive breeds from being introduced into endemic areas, while in the USA the presence of the tick vector means that there is a continual threat of the disease being introduced at some stage (Barré et al., 1988). Regardless, there is no safe and effective vaccine currently available. The only commercially available heartwater vaccine employs infection with virulent Ball3 strain infected sheep blood followed by antibiotic treatment and this method offers limited protection against several common virulent genotypes (Allsopp, 2015).

Research has shown that bacterial pathogens interrupt the hosts' cellular pathways for survival and replication and in turn, the infected host cells respond to the invading pathogen through cascading changes in gene expression (Humphrys et al., 2013). Thus, understanding these complex processes to identify novel bacterial virulence factors and host immune response pathways remains vital, particularly in vaccine development (Allsopp, 2015). The molecular mechanisms by which *E. ruminantium* proteins manipulate the host thereby facilitating infection have not been well defined. We therefore hypothesized that those genes highly expressed in the elementary bodies (EBs) from bovine cells (BovEBs) and tick cells (TicEBs) may be important for infection of mammalian and tick host cells. Similarly, those expressed in the reticulate bodies (RBs) may be central for survival in the host cells. Since this study was used as a platform to select for promising vaccine candidates, our focus was

mainly on the genes from these functional categories: hypothetical genes, membrane-associated, exported or secreted proteins, pathogenicity associated and transporters.

The sequential development of *E. ruminantium* has been described both in vertebrate and invertebrate hosts (Allsopp, 2015). *In vitro* studies revealed that *E. ruminantium* has a complex life cycle that is described as *Chlamydia*-like developmental stages (Jongejan et al., 1991). The *E. ruminantium* life cycle has the smaller extracellular EBs that are infectious but not replicative and the larger intracellular replicative and non-infectious RBs (Jongejan et al., 1991). In the early stages, the EBs adhere to the host target cells and are immersed quickly. These persist within the intracytoplasmic vacuoles where they divide by binary fission to form vegetative forms, RBs and 2-4 days post infection (pi), the intermediate bodies are formed. After 5-6 days the host cells are disrupted and EBs are released and thus initiating a new cycle of infection (Jongejan et al., 1991).

Global analysis of bacterial gene expression has previously been hindered by several factors, some of which include high abundance of ribosomal RNA and/or RNA instability (Neidhardt and Umbarger, 1996; Filiatrault, 2011). Furthermore, mRNA enrichment has been challenging in previous years (Rossetti et al., 2010). However, with the advent of new technology such as next generation sequencing (NGS) in combination with mRNA enrichment and tilling array technology, it has become practicable to understand and analyse the bacterial transcriptome (Sorek and Cossart, 2010). NGS allows an opportunity to obtain millions of reads at a low cost and has opened the door to study microorganisms that cannot be easily purified (Wang et al., 2009). To our knowledge, this is the first study to investigate the global gene expression analysis of *E. ruminantium* Welgevonden in different developmental stages using transcriptome or RNA-sequencing. Available studies are based on microarray analysis (Emboulé et al., 2009; Pruneau et al., 2012). In general, microarray

results are consistent with RNA-seq data, but the RNA-seq technique is more sensitive (Wang et al., 2009). Microarray-based techniques require knowledge of the genome sequence and high background cross hybridisation occurs. Additionally, comparisons of expression levels across different experiments are difficult and requires complex normalisation. Transcriptome sequencing on the other hand, offers an advantage over previously described methods for studying bacterial gene expression in that it provides a more precise measurement of transcripts and their isoforms (Wang et al., 2009). Generally, the extracted RNA from infected cells is a mixture of host and bacterial RNA. Whereas most of the bacterial RNA is ribosomal RNA (rRNA) and tRNA (up to 98%), bacterial mRNA is a typical minor fraction of the infected cells (Humphrys et al., 2013). Thus, RNA-seq offers advantages in that it is sensitive, transcripts can be accurately quantitated and it is not limited to detect transcripts that correspond to the existing genome sequence. This study reports that several *E. ruminantium* genes were differentially expressed in the developmental stages *in vitro* in mammalian and tick cells. The RNA-seq data was validated with reverse transcription quantitative real time (RT-qPCR).

2. Materials and methods

2.1 Strain of *E. ruminantium*

The *E. ruminantium* Welgevonden strain was originally isolated from a mouse infected with a tick homogenate. This male tick of *Amblyomma hebraeum* was collected on the Welgevonden farm (Naboomspruit/Mokgopong) in the Northern Transvaal (Limpopo province) (Du Plessis, 1985). The EBs of the *E. ruminantium* Welgevonden strain, obtained as passage (179), were stored in 500 µl sucrose-potassium phosphate glutamate medium (SPG) (0.218 M sucrose, 3.8 mM KH_2PO_4, 7.1 mM K_2HPO_4, 4.9 mM $C_5H_8NO_4K$) and frozen in liquid nitrogen.

2.2 Cell lines

Bovine aorta endothelial cells (BA886) were used as mammalian host cells and *Ixodes scapularis* embryonic tick cells (IDE8) as vector host cells for culture of *E. ruminantium*. BA886 cells were propagated in the media containing Dulbecco's modified eagle's medium/Ham's nutrient mixture: F12 (1:1 DME/F12) (Sigma-Aldrich) with 10% fetal bovine serum (FBS) (Life technologies), 1.2 g/L sodium bicarbonate and 100 IU penicillin and 100 µg/ml streptomycin (Sigma-Aldrich). The BA886 cells were cultured as monolayers at 37ºC in a SHEL-LAB CO_2 water-jacked incubator (SHEL-LAB). IDE8 cells were propagated in L-15B media (Munderloh and Kurtti, 1989) with 5% FBS (Life technologies), 10% tryptose phosphate broth (TPB) (Sigma-Aldrich), 0.1% bovine lipoprotein concentrate (MP Biomedicals), 2 mM L-glutamine (Sigma-Aldrich) and 100 IU penicillin and 100 µg/ml streptomycin (Sigma-Aldrich). The tick cells were cultured as monolayers at 32ºC in a Labtech incubator (Daihan LabTech). The tick cells were kindly provided by Dr Lesley Bell-Sakyi from the Tick Cell Biobank, Pirbright Institute, UK.

2.3 *In vitro* infection of cell monolayers with *E. ruminantium* (Welgevonden) strain

The BA886 cells were infected as described previously (Zweygarth et al., 1997). Briefly, the BA886 cells were inoculated with the supernatant containing the EBs of *E. ruminantium* Welgevonden (passage 179). The EBs were collected after 5-6 days pi in the supernatant when 90% of the cells were infected and the EBs could be visualized extracellular by Kyro-Quick stain (KYRO). The EBs were purified for transcriptome sequencing as follows: Any intact host cells were lysed by 2 passages through a 25-gauge needle syringe (Terumo Medicals) and host cell debris removed from the EBs by centrifugation at 1 500 xg for 10 min. The supernatant was transferred to a new microcentrifuge tube and the EBs were

collected by centrifugation at 20 000 xg for 30 min. The pellet was resuspended in 1 ml TRI Reagent (Sigma-Aldrich) for RNA extraction according to the manufacturer's instructions.

For the isolation of RBs embedded in the BA886 cells, a new batch of confluent bovine endothelial cells (BA886) were infected. To remove all EBs from the inoculum, all the media was discarded after 24 h and replaced with 5 ml fresh medium. In addition, the level of infectivity was monitored every 6 h by microscopy. When the RBs were observed, the medium was discarded and the infected endothelial cells were washed twice with 2 ml Dulbecco's PBS (Sigma-Aldrich) and fresh media added. Once 70-80% host cells contained RBs (2-3 days pi), the cells were harvested and centrifuged at 1500 xg for 10 min. The supernatant was discarded and the pellet (containing RBs) was collected and resuspended in 1 ml TRI Reagent.

The IDE8 cells were inoculated with the supernatant containing the EBs of *E. ruminantium* (Welgevonden) that were collected from previously infected bovine endothelial cells as previously described (Bell-Sakyi et al., 2000). The inoculated tick cells were maintained with weekly medium changes. The cultures were monitored for growth and infection by weekly preparation of cytospin Kyro-quick (KYRO) stained smears (Zweygarth et al., 1997). Cultures that were not infected after 12-14 weeks were discarded. The EBs were collected as described for the BA cells.

2.4 Transcriptome sequencing

RNA was processed from two biological replicates of *E. ruminantium* EBs from ticks (TicEBs), bovine EBs (BovEBs) and bovine RBs (BovRBs). The TRI Reagent (Sigma-Aldrich) protocol was used for RNA extraction. The Ribo-Zero^{TM} magnetic kit (Gram negative bacteria, Epicentre, Illumina) was used to remove ribosomal RNA. The rRNA depleted samples were further purified using the RNeasy MiniElute Cleanup kit (Qiagen)

according to the manufacturer's instruction. The RNA libraries were prepared using the ScriptSeq v2 kit (Epicentre) and the samples were sequenced with HiScan or MiSeq Illumina technology with 100 bp reads.

2.5 Comparative transcriptional analysis

CLC genomics Workbench 8.0.2 (http://www.clcbio.com/products/clc-genomics-workbench/) from Qiagen was used to process and analyse RNA-seq data. The reads were initially subjected to quality and adaptor trimming prior to mapping to the reference genome. Quality trimming was performed following these parameter settings: the limit of low quality sequence allowed was 0.05, only two ambiguous nucleotides were allowed and finally one terminal nucleotide was removed at the 3' and 5' end, respectively. Subsequent quality trimming, the reads were further subjected to the adapter trimming using the Trim adapter library RNA and Trim RNA library http://www.clcbio.com/products/clc-genomics-workbench/) from Qiagen under the following parameters: one nucleotide was removed at each terminal end of the 5' and 3' end and the read length below 15 and above 1000 was discarded. The resulting reads were mapped to the reference genome, *E. ruminantium* (Welgevonden) (Accession no: NC_005295) with the following mapping settings: minimum similarity fraction of 0.8, minimum length fraction of 0.8, maximum hit for a read was 10 and a type of organism prokaryote. The expression levels of genes were normalized by considering both the library size and gene length effects with respect to the RPKM values (reads per kilo base of gene model per million mapped reads) (Mortazavi et al., 2008; Hammac et al., 2014). Transcripts were considered to be highly expressed if RPKM values were above 3200 ($Log_{(RPKM+1)}$ value ≥ 4); while genes that had RPKM values above > 100 but below < 3200 ($Log_{(RPKM+1)}$ value ≥ 2 but ≤ 4) were considered expressed. Venn diagrams were drawn using the Draw Venn

diagram tool (htpp://bioinformatics.psb.ugent.be/webtools/venn/) using all genes with RPKM values ≥ 100.

