



## OPEN

SUBJECT AREAS:  
GENE EXPRESSION  
PARASITIC INFECTIONReceived  
2 December 2014Accepted  
16 February 2015Published  
13 March 2015Correspondence and  
requests for materials  
should be addressed to  
M.K.  
(mich\_kotsyfakis@  
yahoo.com)

# Tissue- and time-dependent transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on the vertebrate host

Michalis Kotsyfakis<sup>1</sup>, Alexandra Schwarz<sup>1</sup>, Jan Erhart<sup>1</sup> & José M. C. Ribeiro<sup>2</sup><sup>1</sup>Institute of Parasitology, Biology Center of the Czech Academy of Sciences, Branisovska 31, 37005 Budweis, Czech Republic, <sup>2</sup>Section of Vector Biology, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, 12735 Twinbrook Parkway room 2E32D, Rockville MD 20852, USA.

*Ixodes ricinus* is a tick that transmits the pathogens of Lyme and several arboviral diseases. Pathogens invade the tick midgut, disseminate through the hemolymph, and are transmitted to the vertebrate host via the salivary glands; subverting these processes could be used to interrupt pathogen transfer. Here, we use massive *de novo* sequencing to characterize the transcriptional dynamics of the salivary and midgut tissues of nymphal and adult *I. ricinus* at various time points after attachment on the vertebrate host. Members of a number of gene families show stage- and time-specific expression. We hypothesize that gene expression switching may be under epigenetic control and, in support of this, identify 34 candidate proteins that modify histones. *I. ricinus*-secreted proteins are encoded by genes that have a non-synonymous to synonymous mutation rate even greater than immune-related genes. Midgut transcriptome (mialome) analysis reveals several enzymes associated with protein, carbohydrate, and lipid digestion, transporters and channels that might be associated with nutrient uptake, and immune-related transcripts including antimicrobial peptides. This publicly available dataset supports the identification of protein and gene targets for biochemical and physiological studies that exploit the transmission lifecycle of this disease vector for preventative and therapeutic purposes.

*Ixodes ricinus* is a European tick vector responsible for transmitting the microbial agents that cause Lyme disease and several arboviral diseases including the serious tick-borne encephalitis<sup>1,2</sup>. The life cycle of these pathogens within the tick involves invasion of midgut cells, dissemination through the hemolymph, and invasion of the salivary glands, from where they are delivered to their vertebrate hosts within salivary secretions. Tick saliva serves multiple functions during blood feeding including acting as a cement to adhere the tick to the host skin and counteracting the hemostatic, inflammatory, and immune reactions of the host that could disrupt feeding<sup>3</sup>.

Next generation deep transcriptome sequencing (and the subsequent assembly of the sequence reads into transcript contigs) allows the simultaneous identification of many thousands of coding sequences, while comparisons between the different tissue libraries sequenced allow the identification of tissue-specific transcription. Libraries prepared from tissues at different physiological time points can also help to identify the succession of transcripts, for example as the tick midgut or salivary glands change from a resting state to one of full activity. We have previously reported the results of a low-throughput gene discovery project based on normalized salivary gland libraries of *I. ricinus* to characterize the entire life cycle of this tick<sup>4</sup>, and in doing so vastly expanded the known repertoire of salivary gland protein sequences in *I. ricinus*<sup>5</sup>. We also created an *I. ricinus* salivary transcript database to support the simultaneous study of the proteomic and transcriptomic dynamics of almost 1,500 genes and their products<sup>6</sup> which revealed that proteomic and transcriptomic dynamics do not agree in the case of some genes.

Subverting the normal tick feeding process could be exploited to disrupt pathogen transfer between the vector and vertebrate host. The importance of revealing the tissue specificity in gene expression upon an infectious meal of the tick *I. ricinus* with the pathogen *Bartonella henselae* and of the tick *Rhipicephalus (Boophilus) microplus* with the parasite *Babesia bovis* has already been demonstrated<sup>7-9</sup>. Here, we aim to describe the transcriptional



changes occurring in the salivary and midgut tissues of nymphal and adult *I. ricinus* following attachment and blood feeding to the vertebrate host. In doing so, we increase the number of known transcripts in this tick and develop insights into gene regulatory mechanisms. The pattern of early salivary gland transcription reveals possible immunogens that could be used in vaccine development for early tick removal prior to pathogen delivery. The identification of salivary gland transcripts aids the future discovery of biologically active components that account for the properties of tick saliva, while the identification of midgut transcripts furthers our understanding of the receptors and proteins that regulate pathogen uptake, innate immunity, and digestion.

## Results and Discussion

**Overall characteristics of the salivary and midgut transcriptomes of *Ixodes ricinus*.** Ten non-normalized Illumina libraries were prepared from total RNA isolated from *I. ricinus* salivary glands or midguts dissected from female adult ticks (fed on the vertebrate host for either 12, 24, or 36 h) or nymphal ticks (fed on the vertebrate host for either 12 or 24 h). Sequencing of the ten different libraries (see Materials and Methods for details) produced 268,914,130 Illumina sequence reads that were *de novo* assembled into contigs along with 441,381,454 pyrosequencing reads and 67,703,183 Illumina reads from our previous work<sup>4</sup>. Extraction of the coding sequences (CDS) from the raw data assembly generated 25,808 CDS larger than 150 nucleotides (nt), to which the reads of the ten libraries from nymphal and adult salivary glands and midguts were mapped to identify tissue- and time point-specific transcriptional patterns (Supplemental File S1). A total of 148,813,058 reads (55% of total library reads) were thus mapped on the 25,808 CDS. The CDS were classified into five general categories following automatic annotation and manual curation, namely secreted (S), housekeeping (H), transposable elements (TE), viral (V), and unknown (U) (Supplemental Table 1). The S class contained almost one third of the total contigs and reads, while the H class contained half of the total CDS and 60% of the total reads; together, these classes accounted for 85% of the CDS and over 96% of the mapped reads. The S class was further subdivided into 54 functional classes (see Supplemental Table 2) as previously described<sup>4,10,11</sup>, with basic tail, Kunitz, and lipocalin proteins accounting for 6.4% of the 25,508 CDS and 12.1% of the total reads. In the H class (Supplemental Table 2), the protein synthesis machinery class accounted for the majority of reads, followed by unknown conserved, extracellular matrix, signal transduction, and protein modification machinery. Functional classification of the transcriptome by CDS number (Figure 1A) or number of reads (Figure 1B) highlights the complexity and depth of the resulting assembly and helps to visualize the significant metabolic investment occurring in these tissues during feeding. For example, the protein synthesis machinery class has a relatively small number of coding sequences represented, but comparatively greater “mass” as represented by the number of reads. Conversely, the unknown category has a larger number of putative CDS but a relatively small number of accumulated reads (Figure 1). The hyperlinked spreadsheets with transcript annotations and dynamics (Supplemental Files S1, S2, and S3) are available from <http://exon.niaid.nih.gov/transcriptome/Ixric-sgmg/Ixric-sgmg-S1-web.xlsx>, <http://exon.niaid.nih.gov/transcriptome/Ixric-sgmg/Ixric-sgmg-S2-web.xlsx>, and <http://exon.niaid.nih.gov/transcriptome/Ixric-sgmg/Ixric-sgmg-midgut-specific-S3-web.xlsx>. All the sequences can be seen by double clicking on the name of each transcript in column A of the spreadsheets.

