

# SCIENTIFIC REPORTS



OPEN

## Analysis of long non-coding RNA and mRNA expression in bovine macrophages brings up novel aspects of *Mycobacterium avium* subspecies *paratuberculosis* infections

Pooja Gupta<sup>1,6</sup>, Sarah Peter<sup>2</sup>, Markus Jung<sup>2</sup>, Astrid Lewin<sup>3</sup>, Georg Hemmrich-Stanisak<sup>4</sup>, Andre Franke<sup>4</sup>, Max von Kleist<sup>1</sup>, Christof Schütte<sup>1,6</sup>, Ralf Einspanier<sup>5</sup>, Soroush Sharbati<sup>5</sup> & Jennifer zur Bruegge<sup>5</sup>

Paratuberculosis is a major disease in cattle that severely affects animal welfare and causes huge economic losses worldwide. Development of alternative diagnostic methods is of urgent need to control the disease. Recent studies suggest that long non-coding RNAs (lncRNAs) play a crucial role in regulating immune function and may confer valuable information about the disease. However, their role has not yet been investigated in cattle with respect to infection towards Paratuberculosis. Therefore, we investigated the alteration in genomic expression profiles of mRNA and lncRNA in bovine macrophages in response to Paratuberculosis infection using RNA-Seq. We identified 397 potentially novel lncRNA candidates in macrophages of which 38 were differentially regulated by the infection. A total of 820 coding genes were also significantly altered by the infection. Co-expression analysis of lncRNAs and their neighbouring coding genes suggest regulatory functions of lncRNAs in pathways related to immune response. For example, this included protein coding genes such as *TNIP3*, *TNFAIP3* and *NF-κB2* that play a role in NF-κB2 signalling, a pathway associated with immune response. This study advances our understanding of lncRNA roles during Paratuberculosis infection.

*Mycobacterium avium* subspecies *paratuberculosis* (MAP) is the causative agent of Paratuberculosis or Johne's disease, a chronic enteritis in ruminants. Animals suffer from prolonged diarrhoea and progressive wasting<sup>1</sup>. The disease causes major economic losses for the dairy industry worldwide. A causal role of MAP in the pathogenesis of human Crohn's disease is discussed controversially<sup>2</sup>.

It is estimated that around 31–71% of cattle herds in European countries are infected<sup>3–5</sup>. Due to the unique pathogenesis of MAP, diagnosis of Paratuberculosis is difficult, especially in the early phase of infection. Neonates and calves in their first few months of life are most susceptible to MAP<sup>6</sup>. The pathogen is faecal-orally transmitted via ingestion of contaminated colostrum, water or feed. Intestinal epithelial cells are the main port of entry for MAP, which preferentially invades enterocytes and specialized epithelial cells known as M cells<sup>1</sup>. These cells then translocate MAP from the intestinal lumen to the submucosa where MAP is taken up by subepithelial macrophages<sup>1</sup>. The interaction of MAP and macrophages determines whether the host is able to eliminate the pathogen or if infection is being established. Control of MAP infections depends on several factors including an

<sup>1</sup>Department of Mathematics and Informatics, Freie Universität Berlin, Berlin, Germany. <sup>2</sup>Institute for the Reproduction of Farm Animals Schönow Inc, Bernau, Germany. <sup>3</sup>Robert Koch-Institute, Department Infectious Diseases, Berlin, Germany. <sup>4</sup>Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Germany. <sup>5</sup>Institute of Veterinary Biochemistry, Department of Veterinary Medicine, Freie Universität Berlin, Berlin, Germany. <sup>6</sup>Department of Mathematics for Life and Materials Sciences, Zuse Institute Berlin, Berlin, Germany. Correspondence and requests for materials should be addressed to P.G. (email: [gupta@math.fu-berlin.de](mailto:gupta@math.fu-berlin.de))

early protective Th1 cell response promoting INF- $\gamma$  release and activation of antimicrobial mechanisms in macrophages<sup>7</sup>. However, like many other Mycobacteria, MAP is able to modulate macrophage response to enhance its intracellular persistence and survival. For example, it inhibits phago-lysosomal maturation in macrophages in order to avoid acidification and bacterial killing, and it also impairs antigen presentation to T cells<sup>8</sup>. Furthermore, MAP modulates INF- $\gamma$  signalling or induces increased secretion of IL-10 to promote bacterial persistence and establish infection<sup>9</sup>.

Depending on the infection dose and animal health status, the incubation period for Paratuberculosis is long (>2 years). Around 10–15% of infected cattle develop clinical signs<sup>6</sup>. During the incubation time, animals show little or no clinical signs of infection<sup>1</sup>. Nevertheless, infected animals in the subclinical phase may shed MAP and serve as a direct source of transmission. Detection and management of such animals is crucial for the success of Paratuberculosis control measures. However, serum antibodies are usually not detectable in the preclinical phase. Faecal culture is considered as the “gold standard” for diagnosis but sensitivity in early subclinical infections is low and cultivation is protracted. Detection of MAP in faeces via PCR may be an alternative to faecal culture<sup>10</sup>, however, detection rates depend on the shedding status of the animal. Thus, the development of alternative diagnostic tools is of urgent need to control the disease.

The use of non-coding RNAs as a novel and promising diagnostic approach of infectious and non-infectious diseases has recently become a major focus of investigation<sup>11,12</sup>. It was shown that bacteria interfere with the expression of mammalian regulatory RNAs to modify immune signalling, autophagy, or the apoptotic machinery. Recently, a new class of regulatory RNAs, long non-coding RNAs (lncRNAs), was reported to play a crucial role in the regulation of eukaryotic gene expression. A rising amount of literature reports on specific involvement of lncRNAs in the host cell response towards bacterial infections<sup>13</sup>. Based on their size (larger than 200 nt) lncRNAs are distinguished from other non-coding RNAs such as microRNAs (miRNAs). They function as protein scaffolds, activators or inhibitors of transcription, antisense RNA or miRNA sponges<sup>14</sup>. They are known to possess low evolutionary conservation and exhibit lower levels of cellular concentration than protein coding genes but have a higher degree of tissue specificity<sup>15</sup>. Nonetheless, despite having a key role in the regulation of gene expression, lncRNAs remain poorly identified and annotated in domesticated animals in comparison to other species such as humans and mice<sup>16</sup>.

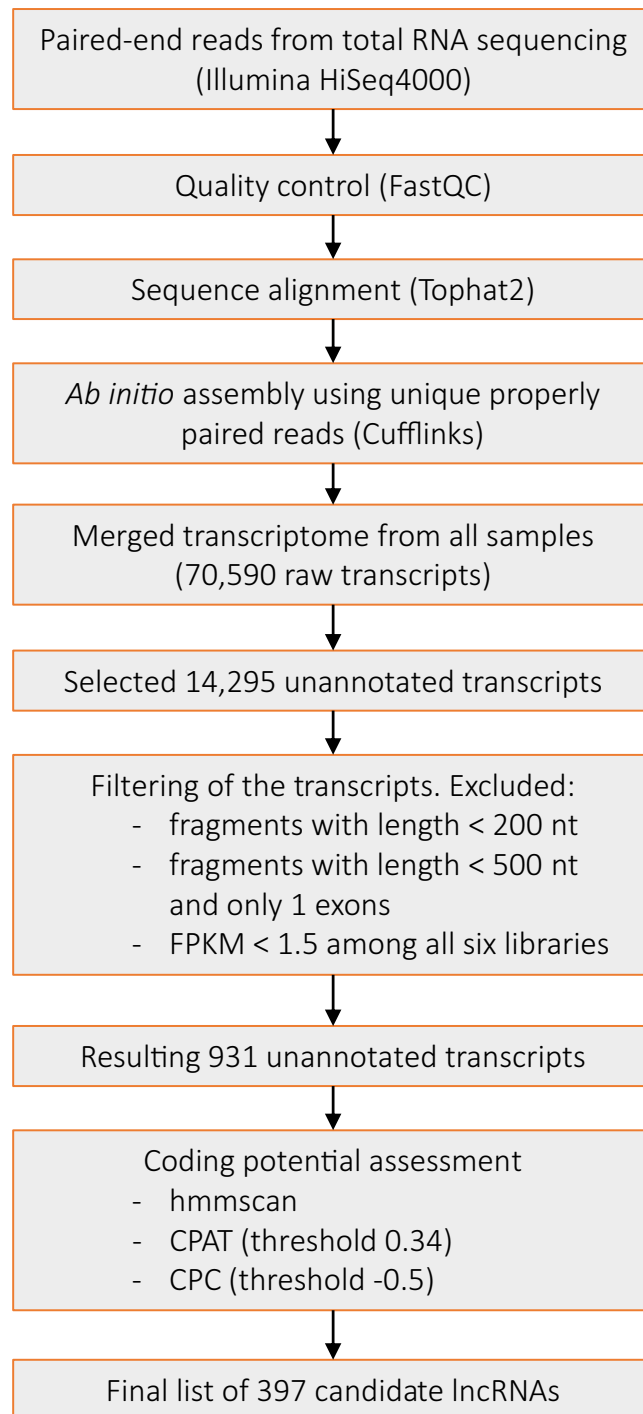
The objective of this study is to deepen the knowledge about the crucial interplay between bovine macrophages and MAP, and furthermore provide novel insight into the regulatory function of lncRNAs in cattle during MAP infection. To this end, we investigated the changes in mRNA expression profiles and the presence of potentially novel lncRNA candidates and their role in bovine macrophages with respect to MAP infection via RNA sequencing.