Next, differential expression was investigated between the BovRBs vs BovEBs, and BovRBs vs TicEBs or TicEBs vs BovEBs data sets and represented as a fold change. The duplicates were combined of bovine EBs (EBsA and EBsB) and RBs (RBs21 and RBs24) and the tick EBs (Wtick1 and Wtick2) and the multi-group analyses tool in CLC was used to set up the experiments. The expression values were normalised in quantiles prior to statistical test analysis. To determine gene regulation across the developmental stages, Baggerly's T-test (Baggerly et al., 2003) or empirical analysis of digital gene expression (EDGE) (Robinson et al., 2010) was used to evaluate differential expression between BovRBs vs BovEBs, BovRBs vs TicEBs, and TicEBs vs BovEBs. The significantly expressed genes ($p \leq 0.01$, FC ≥ 2 or ≥ -2) were grouped according to their functional categories (data not shown). For the purpose of this study, only genes from membrane-associated genes, hypothetical or unknown proteins, pathogenicity-associated, exported or secreted proteins, transporter proteins were considered.

2.6 Reverse transcription quantitative real time PCR (RT-qPCR)

RT-qPCR was used to confirm the RNA-seq results from BovEBs, BovRBs, and TicEBs, and also to determine the transcription levels in these samples. The qPCR was performed as previously described (Kuriakose et al., 2011; Pruneau et al., 2012). For Kuriakose et al. (2011) method, gene expression was measured by cDNA quantification of each gene and normalized by subtracting the threshold value (Ct) from the total number of the PCR cycles. For Pruneau et al. (2012) method, absolute quantification was used to measure gene expression. Briefly, 9.8×10^1 to 9.8×10^7 copies/µl of the genomic DNA of *E. ruminantium* Welgevonden was used for standard calibration and was processed simultaneously the *E. ruminantium* cDNA samples. The *E. ruminantium* genomic copy

numbers were calculated using the pCS20 gene fragment as described previously (Steyn et al., 2008). For normalisation, the cDNA copies of each targeted *E. ruminantium* gene was divided by the number *E. ruminantium* (genomic DNA copies) present in each sample and analysed using the published calculations (Pruneau et al., 2012). The results were then expressed as Log_2(FC) and the positive Log_2(FC) indicated up-regulation in the EBs and the negative Log_2(FC) was vice versa. The complementary DNA (cDNA) was synthesized from total RNA using the Quantitect Reverse Transcriptase kit (Qiagen). The real time PCR was performed in a Rotor-Gene Q lightcycler (Qiagen) using the Rotor-Gene® SYBR® Green PCR kit (Qiagen). The primers used in this study are listed in Table S1. These were designed using the CLC Genomics Workbench 8.02 (http://www.clcbio.com/products/clc-genomics-workbench/) from Qiagen.

3. Results

3.1 Purifying and sequencing RNA from the EBs and RBs of *E. ruminantium* (Welgevonden) in mammalian and tick cell culture

After RNA depletion concentrations were significantly reduced in the rRNA depleted samples. Approximately 11-24 million and 10-75 million reads were recovered from the BovEBs and BovRBs replicates, respectively (Table 1). The reads were slightly reduced in TicEBs replicates; thus 462 000-900 000 reads were obtained. Only 4-5% and 1-13% of the total reads recovered from BovEBs and BovRBs respectively in mammalian cells were mapped to the reference genome. On the other hand, 4.15% of the total reads recovered from the TicEBs were mapped to the reference genome.

Table 1: The concentration of the RNAs and the read coverage of RNA-seq data obtained from the EBs and RBs of *E. ruminantium* (Welgevonden) strain.

Developmental stage	Sample id	Total RNA concentration (ng/µl)	rRNA depleted RNA concentration (ng/µl)	No of reads after trimming	Average read length after trimming	Unique reads	Unmapped reads	% of mapped reads
BovEBs	EBsA	159.5	4.76	24 901 385	92.4	1 183 459	23 717 926	4.8
	EBsB	125.9	7.5	11 564 633	92	626 195	10 919 438	5.4
BovRBs	RBs21	319.5	39.2	75 314 827	91	1 314 846	73 314 827	1.7
	RBs24	244.9	10.6	10 379 810	92.6	1442 848	9 934 632	13.9
TicEBs	Wtick1	46.4	4.63	462460	92	42222	420134	9.2
	Wtick2	76.2	6.0	988367	91	45733	942592	4.6

3.2. Comparative transcriptional analysis of *E. ruminantium* genes in the developmental stages in bovine and tick cells

RNA from the tick and bovine host cells' EBs and RBs was analysed by RNA-seq. Expression levels were quantified by calculating the RPKM of each gene (Figure 1). The non-coding RNAs, including the ribosomal RNAs (Erumr01, Erumr02, Erumr03 coding for 16S, 5S, and 23S rRNA respectively), the 36 tRNAs, *rnpB* (the RNA component of Ribonuclease P) and transfer-messenger RNA (tm-RNA) were detected in all the samples and removed from all subsequent analyses. In addition, reads that were mapped to the 32 pseudo-genes were not included in the final analyses and instead the reads were assigned to the original gene. Of the remaining 888 predicted *E. ruminantium* genes (Collins et al., 2005), reads were mapped to 604-725 (70-81%) and 825 (92%) genes in mammalian and tick cell culture, respectively. A Venn diagram was plotted using *E. ruminantium* expressed genes ($Log_{(RPKM+1\geq 2)}$) in the developmental stages in mammalian and tick cells (Figure 2a). The majority of genes were shared between the datasets of which several genes were highly expressed (Table 2), while others were exclusively expressed at each developmental stage (Table S3). The *E. ruminantium* genes exclusively expressed in each dataset were grouped into their functional categories as described previously (Collins et al., 2005) (Figure 2b). Hypothetical or unknown genes and genes involved in information transfer were amongst the highest categories of expressed genes.

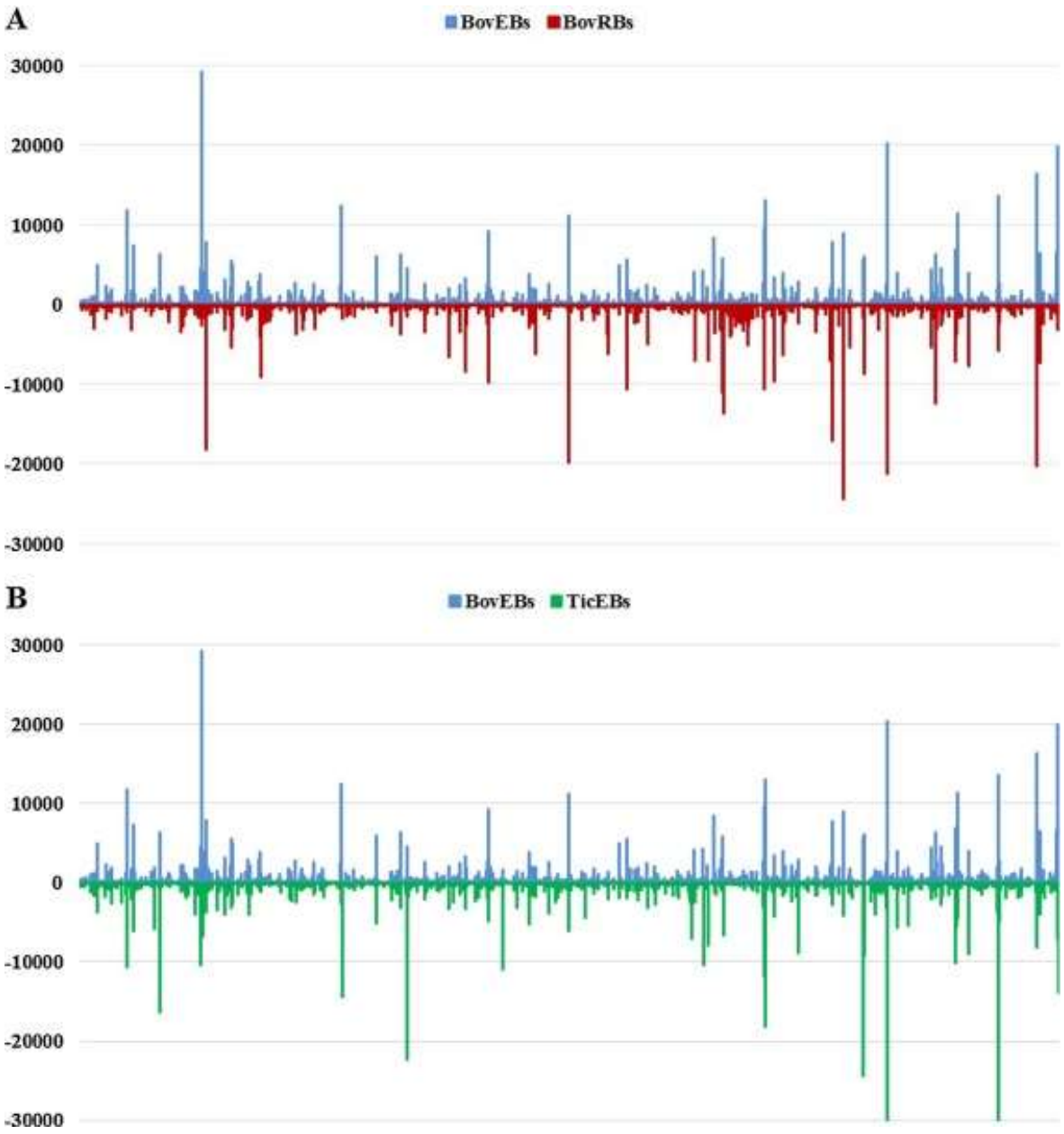


Figure 1. Whole-genome comparison of transcriptional activity of *E. ruminantium* in the bovine and tick cells. The 888 protein coding genes are arranged on the X axis from left to right according to the published *E. ruminantium* (Welgevonden) genome. Normalised expression values are plotted on the Y axis; those for the bovine EBs are shown in blue in the upper part of each graph and values for the bovine RBs are shown in red in the lower part of A and those for Tick EBs in green in the lower part of B.