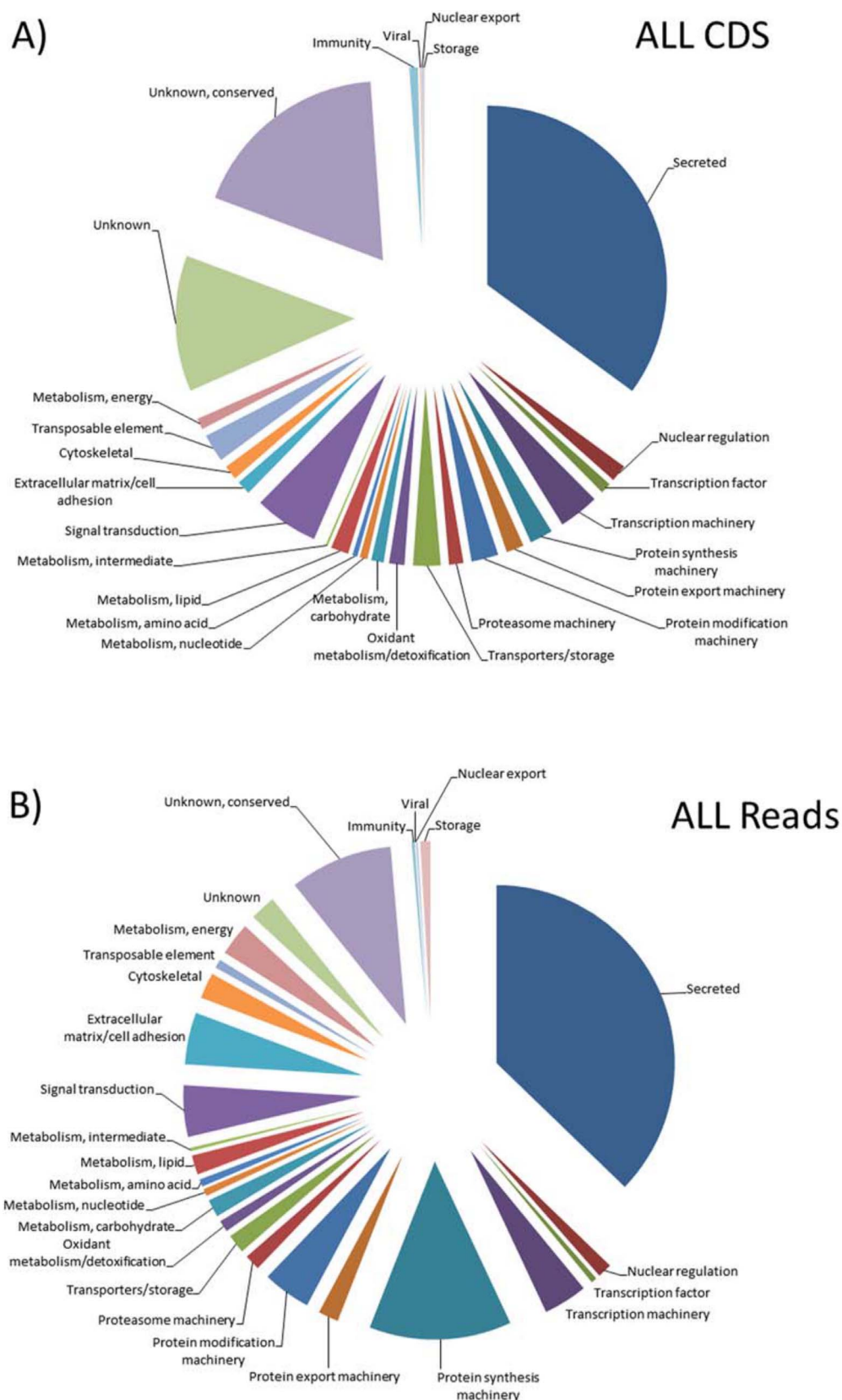
To visualize the full breadth of differential expression between tissues and time points, a box plot and heat map were constructed using normalized reads per kilobase per million (RPKM) values for CDS with a total RPKM of 50 or greater (considering the reads of all ten libraries) to avoid inclusion of poorly expressed contigs. For this

set of 11,775 CDS, RPKM values for each library were divided by the largest value and multiplied by 100. The salivary gland (SG) reads had a smaller average and larger dispersion than the midgut (MG) reads (Figure 2). The corresponding heat map (Figure 3) reveals clear clustering of SG or MG CDS according to expression. The lower left corner of the heat map shows CDS highly expressed in the SG at particular times points post feeding or developmental stage, while the upper right corner indicates midgut transcripts that are transcribed in different tissues or at different time points.

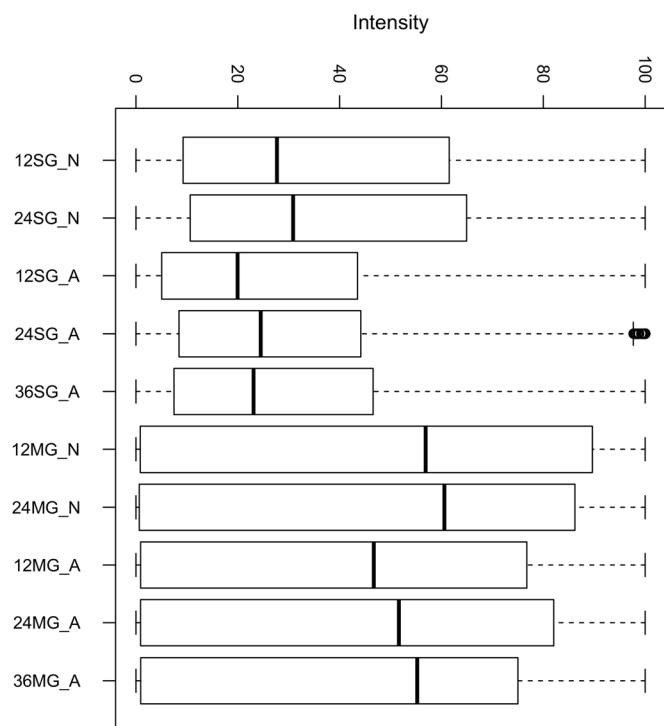
Tissue- and time-specific transcription was also evident by comparing the number of mapped reads from library pairs. Thus, 5,406 CDS were ten times or more overexpressed in the combined SG than MG libraries and 2,943 CDS in the reverse comparison, indicating putative SG- or MG-specific transcripts (Table 1) and corroborating the heat map results (Figure 3). Of the 5,406 transcripts overexpressed in SGs, 4,332 (80%) belonged to the S class (see Supplemental Table S3 for details), compared to only 733 (24%) S class overexpression in the MG (Supplemental Table S4). Other pairwise comparisons are also presented in Table 1. With respect to early expressed SG transcripts, 189 CDS were overexpressed at 12 h vs. 24 h in nymphs, with similar numbers (173) overexpressed at 12 h in adults; however, only 14 were common to both groups (Figures 4A and B) indicating that gene profiles in adults and nymphs are different at early time points. Similarly, a comparison of transcripts overexpressed at 24 h compared to 12 h also revealed a small number of common CDS between nymphs and adults, again indicating specialization of gene expression at each developmental stage. Despite the secreted class representing less than 50% of the total CDS and number of reads, they represented the majority of overexpressed CDS in all paired SG time-dependent comparisons in both nymphs and adults (Supplemental Figure S1). A similar pattern was found for time-dependent MG expressions (Figures 4C and D and Supplemental Figure S2). Furthermore, while hundreds of CDS were overexpressed in all SG library comparisons, the adult MG became more stable at 36 h when only a few CDS appeared uniquely expressed compared to the adult MG at 24 h (Supplemental Figure S3). On the other hand, over 300 CDS were at least ten-times overexpressed at 36 h compared to the 24 h library in the adult SG. The SG appears to be a lot more transcriptionally labile with continuous appearance of new transcripts, particularly of the secreted class, while the adult midgut stabilizes its gene expression at this time point, indicating it reached a functional steady state.

**Functional classification and expression pattern of salivary transcripts associated with time-dependent expression.** A total of 1,447 salivary CDS were at least ten-times overexpressed in all time-dependent pairwise comparisons (Supplemental Figure S3). Their detailed functional classification is shown in Supplemental Table S5. Of these, 1,135 (77%) belonged to the S class and included 212 CDS for Kunitz domain-containing peptides, 97 members of the Salp 15/ixostatin superfamily, 95 lipocalins, 16 metalloproteases, and several other recognizable families of unknown function. Kunitz, lipocalin, and metalloproteases tended to be uniquely expressed at a particular developmental stage (nymph or adult) and at particular times (Figure 5 and Supplemental Figures S4 and S5).