## Results

**Identification of putative lncRNA transcripts.** We analysed high-throughput RNA-Seq data derived from MAP infected and uninfected primary bovine macrophages. The alignment of paired-end reads resulted in concordant mapping rates of >90% for all libraries (Supplementary Table S1). We obtained 70,590 transcripts (Fig. 1) by assembling 240 million uniquely mapped and properly paired reads (35.9 to 43.0 million paired-end reads per sample). A computational pipeline (Fig. 1) adapted from earlier studies<sup>17–21</sup> was implemented to characterize and identify candidate lncRNAs. Initially, 14,295 transcripts with the class codes ‘u’ and ‘x’ that have unknown annotations were retained to nominate putative lncRNAs. We obtained 931 unannotated transcripts (Fig. 1, Supplementary Table S2) after filtering out low quality assemblies such as those with a small fragment size (<200 bp; <500 bp for single-exon transcripts<sup>20</sup>) and low expression (FPKM <1.5 among all six libraries<sup>17,20</sup>). The selected lncRNAs were searched for known protein domains against the Pfam database<sup>22</sup>. In addition, their protein coding ability was assessed using coding potential calculator<sup>23</sup> (CPC) and coding-potential assessment tool<sup>24</sup> (CPAT). We included only those lncRNAs that showed no matching domain in the Pfam database, resulting in 490 candidates. Using a threshold cut-off score of  $-0.5$  for CPC<sup>18</sup>, we obtained 425 non-coding candidate transcripts. Furthermore, CPAT provided 439 non-coding candidate transcripts with a cut-off threshold of 0.34. An intersection of all three approaches (Supplementary Fig. S1) provided a final list of 397 candidate lncRNAs (Supplementary Data 1a,b).

**Genomic characteristics of lncRNAs in cattle.** lncRNAs have been shown to possess distinguishing features in comparison to other RNAs, e.g. lower levels of cellular concentration, fewer exons and shorter length than protein coding genes<sup>16</sup>. Therefore, we analysed the differences in number of exon, length, expression and genomic distribution between lncRNAs and 12,092 mRNAs expressed in our samples (Fig. 2). We observed that lncRNAs have fewer number of exon (median = 2) than mRNAs (median = 8;  $P = 4.12e-178$ , two-tailed Mann-Whitney U test, Fig. 2a). On an average, we found around 71% transcripts with less than three exons and almost 48% transcripts having single exon. We also found that the most frequently observed length of lncRNAs (median = 1,505) was smaller (Fig. 2b) than the length of mRNAs (median = 2,144;  $P = 8.18e-09$ , two-tailed Mann-Whitney U test). Furthermore, the gene expression levels differed significantly between candidate lncRNAs and mRNAs (Fig. 2c). The expression of lncRNAs (mean FPKM = 6.59) is lower compared with mRNAs (mean FPKM = 22.84;  $P = 2.37e-07$ , two-tailed Student's t-test). Figure 2d shows the distribution of candidate lncRNAs across the genome of the cow. The average number of candidate lncRNAs per chromosome per 10 Mb was approximately 1.6. The distribution of candidate lncRNAs across the genome exhibited a pattern similar to the distribution of protein-coding genes expressed in our samples (Supplementary Fig. S2). Similar to mRNA distribution, chromosomes 18, 19, 25 and 29 contained more lncRNAs in comparison to the rest of chromosomes, with each containing >5% of candidate lncRNAs per chromosome per 10 Mb (Fig. 2d, Supplementary Table S3).

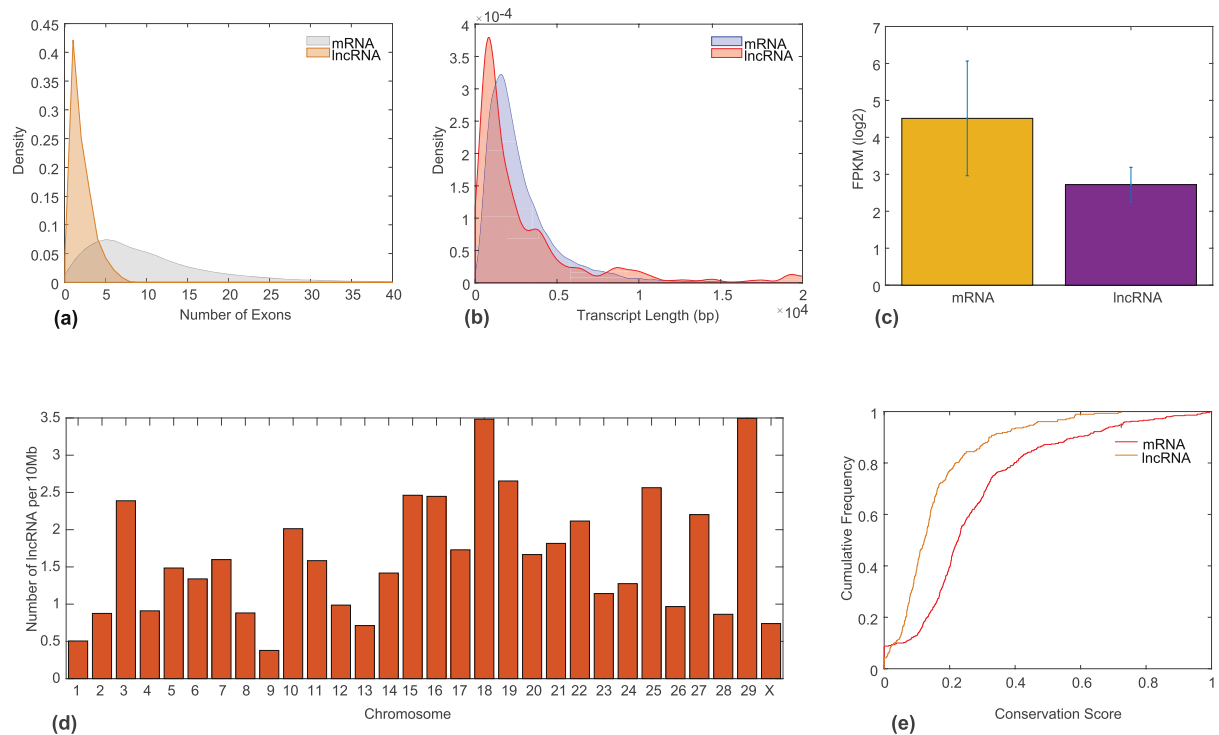
As it has been shown that lncRNAs are evolutionarily less conserved than protein-coding genes<sup>15</sup>, we investigated the conservation scores of our candidate lncRNAs. Using UCSC's ‘liftover’ based conversion of genomic



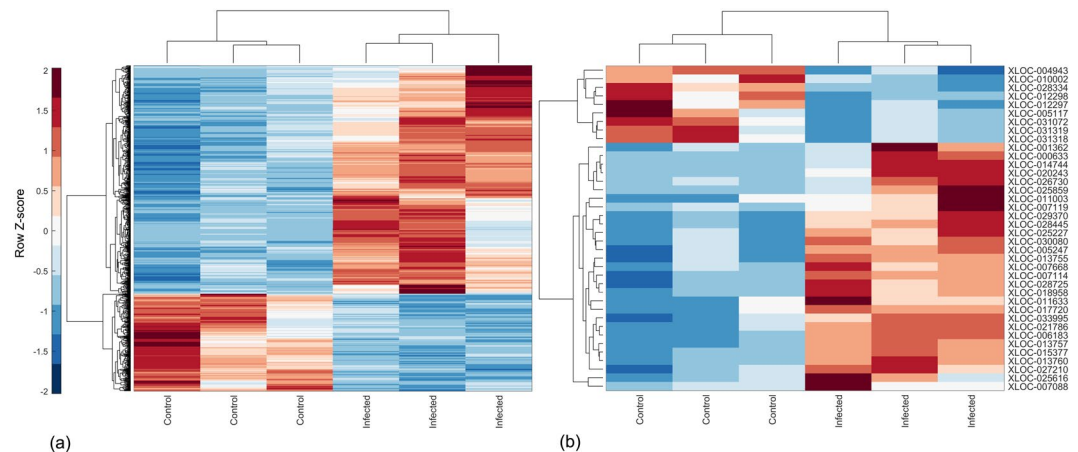
**Figure 1.** Pipeline for selecting lncRNA. A schematic overview of the pipeline employed for selecting candidate lncRNA based on relevant filtering criteria.

coordinates from the bosTau8 to the bosTau4 assembly, we retrieved the PhastCons scores<sup>25</sup> for 282 candidate lncRNAs; the coordinates for remaining lncRNAs did not exist in the bosTau4 assembly. Figure 2e shows the cumulative density plot of the PhastCons scores for the lncRNAs and 500 randomly selected mRNAs (annotated in Ensemble GTF). As expected, we found that lncRNAs are not as conserved as protein-coding transcripts (Fig. 2e,  $P = 2.18 \times 10^{-11}$ , two-tailed Student's t-test). However, despite an overall low conservation score, we observed local regions with reasonable conservation across the transcripts (Supplementary Data 2).

**Differential expression of lncRNA and mRNA.** In total, 820 protein-coding genes (Fig. 3a, Table 1, Supplementary Data 3, Supplementary Figs S3 and S4) and 38 lncRNAs genes (Fig. 3b, Table 2, Supplementary Data 4) were significantly differentially expressed (DE; corrected p-value for multiple testing using



**Figure 2.** Genomic properties of the candidate lncRNA. **(a)** Density plot for the number of exons in mRNA and candidate lncRNA. **(b)** Density plot for the length of mRNA and candidate lncRNA. **(c)** Expression levels (FPKM) of mRNA and lncRNA. **(d)** Distribution of the candidate lncRNA across the cow genome. **(e)** Cumulative density plot for the conservation score of mRNA and candidate lncRNA.