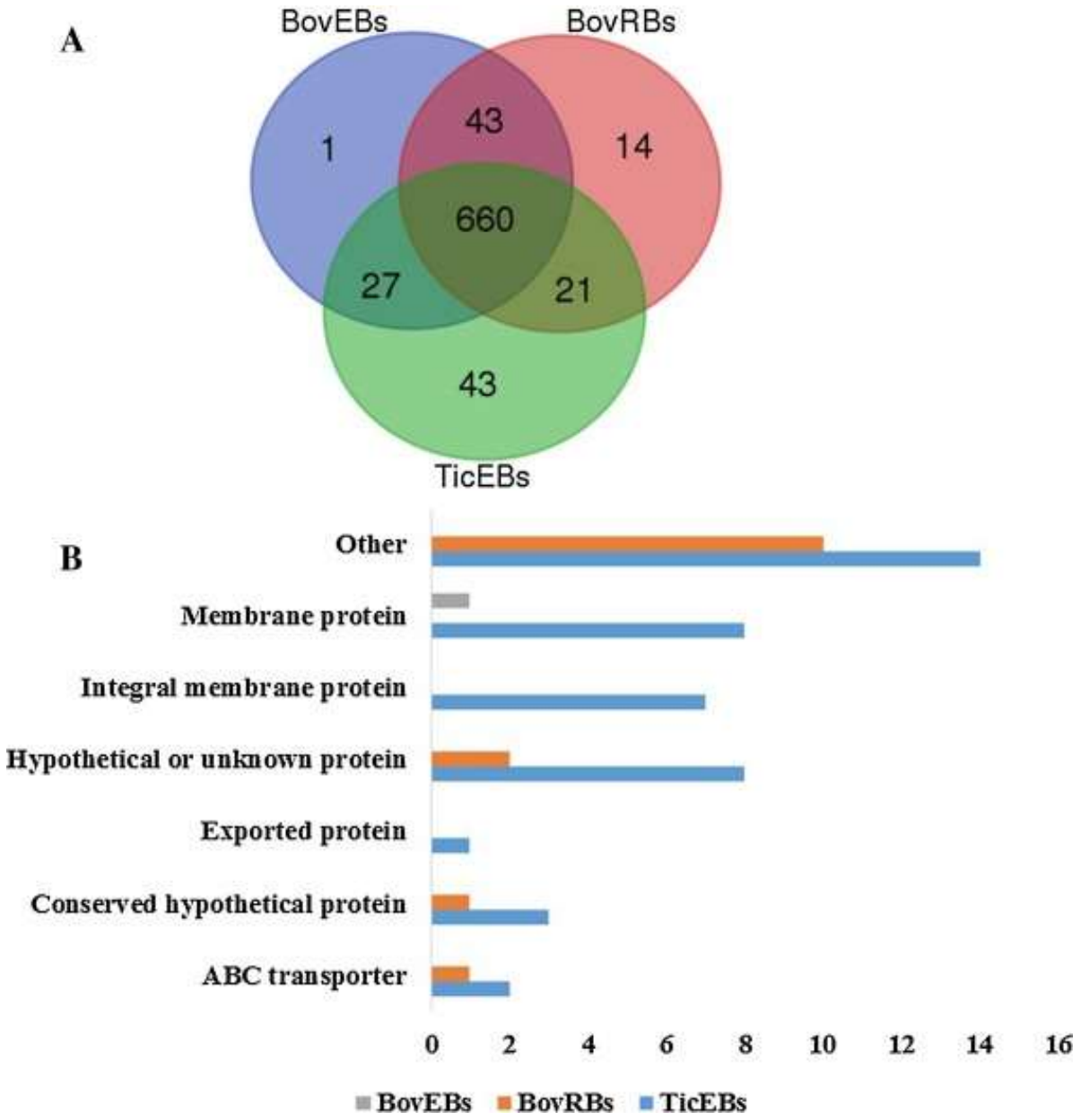


Figure 2: Venn diagram analyses (A) of *E. ruminantium* genes that were expressed (RPKM ≥100) in BovEBs, BovRBs, or TicEBs or shared between datasets. The genes exclusively expressed in each dataset were grouped into their functional categories (B).

Table 2: *E. ruminantium* genes that were highly expressed in all three samples analysed (BovEBs, BovRBs and TicEBs) from mammalian and tick cell cultures.

Erum no	Functional annotation	Gene id	BovEBs RPKM	BovEBs $Log_{(RPKM+1)}$	BovRBs RPKM)	BovRBs $Log_{(RPKM+1)}$	TicEBs RPKM	TicEBs $Log_{(RPKM+1)}$
Erum1150	Hypothetical or unknown protein		7391	4	18220	4	7920	4
Erum2930	DNA-binding protein HU-beta	*hupB*	6296	4	3797	4	3245	4
Erum3530	30S ribosomal protein S15	*rpsO*	3275	4	8405	4	3407	4
Erum3730	Hypothetical or unknown protein		9214	4	9839	4	4869	4
Erum4470	Exported protein		11209	4	19915	4	6156	4
Erum5870	30S ribosomal protein S13	*rpsM*	5835	4	13630	4	6658	4
Erum6230	Exported protein		9631	4	10671	4	11844	4
Erum6320	Hypothetical or unknown protein		3432	4	9670	4	4273	4
Erum6970	Hypothetical or unknown protein		8974	4	24373	4	4150	4
Erum7160	Membrane protein		6098	4	8647	4	9220	4
Erum7380	Membrane protein		20294	4	21198	4	397435	6
Erum7990	Integral membrane protein		6916	4	7175	4	10135	4
Erum8000	Integral membrane protein		3928	4	3774	4	5634	4
Erum8110	Integral membrane protein		3990	4	7779	4	9009	4
Erum8380	Probable ATP synthase B subunit	*atpF*	13630	4	5839	4	31590	4
Erum8740	Major antigenic protein MAP1	*map1*	16351	4	20217	4	8222	4
Erum8770	Hypothetical or unknown protein		6509	4	7256	4	4065	4
Erum8920	Integral membrane protein		6650	4	3200	4	7193	4

3.3 Differential expression analyses

3.3.1 Functional categories of differentially expressed genes

Next, differential expression was investigated between the BovRBs vs BovEBs, TicEBs vs BovEBs and BovRBs vs TicEBs datasets and represented as a fold change. Several *E. ruminantium* genes were significantly differentially expressed in the BovEBs, BovRBs or TicEBs. Analysis of global transcriptional profiles showed that hypothetical or uncharacterised proteins, membrane associated and genes involved in information transfer were amongst the highest categories of the differentially expressed genes (Figure 3).

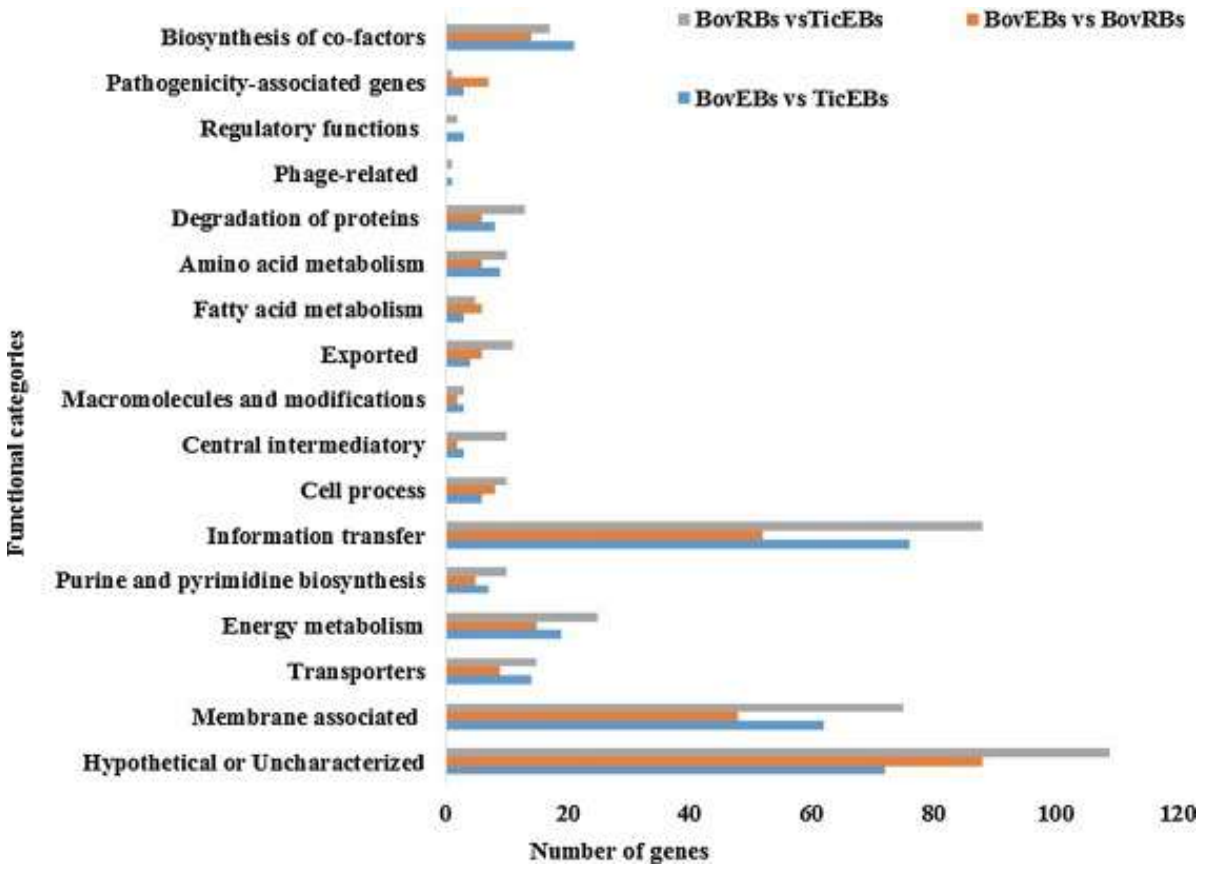


Figure 3: Functional categories of differentially expressed genes in the EBs and RBs of *E. ruminantium* in mammalian and EBs in tick cell culture.