**Time-dependent functional classification and expression patterns of midgut transcripts.** A total of 440 midgut CDS were at least ten-times overexpressed in all time-dependent pairwise comparisons (Supplemental Figure S3). 269 (61%) of these were classified as S class, including 38 Kunitz domain-containing peptides, 28 basic tail peptides, and 11 lipocalin family peptides; these peptides are typically found in tick sialotranscriptomes (Supplemental Table S6)<sup>10</sup>. 83 of the 440 overexpressed CDS were the same in both the SG and MG (Supplemental Figure S6). Of these, 74 (89%) belonged to the S class, including 15 Kunitz members and two lipocalins. These large protein families typically found in sialotranscriptomes also



**Figure 1** | The genes belonging to different functional classes in the salivary gland and midgut transcriptomes of *Ixodes ricinus*. Depending on the predicted function of the polypeptides for which the coding sequences (CDS) encode, two different pie charts are shown. In (A), the proportion of the total number of different CDS encoding a polypeptide of the same predicted function are compared to the total number of CDS found in our transcriptomes. In (B), the proportion of the total number of different sequence reads assigned to CDS encoding a polypeptide of the same predicted function are compared to the total number of mapped sequence reads in our transcriptome.



**Figure 2 | The breadth of tissue- and time-dependent differential expression.** A box plot was constructed using normalized reads per kilobase per million (RPKM) values for CDS with a total RPKM (considering the reads of all ten libraries) of 50 or larger to avoid inclusion of poorly expressed contigs. The normalized RPKM values (maximum 100) for ten different libraries from *Ixodes ricinus* are as follows: the first two numbers indicate the time point of organ collection from 0–12 h (12), 12–24 h (24), or 24–36 h (36). Organs were either salivary glands (SG) or midguts (MG) and developmental stage was either nymphs (N) or adults (A). For more details, please consult the methods.

appear to be differentially expressed in the tick midgut. Indeed, a five Kunitz domain-containing protein showing low salivary expression has been identified in tick midguts, which is important for *Rickettsia* virulence<sup>12,13</sup>.

**Insight into the midgut transcriptome (mialome) of *Ixodes ricinus*.** Although the sialome of ticks is reasonably well described, including in *I. ricinus*, few studies have focused on midgut transcriptome, also known as the ‘mialome’<sup>14</sup>. Here, we highlight CDS that appear to be related to midgut function, namely the 2,943 CDS expressed ten-times or more in the MG compared to the SG (Supplemental File S3).

**Enzymes.** In ticks, meal digestion is intracellular and requires endocytosis and the action of lysosomal enzymes<sup>15</sup>. Proteins are digested via a multi-enzyme pathway that has been characterized by proteomic analyses using peptidase inhibitors<sup>16</sup>. These studies have identified cathepsins B, C, and D, legumain, and serine carboxypeptidase and leucine aminopeptidase as the main enzymes acting in the hemoglobinolytic system of *I. ricinus*. The MG transcriptome libraries were enriched for six cathepsin B enzyme-coding transcripts, which were present in all the MG libraries and had high levels of expression (RPKM between 1,300 and 19,000). Their phylogeny is represented in Figure 6. A single well-expressed CDS coding for a cathepsin D (RPKM 2453) was also identified, 96% identical to the *I. scapularis* homolog. Four papain-like cathepsin L CDS were identified, coding for enzymes of ~335 amino acids. Two of these were overexpressed in MG vs. SG libraries, with only 72% identity suggesting derivation

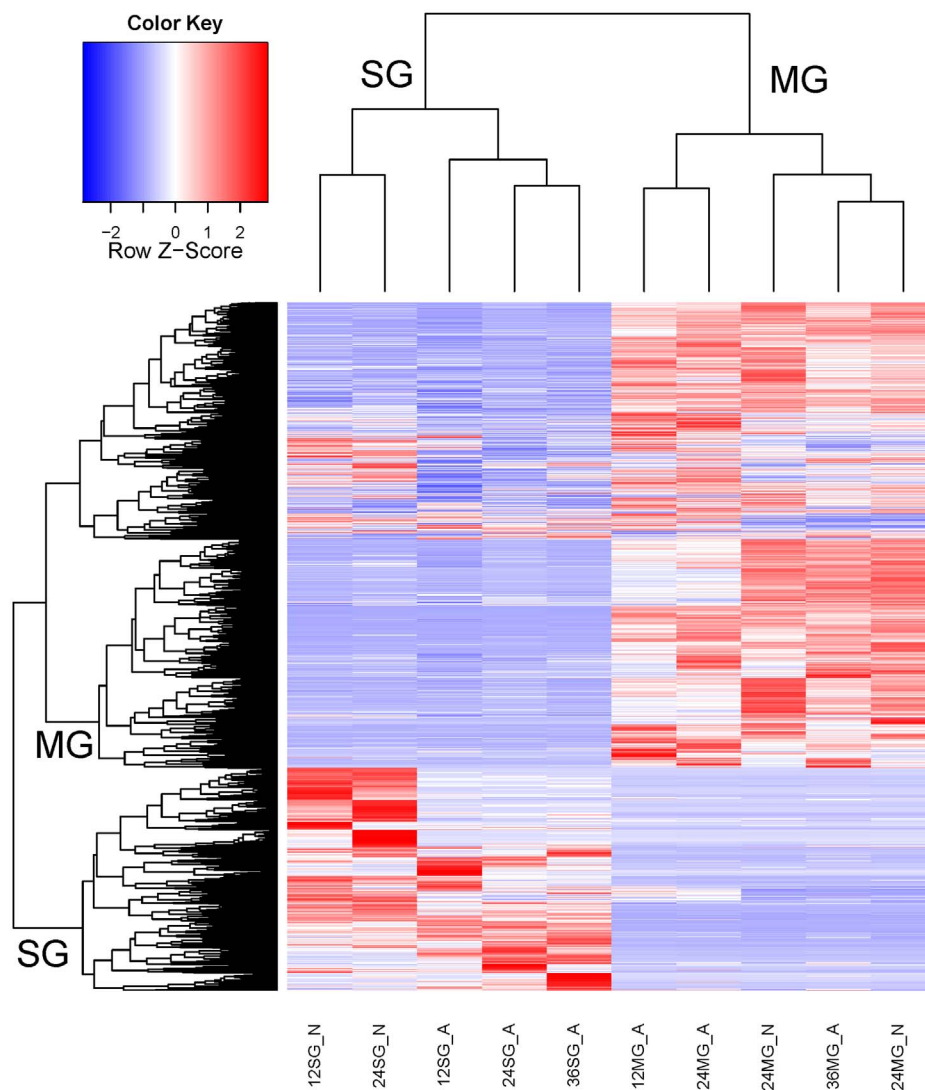
from two genes. Another CDS coded for a 558 amino acid proenzyme with 63% identity to a midgut cysteine proteinase from *R. appendiculatus* (gi|28932702)<sup>17</sup>. Three full-length legumains (RPKM between 250 and 6,500) were also identified; these had only 60% identity, again indicating multiple gene origin. Thirty CDS coding for serine proteases were overexpressed in the MG libraries, 21 of which were over 200 amino acids in length. Of these, six coded for large proteins (>400 amino acids) and included a CUB domain. Three were highly expressed, including SigP-158745 (RPKM 12,500; 387,923 reads), which had 50% identity to a cubilin serine protease in *Haemaphysalis longicornis* (gi|47847115)<sup>18</sup>. These *I. ricinus* cubilins appear to be derived from multiple genes, since they shared only 44% identity. Six additional serine proteases were highly expressed in the tick MG (RPKM between 600 and 4,300) and also appeared to be derived from multiple genes, since they had less than 70% identity. The MG was also enriched for large metalloproteases, including an astacin-like metalloprotease with an RPKM of 3,013. A full-length dipeptidyl peptidase (cathepsin C) was identified (RPKM 2,963). Several carboxypeptidases were also found; for example, SigP-149623 codes for a zinc carboxypeptidase with an RPKM of 9,305 (from 94,159 sequence reads). Several serine carboxypeptidases were identified, six of which were full length or near full length (based on multiple sequence alignments, data not shown). Their alignment and dendrogram indicate that they are derived from multiple genes (data not shown).