**Figure 3.** Cluster analysis of the differentially expressed genes. **(a)** A hierarchical heatmap showing differentially expressed mRNA in control and infected samples. **(b)** A hierarchical heatmap showing differentially expressed lncRNA candidates in control and infected samples. Red shows higher expression and blue shows lower expression.

Benjamini-Hochberg technique<sup>26</sup>  $p_{BH} < 0.05$ ). Among them, 347 upregulated and 156 downregulated genes as well as 25 upregulated and 7 downregulated lncRNAs showed distinct dysregulation with a  $\log_2$  fold-change (L2FC)  $\geq 1$  or  $\leq -1$ . For 20 out of these 38 DE lncRNAs, the closest neighbouring mRNAs were also significantly DE (11 upregulated and 1 downregulated for  $L2FC \geq 1$  or  $\leq -1$  and  $p_{BH} < 0.05$ , Table 3).

**Analysis of neighbouring protein-coding genes.** The average distance between lncRNAs and their closest neighbour was 34 kb (Supplementary Data 5). DE neighbouring mRNAs were present within shorter distances (mean = 7.3 kb, Table 3). We also analysed the expression levels of DE neighbours. Interestingly, a high positive correlation (Pearson correlation  $r > 0.80$ ,  $P < 0.05$  (two-tailed Student's t-test); performed in MATLAB (version R2017b) using the function *corr*) was observed between the pair of DE lncRNA and their corresponding

Gene symbol/Ensemble ID	Chromosome	Base mean	L2FC*	p.BH <sup>#</sup>
<i>CCL8</i>	19	11904.20	6.53	3.02E-05
<i>CSF2</i>	7	2046.87	5.85	2.75E-03
<i>IFIT2</i>	26	2211.50	5.18	1.68E-04
<i>IL9</i>	7	28.77	4.97	1.12E-06
<i>LIF</i>	17	62.70	4.90	6.00E-04
<i>RSAD2</i>	11	35769.02	4.87	3.48E-10
<i>IFIT1</i>	26	34648.49	4.51	5.04E-04
<i>CSF3</i>	19	2155.09	4.44	2.06E-03
<i>PTX3</i>	1	1027.85	4.37	1.62E-02
<i>IL6</i>	8	12.19	4.32	8.55E-04
<i>TSG-6</i>	2	187.69	4.31	1.39E-02
<i>TACSTD2</i>	3	9.59	4.22	2.44E-02
<i>SLFN11</i>	19	71.56	4.21	4.89E-06
<i>TNF</i>	23	5526.87	4.21	1.41E-12
<i>ISG20</i>	21	255.51	3.99	1.98E-07
<i>CXCL2</i>	6	34701.65	3.87	2.29E-02
<i>CMPK2</i>	11	5036.52	3.82	3.63E-06
<i>INHBA</i>	4	640.91	3.80	1.52E-02
<i>MEFV</i>	25	2935.40	3.74	6.77E-08
<i>IL1B</i>	11	52671.40	3.73	1.45E-02
<i>ENSBTAG00000031795</i>	1	36.15	-3.87	1.56E-02
<i>GAS1</i>	8	39.31	-3.47	1.75E-04
<i>RAB42</i>	2	71.41	-3.42	5.73E-04
<i>PIK3IP1</i>	17	745.19	-3.40	9.17E-08
<i>SRMS</i>	13	35.83	-2.93	2.59E-04
<i>LPAR5</i>	5	35.97	-2.78	2.54E-02
<i>RARG</i>	5	93.86	-2.78	3.48E-10
<i>C19orf35</i>	7	93.20	-2.74	1.41E-02
<i>SGSM1</i>	17	157.99	-2.69	2.23E-09
<i>SMTNL2</i>	19	40.68	-2.69	1.75E-02
<i>TFAP4</i>	25	58.19	-2.60	6.61E-03
<i>SMAD6</i>	10	83.82	-2.54	8.08E-06
<i>CRACR2B</i>	29	30.58	-2.52	2.16E-02
<i>CBFA2T3</i>	18	177.92	-2.49	7.71E-05
<i>TRERF1</i>	23	164.64	-2.42	3.76E-04
<i>DBP</i>	18	174.75	-2.35	7.00E-06
<i>MEF2C</i>	7	66.10	-2.32	6.92E-03
<i>CBX8</i>	19	134.95	-2.27	2.14E-04
<i>NRADD</i>	22	310.52	-2.27	7.48E-07
<i>ENSBTAG00000046828</i>	21	33.24	-2.24	6.64E-03

**Table 1.** Top 20 upregulated and downregulated mRNAs. \*L2FC denotes the log<sub>2</sub> fold-change. <sup>#</sup>p.BH denotes the p-value corrected for multiple testing using Benjamini-Hochberg method.

DE neighbouring protein-coding gene (except the pair *XLOC\_012298-MAF*; Fig. 4). The pair exhibited a congruent co-expression, i.e. the change in expression was in the same direction. As shown in Fig. 4, *TNRC6B* and *MAF* were downregulated similar to their corresponding non-coding RNA neighbours (Table 3). On the other hand, for the upregulated lncRNAs, their corresponding neighbouring genes *ZMIZ2*, *TNIP3*, *TNFAIP3*, *RNF213*, *RHBDF2*, *RELL1*, *RASSF3*, *PIM3*, *PIM1*, *NFKB2*, *MX2*, *IL17REL*, *IFI44L*, *DCSTAMP* and *CSF2RA* were also upregulated (Fig. 4, Table 3).

**Enrichment analysis.** The enrichment analysis of 820 DE mRNAs pointed out 26 significantly enriched KEGG pathways (EASE score<sup>27</sup> <0.05, p.BH <0.05) with ‘TNF signalling pathway’, ‘Influenza A’, ‘NF-κB signalling pathway’, ‘Herpes simplex infection’, ‘Cytosolic DNA-sensing pathway’, ‘Toll-like receptor signalling pathway’, ‘Jak-STAT signalling pathway’, ‘NOD-like receptor signalling pathway’, ‘Cytokine-cytokine receptor interaction’ and ‘Measles’ being the ten most enriched pathways (Fig. 5). Similarly, the significant gene ontology terms (EASE score <0.05, p.BH <0.05) associated with biological processes included ‘cellular response to lipopolysaccharide’, ‘negative regulation of viral genome replication’, ‘positive regulation of NF-κB transcription factor activity’, ‘innate immune response’, ‘defence response to virus’, ‘inflammatory response’, ‘immune response’, ‘positive regulation of interleukin-6 production’, ‘lipopolysaccharide-mediated signalling pathway’ and ‘regulation of transcription from

Gene ID	Chromosome	Base mean	L2FC*	p.BH <sup>‡</sup>
XLOC_000633	1	613.81	4.68	1.57E-12
XLOC_014744	19	86.99	3.27	3.80E-07
XLOC_025859	3	37.60	3.15	6.31E-05
XLOC_007088	14	94.29	2.72	1.45E-02
XLOC_021786	26	63.96	1.96	2.26E-03
XLOC_026730	4	472.66	1.89	4.70E-03
XLOC_007668 <sup>§</sup>	14	309.43	1.76	4.86E-03
XLOC_033995	9	604.78	1.70	1.90E-02
XLOC_001362	1	2492.84	1.64	2.00E-02
XLOC_018958	23	265.57	1.56	1.24E-04
XLOC_005117	12	607.08	-2.24	2.26E-02
XLOC_012297	18	42.87	-1.83	8.13E-03
XLOC_012298	18	34.84	-1.77	7.81E-03
XLOC_031319	7	239.91	-1.60	1.20E-02
XLOC_031318	7	141.73	-1.56	1.52E-02
XLOC_004943	12	574.63	-1.35	1.40E-03
XLOC_028334	5	1178.68	-1.02	4.24E-05
XLOC_010002	16	399.07	-0.85	4.00E-02
XLOC_031072	7	486.79	-0.81	4.43E-02

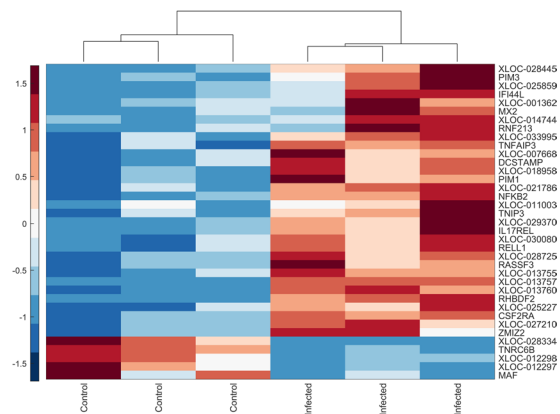
**Table 2.** Top 10 upregulated and top 9 downregulated lncRNAs. \*L2FC denotes the log<sub>2</sub> fold-change. <sup>‡</sup>p.BH denotes the p-value corrected for multiple testing using Benjamini-Hochberg method. <sup>§</sup>The region XLOC\_007668 assembled by Cuffmerge contained both lncRNA and mRNA DCSTAMP loci. Therefore, here for we provide L2FC and p.BH estimates for region encompassed by the lncRNA transcripts within the XLOC\_007668 region.