3.3.2 The *map1* family

Transcripts for all the *map1* paralogs could be detected with the highest levels of expression RPKM values detected for *map1* and *map1+1* (Figure S1). However, these were differentially transcribed when the data sets were compared (Figure 4). In BovEBs *map1-8* was significantly up-regulated when compared to BovRBs and *map1, map1+1, map1-5* and *map1-9,* were significantly up-regulated when compared TicEBs. In BovRBs *map1-2* and *map1-3* were up-regulated when compared to BovEBs, and *map1-5* and *map1+1* when compared to TicEBs. Likewise, *map1-1* and *map1-14* were significantly up-regulated in TicEBs when compared to both BovEBs and BovRBs.

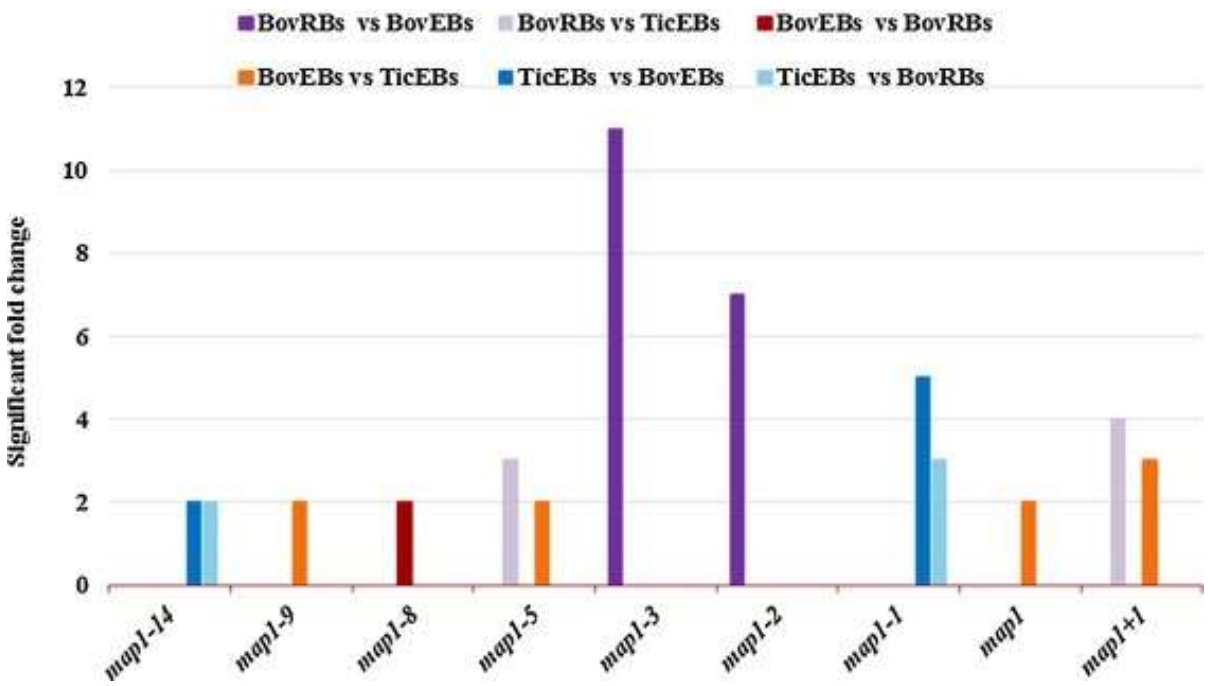


Figure 4: Up-regulated transcripts of the *map1* family genes determined by comparison between all three data sets. Only genes with significant fold changes are shown.

3.3.3 Type IV secretion system proteins

Transcripts for all the genes that encoded for the type IV secretion system protein (T4SS) were detected at RPKM levels above 100 (Figure S2). The expression levels was much

higher in the EBs originating both from tick and bovine cells in comparison with the BovRBs. There was also no significantly up-regulated T4SS transcripts detected for BovRBs in comparison with either BovEBs or TicEBs (Figure 5). In contrast, *virB3, virB6,* Erum 4410 and Erum7980, were up-regulated in BovEBs and TicEBs when compared to BovRBs. In addition, *virD4, virB4, virB9* and v*irB10* were expressed at much higher levels in BovEBs when compared to TicEBs and/or BovRBs. Erum5230 was only up-regulated in TicEBs.

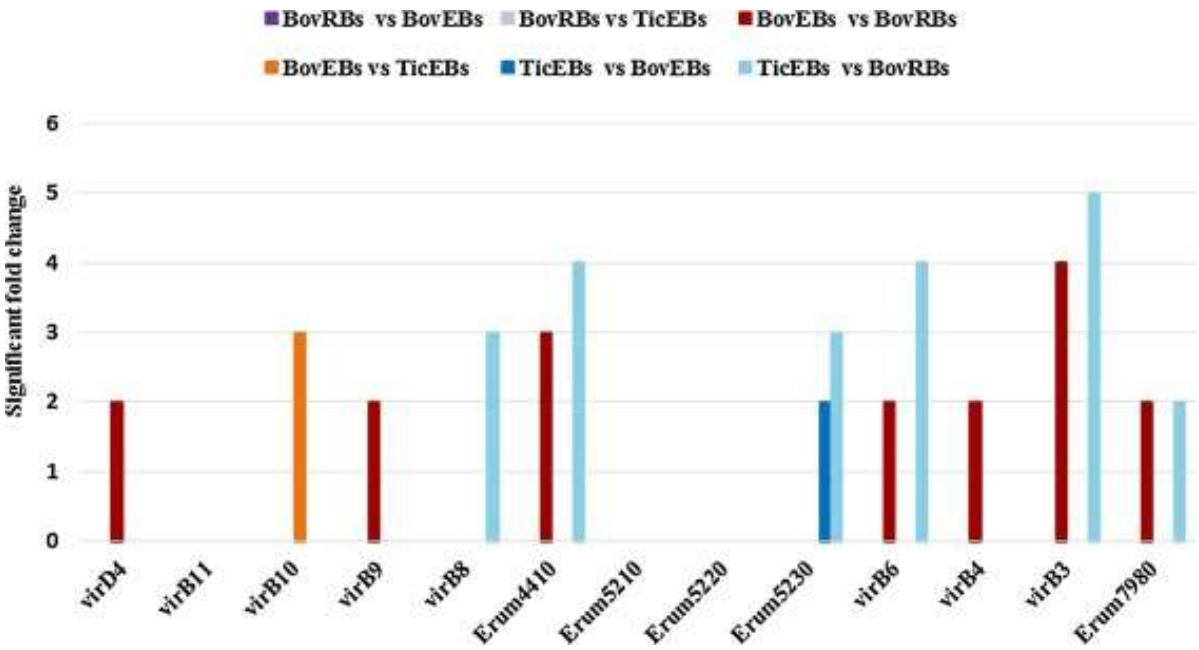


Figure 5: Up-regulated transcripts of the Type IV secretion system determined by comparison with all three data sets. Only genes with significant fold changes are shown.

3.3.4 Genes encoding proteins of unknown function and containing ankyrin or tandem repeats

Four genes that contain tandem repeats (TR) or ankyrin repeats (Erum0250, Erum1110, Erum3730, and Erum8770) had higher RPKM values compared with other genes containing these repeats (Figure S3). Erum0250, Erum1110, Erum3750 and Erum8770 were significantly up-regulated in BovEBs when compared to BovRBs or TicEBs (Figure 6).

Erum0320, Erum0660 and Erum1040 were specifically up-regulated in TicEBs when compared to both BovEBs and BovRBs.

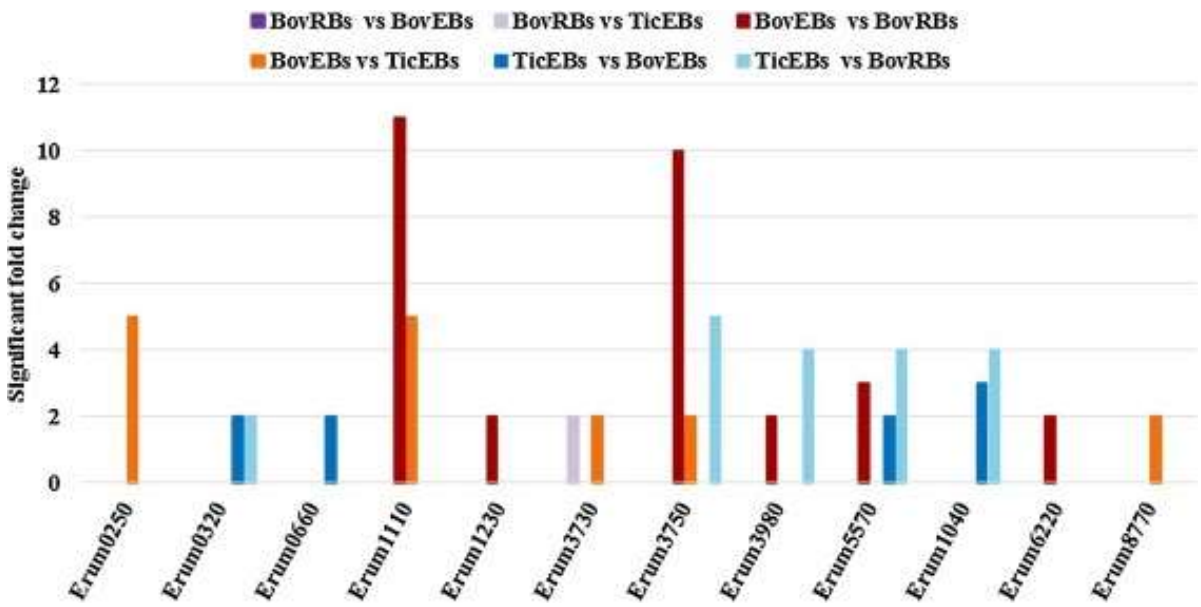


Figure 6: Up-regulated transcripts of genes encoding proteins of unknown function and containing ankyrin or tandem repeats or TPRs, determined by comparison with all three data sets. Only genes with significant fold changes are shown.

3.3.5 Exported proteins

Exported proteins Erum7110 and Erum8090 were significantly up-regulated in BovEBs (Figure S4). Erum1540, Erum3450, Erum4470, Erum5010 and Erum7970 were significantly up-regulated in BovRBs and BovEBs when compared to TicEBs. On the other hand, Erum1070, Erum1460, Erum2320, Erum5000, Erum6650, Erum7120 and Erum8220 were up-regulated in TicEBs.