The exact mechanism of lipid digestion in ticks remains elusive, although ultrastructural work has shown lipid accumulation in MG cells as blood digestion progresses<sup>19,20</sup>. Forty-four CDS coding for lipases were overexpressed in *I. ricinus* MG libraries, 15 of which with an RPKM greater than 500 and 12 greater than 1,000. These included several pancreatic lipase-like enzymes, phospholipase A2, and sphingomyelinases. Similarly, the enzymes involved in tick carbohydrate digestion are unknown; several putative carbohydrate-metabolizing enzymes with elevated RPKMs are shown in Supplemental File S3 and include alpha-amylase, beta-N-acetylhexosaminidase, beta-galactosidases, alpha-fucosidase, alpha-glucosidase, and alpha-galactosidase. A chitinase and chitolectin-chitotriosidase were also found in the transcriptomes, which may be involved with peritrophic matrix modeling. With respect to DNA digestion, a CDS coding for a full-length DNase II was highly MG specific (MG RPKM of 4,413 vs. SG RPKM of 10).

**Immunity-related products.** Transcripts coding for antimicrobial peptides of the defensin, microplusin, and lysozyme families were overexpressed in the MG libraries, along with transcripts coding for peptidoglycan recognition proteins and ML-domain containing peptides; these may function as pathogen recognition receptors or as lipid carriers in lysosomes. Transcripts coding for cytokines and cytokine receptors were also identified.

**Protease inhibitors.** Peptides with TIL, Kunitz, and cystatin domains and serpins were highly expressed in MGs, including many with high RPKMs and over 1,000-fold overexpression in relation to the SG. As mentioned above, this suggests that some families classically thought to be SG-restricted, as previously reviewed in Ref. 10 are also MG abundant, with some members being MG specific, at least at the transcriptional level.

**Cytotoxin and 13 kDa family.** Several MG-specific transcripts coding for proteins with an ETX\_MTX2 PFAM domain derived from *Clostridium epsilon* mosquitoicidal toxin MTX2 were identified with high RPKMs and several hundred-fold overexpression compared to the SG. Although their function is unknown, they may be associated with the hemolytic activity previously described in *I. scapularis* midguts<sup>21</sup>. Similarly, many transcripts coding for members of the 13 kDa family of salivary proteins previously found in *Ixodes* salivary glands were overexpressed in the MG.



**Figure 3** | Heat map of normalized RPKM data from the salivary glands (SG) and midguts (MG) of nymphal and adult *Ixodes ricinus* fed for different periods of time. The Z score represents the deviation from the mean by standard deviation units. For other details, please see the Figure 2 legend.

Comparison	Number of sequences
SG > MG	5,406
MG > SG	2,943
12SGN > 24SGN	189
24SGN > 12SGN	504
12SGA > 24SGA	173
24 SGA > 12 SGA	278
24 SGA > 36 SGA	255
36SGA > 24SGA	329
12 MGN > 24 MGN	93
24 MGN > 12 MGN	121
24 MGA > 36 MGA	67
36 MGA > 24 MGA	5

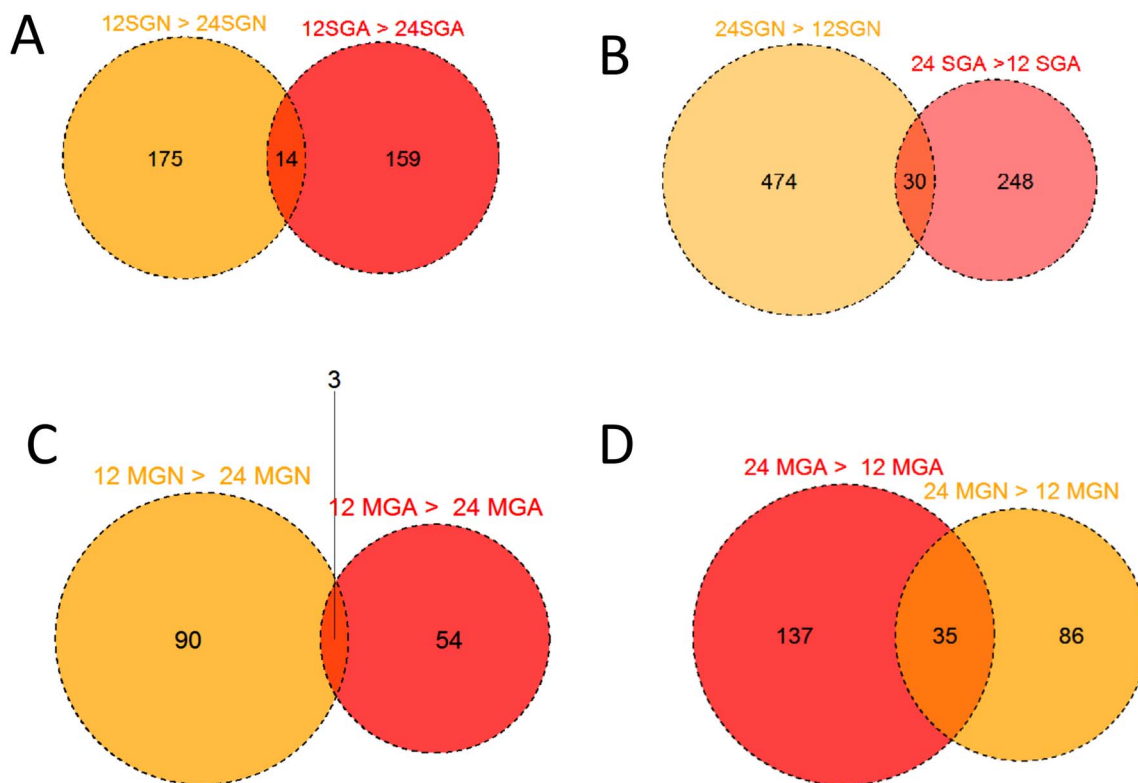
SG = salivary glands, MG = midgut, A = adult, N = nymph, numbers indicate the number of hours post host attachment that the tissues were collected. SG > MG indicates comparison for SG transcripts significantly 10 fold or more expressed in the SG than in MG tissues, and so on.

**Other secreted proteins.** An additional 583 transcripts coding for putative secreted proteins and overexpressed in the MG are shown in Supplemental File S3. Many of these were short in size and may therefore represent artifactual CDS derived from non-coding RNAs.

**Chitin deacetylase and chitin-binding domains.** One predicted peptide was found to contain the PFAM CBM\_14 chitin-binding peritrophins-A domain, two coded for chitin synthases, and 42 had the CDD CE4\_CDA\_like\_2 domain (the putative catalytic domain of chitin deacetylase-like proteins). These proteins may be involved in peritrophic matrix formation and maintenance.

**Detoxification.** Over 90 transcripts overexpressed in the MG coded for members of the cytochrome P450 family, at least 30 of which appeared to be full length with lengths over 400 amino acids. Glutathione transferases, selenoproteins, and sulfotransferases were also found.

**Signal transduction.** This functional class was highly represented in MG-overexpressed transcripts, with 325 members ten-times or more overexpressed in the MG vs. SG libraries. Twelve code for 7 transmembrane domain G-coupled receptors that might mediate agonist hormone activity on MG cells or their musculature.



**Figure 4** | Venn diagrams representing pairwise comparisons of transcripts at least ten-times ( $10\times$ ) overexpressed in the salivary glands (SG) or midguts (MG) of nymphal (N) or adult (A) *Ixodes ricinus* ticks fed for up to 12, or from 12 to 24 hours, on a host. (A) Comparison of transcripts  $10\times$  overexpressed at 12 h vs. 24 h of feeding in nymphs and adults. (B) Same as A, but transcripts overexpressed at 24 h vs. 12 h. (C) Same as A, but using MG instead of SG. (D) Same as B, but using MG instead of SG.