Gene symbol	L2FC*	p.BH <sup>‡</sup>	lncRNA neighbour	Distance (bp)	L2FC*	p.BH <sup>‡</sup>
MAF	-1.89	2.26E-03	XLOC_012297	14863	-1.83	8.13E-03
			XLOC_012298	14231	-1.77	7.81E-03
TNRC6B	-0.55	2.11E-02	XLOC_028334	3506	-1.02	4.24E-05
ZMIZ2	0.78	3.09E-02	XLOC_027210	4650	1.27	2.11E-02
CSF2RA	0.78	2.40E-03	XLOC_025227	23467	1.04	6.46E-03
RHBDF2	0.86	5.02E-05	XLOC_013760	9945	0.94	4.54E-03
			XLOC_013755	14339	1.04	1.20E-02
			XLOC_013757	12713	1.08	9.79E-05
RASSF3	0.94	2.56E-02	XLOC_028725	2463	1.04	7.35E-03
RELL1	1.18	1.09E-02	XLOC_030080	14289	1.43	4.41E-04
IL17REL	1.30	3.02E-05	XLOC_029370	791	1.27	6.70E-05
TNIP3	1.42	1.68E-02	XLOC_011003	5602	1.27	3.64E-02
NFKB2	1.50	7.94E-05	XLOC_021786	325	1.96	2.26E-03
PIM1	1.55	2.35E-05	XLOC_018958	515	1.56	1.24E-04
DCSTAMP	1.57	1.72E-02	XLOC_007668	6809	1.76	4.86E-03
TNFAIP3	1.76	5.25E-03	XLOC_033995	1049	1.70	1.90E-02
RNF213	1.91	1.41E-02	XLOC_014744	24052	3.27	3.80E-07
MX2	2.25	2.56E-02	XLOC_001362	96	1.64	2.00E-02
IFI44L	2.57	3.72E-05	XLOC_025859	2065	3.15	6.31E-05
PIM3	2.63	7.62E-09	XLOC_028445	2456	1.18	2.72E-04

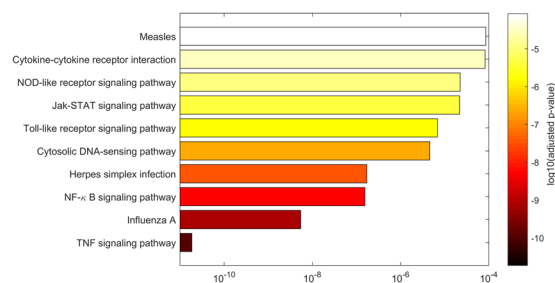
**Table 3.** List of differentially expressed mRNA neighbours to differentially expressed lncRNAs. \*L2FC denotes the log<sub>2</sub> fold-change. <sup>‡</sup>p.BH denotes the p-value corrected for multiple testing using Benjamini-Hochberg method.

RNA polymerase II promoter' (Supplementary Fig. S5). Significantly enriched UniProtKB keywords (EASE score <0.05, p.BH <0.05) included 'Cytokine', 'Activator', 'Transcription regulation', 'Innate immunity', 'Inflammatory response' and 'Transcription' (Supplementary Fig. S5). We also performed an enrichment analysis of the neighbouring mRNA genes of DE lncRNAs (Supplementary Fig. S6). Pathways and functions associated with immune response such as NF- $\kappa$ B signalling (*TNFRSF11A*, *NFKB2*, *TNFAIP3*), organelle fission (*MX1*, *MX2*) and cytokine production (*MAF*, *NFAT5*) were enriched (EASE score <0.05).





**Figure 4.** Co-expression analysis. A hierarchical heatmap showing the co-expression of differentially expressed lncRNA candidates and its corresponding differentially expressed neighbour mRNA. Red shows higher expression and blue shows lower expression.



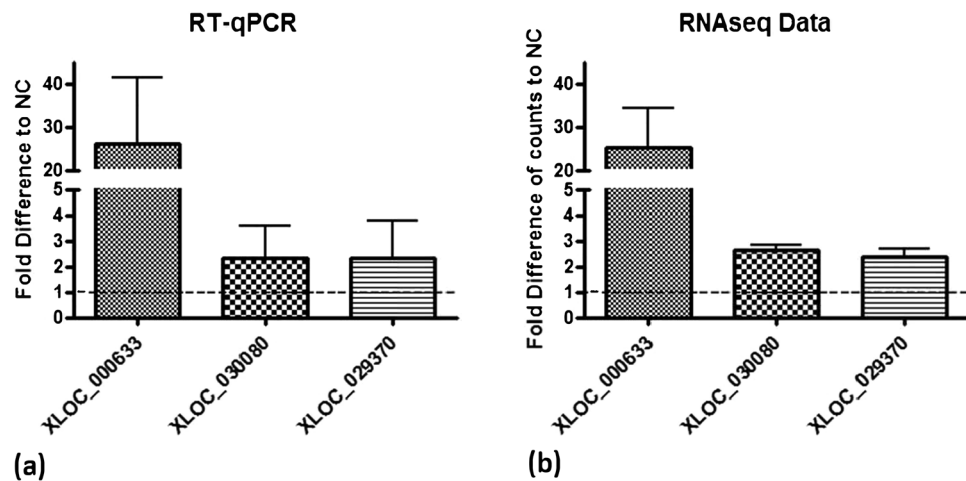
**Figure 5.** Top ten KEGG terms for the differentially expressed mRNA. An overview of the significantly enriched KEGG terms ( $p.BH < 0.05$ ) for the differentially expressed mRNA in the form of a bar plot. The length of the bar plot indicates p.BH i.e. the adjusted p-value obtained using Benjamini-Hochberg method for multiple test correction.

**Overlap with previously published lncRNAs in cattle.** In total, we found 172 of 397 candidate lncRNA transcripts in our dataset to have an overlap (i.e. the number of nucleotides shared between the two overlapping features on the same strand) with previously published transcripts (Supplementary Data 6) in cattle from five different sources mentioned below. 225 of our candidate transcripts were unique, having no match to any transcript from the sources investigated. We identified 85 and 63 transcripts having a match with a non-coding RNA in the NONCODE (NONCODE2016\_bosTau6) and ALDB (ALDB.cow.lincRNAs.v1.0) databases (Supplementary Table S4). 36, 53 and 44 transcripts found a match with a non-coding RNA listed in the publications from Weikard *et al.*<sup>18</sup>, Billerey *et al.*<sup>19</sup> and Koufariotis *et al.*<sup>28</sup>, respectively (Supplementary Table S4). No intersection was found among the five sources (Supplementary Fig. S7). 80 transcripts had an overlap of more than 90% with a previously published non-coding RNA (Supplementary Data 6, Sheet 1, Column 20).

**RT-qPCR analysis reflects expression of selected lncRNAs determined by RNA sequencing.** To validate the RNA sequencing data and the implemented lncRNA nomination pipeline, expression of three randomly selected lncRNAs was analysed by means of RT-qPCR. Expression patterns determined by RT-qPCR reflects the RNA sequencing data (Fig. 6a,b). *XLOC\_000633* was highly upregulated in infected samples compared to the controls (analysis with RT-qPCR resulted in a mean fold difference = 26.09, the mean fold difference calculated for sequencing counts compared to negative controls = 25.24). *XLOC\_030080* and *XLOC\_029370* were moderately upregulated by the infection compared to the controls. RT-qPCR resulted in a mean fold difference = 2.32 for *XLOC\_030080* and a mean fold difference = 2.34 for *XLOC\_029370*. The mean fold difference calculated for sequencing counts was 2.64 for *XLOC\_030080* and 2.38 for *XLOC\_029370*.

## Discussion

Paratuberculosis is being recognized as a major problem affecting animal health, farming and the dairy industry worldwide. Because of specific virulence mechanisms employed by the pathogen, infection may result in a long-term subclinical carriage and may finally lead to diarrhoea and progressive wasting. In this study, we used RNA-Seq data to investigate the alteration in genomic expression profiles of mRNA and lncRNA in bovine macrophages upon MAP infection. A comprehensive computational pipeline adapted from earlier studies<sup>17–21</sup> was implemented for the identification of lncRNAs in cattle. We inferred the putative biological roles of the identified candidate lncRNAs by assessing the function of neighbouring protein-coding genes and their co-expression



**Figure 6.** RT-qPCR validation. (a) Expression of selected lncRNAs validated with RT-qPCR and (b) in comparison fold differences of transcript counts to negative control (NC) determined by RNA sequencing.

patterns. The current study brings up novel information about the early immune response of bovine macrophages upon MAP infection and points out the importance of regulatory RNAs.

### Analysis of mRNAs confirms and broadens existing knowledge on early immune response of macrophages upon MAP infection.

One aim of the present study was the analysis of mRNAs expressed in bovine macrophages in response to infection with MAP. As expected, pathway analysis revealed the highest enrichment of identified coding genes in general immune related functions such as TNF signalling, NF- $\kappa$ B signalling and TLR signalling. The most upregulated mRNA (ranked by L2FC) was C-C motif chemokine ligand 8 (*CCL8*, Table 1), also known as monocyte chemoattractant protein 2 (*MCP2*) – a pro-inflammatory chemokine which is primarily expressed in monocytes/macrophages e.g. after Interferon (INF) stimulation from T cells or upon recognition of antigens. *CCL8* was demonstrated to be highly expressed in pleural effusions in *M. tuberculosis* infected patients as well as in *M. bovis* BCG and *M. tuberculosis* infected human and mouse macrophages<sup>29</sup> and is a promising biomarker for detection of tuberculosis<sup>30</sup>. *CCL8* was also among the most upregulated genes in macrophages from red deer and in bovine ileal tissue infected with MAP<sup>31,32</sup>. Furthermore, *CCL8* has been linked to granuloma formation<sup>33</sup> – a key event in paratuberculosis pathogenesis. In our study, further chemotactic molecules such as *CXCL2*, *CSF2*, *CSF3* and pro-inflammatory cytokines such as *TNF* and *IL-1 $\beta$*  were among the top 20 upregulated coding genes reinforcing that chemokine and cytokine signalling is a central event in early mycobacterial infection *in vitro*<sup>34</sup>.

