

RESEARCH

Open Access



RNA-Seq analysis reveals the long noncoding RNAs associated with immunity in wild *Myotis myotis* bats

Sebastien Riquier^{1†}, Samuel Carthy^{1†}, Graham M. Hughes¹, Frederic Touzalin^{1,2}, Wilfried Haerty^{3,4}, Zixia Huang¹ and Emma C. Teeling^{1*}

Abstract

Background Bats possess a uniquely adapted immune system that enables them to live with viral infections without the expected maladies. The molecular basis and regulation of bats' immune response is still not fully understood. Long non-coding RNAs (lncRNAs) represent an emerging class of molecules with critical regulatory roles in multiple biological processes, including immunity. We hypothesise that lncRNA-based regulation in bats may enable them to limit disease and live with viral pathogens.

Results We developed a lncRNA prediction pipeline to annotate the long non-coding transcriptome across multiple bat tissues and at the population level. Characterisation of our lncRNA dataset based on 100 blood transcriptomes from wild *Myotis myotis* bats revealed lower and more tissue-specific expression compared with coding genes, reduced GC content and shorter length distributions, consistent with lncRNA profiles observed in other species. Using WGCNA network analyses and gene ontology, we identified two mRNA-lncRNA co-expression modules in *Myotis myotis* associated with distinct immune response: one linked to T-cell activation and viral processes, and the other to inflammation. From these immune-related lncRNAs, we selected four candidates with high translational potential for regulating viral infections and inflammation. These include a newly identified lncRNA, *BatLnc1*, with potential anti-viral functions; the *M. myotis* ortholog of *TUG1*, implicated in viral-host interactions; and well-known lncRNAs *MALAT1* and *NEAT1*, recognised for their roles in inflammatory regulation.

Conclusions We conducted the first ab initio prediction of lncRNAs in a non-model bat species, the wild-caught *M. myotis*. Our network analysis revealed significant variation in immune status among a subset of individuals, potentially due to pathogenic conditions. From these variations, we identified lncRNAs most likely associated with immune response in bats. This initial exploration lays the groundwork for future experimental validations of lncRNA functions, offering promising insights into their role in bat immunity.

Keywords Chiroptera, Transcriptomics, Bioinformatics, Immunity, Blood RNA, Virus, Inflammation

[†]Sebastien Riquier and Samuel Carthy contributed equally to this work.

*Correspondence:

Emma C. Teeling

emma.teeling@ucd.ie

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Background

Bats (Order Chiroptera) are the only mammals capable of true self-powered flight and have evolved unique adaptations, including laryngeal-echolocation [1], unusual longevity [2], and remarkable disease resistance [3]. Furthermore, bats can harbour some of the most diverse virospheres [4] without displaying the expected signs of disease or mortality. This suggests that bats have evolved a uniquely adapted immune system allowing them to live with these pathogens without the associated maladies [5]. This resilience has prompted scientists to investigate the immune adaptations that bats have evolved, both to identify novel therapeutic targets and to better understand how to limit infection-driven disease [6, 7].

Based on comparative genomics and initial functional assays [3], it has been suggested that bats have evolved two distinct immune strategies to limit disease: viral tolerance and enhanced antiviral defence. Their viral tolerance mechanisms arguably stem from dampened viral recognition pathways and a reduced pro-inflammatory response, effectively minimising excessive inflammation. This is achieved through adaptations in bat viral inflammatory responses, such as: different isoform expression of the intracellular sensor *NLRP3* [8]; dampened activity of the interferon-response regulator STING due to a mutation of its serine residue (S358) [9]; the bat-specific loss of the entire *PYHIN* inflammasome gene family [10]; and positive selection of *c-REL* [11], a subunit of NF- κ B and transcriptional repressor of TNF α -dependent inflammation [12]. In parallel, some bats also exhibit a highly adapted antiviral defence system. This includes a robust and constant baseline expression of interferon and interferon-response genes without stimulation [13–16], the expansion and selection of viral targeting genes such as the antiviral APOBEC3 enzyme family [17], the Mx proteins lowering viral polymerase activity [18], and duplication of *TRIM* and tetherin antiviral proteins [19, 20]. These adaptations drive a finely tuned immune system capable of both tolerating and suppressing viral infections, contributing to bats' unique ability to coexist with diverse viral pathogens without expected disease [21]. However, the regulatory mechanisms underlying bats' viral tolerance and antiviral mechanisms remain poorly understood.