**Storage.** Ferritin, vitellogenins, and vitellogenin receptors were overexpressed in MG libraries, including SigP-192477, which codes for a full-length ferritin and was assembled from 118,002 reads, nearly all originating from the MG libraries. Some vitellogenins were well expressed in both adults and nymphs and could be linked to lipid transport rather than acting as an egg yolk protein.

**Transcription factors.** The depth of sequencing allowed for full-length retrieval of several transcription factors overexpressed in the MG, including GATA, MYB, E2F7, forkhead, and several others.

**Transporters.** As expected, a large number (132) of transporters and channels were overexpressed in MG tissues. The most overexpressed included an organic cation/carnithine transporter (MG RPKM = 767, SG RPKM = 3.4), an ABC transporter (MG RPKM = 725, SG RPKM = 1.1), and an epithelial chloride channel (MG RPKM = 577, SG RPKM = 0.7). RNAi experiments against some of these proteins could identify targets for digestion disruption in *I. ricinus*.

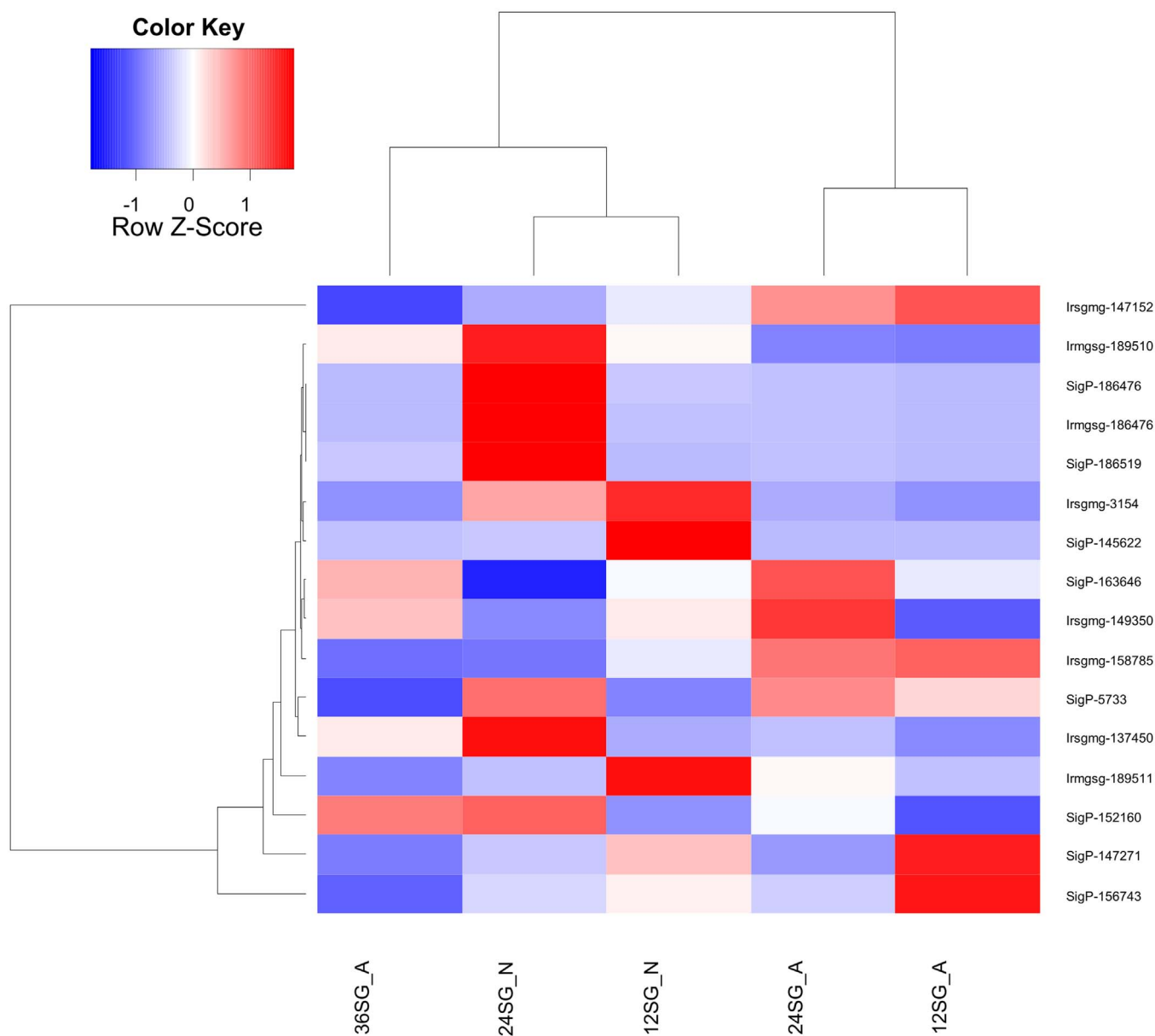
**Transposable elements.** Two NLTR retrotransposons were highly expressed in the MG (MG RPKMs 906 and 1021, SG RPKM < 5), which may represent domesticated retrotransposons associated with gene expression control<sup>22</sup>. Sixty-five other transposons were overexpressed in the MG and mostly represented fragments with small RPKMs (5–50).

**Viral sequence.** A predicted 611 amino acid protein had 35% identity and 57% similarity to the *Aguaate virus* glycoprotein. This protein also matches with the PFAM domain Phlebovirus\_G2 with an E value of  $1e-127$ . This contig was assembled from 50 reads from the MG libraries and none from the SG and had a small RPKM = 0.86. Whether this viral sequence is inserted in the tick genome remains to be elucidated.

### Possible gene switching mechanisms underlying differential gene expression within gene families.

It is possible that epigenetic mechanisms, such as histone methylation or acetylation, regulate the time- or stage-dependent transcriptional switching of genes as in embryonal development<sup>23,24</sup> or switching of endothelial adhesive proteins in *Plasmodium*<sup>25</sup>. The tick genes associated with salivary gland function may have increased over time via episodes of genome duplication in addition to the more common mechanism of tandem repeat duplications, allowing different genomic regions to contain subsets of functionally similar but divergent genes with different antigenic properties. Indeed, tick genomes are large: *I. scapularis*'s genome is larger than that of the chicken<sup>26</sup>, and metastriate ticks have larger genomes than humans<sup>27</sup>. Histone modifications at different genomic regions may, therefore, regulate gene transcription. Here, 34 transcripts coding for proteins associated with histone modifications were identified, including two near full-length histone deacetylase components (RPD3) and six sirtuins (NAD-dependent protein deacetylases). Seven transcripts coded for proteins with SET domains characteristic of histone lysine methyltransferases, two of which were ten-times or more overexpressed in the MG. Although some of these CDS were not full length, their sequences could still be used to design RNAi experiments to investigate the functional role of these genes in the control of gene transcription in *I. ricinus*. The CDS and protein sequences are available on Supplemental File S1 (worksheet named 'Epigenetics').

**Polymorphisms analysis.** Previous work in mosquitoes has indicated that genes encoding salivary proteins in hematophagous insects have higher numbers of polymorphisms and non-synonymous substitutions than other gene classes<sup>28,29</sup>. Mapping the reads on the assembled contigs allowed us to determinate single



**Figure 5** | Genes encoding salivary metalloproteases in *Ixodes ricinus* that are at least ten-times differentially expressed at different tick developmental stages (nymphs or adults) and times post feeding (12, 24, or 36 hours).