The data suggests a MAP driven immune response favouring the establishment of infection, as a strong induction of anti-inflammatory mediators such as *IL-10*, *IL-27* and Suppressor of Cytokine Signalling (SOCS) 1 and 3 was observed (Supplementary Data 3). These factors have been previously identified as potential key players in the pathogenesis of MAP<sup>6</sup>. As an example, *IL-10* inhibits production of pro-inflammatory cytokines and decreases antigen presentation by macrophages. Mycobacteria-mediated upregulation of *IL-10* is a well-known virulence mechanism to regulate antimicrobial activity resulting in enhanced pathogen survival<sup>35</sup>. Further confirming previous findings, we found a decreased expression of several *RAB* (Ras-related in brain) family members, such as *RAB3A*, *RAB42* or *RAB11* family-interacting protein 5 (Supplementary Data 3). These proteins are mainly involved in vesicular trafficking and the coordination of endosomal transport<sup>36</sup>. *RAB3A* was the most downregulated gene in bovine macrophages 2 h after infection with MAP, likely representing a strategy of MAP to inhibit phagosomal maturation to promote its own intracellular survival<sup>37</sup>.

Interestingly, RNA analysis in our study reflects that MAP seems to have a striking impact on INF induced signalling. INFs are cytokines that are widely produced and are involved in the inflammatory response towards several infectious agents and surveillance of malignant cells<sup>38</sup>. We found that several Type I INF response genes were DE by MAP infection, including *ISG20*, *CMPK2* and *IFIT 1* and 2<sup>39</sup> (which were also among the top 20 upregulated mRNAs), supporting the relevance of Type I INF signalling in the MAP-macrophage interplay.

In the present study, we identified several genes that have not yet been clearly linked to mycobacterial infections (e.g. leukaemia inhibitory factor, interleukin 6 family cytokine (*LIF*), the schlafen family member 11 (*SLFN11*) or Growth arrest specific 1 (*GAS1*)). *LIF* was among the top five upregulated transcripts. However, the role of *LIF* in bacterial infection is not clear. It was shown to stimulate the production of inflammatory cytokines *IL-1 $\beta$* , *IL-6*, and *IL-8*<sup>40</sup>. In case of mycobacterial infection, it was induced in macrophages infected with *M. smegmatis* but not with pathogenic *M. tuberculosis* or *M. avium*. The authors concluded that it plays a role in clearing non-pathogenic mycobacteria<sup>41</sup>. To our knowledge, it was yet not associated with Paratuberculosis. Another example is *SLFN11*, which has been described as an inhibitor of HIV-1 by interfering with viral protein synthesis<sup>42</sup>. Primary immune cells and other tissue express human *SLFN11* in response to INF Type I signalling<sup>43</sup>. *SLFN11* and other members of the schlafen family have also been shown to be induced by TLR ligands<sup>43</sup> or bacterial infection pointing to a role in immune processes. However, their role in bacterial infection is not clear. The most downregulated DE coding gene in bovine macrophages in response to MAP was *GAS1*, which plays a role



in growth suppression and is a putative tumour suppressor gene<sup>44</sup>. Overexpression induces apoptosis in tumour cells<sup>44</sup> suggesting a role in the cell cycle or apoptosis. However, it has not been associated to infectious diseases.

**Analysis of lncRNAs point to regulatory functions in the immune response towards MAP infection.** In the present study, we were interested in the expression of lncRNAs since they were reported to play a fundamental role in the regulation of the immune response towards several bacterial pathogens<sup>15</sup> and are known to be influenced by mycobacterial infection<sup>12</sup>. The identification and functional annotation of lncRNAs in cattle is scarce<sup>16</sup>, therefore, one of the main challenges of the study was to identify lncRNAs and infer their biological role. For this, we implemented a computational pipeline based on earlier studies<sup>17–21</sup> that allowed nominating candidate lncRNAs. As described in these studies, we utilized Tophat2/Cufflinks for the initial alignment and assembly steps. This allowed us to obtain comparable results, especially with respect to previous studies on cattle<sup>28,45</sup>. Moreover, we obtained very good alignment statistics (>90% concordant alignment for all samples, Table S1) with Tophat2, which in comparison to its succeeding software<sup>46</sup> with faster search algorithm has been shown to provide similar accuracy<sup>46,47</sup>. Additionally, many layers of different information from each step in the pipeline synergistically aided in the identification of potentially functional lncRNA transcripts. As recently shown, ‘the choice of differential expression analysis approach exhibits the strongest impact on results’, with more modest influences from the read aligner and expression modeller<sup>48,49</sup>, we utilized DESeq2 that has been shown to perform robustly in comparison to other existing differential expression tools<sup>50</sup>. Thus, in conjunction with the chosen filtering parameters, coding potential assessment and bioinformatics analysis, we generated a reliable dataset in our study. Nonetheless, we strongly suggest testing the role newer tools, such as pseudoaligners, in future studies with respect to the identification and discovery of unannotated and novel lncRNAs.

Following stringent filtering criteria based on the genomic features comprising exon number, length, expression levels and coding potential, we obtained 397 candidate lncRNAs of which 225 were novel. The genomic features of our candidate lncRNAs were in concordance with earlier observations in cattle and other species<sup>17–21,28,51</sup>, which further establishes the validity of our pipeline. For instance, the candidate lncRNAs possessed lower expression levels (Fig. 2c), fewer exons (Fig. 2a) and the length was typically shorter (Fig. 2b) relative to mRNA, likely attributed to the fewer number of exons in the transcript. The identified lncRNAs exhibited lower average conservation across the transcripts in comparison to protein-coding mRNA, which may provide them the ability to undergo rapid functional diversification. For instance, the validated non-coding RNA gene locus *XLOC\_000633* had a very low average conservation score of 0.0331. However, in agreement to earlier studies<sup>20,52</sup>, we observed local regions displaying reasonable conservation (Supplementary Data 2) relative to the rest of the transcript, showing that lncRNAs display more positional conservation than sequence conservation.

We observed an overlap between few predicted lncRNA candidates and the previously published non-coding RNAs from five different sources for cattle. Although, the previously published non-coding RNAs were identified in different types of tissue, it is likely that some of the candidate lncRNAs share a common functionality and expression with their overlapping counterpart, pointing towards the fact that the expression of some of these lncRNAs may not be cell-specific. However, to the best of our knowledge, inadequate information existed regarding the functionality of these previously published non-coding RNAs. Therefore, a fundamental question raised following the identification of novel lncRNAs is regarding their relevance in biological function. To address this question, we utilized the fact that lncRNA fine-tune the expression of neighbouring genes<sup>53</sup>. Interestingly, on comparing the distribution of lncRNAs in relation to mRNAs across the genome (as described earlier in the results section), we observed that several DE lncRNAs were present on chromosomes 18 and 19, a pattern analogous to DE mRNA. Interestingly, chromosomes 18 and 19 harbour some of the protein-coding genes associated with signalling pathways for innate immune response. For instance, DE genes on chromosomes 18 and 19 are involved with TNF signalling pathway (*CCL5*, *CCL2*, *MAP2K3*, *MAP2K4*, *MAP3K14*, *MLKL*, *PIK3R5* and *SOCS3*), chemokine signalling pathway (*CCL5*, *NFKB1B*, *ARRB2*, *CCL2*, *CCL4*, *CCL8*, *PIK3R5*), NOD-like receptor signalling pathway (*CCL5*, *NFKB1B*, *NLRP1*, *CCL2*) and cytokine-cytokine receptor interaction (*CCL5*, *CCL2*, *CCL4*, *CCL8*, *CSF3*). To investigate the functional aspect in more detail, we analysed the neighbouring genes of the lncRNAs. Interestingly, we observed differential expression of corresponding mRNA neighbours for 20 DE lncRNAs. The co-expression analysis indicated that these neighbouring lncRNA-mRNA pairs showed changes in expression along the same direction (Fig. 4). Notably, the distance between DE lncRNA-mRNA gene pairs was much shorter than the commonly observed distance between lncRNA-mRNA pairs (as described in the results section), which may support the hypothesis that these lncRNA participated in cross-regulation. Enrichment analysis clearly showed that neighbouring DE mRNAs were involved in pathways related to immune response (Supplementary Fig. S6). For example, this included protein coding genes that play a role in NF- $\kappa$ B signalling such as TNFAIP3 interacting protein 3 (*TNIP3*), TNF alpha induced protein 3 (*TNFAIP3*) and NF- $\kappa$ B2.

TNIP3 and TNFAIP3 negatively influence NF- $\kappa$ B signalling. *TNFAIP3* is induced by TNF and limits inflammation by terminating TNF induced NF- $\kappa$ B response<sup>54</sup>. The murine lncRNA *lincRNA-TNFAIP3*, located proximal to the *TNFAIP3* gene, was shown to be an early-primary response gene controlled by NF- $\kappa$ B signalling in murine macrophages<sup>55</sup>. The authors demonstrated that *lincRNA-TNFAIP3* seems to assemble a NF- $\kappa$ B/HMGB1/*lincRNA-TNFAIP3*-complex in response to lipopolysaccharides which leads to epigenetic chromatin remodelling and transactivation of inflammatory genes. It was recently shown that a lncRNA adjacent to *TNFAIP3* was among eight lncRNAs that were exclusively expressed in *Campylobacter concisus* infected THP-1 macrophages accompanied by co-expression of *TNFAIP3*<sup>56</sup>. Strikingly, in our study the TNFAIP3 interacting protein (*TNIP3*) was also differentially regulated as well as the adjacent lncRNA further suggesting that MAP-induced lncRNAs influence macrophage response towards infection. Regulation of *TNFAIP3* during mycobacterial infection has been shown for miRNAs leading to limitation of NF- $\kappa$ B inflammatory signalling which favours mycobacterial survival<sup>57</sup>. Thus, regulation of NF- $\kappa$ B signalling by *TNFAIP3* may be orchestrated by multiple classes of non-coding RNAs including lncRNAs.