3.3.6 ABC transporters

Five genes encoding ABC transporter proteins were differentially transcribed. Erum2550 was significantly up-regulated in BovRBs and TicEBs (Figure S4), while Erum2580 was

up-regulated in TicEBs. Erum1490 was significantly up-regulated in TicEBs, while Erum5760 and Erum6270 were up-regulated in BovRBs.

3.4 RT-qPCR validation of RNA-seq data

Differential gene expression was validated by qPCR using the two protocols described previously (Kurikose et al., 2011; Pruneau et al., 2012). qPCR was performed on a set of 8 *E. ruminantium* genes that were significantly expressed according to the RNA-seq data (Erum0440 (*dksA*) Erum0500, Erum1110, Erum2380, Erum6240, Erum8010, Erum8380 (*atpF)* and Erum8930). These genes were up-regulated in BovEBs and were down-regulated in BovRBs. The qPCR data was converted by subtracting the Ct (threshold cycle) value within and between different developmental stages from forty (total PCR cycles), since lower threshold correspond to higher transcript levels. Similarly, the Log_2 (FC) for the selected genes correlated with RNA-seq data (Table 3).

Table 3: *E. ruminantium* genes that were differentially expressed in BovEBs and BovRBs by qPCR compared to RNA-seq.

	qPCR		RNA-seq		
Gene id	**Fold change (FC) (R_{EBs}/R_{RBs})**	**Log_2(FC)**	**BovEBs normalized expression value**	**BovRBs normalized expression value**	**Fold change**
Erum0440	92	7	11808	973	9
Erum0500	8.6	3	7391	456	12
Erum1110	45.4	6	29173	2638	14
Erum2380	9.47	3	12451	1847	6
Erum6240	13.5	4	13069	865	11
Erum8010	14.5	4	11383	1584	5
Erum8380	9.5	3	13630	5839	3
Erum8930	23	5	19915	1362	9

4. Discussion

The use of bioinformatics tools to search for additional vaccine candidates against heartwater has been previously demonstrated (Liebenberg et al., 2012). Although, some of the predicted antigens have been tested previously as cell-mediated DNA vaccines (Pretorius et al., 2007,

2008; Sebatjane et al., 2010; Thema et al., 2016), these have displayed limited or no protection against a natural tick challenge. Our current approach focused on studying gene expression using transcriptome sequencing at both the developmental stage *in vitro* in mammalian and tick cells. This will be beneficial in the future vaccine design because it will focus on knowledge of proteins or genes highly expressed at all growth stages and in different hosts, as well as those expressed exclusively in each.

BA886 and IDE8 cells were successfully infected with the *E. ruminantium* Welgevonden strain. The EBs and RBs were collected from infected bovine and EBs from tick cells and the RNA extracted from these was subjected to transcriptome sequencing. Less than 13% of the total reads obtained from the EBs and RBs were mapped to the reference genome, thus indicating the presence of host RNA in the RNA-seq data. Moreover, non-coding RNAs (rRNA, tmRNA, tRNA) were also detected despite rRNA depletion treatment. Regardless, EB-specific and/or RB-specific reads were mapped to 70-90% of *E. ruminantium* genes in bovine and tick cells thus indicating the reliability of using RNA-seq to measure gene expression (Wang et al., 2009). When the RPKM values were considered, majority of the genes were shared between the datasets, however there seemed to be a percentage of *E. ruminantium* genes that were uniquely expressed in the TicEBs. This finding suggested that *E. ruminantium* expresses different genes in different host cells.

Membrane-associated genes expressed in the EBs (BovEBs and TicEBs) may be essentially required for host cell invasion and these can be targeted with antibodies. The MAP1 protein is the most studied immunodominant protein of *E. ruminantium*. In depth transcriptome analysis of the MAP1 protein family revealed that transcripts for all the *map1* family paralogs were detected in the BovEBs, BovRBs and TicEBs. This is in contrast to microarray analyses which only reported the detection of *map1* (Emboulé et al., 2009) and *map1-6* (Pruneau et al.,

2012). In our study, *map1*, *map1+1*, *map1-13* and *map1-14* transcripts were expressed at higher RPKM levels than the other paralogs. This correlated with EB proteome studies were these MAP1 proteins were also detected in *E. ruminantium* Gardel cultured in endothelial cells (Marcelino et al., 2015; Moumène et al., 2015). Differential expression analyses revealed that *map1*, *map1-8* and *map1-9* genes were up-regulated in the BovEBs when compared to BovRBs or TicEBs. In contrast, *map1+1*, *map1-2*, *map1-3*, and *map1-5* were up-regulated in BovRBs when compared to BovEBs or TicEBs. Finally, *map1-1* and *map1-14* were up-regulated in TicEBs when compared to BovEBs or TicEBs. *Map1-1* mRNA was previously detected in *Amblyomma* ticks during feeding (Postigo et al., 2008). The paralogs of MAP1 have been shown to be differentially expressed *in vivo* or *in vitro* in mammalian or tick host cells (Bekker et al., 2002; 2005; Postigo et al., 2007; 2008, Van Heerden et al., 2004). Here and in previous studies different sets of the paralogs were expressed at a given time point or by various host cells confirming the notion that differential expression of *map1* genes may play a role in immune evasion. Thus, *E. ruminantium* induces high expression of MAP1 related proteins that in turn induces high but unprotective antibodies (Marcelino et al., 2015). The MAP1 protein family seems to be essential for *E. ruminantium* intracellular survival and adaptation, however, their functions are still unknown to date and need to be further investigated. It should be noted that, in this study, an *I. scapularis* tick cell line was used. Although, these ticks are susceptible to *E. ruminantium* infection, these are not vectors of *E. ruminantium*. Thus, it remains to be determined whether these genes will be up-regulated in *Amblyomma* cells.

Erum3750 and Erum1110 were up-regulated in the EBs (BovEBs and TicEBs). Erum3750 has ankyrin repeats and these have been implicated with attachment and adhesion and thus is a good vaccine candidate (Meunier et al., 2016). A homolog of Erum1110 was previously implicated with attachment to tick cells (De la Fuente et al., 2004). Previous studies have

reported that TR-proteins elicit a strong protective antibody response during infection that is directed at continuous species-specific epitopes located at the TR region (Lina et al., 2016). Ankyrin-domain containing gene *AnkA* in *E. chaffeensis* is involved with host-pathogen interaction (Wakeel et al., 2010). Thus, it is possible that the TR or ankyrin-repeat containing genes that were up-regulated in both BovEBs and TicEBs are essentially required for attachment to mammalian or tick host cells. Consequently, antibodies directed to these proteins may impede attachment and infection. Another hypothetical gene, Erum8770 was highly expressed in all the developmental stages in bovine and/or tick cells. Recombinant rErum8770 induced interferon (IFN)-γ protein production in the PBMC collected from previously immunised sheep but failed to induce the expression of other T-helper1 (Th1) cytokines (Liebenberg et al., 2012). IFN-γ is a potent inhibitor of *E. ruminantium* growth *in vitro* (Totté et al., 1997*).* Thus, it was postulated that the overexpression of this gene in both the EBs and RBs could possibly suggest that it may be involved in pathogenicity and adaptation into the host cells and is a good vaccine candidate.

Exported or secreted proteins exclusively expressed in BovRBs may be ideal targets for CTL or cell mediated immunity. RBs are non-infectious replicative forms embedded inside the host cytoplasmic vacuoles. For the host immune response to recognise antigens at this stage, these must be secreted outside cells or expressed on the surface of the infected cells. Eight exported proteins were identified that were up-regulated in BovRBs of which Erum1960 and Erum2310 were exclusive. Thus, these genes may possibly be ideal targets for CTL immunity.

Lina et al. (2016) reviewed that pathogenicity-associated genes (e.g. type IV secretion system proteins (T4SS)) play an important role in the in *E. chaffeensis* growth and virulence. Thus, understanding the development as well as pathogenicity of *E. ruminantium* may play a crucial

role in vaccine development. Transcripts for the pathogenicity-associated T4SS proteins VirB9, VirB10 and VirB4 were up-regulated in BovEBs and correlates to protein detected on the outer membrane protein fraction of EBs (Moumène et al. 2015). Marcelino et al. (2015) identified VirB10, VirB11, VirB3, VirB6, VirB8 and VirD4 in the EBs proteome of the virulent Gardel strain and all these transcripts except *virB11* were detected in our study. Another gene that was implicated previously with pathogenicity in *Salmonella typhimurium* (Nakanishi et al., 2006), *dksA,* was up-regulated in the EBs (BovEBs and TicEBs) when compared to the RBs (results not shown). This gene was also found to be up-regulated in *E. ruminantium* Gardel using microarray analyses (Pruneau et al., 2012). Thus, these findings suggest that *E. ruminantium* expresses virulence factors such as T4SS and *dksA* in the tick and mammalian host cells.

Three genes coding for ABC transporters were up-regulated in TicEBs. Two of these were components of the 1H12 DNA recombinant vaccine (Erum2550 and Erum2580) that induced complete protection in sheep after a lethal homologous needle challenge (Pretorius et al., 2008). Thus, the additional ABC transporter shown to be up-regulated in TicEBs should also be investigated as a vaccine candidate.

The expression profiles were validated on selected genes by qPCR. Several genes (*16S, ffh, rpoD*) to be used as internal reference for qPCR for other pathogens have been described (Rocha et al., 2015). However, similar to Pruneau et al. (2012) since we used two different developmental stages, constitutively expressed *E. ruminantium* genes to use as normalisers for the relative quantification of gene expression could not be identified. As a result, gene expression was measured in two ways: by cDNA quantification of each gene and normalised by subtracting the threshold (Ct) value from the total number of the PCR cycles as described previously (Kuriakose et al., 2011) and by using the serial dilution of *E. ruminantium*

genomic DNA copies as a standard calibrator to determine each copy of the expressed gene as previously described (Pruneau et al., 2012). Both methods proved to be effective in validating the RNA-seq data. However, in our view, the method described in Pruneau et al. (2012) seems to be the more accurate of the two. This method not only calculates the copy numbers of the targeted gene, but also provides an estimate transcriptional fold change. Future studies should focus on identifying stably expressed *E. ruminantium* genes using other web-based tools such as RefFinder (https://omictools.com/reffinder-tool).