Long non-coding RNA (lncRNAs) are transcripts that can regulate biological processes [22], and can serve as biomarkers for tissues, pathologies [23, 24] and cell types [25]. To date, research into lncRNA in bat species has been limited, with only one published study [26] detailing bat lncRNA prediction from genomic sequences. In mice and humans, some lncRNAs can modulate key immune processes such as inflammation, pathogen recognition, and interferon signalling [27–29]. Several of the genes

targeted by these lncRNAs are of particular interest in bats. For example, the lncRNA *NEAT1* promotes the assembly of the NLRP3 inflammasome in mouse macrophages by interacting with its caspase-1 p20 domain [30]. Similarly, *MALAT1* increases *NLRP3* expression by acting as a molecular decoy for inhibitor miRNA [31, 32]. “*LncRNA#32*”, or *LUARIS*, stimulates an antiviral response by increasing the transcription of interferon-stimulated genes, including *APOBEC3A* and *APOBEC3G* [33]. Other lncRNAs have also been reported as direct viral restriction factors, such as “*LncRNA#61*”, identified as anti-IAV (Influenza A virus) lncRNA by interacting with viral polymerases [34]. Therefore, lncRNA could contribute to the regulation of bat's unique immune response.

To uncover lncRNAs in bats and predict their potential function, we developed a comprehensive bioinformatic pipeline to annotate and quantify lncRNAs from RNA-Seq datasets encompassing five tissues of the greater mouse-eared bats (*Myotis myotis*). By analysing mRNA-lncRNA co-expression networks in 100 blood transcriptomes ($n=70$) of *M. myotis*, we further identified key lncRNA candidates that may shape bat-specific immunity, distinguishing them from other mammals. These findings not only enhance our understanding of bat immunology but also pave the way for potential novel therapeutic targets in the fight against emerging zoonotic diseases.

Methods

Step 1: Ab initio prediction of lncRNAs

Tissue samples, RNA-Seq datasets and metadata

The dataset used for lncRNA prediction consisted of stranded, ribo-depleted paired-end RNA sequencing public data from two main sources: 1) Seven tissue samples from an injured bat that required euthanasia as detailed in Jebb et al. 2020 study [35], specifically kidney ($n=2$), liver ($n=2$), heart ($n=2$), brain ($n=1$); 2) 100 globin-depleted whole blood RNA-Seq samples from 70 female *M. myotis*, collected annually from four colonies in Brittany, France, with an age range of 0 to 7+ years (where “X+” indicates a minimum age of X years) [36]. All bats were marked with transponders enabling us individual identification on recapture and passive monitoring of roost entry and exit using antennas placed at the roost exit from 2014 to 2023 (methods as detailed in Touzalin et al. 2023 [37]). This enabled us to ascertain the survival of individuals year to year. The sample information, RNA isolation and purification, and RNA sequencing have been described thoroughly in our previous studies [36, 38] [Sup File 1]. *M. myotis* genome sequences and annotation (GFF files) were obtained from Jebb et al. 2020. The mitochondrial genome and the corresponding

mitochondrial genes predicted by Jebb et al. 2017 (HLmyoMyo6, [35, 39]), Genbank reference KT901455.1, were downloaded and added to the FASTA and GFF files.

Read trimming & mapping

Illumina universal adapter sequences and low-quality regions were trimmed from the raw FASTQ files using TrimGalore v0.6.8 [40], with a default Phred quality threshold of 20. Sequencing quality and trimming efficiency were assessed using FastQC [41] and MultiQC [42]. The trimmed paired-end FASTQ files were then mapped to the indexed *M. myotis* genome with STAR v2.7.10b-alpha [43] in 2-pass mode (`-twopassMode Basic`) with XS strand attributes generated (`-outSAMstrandField intronMotif`). For each sample, the read mapping file in 'Binary Alignment Map' (BAM) format was generated for downstream analyses.