Interestingly, in our study, two protein coding genes (*IFI44F* and *MX2*) that have been implicated in the cellular response to type I interferons<sup>58,59</sup> as well as their neighbouring lncRNAs were significantly upregulated in response to infection suggesting a supportive role of lncRNAs in type I INF signalling during MAP infection. In accordance with that assumption, it has recently been shown that type I INFs regulate the expression of lncRNAs and vice versa<sup>60</sup>. However, to which extent the identified lncRNAs in our study contribute to the regulation of INF signalling during MAP infection has to be analysed in more depth in the future.

Among the possible mode of actions of lncRNAs, their regulatory role may be either inhibiting or stimulating gene expression. In our study, several lncRNAs were highly regulated by the infection, however, their functional role remains to be investigated and validated in terms of regulatory mechanisms. The potential of lncRNAs to serve as a biomarker has previously been demonstrated in *M. tuberculosis* infection<sup>12</sup> and may also be a promising approach in case of Paratuberculosis.

In conclusion, we have shown that MAP induces an expression profile of genes related to immune response and bactericidal defence mechanisms in bovine macrophages. Our data support existing knowledge about the ability of MAP to subvert the host immune defence. However, we also identified DE genes, which have not yet been associated with Paratuberculosis infections, opening new insight into the pathogenic mechanisms employed by MAP or yet unknown defence strategies used by the host cell. To our knowledge, this is the first study that provides novel insights into the role of lncRNAs and their interplay with mRNAs during MAP infection in cattle.

Whether identified lncRNAs are specific to MAP infection and may serve as potential biomarkers to detect infected animals has to be the matter of future investigations.

## Materials and Methods

**Isolation and Cultivation of Primary Bovine Macrophages.** Whole blood of three 15 month old Holstein Friesian cows was obtained from the jugular vein under sterile conditions and kept in Na-citrate to prevent coagulation. Animals were tested negative for MAP infection by ELISA and fecal culture (Laboklin, Bad Kissingen, Germany). Blood mononuclear cells were isolated by Ficoll-Paque Plus (GE Healthcare, Sigma-Aldrich, Taufkirchen, Germany) centrifugation, three washing steps with PBS and subsequent attachment to cell-culture plate surface. Cells were cultivated at 37 °C and 5% CO<sub>2</sub> for 6 days in six well plates in 1.5 ml IMDM medium containing 10% FCS, 10 µg/ml gentamycin (Biochrom, Berlin, Germany) and 30 ng/ml M-CSF (PAN Biotech, Aidenbach, Germany) and 1 x Penicillin-Streptomycin-Amphotericin B (Sigma-Aldrich, Taufkirchen, Germany). The media was changed one to two times during cultivation.

Flow cytometry was performed to analyse CD14 expression, a cell surface marker for monocytes and macrophages, to reveal the purity of the isolated macrophage population. For that purpose, part of the isolated cells were stained with a CD14 antibody as well as a secondary fluorochrome-labelled antibody and analysed with a FACSCalibur flowcytometer (Becton Dickinson GmbH, Heidelberg, Germany) as described previously<sup>61</sup>. The fluorochrome-stained cells were gated and analysed for CD14 antibody staining. If FACS analysis proved that the analysed cell population contained around 90% CD14 + cells (Supplementary Fig. S8), cells were washed with PBS and fresh IMDM media was supplied containing 10% FCS. For quantification, two wells were counted by detaching the cells with Accutase (PAA, Pasching, Austria) and cell number was determined by staining the cells with trypan blue and counting in a Neubauer-Chamber. Subsequently, cells were infected as described below.

**Bacterial strain and culture condition.** MAP (ATCC 19698) was cultured at 37 °C on Herrold's Egg yolk slants (BD Life Sciences, Heidelberg, Germany) or Agar plates (1 l containing 900 ml Aqua dest, 4.7 g Middlebrook-Bouillon 7H9 (BD Life Sciences, Heidelberg, Germany), 2 ml Glycerin (Roth, Karlsruhe, Germany), 0.5 g Tween 80 (Roth, Karlsruhe, Germany), 16 g Agar-Agar (Roth, Karlsruhe, Germany), 100 ml Middlebrook OADC-Supplement (BD Life Sciences, Heidelberg, Germany) and 2 ml Mycobactin J (ID Vet, Grabels, France). Cultures were transferred from slants or plates to Middlebrook-Bouillon (1 l containing 900 ml Aqua dest, 4.7 g Middlebrook-Bouillon 7H9, 2 ml Glycerin, 0.5 g Tween 80, 100 ml Middlebrook OADC-Supplement and 2 ml Mycobactin J) and kept at 37 °C until the culture reached OD<sub>600</sub> = 1. Bacteria were frozen in PBS containing 10% Glycerol until used for infection experiments. For quantification of bacteria, the number of colony forming units was determined by plating serial dilutions on 7H9 agar plates which were incubated at 37 °C until cultures were visible.

**Infection experiments.** Approximately  $2 \times 10^7$  bacterial cells were inoculated on  $2 \times 10^5$  primary bovine macrophages (multiplicity of infection [MOI] = 100) and incubated at 37 °C and 5% CO<sub>2</sub>. Non-infected cells served as a negative control. After 2 h of infection, cells were washed three times with PBS and incubated in fresh media containing 10% FCS for another 4 h to remove remaining extracellular bacteria. Samples were taken 6 h after infection. For RNA extraction, cells were washed three times with PBS, lysed with RNA lysis buffer (mirVANA, Life Technologies, Darmstadt, Germany) and total RNA was isolated according to the manufacturer's instruction.

**Quality control of isolated RNA.** Quantity and quality of RNA were first determined by measuring absorbance at 260 and 280 nm with a Nano Drop 1000 spectrophotometer according to the manufacturer's instructions (Thermo Fisher Scientific, Germany). Samples were further analysed for their RNA integrity with an Agilent 2100 BioAnalyzer and RNA 6000 Nano Kits (Agilent, Waldbronn, Germany) according to the manufacturer's protocol. RNA with integrity value (RIN) of  $\geq 8.8$  was used for further investigation.

**RNA library preparation and sequencing.** Sequencing of total RNA was performed at the Institute of Clinical Molecular Biology at Christian-Albrechts-University Kiel for six samples, where both infected and

control samples consisted of three biological replicates. mRNA and lncRNA expression were derived from same total RNA sequencing. Input RNA quantity was assessed using a Qubit 2.0 Fluorometer and the Qubit RNA BR Assay Kit according to manufacturer's instructions. Sequencing libraries were constructed using the Illumina TruSeq stranded total RNA kit according to manufacturer's instructions with an input of 500 ng total RNA per sample. This kit includes ribosomal RNA removal by Ribozero treatment. The quality of final libraries was checked on a TapeStation2200 using the D1000 Screen Tape. Libraries were sequenced on an Illumina HiSeq4000 (Illumina) with  $2 \times 75$  bp paired end reads and six samples per lane.

**Data processing, mapping and transcriptome assembly.** For the initial steps of the analysis, mainly involving sequence alignment and assembly, we utilized the main cloud-computing server of Galaxy (usegalaxy.org), which provides an open source, web-based platform for the next generation sequencing specific analysis. Prior to mapping and assembly, the raw RNA-Seq fastq files were transformed (FASTQ Groomer, Galaxy version 1.0.4) to a format compatible with Galaxy and assessed for quality using FastQC (Galaxy version 0.68). The paired-end reads from each sample were aligned to the cow genome (bosTau8 June 2014) using the spliced-read aligner Tophat2<sup>62</sup> (Galaxy version 2.1.0). An *ab initio* transcriptome assembly was performed with Cufflinks<sup>63</sup> (Galaxy version 2.2.1) using uniquely mapped and properly paired reads to identify annotated and novel transcripts in the data. Finally, using Cuffmerge<sup>63</sup> (Galaxy version 2.2.1.0) we created a single assembly GTF file from the individual Cufflinks assemblies. In order to maximize the overall assembly quality with Cuffmerge, we provided the cow genome and annotation files (Ensemble UMD3.1, [https://support.illumina.com/sequencing/sequencing\\_software/igenome.html](https://support.illumina.com/sequencing/sequencing_software/igenome.html)) as inputs.

**Filtering of transcripts.** To select potential long non-coding RNAs from the transcripts generated by Cufflinks, firstly, we retained transcripts with no known annotation (novel intergenic: class code u and novel antisense: class code x). In the next step, we discarded transcripts with a length less than 200 bp. It has been shown that many lncRNAs consist of single-exon transcripts<sup>18,28,64</sup>. However, often single-exon transcripts are discarded as they can potentially contribute to background noise resulting from erroneous transcript assembly, genomic contamination in the RNA-Seq library or experimental artifacts. An earlier study in humans<sup>20</sup> applied a stringent size threshold of 1 kb for selecting single exon transcripts to overcome the above-mentioned problems. We applied a size threshold of 500 bp (similar to that reported in a recent study by Bush *et al.*<sup>65</sup>) to single-exon transcripts, considering the above-mentioned limitations and the fact that the average length of a single-exon transcripts in cattle is reported to be 286 bp<sup>64</sup>. We estimated number of fragments in all paired-end libraries using featureCounts<sup>66</sup> (version 1.5.2), and normalized the expression in terms of fragments per kilobase of transcript per million mapped reads (FPKM). Finally, to select expressed versus background transcripts, we discarded any transcript having an abundance of less than 1.5 FPKM in all six samples based on earlier studies<sup>17,20,67</sup>. However, we strongly recommend investigating the distribution of transcript abundance for selecting a reasonable cut-off.