RNA- or transcriptome-seq has proven to be a useful tool in studying gene expression in *E. ruminantium in vitro.* Although this technique is sensitive and high throughput, it has limitations. For example, low quantity of bacterial mRNA and eukaryotic mRNA contamination were observed in this study. Thus, future studies should consider these factors when using this approach in determining gene expression in prokaryotes. Perhaps, more effort should be focused on the development of new capture methods or improvement of current bacterial mRNA enrichment techniques.

In conclusion, our study has revealed that several *E. ruminantium* genes were differentially expressed *in vitro* during the developmental stages in bovine and tick cell culture. Furthermore, pathways that might be important for *E. ruminantium* development and pathogenicity were also identified. Those genes exclusively expressed in BovEBs or TicEBs may potentially be used for attachment to host cells and stimulate both cell mediated and humoral immune responses. Genes responsible for adaption of bacterial pathogens in the host cells were identified in the RBs. Our study also revealed high energy metabolism in the EBs in both tick and bovine host cells. The role of these genes in *E. ruminantium* survival requires further investigation. Further analysis of these differentially expressed gene functions *in vivo,*

in combination with comparative genomic and proteomic approaches might provide a better view of mechanisms of *E. ruminantium* infection.

Acknowledgements

The authors acknowledge the financial support received from the Economic Competitiveness Support Programme and the ARC-OVI, South Africa (2012-2015). We thank Dr Lesley Bell-Sakyi of the Centre for Tropical Veterinary Medicine, Royal School of Veterinary studies, University of Edinburg for providing the tick cells (IDE8).

Author's contribution

MA Tjale planned the study and did most of the experimental lab work, the data analyses as well as the preparation of manuscript. A. Pretorius, J. Liebenberg and M. van Kleef were the study supervisors, contributed to the planning of the study and the editing of the manuscript. A. Josemans provided training for endothelial cell culture work and assisted with the tick cell culture work as well as preparation of the manuscript.

References

Allsopp, B.A., 2015. Heartwater - *Ehrlichia ruminantium* infection. Rev. Sci. Tech. 34, 557-556.

Baggerly, K., Deng, L., Morris, J., Aldaz, C., 2003. Differential expression in SAGE: accounting for normal between-library variation**.** Bioinformatics. 19, 1477-1483.

Barré, N., Garris, G.I., Borel, G., Camus, E., 1988. Hosts and population dynamics of *Amblyomma variegatum* (Acari: Ixodidae) on Guadeloupe, French West Indies. J. Med. Entomol. 25, 111-115.

Bekker, C.P., Bell-Sakyi, L., Paxton, E.A., Martinez, D., Bensaid, A., Jongejan, F., 2002. Transcriptional analysis of the major antigenic protein 1 multigene family of *Cowdria ruminantium.* Gene. 285, 193-201

Bekker, C.P., Postigo, M., Taoufik, A., Bell-Sakyi, L., Ferraz, C., Martinez D., Jongejan, F., 2005. Transcription analysis of the major antigenic protein 1 multigene family of three *in vitro*-cultured *Ehrlichia ruminantium* isolates. J. Bacteriol. 187, 4782-4791.

Bell-Sakyi, L., Paxton, E.A., Munderloh, U.G., Sumption K.J., 2000. Growth of *Cowdria ruminantium*, the causative agent of heartwater, in a tick cell line. J. Clin. Microbiol. 38, 1238-1240.

Collins, N.E., Liebenberg, J., de Villiers, E.P., Brayton, K.A., Louw, E., Pretorius, A., Faber, F.E., Van Heerden, H., Josemans, A., Van Kleef, M., Steyn, H.C., van Strijp, M.F., Zweygarth, E., Jongejan, F., Maillard, J.C., Berthier, D., Botha, M., Joubert, F., Corton, C.H., Thomson, N.R., Allsopp, M.T., Allsopp, B.A., 2005. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. Proc. Natl. Acad. Sci. 102, 838-843.

De la Fuente, J., Garcia-Garcia, J.C., Barbet, A.F., Blouin, E.F., Kocan, K.M., 2004. Adhesion of outer membrane proteins containing tandem repeats of *Anaplasma* and *Ehrlichia* species (Rickettsiales; Anaplasmataceae) to tick cells. Vet. Microbiol. 98, 313-322.

Du Plessis, J.L., 1985. A method of determining the *Cowdria ruminantium* infection rate of *Amblyomma hebraeum*: effects in mice injected with tick homogenates. Onderstepoort. J. Vet. Res. 52, 55-61.

Emboulé, L., Dailgle, F., Meyer, D.F., Mari, B., Pinarello, V., Sheikboudou, C., Magnone, V., Frutos, R., Viari, A., Barbry, P., Martinez, D., Lefrançois, T., Vachiéry, N., 2009. Innovative approach for transcriptomic analysis of obligate intracellular pathogen: selective capture of transcribed sequences of *Ehrlichia ruminantium.* BMC Mol. Biol. 10, 111-118.

Filiatrault, M.J., 2011. Progress in prokaryotic transcriptomics. Curr. Opin. Microbiol. 14, 579-586.

Hammac, G.K., Pierle, S.A., Cheng, X., Scoles, G.A., Brayton, K.A., 2014. Global transcriptional analysis reveals surface remodeling of *Anaplasma marginale* in the tick vector. Parasit. Vectors. 7, 193-198.

Humphrys, M.S., Creasy, T., Sun, Y, Shetty, A.C., Chibucos, M.C., Drabek, E.F., Fraser, C.M., Farooq, U., Sengamalay, N., Ott, S., Shou, H., Bavoil, P.M., Mahurkar, A., Myers, G.S., 2013. Simultaneous transcriptional profiling of bacteria and their host cells. PloS One. 48, e8059.

Jongejan, F., Zandbergen, T.A., van de Wiel, P.A., de Groot, M., Uilenberg, G., 1991. The tick-borne rickettsia *Cowdria ruminantium* has a *Chlamydia*-like developmental cycle Onderstepoort. J. Vet. Res. 58, 227-237.

Kuriakose, J.A., Miyashiro, S., Luo, T., Zhu, B., McBride, J.W., 2011. *Ehrlichia chaffeensis* transcriptome in mammalian and arthropod hosts reveals differential expression and post transcriptional regulation. PloS One. 6, e24136.

Liebenberg, J., Pretorius, A., Faber, F.E., Collins, N.E., Allsopp, B.A., Van Kleef, M., 2012. Identification of *Ehrlichia ruminantium* proteins that activate cellular immune responses using a reverse vaccinology strategy. Vet. Immunol. Immunopathol. 145, 340-349.

Lina, T.T., Farris, T., Luo, T., Mitra, S., Zhu, B., McBride, J.W., 2016. Hacker within! *Ehrlichia chaffeensis* effector driven phagocyte reprogramming strategy. Front. Cell. Infect. Microbiol. 6, 58.

Marcelino, I., Ventosa, M., Pires, E., Muller, M., Lisacek, F., Lefrançois, T., Vachiery, N., Coelho, A.V., 2015. Comparative proteomic profiling of *Ehrlichia ruminantium* pathogenic strain and its high-passaged attenuated strain reveals virulence and attenuation-associated proteins. PloS One. 10, e0145328.

Meunier, M., Guyard-Nicodème, M., Hirchaud, E., Parra, A., Chemaly, M., Dory, D., 2016. Identification of novel vaccine candidates against Campylobacter through reverse vaccinology. J. Immunol. Res. e5715790.

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.

Moumène, A., Marcelino, I., Ventosa, M., Gros, O., Lefrançois, T., Vachiéry, N., Meyer, D.F., Coelho, A.V., 2015. Proteomic profiling of the outer membrane fraction of the obligate intracellular bacterial pathogen *Ehrlichia ruminantium.* PloS One. 10, e0116758.

Munderloh, U.G., Kurtti, T.J., 1989. Formulation of medium for tick cell culture. Exp. Appl. Acarol. 7, 219-229.

Nakanishi, N., Abe, H., Ogura, Y., Hayashi, T., Tashiro, K., Kuhara, S., Sugimoto, N., Tobe, T., 2006. ppGpp with DksA controls gene expression in the locus of enterocyte effacement (LEE) pathogenicity island of enterohaemorrhagic *Escherichia coli* through activation of two virulence regulatory genes. Mol. Microbiol. 61, 194–205.

Neidhardt, F.C., Umbarger, H.E., 1996. Chemical composition of *Escherichia coli*, In: Neidhardt, F.C. (Ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology 1, ASM Press, Washington DC, pp. 13–16.

Postigo, M., Taoufik, A., Bell-Sakyi, L., de Vries, E., Morrison, W.I., Jongejan, F., 2007. Differential transcription of the major antigenic protein 1 multigene family of *Ehrlichia ruminantium* in *Amblyomma variegatum* ticks. Vet. Microbiol. 122, 298-305.

Postigo, M., Taoufik, A., Bell-Sakyi L., Bekker, C.P., de Vries, E., Morrison, W.I., Jongejan, F., 2008. Host cell-specific protein expression *in vitro* in *Ehrlichia ruminantium*. Vet. Microbiol. 128, 136–147.

Pretorius, A., Collins, N.E., Steyn, H.C., Van Strijp, F., Van Kleef, M., Allsopp, B.A., 2007. Protection against heartwater by DNA immunisation with four *Ehrlichia ruminantium* open reading frames. Vaccine. 25, 2316–2324.

Pretorius, A., Van Kleef, M., Collins, N.E., Tshikudo, N., Louw, E., Faber, F.E., Van strijp, M.F., Allsopp, B.A., 2008. A heterologous prime/boost immunisation strategy protects against virulent *E. ruminantium* Welgevonden needle challenge but not against tick challenge. Vaccine. 26, 4363-4371.

Pruneau, L., Emboulé, L., Gely, P., Marcelino, I., Mari, B., Pinarello, V., Sheikboudou, C., Martinez, D., Diagle, F., Lefrancois, T., Meyer, D.F., Vachiery, N., 2012. Global gene expression profiling of *Ehrlichia ruminantium* at different stages of development. FEMS. Immunol. Med. Microbiol. 64, 66-73.

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

Rocha, D.J.P., Santos, C.S., Pacheco, L.G.C., 2015. Bacterial reference gene expression studies by RT-qPCR: survey and analysis. Antonie Leeuwenhoek. 108, 685-693.