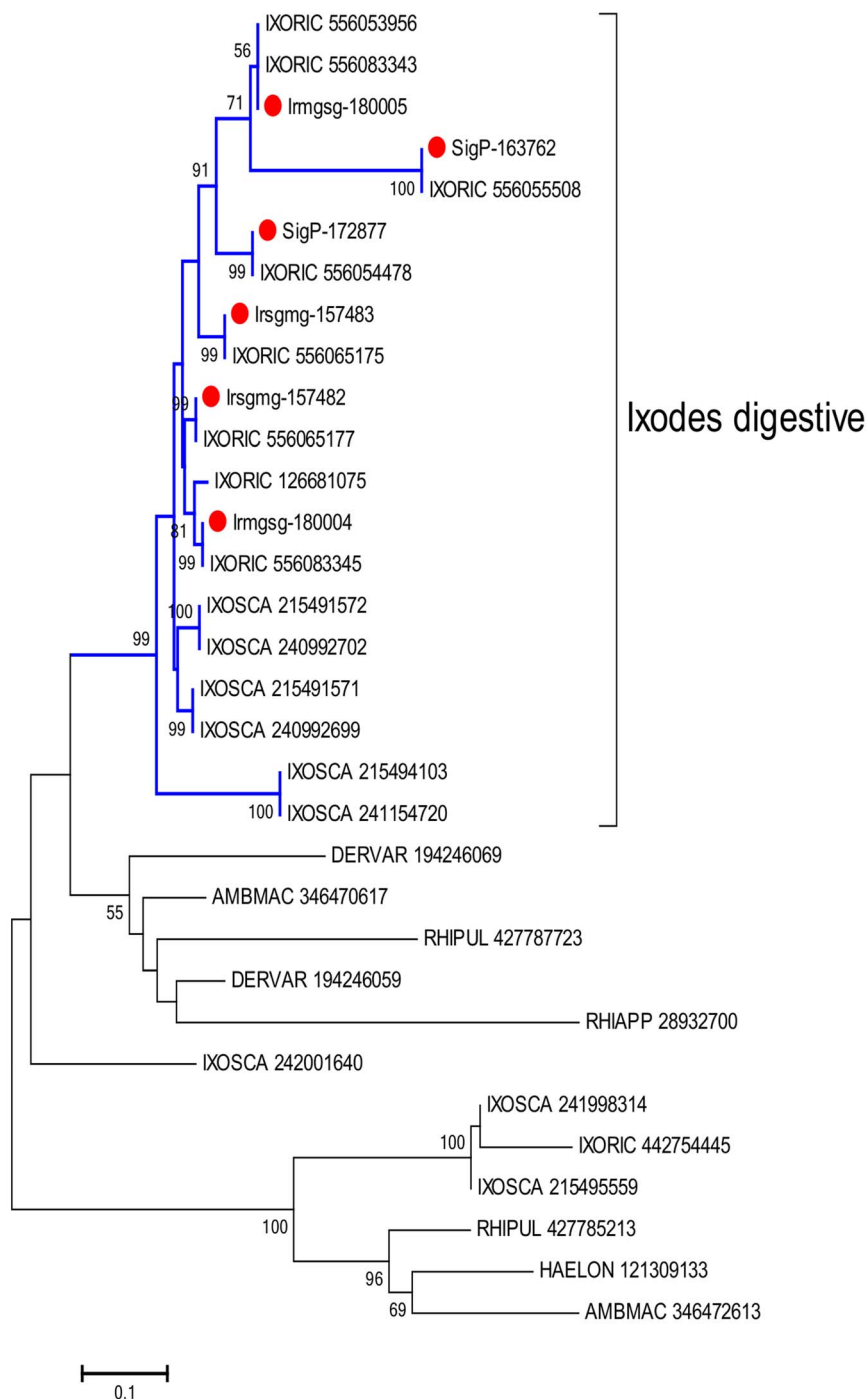
nucleotide polymorphisms (SNPs) in the CDS. For this analysis, a subset of 13,910 CDS with an average read depth of 50 or more was used to avoid inclusion of low read-depth assembled sequences. Immunity, secreted, and unknown classes had the highest rate of non-synonymous substitutions per codon and the highest non-synonymous/synonymous ratios, suggesting accelerated evolution of these gene classes (Table 2).

## Conclusions

In this paper, we *de novo* assembled 441,381,454 pyrosequencing reads and 67,703,183 Illumina reads obtained from normalized libraries of nymphal and adult *I. ricinus* from our previous work<sup>4</sup> together with 268,914,130 reads from ten non-normalized libraries from nymphal and adult salivary glands and midguts. 25,808 CDS larger than 150 nt were extracted in total. Mapping only the non-normalized reads coming from this work (the reads from Ref. 4 were not mapped) to the deduced CDS allowed detailed analysis of tissue- and time-specific gene expression. Tick salivary proteins contain many protein families including the lipocalins, Kunitz-domain con-

taining proteins, and metalloproteases, and members of these families show stage and time-specific expression (Figure 5 and Supplemental Figures S4 and S5). Members of these families can be very diverse, and it has been suggested that such selective expression may represent a mechanism of antigenic variation<sup>30</sup>. We hypothesize that gene expression switching may be under epigenetic control, and the identification of 34 candidate proteins involved in histone modification supports this notion. Secreted proteins are coded for by genes with a high rate of non-synonymous to synonymous mutations, even greater than immune-related genes, indicating fast evolution of this class.

We analyzed the mialome of *I. ricinus* in detail, and in doing so describe several enzymes associated with protein, carbohydrate, and lipid digestion, along with transporters and channels that might be related to nutrient intake. This work is limited by the focus on only two tick tissues and the lack of biological replication due to cost constraints, and future work will first require validation of these findings. Nevertheless, the reported results are biologically intuitive: Transcripts known to be associated with salivary gland function were



**Figure 6 | Phylogeny of the cathepsin B family of tick proteases.** Multiple protein sequence alignments of all the cathepsin B proteins from ticks was used to construct a phylogenetic tree of the specific family. The red symbols indicate the six enzymes overexpressed in the midgut of *Ixodes ricinus* (found as overexpressed in this work) beside their GenBank deposited sequence references. GenBank sequences are recognizable by the three first letters of their genus name followed by the first three letters of the species name followed by their GenBank accession number (IXORIC: *Ixodes ricinus*, IXOSCA: *Ixodes scapularis*, DERVAR: *Dermacentor variabilis*, AMBMAC: *Amblyomma maculatum*, RHIPUL: *Rhipicephalus pulchellus*, RHIAPP: *Rhipicephalus appendiculatus*, HAELON: *Haemaphysalis longicornis*). The numbers near the branches indicate the percentage of bootstrap support using the neighbor joining algorithm. Values lower than 50 are not shown. The bar at the bottom indicates 10% amino acid substitution rate.

expressed at hundred-fold higher levels than midgut transcripts, and known digestive enzymes were expressed at orders of magnitude higher in the midgut compared to the salivary glands. This study represents a valuable mining platform for future studies involving the salivary glands and midgut biochemistry and physiology of *I. ricinus* and helps broaden our understanding of the molecular events mediating the transmission lifecycle of this important disease vector.

## Methods

**Ethics statement.** All animal experiments were carried out in accordance with the Animal Protection Law of the Czech Republic (§17, Act No. 246/1992 Sb) and with the approval of the Akademie Ved České Republiky (approval no. 161/2010).

**Ticks, tissue dissection, total RNA/protein isolation, and sequencing.** Briefly, 1,080 nymphs and 420 adult females and males were attached to experimental animals for feeding. mRNA was extracted from pools of tissues dissected from female





**Table 2 | Rate of synonymous (Syn) and non-synonymous (NS) nucleotide substitutions per 1,000 codons in different functional classes for 13,910 coding sequences with a read depth greater than 50 per nucleotide site**