**Characterization of the coding potential.** We used coding potential calculator<sup>23</sup> (CPC; version cpc-0.9-r2), Coding-Potential Assessment Tool<sup>24</sup> (CPAT; version 1.2.2) and hmmscan algorithm (version HMMER 3.1b2, <http://hmmer.org/>) to examine the protein coding ability of the shortlisted transcripts. CPC and CPAT utilize sequence features to distinguish coding from non-coding RNAs, therefore, we extracted DNA sequences of the transcripts using Cufflinks' gffread<sup>63</sup>. To classify a transcript as non-coding, we chose a stringent threshold of  $-0.5$  for CPC to reduce possible false positive classification<sup>18</sup>. Next, we used CPAT to assess the protein coding potential of candidate transcripts. In order to choose the cut-off threshold for cattle, we prepared a training dataset consisting of 10,000 known 'coding sequences' and 10,000 'non-coding sequences' (UMD3.1; [ftp://ftp.ensembl.org/pub/release-91/fasta/bos\\_taurus/](ftp://ftp.ensembl.org/pub/release-91/fasta/bos_taurus/)), similar to Billerey *et al.*<sup>19</sup>. As only 3,801 sequences were available for the non-coding RNAs, we used 6,199 intron sequences for creating a balanced training dataset. Both the coding and intronic sequences were selected randomly. We performed 10-fold cross-validation to obtain the optimal cut-off value from the two-graph ROC curve. In comparison to other studies, we also arrived at a similar<sup>19</sup> cut-off value of 0.34 (Supplementary Fig. S9). Lastly, we used the hmmscan algorithm to scan the Pfam protein domains in the candidate lncRNAs. The sequences were translated into the three open reading frames using EMBOSS transeq<sup>68</sup> (EMBOSS: 6.5.7.0) and queried against the Pfam protein profile database<sup>22</sup> (version Pfam31.0) using hmmscan. Candidate lncRNAs with a match to a domain in a known protein sequence were discarded. Finally, we chose only those transcripts as non-coding RNAs that did not possess coding potential with respect to any of the three above-mentioned tools.

**Conservation analysis.** We downloaded the available PhastCons scores<sup>25</sup> from UCSC (<http://hgdownload.soe.ucsc.edu/goldenPath/bosTau4/phastCons5way/>) for multiple alignments of four vertebrate genomes (dog, human, mouse and platypus) to the cow genome. For extracting the conservation scores for our genomic region of interest, we converted the file to BED format using bedops<sup>69</sup> (version 2.4.26). Since the PhastCons scores are only available for the bosTau4 assembly, we used the 'liftover' functionality of UCSC to convert the bosTau8 assembly based genome coordinates of our candidate lncRNAs to bosTau4. Finally, we computed an average conservation score from the base-by-base conservation scores for our candidate lncRNA transcripts.

**Nearest neighbour analysis to predict putative target genes.** Using bedtools closest<sup>69</sup> (version 2.26.0), we searched for the nearest cis/trans and upstream/downstream non-overlapping neighbours of the lncRNAs with respect to known annotated genes (Ensemble UMD3.1). Thereafter, we deduced the biological functionality of our candidate lncRNAs by investigating the properties (such as biological role and level of expression) of its closest identified gene.

**Comparison to known lncRNAs.** We compared the overlap of our lncRNAs to previously published lncRNAs<sup>18,19,28</sup> and lncRNAs in the NONCODE (NONCODE2016\_bosTau6) and ALDB (ALDB.cow.lincRNAs.v1.0) databases for cattle. Using bedtools intersect<sup>69</sup> (version 2.26.0), we reported the ‘degree of overlap’ as the number of nucleotides shared between the two overlapping features on the same strand. Since the annotation BED file in the NONCODE database is based on the bosTau6 assembly, we used the ‘liftover’ functionality of UCSC to convert the BED file from the bosTau6 to the bosTau8 assembly.

**Differential expression, gene set enrichment and pathway analysis.** Differential expression analysis was performed using the count-based method DESeq2<sup>70</sup> (version 1.16.1) that utilizes a negative binomial distribution to model the count. A gene was considered significantly DE if  $p_{BH} < 0.05$ .

To infer the biological function of the gene expression data, we performed gene set enrichment analysis for the DE genes using Functional Annotation Tool in DAVID<sup>27</sup> (version 6.8). Using the DE protein-coding gene list as an input to DAVID and *Bos tau* as background, we obtained annotation terms pertinent to MAP infection. DAVID provided EASE score (a modified Fisher Exact p-value) as well as  $p_{BH}$  signifying enriched biological annotations. An EASE score  $< 0.05$  was used as a cut-off.

**Validation of selected lncRNAs using RT-qPCR.** cDNA was synthesized from two RNA samples (the same that were used for RNA-Seq) of infected as well as non-infected macrophages with the Maxima First Strand cDNA synthesis kit (Thermo Scientific) as described in the manufacturer’s protocol. Pooled cDNA was taken as a template for testing the primers listed in Supplementary Table S5 via gradient RT-qPCR and gel electrophoresis of PCR products. Amplicons showing the expected size were isolated and sequenced to ensure specificity. Expression analysis was performed by means of SYBR Green detection chemistry using the SensiMix SYBR Hi-ROX Kit (Bioline GmbH) as described before<sup>71</sup>. A PikoReal Cycler was used instead of a Step one plus. Expression of housekeeping genes *GAPDH* and *SDHA* was analysed in parallel to normalize the data. The  $\Delta\Delta CT$  method<sup>72</sup> was used to calculate the relative fold difference of RNA expression levels compared to the negative control.

**Ethics.** The study was approved by the local animal welfare committee of the regional authorities at the federal state Brandenburg, Germany (No. 2347-A-3-1-2016) and carried out in accordance with the approved guidelines and regulations.

## Data Availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files). The raw RNA-Seq fastq data discussed in this publication have been deposited in NCBI’s Gene Expression Omnibus<sup>73</sup> and are accessible through GEO Series accession number GSE122933 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE122933>).

## References

- Sweeney, R. W. Pathogenesis of paratuberculosis. *Vet Clin North Am Food Anim Pract* **27**, 537–546 (2011).
- McNees, A. L., Markesich, D., Zayyani, N. R. & Graham, D. Y. *Mycobacterium paratuberculosis* as a cause of Crohn’s disease. *Expert Rev Gastroenterol Hepatol* **9**, 1523–1534 (2015).
- Muskens, J. *et al.* Prevalence and regional distribution of paratuberculosis in dairy herds in The Netherlands. *Vet Microbiol* **77**, 253–261 (2000).
- Good, M. *et al.* Prevalence and distribution of paratuberculosis (Johne’s disease) in cattle herds in Ireland. *Ir Vet J* **62**, 597–606 (2009).
- Pozzato, N. *et al.* Prevalence of paratuberculosis infection in dairy cattle in Northern Italy. *Prev Vet Med* **102**, 83–6 (2011).
- Arsenault, R. J. *et al.* From mouth to macrophage: mechanisms of innate immune subversion by *Mycobacterium avium* subsp. *Paratuberculosis*. *Vet Res* **45**, 54 (2014).
- Rathnaiah, G. *et al.* Pathogenesis, Molecular Genetics, and Genomics of *Mycobacterium avium* subsp. *paratuberculosis*, the Etiologic Agent of Johne’s Disease. *Front Vet Sci* **4**, 187 (2017).
- Hostetter, J., Steadham, E., Haynes, J., Bailey, T. & Cheville, N. Phagosomal maturation and intracellular survival of *Mycobacterium avium* subspecies *paratuberculosis* in J774 cells. *Comp Immunol Microbiol Infect Dis* **26**, 269–283 (2003).
- Arsenault, R. J. *et al.* *Mycobacterium avium* subsp. *paratuberculosis* inhibits interferon gamma-induced signaling in bovine monocytes: insights into the cellular mechanisms of Johne’s disease. *Infect Immun* **80**, 3039–3048 (2012).
- Clark, D. L. *et al.* Detection of *Mycobacterium avium* subspecies *paratuberculosis*: comparing fecal culture versus serum enzyme-linked immunosorbent assay and direct fecal polymerase chain reaction. *J Dairy Sci* **91**, 2620–2627 (2008).
- Fradet, Y. *et al.* uPM3, a new molecular urine test for the detection of prostate cancer. *Urology* **64**, 311–315 (2004).
- Yang, X. *et al.* Microarray analysis of long noncoding RNA and mRNA expression profiles in human macrophages infected with *Mycobacterium tuberculosis*. *Sci Rep* **6**, 38963 (2016).
- zur Bruegge, J., Einspanier, R. & Sharbati, S. A Long Journey Ahead: Long Non-coding RNAs in Bacterial Infections. *Front Cell Infect Microbiol* **7**, 95 (2017).
- Cech, T. R. & Steitz, J. A. The noncoding RNA revolution—trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
- Pang, K. C., Frith, M. C. & Mattick, J. S. Rapid evolution of noncoding RNAs: lack of conservation does not mean lack of function. *Trends in Genetics* **22**(Suppl 1), 1–5 (2006).
- Weikard, R., Demasius, W. & Kuehn, C. Mining long noncoding RNA in livestock. *Anim Genet* **48**, 3–18 (2017).
- Prensner, J. R. *et al.* Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. *Nat Biotechnol* **29**, 742–749 (2011).
- Weikard, R., Hadlich, F. & Kuehn, C. Identification of novel transcripts and noncoding RNAs in bovine skin by deep next generation sequencing. *BMC Genomics* **14**, 789 (2013).
- Billerey, C. *et al.* Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics* **15**, 499 (2014).
- Sun, M., Gadad, S. S., Kim, D. S. & Kraus, W. L. Discovery, Annotation, and Functional Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells. *Mol Cell* **59**, 698–711 (2015).
- Tsoi, L. C. *et al.* Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. *Genome Biol* **16**, 24 (2015).