Rossetti, C.A., Galindo, C.L., Garner, H.R., Adams, L.G., Selective amplification of ***Brucella melitensis*** mRNA from a mixed host-pathogen total RNA. BMC. Res. Notes. 3, 244.

Sebatjane, S.I., Pretorius, A., Liebenberg, J., Steyn, H., Van Kleef, M., 2012. *In vitro* and *in vivo* evaluation of five low molecular weight proteins of *Ehrlichia ruminantium* as potential vaccine components. Vet. Immun. Immunopathol. 137, 217-225.

Sorek, R., Cossart, P., 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. Nat. Rev. Genet 11, 9-16.

Steyn, H.C., Pretorius, A., McCrindle, C.M.E., Steinmann, C.M.L., Van Kleef, M., 2008. A quantitative real-time PCR assay for *Ehrlichia ruminantium* using the pCS20. Vet. Microbiol. 131, 258-26.

Thema, N., Pretorius, A., Tshilwane, S.I., Liebenberg, J., Steyn, H., Van Kleef, M., 2016. Cellular immune response induced *in vitro* by *Ehrlichia ruminantium* secreted proteins and identification of vaccine candidate peptides. Onderstepoort. J. Vet. Res. 83, a1170.

Totté, P., McKeever, D., Martinez, D., Bensua, A., 1997. Analysis of T-cell responses in cattle immunized against heartwater by vaccination with killed elementary bodies of *Cowdria ruminantium*. Infect. Immun. 65, 236-239.

Van Heerden, H., Collins, N.E., Brayton, K.A., Rademeyer, C., Allsopp, B.A., 2004. Characterization of a major outer membrane protein multigene family in *Ehrlichia ruminantium.* Gene. 330, 159-168.

Wakeel, A., Zhu, B., Yu, X.J., McBride, J.W., 2010. New insights into molecular *Ehrlichia chaffeensis*-host interactions. Microbes. Infect. 12, 337-345.

Wang, Z., Gerstein, V., Snyder, V., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet 10, 57-63.

Walker, J.B., Olwage, A., 1987. The tick vectors of *Cowdria ruminantium* (Ixodoidea, Ixodidae, genus *Amblyomma*) and their distribution. Onderstepoort. J. Vet. Res. 54, 353-379.

Zweygarth, E., Vogel, S.W., Josemans, A.I., Horn, E., 1997. *In vitro* isolation and cultivation of *Cowdria ruminantium* under serum-free culture conditions. Res. Vet. Sci. 63, 161-164.

Supplementary Material

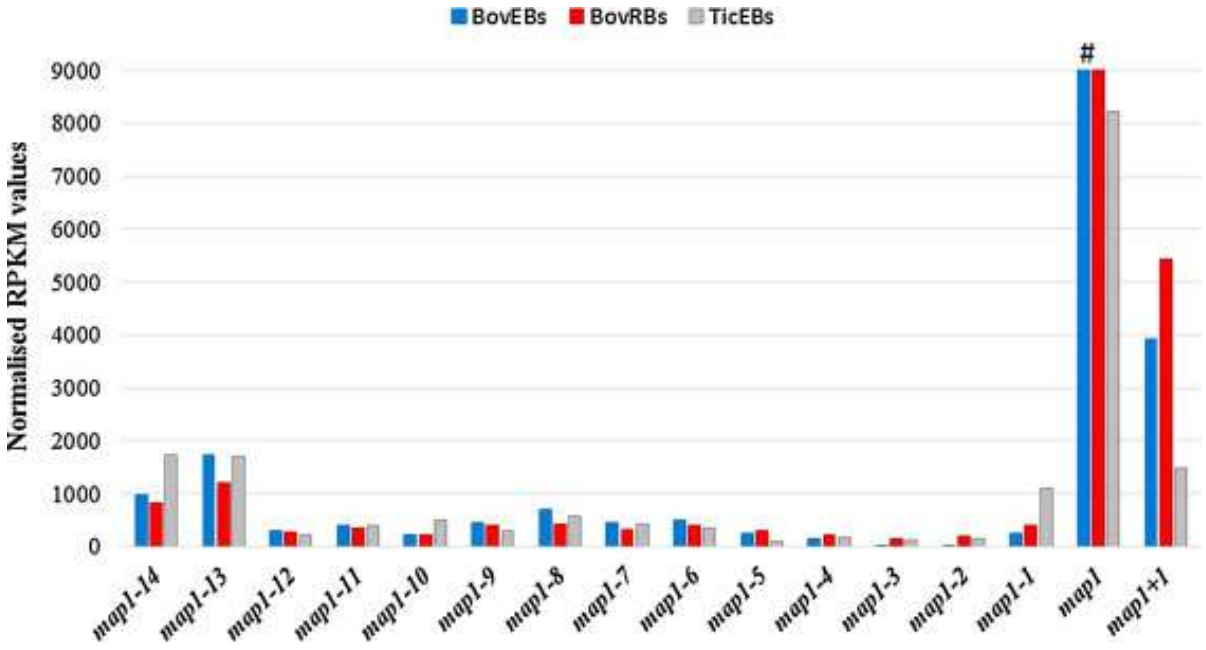
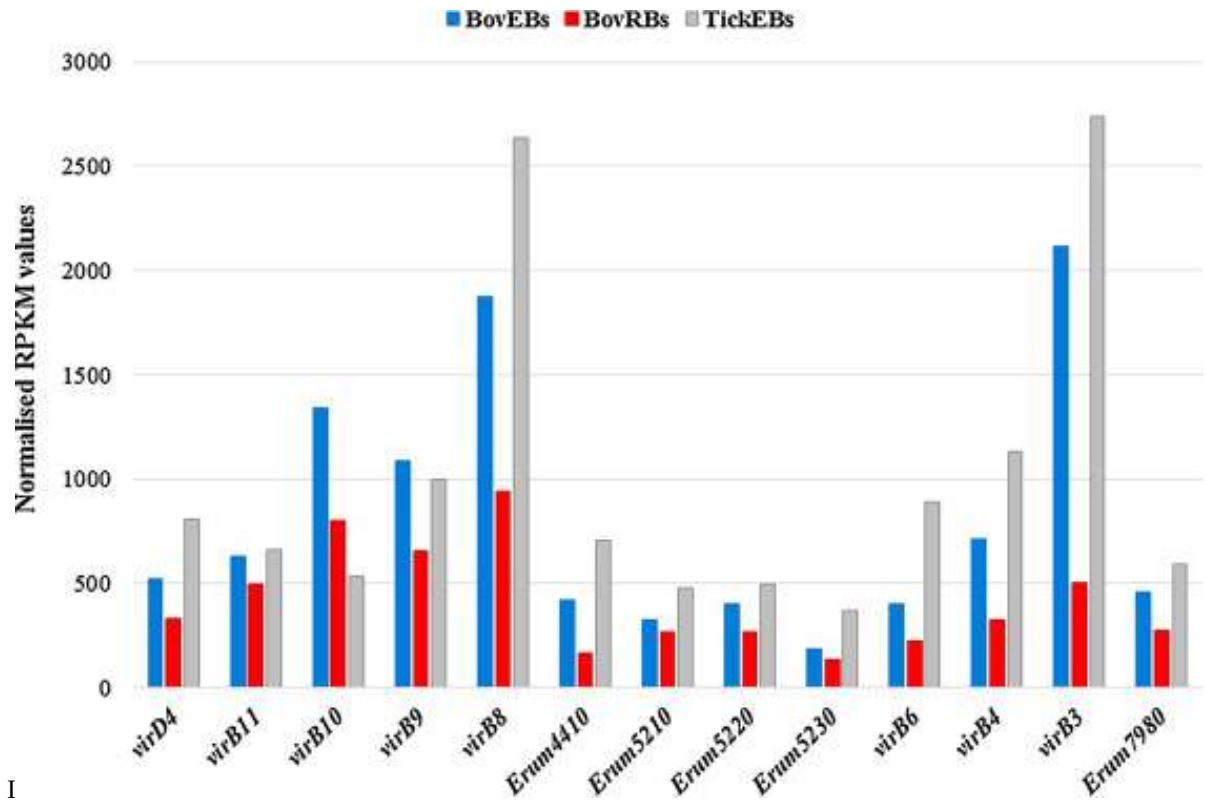


Figure S1: Normalised RPKM values of the *map1* family genes. #*map1* RPKM values for BovEBs = 16351 and BovRBs = 20217.



I

Figure S2: Normalised RPKM values of the genes coding for Type IV secretion system proteins.

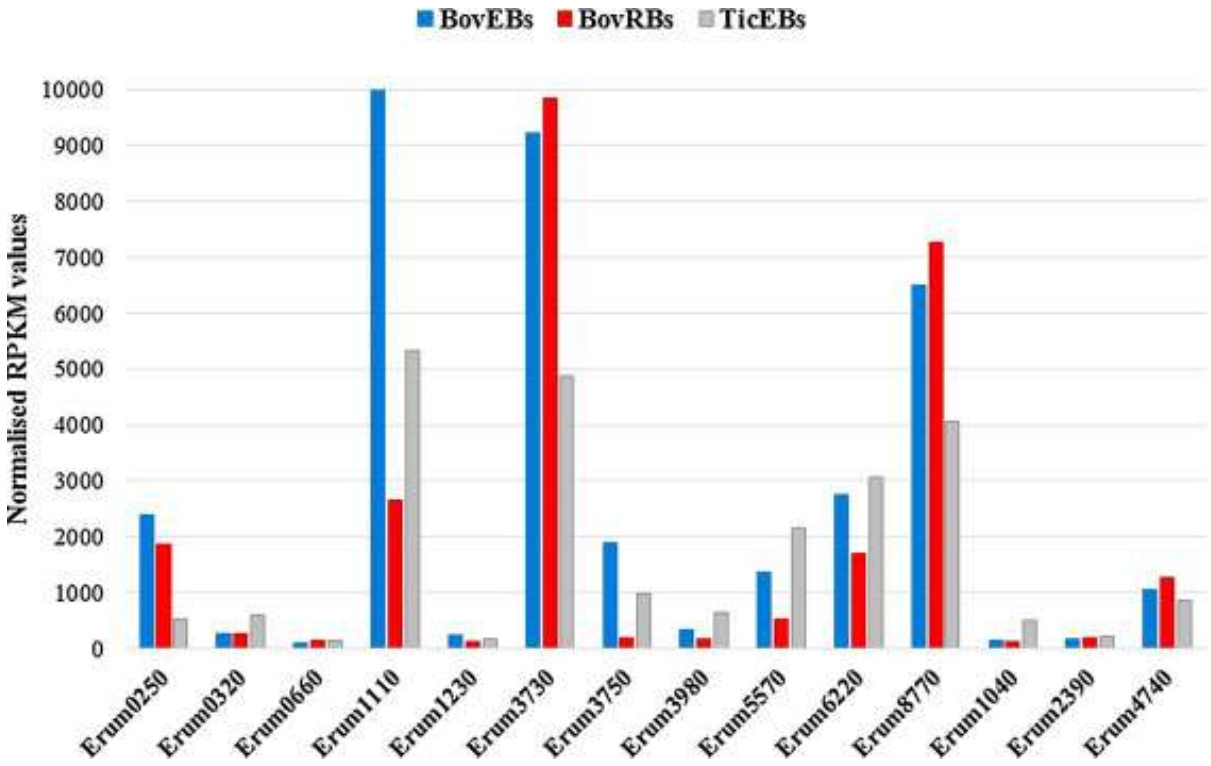


Figure S3: Normalised RPKM values of genes encoding proteins of unknown function and containing ankyrin or tandem repeats or TPRs. #RPKM values of Erum1110 for BovEBs = 29173.