Class	Average Syn/ 1000 codons	SE	Average NS/ 1000 codons	SE	Average NS/ Syn+1	SE	Average coverage depth	SE	N
Proteasome machinery	3.0477	0.2157	0.6539	0.0962	0.2011	0.0403	1,284	276	280
Nuclear regulation	4.0364	0.2516	0.8454	0.0854	0.2038	0.0328	972	227	332
Signal transduction	3.0428	0.1206	0.7344	0.0447	0.2293	0.0202	581	47	1140
Transporters and channels	2.3329	0.1688	0.6717	0.0794	0.2490	0.0416	621	69	417
Cytoskeletal	3.1665	0.2434	0.7611	0.0936	0.2547	0.0486	1,332	219	264
Protein export	3.6622	0.2639	0.7425	0.0850	0.2682	0.0445	1,284	276	342
Transcription Factor	2.9252	0.2712	1.1889	0.1877	0.2931	0.0516	374	70	170
Detoxification	2.3616	0.2578	1.0858	0.1382	0.3107	0.0565	915	121	210
Extracellular matrix	2.1527	0.2173	0.8436	0.1186	0.3387	0.0658	2,596	540	239
Transcription machinery	3.7351	0.1608	1.1853	0.0838	0.3430	0.0321	824	92	751
Metabolism	3.3969	0.1407	1.1396	0.0607	0.3525	0.0305	1,934	199	1016
Protein synthesis	2.5993	0.2237	1.1270	0.1221	0.4475	0.0783	11,566	1,251	383
Protein modification	3.4352	0.2075	1.1842	0.1660	0.5022	0.0530	2,308	468	383
Transposable elements	1.9966	0.2085	1.8702	0.2045	0.6045	0.0701	1,394	636	182
Unknown conserved	3.0053	0.0835	1.7034	0.0644	0.7092	0.0417	1,035	101	2330
Immunity	2.5761	0.3077	2.8458	0.4407	0.8456	0.1821	697	308	100
Secreted	1.8015	0.0488	2.7768	0.0691	1.3630	0.0453	2,770	140	4419
Unknown	1.2875	0.0859	2.8527	0.1544	1.6300	0.1123	1,058	192	952

adult ticks and nymphal ticks feeding for 3 hour periods up to 24 hours (nymphs) or 36 hours (adults), providing ten samples as follows: four samples for 0–12 h and 12–24 h for nymphal salivary glands (SG) and midguts (MG), and six samples for 0–12 h, 12–24 h, and 24–36 h for adult SG and MG. These non-normalized libraries were sequenced using an Illumina HiSeq 2000 machine and the raw data deposited at the Sequence Read Archives (SRA) of the National Center for Biotechnology Information (NCBI) under accessions SRR641305, SRR641306, SRR641307, SRR641308, SRR641309, SRR641327, SRR641328, SRR641329, SRR641330, and SRR641331 (Supplemental Table S1).

**Read assembly and bioinformatics.** The reads indicated above, plus reads derived from our previous work (SRA accessions SRR592662, SRR592663, SRR592664, SRR592665, SRR592674, SRR592675, SRR592676, and SRR592677)<sup>4</sup>, were reassembled using Abyss<sup>31,32</sup> and SOAPdenovo-Trans<sup>33</sup> with various k values, and the assembly of the assemblies was performed with 15 iterations of a blastn and cap3 pipeline<sup>34</sup> to generate 198,504 contigs larger than 150 nt with an average size of 662 nt and L50 = 501 (Supplemental Table S2). Coding sequences (CDS) were extracted based on larger open reading frames (ORFs) containing a signal peptide and by translated contigs that matched protein sequences derived from the Swiss-Prot, invertebrate RefSeq, and GenBank-extracted protein sequences with “acari” in their organism name. From these CDS, 16,002 have been deposited via the Transcriptome Shotgun Assembly (TSA) portal of the NCBI to DDBJ/EMBL/GenBank under the accession GANP00000000 and BioProject ID PRJNA217984. The version described in this paper is the first version, GANP01000000. The non-redundant set of CDS was mapped to a hyperlinked Excel spreadsheet that charts these sequences to several database comparisons and the number of reads derived from each of the ten libraries<sup>28–34</sup>. RPKM (reads per thousand nucleotides per million reads) values for each sequenced library were calculated for each CDS and included in the spreadsheet as defined by the TopHat/Cufflinks program<sup>35</sup>. The number of reads for each library was compared using the  $\chi^2$  test, and the ratio of reads was calculated by dividing the number of reads derived from one library by the number of reads from the second library (plus one to avoid division by zero) and further multiplying or dividing the result by the ratio of the total reads from each library to normalize the results. For visualization of synonymous and non-synonymous sites within coding sequences, the BWA aln tool<sup>36</sup> was used to map the reads to the CDS, producing SAI files that were joined with the BWA sampe module, converted to BAM format, and sorted. The sequence alignment/map tools (samtools) package<sup>37</sup> was used to perform the mpileup of the reads (samtools mpileup), and the binary call format tools (bcftools) program from the same package was used to make the final vcf file containing the single-nucleotide polymorphic (SNP) sites using -D100 and base quality 13 or better (default). Determination of whether the SNPs led to synonymous or non-synonymous codon changes was achieved using a Visual Basic program written by JMCR, the results of which were mapped into the Excel spreadsheet and color visualized in hyperlinked rtf files within Supplemental File S1. Finally, heatmaps were produced with the gplots and heatmap.2 programs in R.

**Availability of supporting data.** The raw non-normalized library data were deposited in the Sequence Read Archives (SRA) of the National Center for Biotechnology Information (NCBI) under accessions SRR641305, SRR641306, SRR641307, SRR641308, SRR641309, SRR641327, SRR641328, SRR641329, SRR641330, and SRR641331. Following assembly, a total of 16,002 coding sequences and their protein translations were deposited in DDBJ/EMBL/GenBank under the

accession GANP00000000 and BioProject PRJNA217984. The version described in this paper is the first version, GANP01000000. The hyperlinked spreadsheets with transcript annotations and dynamics (Supplemental Files S1, S2, and S3) are available from <http://exon.niaid.nih.gov/transcriptome/Ixric-sgmg/Ixric-sgmg-S1-web.xlsx>, <http://exon.niaid.nih.gov/transcriptome/Ixric-sgmg/Ixric-sgmg-S2-web.xlsx>, and <http://exon.niaid.nih.gov/transcriptome/Ixric-sgmg/Ixric-sgmg-midgut-specific-S3-web.xlsx>.

- Mannelli, A., Bertolotti, L., Gern, L. & Gray, J. Ecology of *Borrelia burgdorferi* sensu lato in Europe: transmission dynamics in multi-host systems, influence of molecular processes and effects of climate change. *FEMS microbiology reviews* **36**, 837–861, doi:10.1111/j.1574-6976.2011.00312.x (2012).
- Pettersson, J. H., Golovljova, I., Vene, S. & Jaenson, T. G. Prevalence of tick-borne encephalitis virus in *Ixodes ricinus* ticks in northern Europe with particular reference to Southern Sweden. *Parasit Vectors* **7**, 102, doi:10.1186/1756-3305-7-102 (2014).
- Francischetti, I. M. B., Sá-Nunes, A., Mans, B. J., Santos, I. M. & Ribeiro, J. M. C. The role of saliva in tick feeding. *Frontiers in Biosciences* **14**, 2051–2088 (2009).
- Schwarz, A. *et al.* De novo *Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. *Faseb J* **27**, 4745–4756, doi:10.1096/fj.13-232140 (2013).
- Chmelar, J. *et al.* Insight into the sialome of the castor bean tick, *Ixodes ricinus*. *BMC Genomics* **9**, 233 (2008).
- Schwarz, A. *et al.* A systems level analysis reveals transcriptomic and proteomic complexity in *Ixodes ricinus* midgut and salivary glands during early attachment and feeding. *Mol. Cel. Proteomics In Press* (2014).
- Liu, X. Y. *et al.* IRsPI, a Tick Serine Protease Inhibitor Involved in Tick Feeding and *Bartonella henselae* Infection. *PLoS neglected tropical diseases* **8**, e2993, doi:10.1371/journal.pntd.0002993 (2014).
- Rachinsky, A., Guerrero, F. D. & Scoles, G. A. Proteomic profiling of *Rhipicephalus (Boophilus) microplus* midgut responses to infection with *Babesia bovis*. *Veterinary parasitology* **152**, 294–313, doi:10.1016/j.vetpar.2007.12.027 (2008).
- Heekin, A. M. *et al.* Gut transcriptome of replete adult female cattle ticks, *Rhipicephalus (Boophilus) microplus*, feeding upon a *Babesia bovis*-infected bovine host. *Parasitology research* **112**, 3075–3090, doi:10.1007/s00436-013-3482-4 (2013).
- Francischetti, I. M., Sa-Nunes, A., Mans, B. J., Santos, I. M. & Ribeiro, J. M. C. The role of saliva in tick feeding. *Front Biosci* **14**, 2051–2088 (2009).
- Ribeiro, J. M. *et al.* An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. *Insect Biochem. Mol. Biol.* **36**, 111–129 (2006).
- Ceraul, S. M. *et al.* A Kunitz protease inhibitor from *Dermacentor variabilis*, a vector for spotted fever group rickettsiae, limits *Rickettsia montanensis* invasion. *Infection and immunity* **79**, 321–329, doi:10.1128/IAI.00362-10 (2011).
- Ceraul, S. M., Dreher-Lesnick, S. M., Mulenga, A., Raham, M. S. & Azad, A. F. Functional characterization and novel rickettsiostatic effects of a Kunitz-type serine protease inhibitor from the tick *Dermacentor variabilis*. *Infection and immunity* **76**, 5429–5435, doi:10.1128/IAI.00866-08 (2008).
- Anderson, J. M., Sonenshine, D. E. & Valenzuela, J. G. Exploring the mialome of ticks: an annotated catalogue of midgut transcripts from the hard tick,