LncRNA prediction

The ab initio transcriptome assembly was constructed for each sample from its corresponding BAM file using Stringtie v2.2.1 [44]. The resulting Gene Transfer Format (GTF) files were merged through a two-step procedure using Portcullis v1.2.2 [45] and Mikado v2.3.4 [46] to control isoform numbers and extract primary isoforms. The canonical splice junctions were first extracted from each BAM file using Portcullis, and the resulting Browser Extensible Data (BED) files were merged with Portcullis' 'junctools' command using the "set union" parameter. The GTF files were then processed using Mikado, with two key customisations: 1) a custom parameter file that allowed mono-exonic isoforms to be classified as splice variants; and 2) a custom scoring file that excluded coding DNA sequences (CDS), ensuring that noncoding candidates were retained as principal isoforms.

LncRNA filtering and categorisation

The resulting GTF files were parsed using BEDTools [47] and custom in-house scripts (see github link in data availability) to separate the annotated coding transcripts from unannotated candidates. Genes with unknown function, designated with "LOC" in their FASTA header within the *M. myotis* annotation, were removed from the annotated coding intervals. Two categories of lncRNAs were identified from the unannotated candidates: 1) long intergenic lncRNAs or "lincRNAs", which do not overlap with any annotated genes; and 2) overlapping antisense lncRNAs, or "loancRNA", which overlap annotated genes on the antisense strand. To account for the extended 5' and 3' expression ranges of coding genes that extend beyond their genomic intervals, and address potential misleading predictions arising from UTRs, a two-step selection was applied. First, predicted lncRNAs located within

500 nucleotides (nt) of coding genes, in the same orientation were considered as extended coding intervals. Then, lncRNA within 5 kb of identically oriented coding intervals were removed. To distinguish true noncoding lncRNA from previously unannotated coding genes, the non-coding nature of the predicted principal isoforms, as indicated by Mikado, was validated using the consensus of three tools: CPC2 v1.79 with a coding probability cutoff < 0.5 [48]; CPAT v3.0.4 using the mouse model with a coding cutoff < 0.44 [49]; and PLEK v 1.2 with negative coding score [50]. Candidates were classified as non-coding if identified as such by at least two of the three tools.

All steps of lncRNA characterisation are detailed in (Fig. 1).

Step 2: Quantification and characterisation of annotated genes and lncRNA

Annotated coding transcripts, predicted lincRNA, and loancRNA sequences from each sample were merged into a single reference file, then quantified at the gene level from the BAM files using FeatureCounts v2.0.3 [51] with the following parameters: `-countReadPairs`, `-p` and `-s 2`. lncRNAs and annotated genes with an average expression level above 20 read counts in any tissue were considered to have significant expression. Only coding genes covered by reads on at least 20% of their full gene length were retained to minimise false-positive predictions from partial gene coverage. Gene expression normalisation was performed using Trimmed Mean of M-values (TMM) using the NOISeq package [52]. To characterise previously identified lncRNAs, the principal isoform of each lncRNA, as indicated by Mikado, was compared against the NONCODE database v6 [53], and the RefSeq viral genomes database [54] using BLASTn [55]. Valid matches were selected based on the thresholds of e-value below 0.001 and query coverage above 60%. lncRNAs were annotated as overlapping an endogenous retrovirus (ERV) if more than 50% of their length overlapped an ERV region annotated in the *M. myotis* genome browser. The effect of factors including sampling year, site, sequencing batch, and individual bat was analysed from TMM-normalised expressions using the variance Partition R package [56].