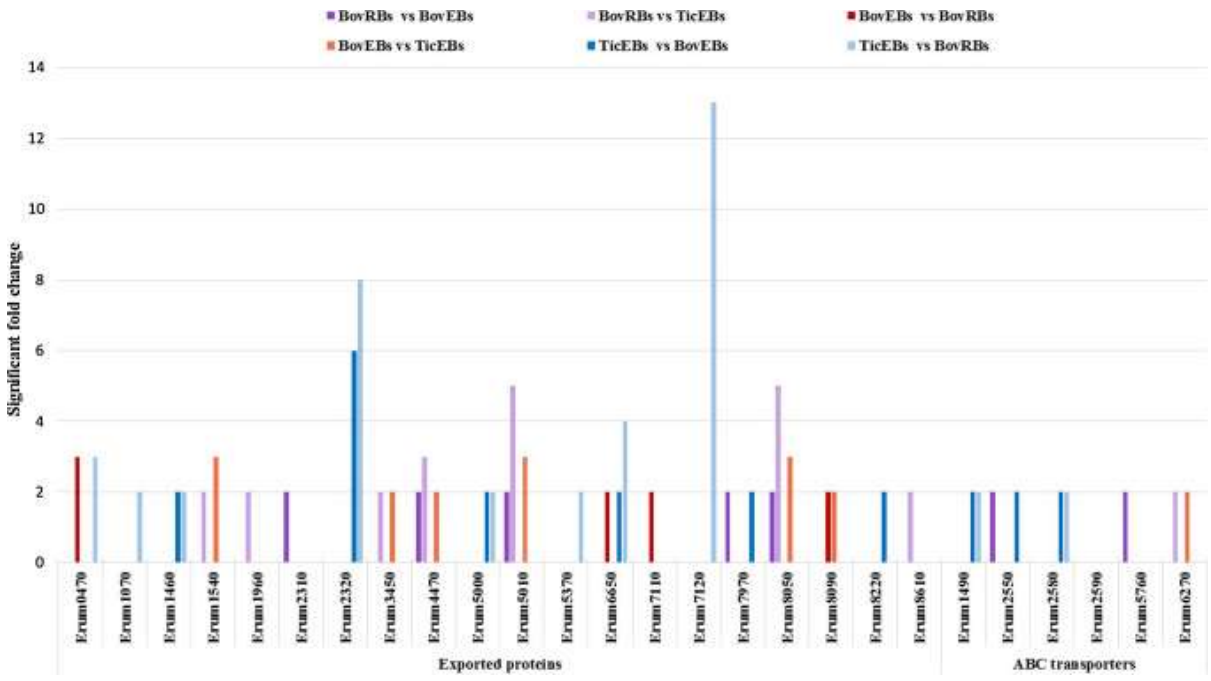


Figure S4: Up-regulated transcripts of the genes predicted to be exported genes or ABC transporters determined by comparison with all three data sets. Only genes with significant fold changes are shown.

Table S1: List of the oligonucleotide primers that were used to validate the expression of *E. ruminantium* genes at the EBs and RBs by qPCR.

Gene	Primer id	Sequence (5-3')
Erum0440 (*dksA*)	dksAF	CAGACCTAACAGACATGGCAATA
	dksAR	CTCCTGTTTCTTCACAGTAGCC
Erum0500	Erum0500F	AAGAAAAGTGTAACGAGGA
	Erum0500R	GCTGTTTTGTTTCAGATG
Erum1110	Erum1110F	AAGGTCATGAAGAGGG
	Erum1110R	AATGGATGAAGATTAAGAGAAAA
Erum2380	Erum2380F	TGCAAGAGATGTACAAGA
	Erum2380R	GTGCAAGTGATTGTGG
Erum6240	Erum6240F	GTGTTAGTTAGCTTTATTAGTG
	Erum6240R	ATCAAAACCTTCTCCAGTA
Erum7380	Erum7380F	GTTTATTGTTGTTGTTCTGTT
	Erum7380R	TCTCCTTTATTTGGTATTTCTT
Erum8010	Erum8010F	ATGTATCTACTGCTGCT
	Erum8010R	ACAAACTTCTCCTGTCT
Erum8380 (*atpF*)	atpF	ACTGCAGTACAAGCAGAAGTAG
	atpR	CCTGGTCAAGTGTAGCCATAAG
Erum8930	Erum8930F	AGAGAAACAAAGAAGGAAAA
	Erum8930R	AAATACACAATCAGCAGG

F: forward primer, R: Reverse prime

Table S2: *E. ruminantium* genes that were exclusively expressed in the BovEBs or BovRBs or TicEBs.

Sample id	Erum no	Gene id	Functional annotation	RPKM values
TicEBs	Erum0330		Integral membrane protein	171
	Erum0580		ABC transporter, ATP binding protein	137
	Erum1780		Na+/H+ antiporter subunit	110
	Erum2050		Conserved hypothetical protein	129
	Erum1760	*rnhB*	Ribonuclease HII	181
	Erum1851		Hypothetical or unknown protein	218
	Erum2140	*smf*	DNA processing protein chain A	117
	Erum2170		Hypothetical or unknown protein	161
	Erum2240		Membrane protein	356
	Erum2280		Membrane protein	199
	Erum2340		Membrane protein	150
	Erum2400		Membrane protein	212
	Erum2480		Integral membrane protein	563
	Erum2520		Biotin--[acetyl-CoA-carboxylase] synthetase	105
	Erum2610		Integral membrane protein	153
	Erum2630		Hypothetical or unknown protein	129
	Erum2640		Conserved hypothetical protein	372
	Erum2710	*nadE*	Probable glutamine-dependent NAD(+) synthetase	149
	Erum2770		Membrane protein	166
	Erum2950		Conserved hypothetical protein	163
	Erum3701		Hypothetical or unknown protein	143
	Erum3820		Integral membrane protein	127
	Erum3150		Integral membrane transport protein	113
	Erum3570		Integral membrane protein	115
	Erum3580		Integral membrane protein	215
	Erum3630		Membrane protein	146
	Erum3790		Exported protein	108
	Erum4030	*ksgA*	Dimethyladenosine transferase	160
	Erum4240	*truA*	tRNA pseudouridine synthase A	158
	Erum4320		Hypothetical or unknown protein	428
	Erum4330	*mutM*	Formamidopyrimidine-DNA glycosylase	169
	Erum4390		Hypothetical or unknown protein	165
	Erum4510		Sodium:dicarboxylate symporter (glutamate)*	103
	Erum4980	*thiL*	Probable thiamine-monophosphate kinase	122
	Erum5480	*thiL*	Probable thiamine-monophosphate kinase	122
	Erum6440	*radC*	DNA repair protein RadC	111
	Erum6760	*ruvA*	Probable junction DNA helicase RuvA	110
	Erum6820		ABC transporter, ATP-binding and membrane-spanning protein	108
	Erum7140		Membrane protein	354
	Erum7280		Membrane protein	126
	Erum7510		Hypothetical or unknown protein	105
	Erum8180		Hypothetical or unknown protein	300
	Erum8580		Possible transcriptional regulator	666

BovRBs	Erum1570		Cytochrome b561	150
	Erum2590		ABC transporter, ATP-binding protein	144
	Erum2600	*ubiB*	Probable ubiquinone biosynthesis protein UbiB	161
	Erum2690		Hypothetical or unknown protein	905
	Erum2970	*thiC*	Thiamine biosynthesis protein ThiC	104
	Erum3310	*dnaG*	Probable DNA primase	145
	Erum4720	*tatC*	Sec-independent protein translocase protein TatC	108
	Erum4850		Conserved hypothetical GTP-binding protein	123
	Erum4880	*ileS*	Bacterioferritin comigratory protein	186
	Erum6110	cmk	Cytidylate kinase	175
	Erum6180	hemH	Ferrochelatase	133
	Erum7060		Hypothetical or unknown protein	115
	Erum8080	ctaG	Cytochrome c oxidase assembly protein	137
	Erum8300	*pgsA*	Probable CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	107
BovEBs	Erum7180		Membrane protein	136

Abbreviations

BA886	Bovine aorta endothelial cell lines
BovEBs	Bovine cell derived EBs
BovRBs	Bovine cell derived RBs
cDNA	Complementary DNA
Ct	Threshold value
CTL	Cytotoxic T-lymphocytes
DNA	Deoxyribonucleic acid
EB	Elementary body
EDGE	Empirical analysis of digital gene expression
FC	Fold change
IDE8	*Ixodes scapularis* embryonic tick cell line
IFN	Interferon
NGS	Next generation sequencing
Pi	Post infection
qPCR	Quantitative PCR
RB	Reticulate body
RNA	Ribonucleic acid
rRNA	ribosomal RNA
RNA-seq	RNA sequencing
RPKM	Reads per kilo base of gene model per million mapped reads
RT-qPCR	Reverse transcription qPCR

T4SS	Type IV secretion system
Th1	T-helper 1
tRNA	transfer RNA
TicEBs	Tick cell derived EBs
tmRNA	Transfer messenger RNA
TR	Tandem repeats
TRPs	Tandem repeat proteins