22. Bateman, A. *et al.* The Pfam Protein Families Database. *Nucleic Acids Res* **28**, 263–266 (2000).
23. Kong, L. *et al.* CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* **35**, W345–W459 (2007).
24. Wang, L. *et al.* CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res* **41**, e74 (2013).
25. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**, 1034–1050 (2005).
26. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**, 289–300 (1995).
27. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44–57 (2009).
28. Koufariotis, L. T., Chen, Y. P., Chamberlain, A., Vander Jagt, C. & Hayes, B. J. A catalogue of novel bovine long noncoding RNA across 18 tissues. *PLoS One* **10**, e0141225 (2015).
29. Liu, H. *et al.* Induction of CCL8/MCP-2 by mycobacteria through the activation of TLR2/PI3K/Akt signaling pathway. *PLoS One* **8**, e56815 (2013).
30. Ruhwald, M. *et al.* Evaluating the potential of IP-10 and MCP-2 as biomarkers for the diagnosis of tuberculosis. *Eur Respir J* **32**, 1607–1615 (2008).
31. Marfell, B. J., O'Brien, R. & Griffin, J. F. Global gene expression profiling of monocyte-derived macrophages from red deer (*Cervus elaphus*) genotypically resistant or susceptible to *Mycobacterium avium* subspecies *paratuberculosis* infection. *Dev Comp Immunol* **40**, 210–217 (2013).
32. Aho, A. D., McNulty, A. M. & Coussens, P. M. Enhanced expression of interleukin-1 $\alpha$  and tumor necrosis factor receptor-associated protein 1 in ileal tissues of cattle infected with *Mycobacterium avium* subsp. *paratuberculosis*. *Infect Immun* **71**, 6479–6486 (2003).
33. Tsuneyama, K. *et al.* Monocyte chemotactic protein-1, -2, and -3 are distinctively expressed in portal tracts and granulomata in primary biliary cirrhosis: implications for pathogenesis. *J Pathol* **193**, 102–109 (2001).
34. MacHugh, D. E. *et al.* Pan-genomic analysis of bovine monocyte-derived macrophage gene expression in response to *in vitro* infection with *Mycobacterium avium* subspecies *paratuberculosis*. *Vet Res* **43**, 25 (2012).
35. Weiss, D. J., Evanson, O. A. & Souza, C. D. Increased expression of interleukin-10 and suppressor of cytokine signaling-3 associated with susceptibility to John's disease. *Am J Vet Res* **66**, 1114–1120 (2005).
36. Schwartz, S. L., Cao, C., Pylpyenko, O., Rak, A. & Wandinger-Ness, A. Rab GTPases at a glance. *J Cell Sci* **120**, 3905–3910 (2007).
37. Casey, M. E. *et al.* Analysis of the Bovine Monocyte-Derived Macrophage Response to *Mycobacterium avium* Subspecies *Paratuberculosis* Infection Using RNA-seq. *Front Immunol* **6**, 23 (2015).
38. Parkin, J. & Cohen, B. An overview of the immune system. *Lancet* **357**, 1777–1789 (2001).
39. Diamond, M. S. & Farzan, M. The broad-spectrum antiviral functions of IFIT and IFITM proteins. *Nat Rev Immunol* **13**, 46–57 (2013).
40. Villiger, P. M., Geng, Y. & Lotz, M. Induction of cytokine expression by leukemia inhibitory factor. *J Clin Invest* **91**, 1575–1581 M (1993).
41. McGarvey, J. A., Wagner, D. & Bermudez, L. E. Differential gene expression in mononuclear phagocytes infected with pathogenic and non-pathogenic mycobacteria. *Clin Exp Immunol* **136**, 490–500 (2004).
42. Li, M. *et al.* Codon-usage-based inhibition of HIV protein synthesis by human schlafen 11. *Nature* **491**, 125–128 (2012).
43. Puck, A. *et al.* Expression and regulation of Schlafen (SLFN) family members in primary human monocytes, monocyte-derived dendritic cells and T cells. *Results Immunol* **5**, 23–32 (2015).
44. Zarco, N., González-Ramírez, R. & González, R. O. & Segovia, J. GAS1 induces cell death through an intrinsic apoptotic pathway. *Apoptosis* **17**, 627–635 (2012).
45. Kern, C. *et al.* Genome-wide identification of tissue-specific long non-coding RNA in three farm animal species. *BMC Genomics* **19**, 684 (2018).
46. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357–360 (2015).
47. Baruzzo, G. *et al.* Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods* **14**, 135–139 (2017).
48. Teng, M. *et al.* A benchmark for RNA-seq quantification pipelines. *Genome Biol* **17**, 74 (2016).
49. Williams, C. R., Baccarella, A., Parrish, J. Z. & Kim, C. C. Empirical assessment of analysis workflows for differential expression analysis of human samples using RNA-Seq. *BMC Bioinformatics* **18**, 38 (2017).
50. Merino, G. A., Conesa, A. & Fernández, E. A. A benchmarking of workflows for detecting differential splicing and differential expression at isoform level in human RNA-seq studies. *Briefings in Bioinformatics*, bbx122 (2017).
51. Wang, Y. *et al.* Analyses of Long Non-Coding RNA and mRNA profiling using RNA sequencing during the pre-implantation phases in pig endometrium. *Sci Rep* **6**, 20238 (2016).
52. Ulitsky, I., Shkumatava, A., Jan, C. H., Sive, H. & Bartel, D. P. Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. *Cell* **147**, 1537–50 (2011).
53. Engreitz, J. M. *et al.* Local regulation of gene expression by lincRNA promoters, transcription and splicing. *Nature* **539**, 452–455 (2016).
54. Song, H. Y., Rothe, M. & Goeddel, D. V. The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF- $\kappa$ B activation. *Proc Natl Acad Sci* **93**, 6721–6725 (1996).
55. Ma, S. *et al.* A long noncoding RNA, lincRNA-Tnfaip3, acts as a coregulator of NF- $\kappa$ B to modulate inflammatory gene transcription in mouse macrophages. *FASEB J* **31**, 1215–1225 (2016).
56. Kaakoush, N. O. *et al.* Transcriptomic and proteomic analyses reveal key innate immune signatures in the host response to the gastrointestinal pathogen *Campylobacter concisus*. *Infect Immun* **83**, 832–845 (2015).
57. Kumar, M. *et al.* MicroRNA let-7 modulates the immune response to *Mycobacterium tuberculosis* infection via control of A20, an inhibitor of the NF- $\kappa$ B pathway. *Cell Host Microbe* **17**, 345–356 (2015).
58. Baechler, E. C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci* **100**, 2610–2615 (2003).
59. Kane, M. *et al.* MX2 is an interferon-induced inhibitor of HIV-1 infection. *Nature* **502**, 563–566 (2013).
60. Kambara, H. *et al.* Negative regulation of the interferon response by an interferon-induced long non-coding RNA. *Nucleic Acids Res* **42**, 10668–10680 (2014).
61. zur Brügge, J. *et al.* MicroRNA response of primary human macrophages to *Arcobacter butzleri* infection. *Eur J Microbiol Immunol* **6**, 99–108 (2016).
62. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
63. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28**, 511–515 (2010).
64. Jorquera, R. *et al.* SinEx DB: a database for single exon coding sequences in mammalian genomes. *Database* **2016**, baw095 (2016).

65. Bush, S. J. *et al.* Cross-species inference of long non-coding RNAs greatly expands the ruminant transcriptome. *Genet Sel Evol* **50**, 20 (2018).
66. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
67. Roux, B. T., Heward, J. A., Donnelly, L. E., Jones, S. W. & Lindsay, M. A. Catalog of Differentially Expressed Long Non-Coding RNA following Activation of Human and Mouse Innate Immune Response. *Front Immunol* **8**, 1038 (2017).
68. Rice, P., Longden, I. & Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* **16**, 276–277 (2000).
69. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
71. Pawar, K., Hanisch, C., Palma Vera, S. E., Einspanier, R. & Sharbati, S. Down regulated lncRNA MEG3 eliminates mycobacteria in macrophages via autophagy. *Sci Rep* **6**, 19416 (2016).
72. Pfaffl, W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45 (2001).
73. Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207–210 (2002).

## Acknowledgements

We thank the Forschungskommission der Freien Universität Berlin for funding this study as well as Barbara Kutz-Lohroff for excellent technical assistance and Robert Pieper for assisting with the preparation of the manuscript. We acknowledge support by the German Research Foundation and the OpenAccess Publication Fund of Freie Universität Berlin.

## Author Contributions

J.z.B. conceived and designed the experiments. P.G. performed the bioinformatics analysis and prepared the figures. P.G. and J.z.B. wrote the manuscript. S.S., M.v.K. and C.S. participated in the discussions, S.S., R.E., S.P., A.L. and M.J. contributed essential reagents, G.H.-S. and A.F. performed the sequencing. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-38141-x>.

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

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019