- Dermacentor variabilis* (Acari: Ixodidae). *BMC genomics* **9**, 552, doi:10.1186/1471-2164-9-552 (2008).
15. Sojka, D. *et al.* New insights into the machinery of blood digestion by ticks. *Trends in parasitology* **29**, 276–285, doi:10.1016/j.pt.2013.04.002 (2013).
  16. Horn, M. *et al.* Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. *Chemistry & biology* **16**, 1053–1063 (2009).
  17. Mulenga, A., Misao, O. & Sugimoto, C. Three serine proteinases from midguts of the hard tick *Rhipicephalus appendiculatus*; cDNA cloning and preliminary characterization. *Experimental & applied acarology* **29**, 151–164 (2003).
  18. Miyoshi, T., Tsuji, N., Islam, M. K., Kamio, T. & Fujisaki, K. Cloning and molecular characterization of a cubilin-related serine proteinase from the hard tick *Haemaphysalis longicornis*. *Insect biochemistry and molecular biology* **34**, 799–808, doi:10.1016/j.ibmb.2004.04.004 (2004).
  19. Matsuo, T. *et al.* Morphological studies on the extracellular structure of the midgut of a tick, *Haemaphysalis longicornis* (Acari: Ixodidae). *Parasitology research* **90**, 243–248, doi:10.1007/s00436-003-0833-6 (2003).
  20. Caperucci, D., Bechara, G. H. & Camargo Mathias, M. I. Ultrastructure features of the midgut of the female adult *Amblyomma cajennense* ticks Fabricius, 1787 (Acari: Ixodidae) in several feeding stages and subjected to three infestations. *Micron* **41**, 710–721, doi:10.1016/j.micron.2010.05.015 (2010).
  21. Ribeiro, J. M. The midgut hemolysin of *Ixodes dammini* (Acari: Ixodidae). *J. Parasitol.* **74**, 532–537 (1988).
  22. de Parseval, N. & Heidmann, T. Human endogenous retroviruses: from infectious elements to human genes. *Cytogenetic and genome research* **110**, 318–332, doi:10.1159/000084964 (2005).
  23. Inbar-Feigenberg, M., Choufani, S., Butcher, D. T., Roifman, M. & Weksberg, R. Basic concepts of epigenetics. *Fertility and sterility* **99**, 607–615, doi:10.1016/j.fertnstert.2013.01.117 (2013).
  24. Iovino, N. Drosophila epigenome reorganization during oocyte differentiation and early embryogenesis. *Briefings in functional genomics* **13**, 246–253, doi:10.1093/bfpg/elu007 (2014).
  25. Jiang, L. *et al.* PfSETvs methylation of histone H3K36 represses virulence genes in *Plasmodium falciparum*. *Nature* **499**, 223–227, doi:10.1038/nature12361 (2013).
  26. Geraci, N. S., Spencer Johnston, J., Paul Robinson, J., Wikel, S. K. & Hill, C. A. Variation in genome size of argasid and ixodid ticks. *Insect biochemistry and molecular biology* **37**, 399–408 (2007).
  27. Ullmann, A. J., Lima, C. M., Guerrero, F. D., Piesman, J. & Black, W. C. T. Genome size and organization in the blacklegged tick, *Ixodes scapularis* and the Southern cattle tick, *Boophilus microplus*. *Insect molecular biology* **14**, 217–222 (2005).
  28. Chagas, A. C. *et al.* A deep insight into the sialotranscriptome of the mosquito, *Psorophora albipes*. *BMC genomics* **14**, 875, doi:10.1186/1471-2164-14-875 (2013).
  29. Arca, B. *et al.* Positive selection drives accelerated evolution of mosquito salivary genes associated with blood-feeding. *Insect molecular biology* **23**, 122–131, doi:10.1111/imb.12068 (2014).
  30. Kotsyfakis, M., Karim, S., Andersen, J. F., Mather, T. N. & Ribeiro, J. M. Selective cysteine protease inhibition contributes to blood-feeding success of the tick *Ixodes scapularis*. *The Journal of biological chemistry* **282**, 29256–29263 (2007).
  31. Birol, I. *et al.* De novo transcriptome assembly with ABySS. *Bioinformatics (Oxford, England)* **25**, 2872–2877, doi:10.1093/bioinformatics/btp367 (2009).
  32. Simpson, J. T. *et al.* ABySS: a parallel assembler for short read sequence data. *Genome research* **19**, 1117–1123, doi:10.1101/gr.089532.108 (2009).
  33. Yang, Y. & Smith, S. A. Optimizing de novo assembly of short-read RNA-seq data for phylogenomics. *BMC genomics* **14**, 328, doi:10.1186/1471-2164-14-328 (2013).
  34. Karim, S., Singh, P. & Ribeiro, J. M. A deep insight into the sialotranscriptome of the Gulf Coast tick, *Amblyomma maculatum*. *PLoS ONE* **6**, e28525, doi:10.1371/journal.pone.0028525 (2011).
  35. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols* **7**, 562–578, doi:10.1038/nprot.2012.016 (2012).
  36. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* **26**, 589–595, doi:10.1093/bioinformatics/btp698 (2010).
  37. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* **25**, 2078–2079, doi:10.1093/bioinformatics/btp352 (2009).

## Acknowledgments

This work was supported by the Grant Agency of the Czech Republic (GACR grant P502/12/2409 to M.K.), the Czech Academy of Sciences (grant no. RVO 60077344 to the Institute of Parasitology), and the 7th Framework Program of the European Union (EU FP7; Marie Curie Reintegration grant PIRG07-GA-2010-268177 to M.K.). Library construction and Illumina sequencing of all the samples was supported by the European Union, 7th Framework Program, project TRANSVAC (FP7-INFRASTRUCTURES-2008-228403 to M.K.). This publication reflects only the authors' views and the European Union is not liable for any use that may be made of the information contained herein. A.S. was funded by the Alexander von Humboldt Foundation (Feodor Lynen Research Fellowship). J.M.C.R. was supported by the Intramural Research Program of the Division of Intramural Research, National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH; USA). M.K. received support from the Czech Academy of Sciences (Jan Evangelista Purkyně Fellowship) and the European Molecular Biology Organization [EMBO; Application Short Term Fellowship (ASTF) 482-2012]. The authors are also grateful to Andrea Gocke (National Center for Biotechnology Information, USA) for helping with sequence submission. The NIH reserves the right to provide the work to PubMedCentral for display and use by the public, and PubMedCentral may tag or modify the work consistent with its customary practices. Rights can be established outside of the United States subject to a government use license. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

## Author contributions

M.K. conceived the study, participated in the design and coordination of the study, and helped to draft the manuscript. A.S. participated in the experimental design, tick dissections, preparation of libraries, and approval of the final manuscript. J.E. participated in the preparation of libraries and approval of the final manuscript. J.M.C.R. performed the bioinformatics analysis and wrote the majority of the manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Kotsyfakis, M., Schwarz, A., Erhart, J. & Ribeiro, J.M.C. Tissue- and time-dependent transcription in *Ixodes ricinus* salivary glands and midguts when blood feeding on the vertebrate host. *Sci. Rep.* **5**, 9103; DOI:10.1038/srep09103 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>