Step 3: mRNA-lncRNAs co-expression network creation

Weighted Gene Co-expression Network Analysis

Weighted Gene Co-expression Network Analysis (WGCNA) was performed on TMM-normalised expression counts using the WGCNA R package [57]. The optimal soft threshold power for the analysis was determined as 9 using the `pickSoftThreshold` function, and this threshold was applied to construct an adjacency matrix [Sup file 2]. This adjacency matrix was then used

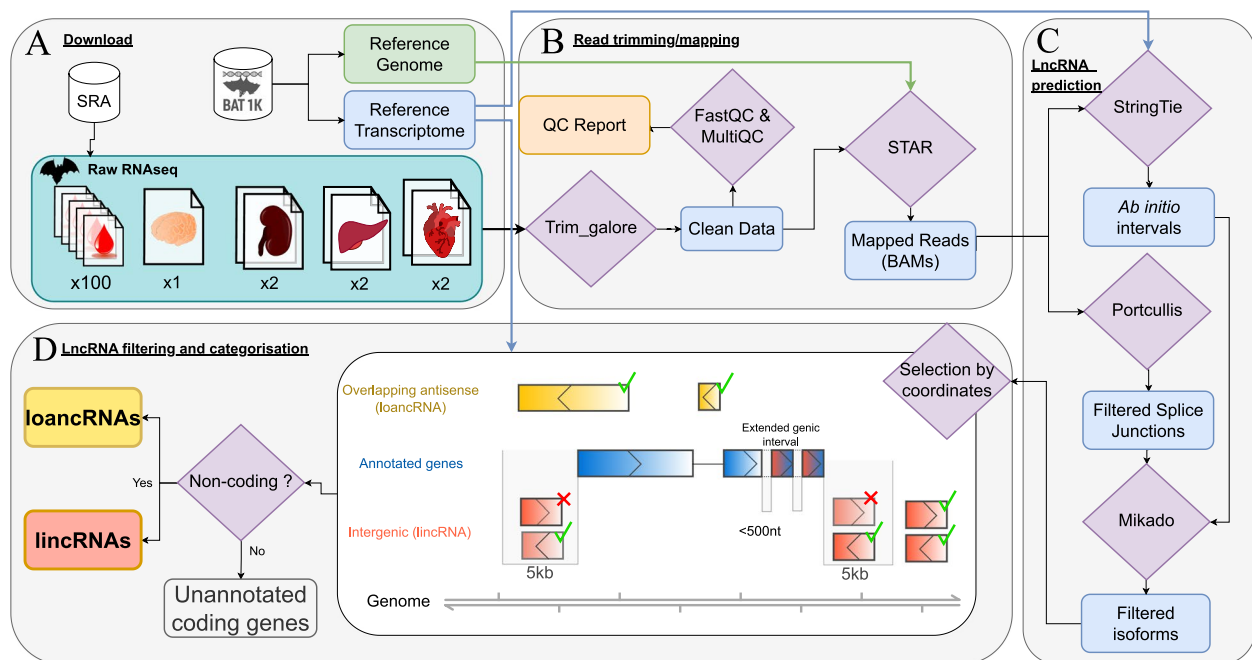


Fig. 1 The ab initio lncRNA prediction pipeline. **A**) 100 whole-blood RNA-Seq datasets and 7 solid samples (2 kidneys, 2 livers, 2 hearts, 1 brain) from *M. myotis* are downloaded for this study; **B**) Quality control of RNA-Seq and reads mapping; **C**) The core lncRNA prediction steps; **D**) lncRNA filtering steps

to construct a Topological Overlap Matrix (TOM). Hierarchical clustering (“hclust” function with “average” method) was performed on the TOM to identify co-expression modules, grouping genes into modules based on their connectivity. Modules were determined using the dynamic tree cut “hybrid” method (cutreeDynamic function), which integrated hierarchical clustering and the dissimilarity matrix. Modules with high similarity (dissimilarity below 0.2) were merged into a new module.

Gene ontology analyses and unsupervised exploration of co-expression modules

Gene ontology and enrichment analyses were conducted on the coding gene content of every module using the ClusterProfiler R package [58]. The predominant cell types in the modules were identified with ‘marker genes’ from the CellMarker2 database (limited to normal cells and excluding scRNA-Seq-derived data) via the *compareCluster* function, applying a p-value cutoff of 0.01. To uncover representative pathways in each module, we utilised the EnrichGO function with the following parameters: minGSSize=50, maxGSSize=200, pvalueCutoff=0.01, ont="BP" (Biological Process only). To reduce pathway redundancy, we employed the simplify function from the same package, using the following parameters: cutoff=0.6, by="p.adjust", select_fun=min, measure="Wang", and semData=NULL.

Principal Component Analysis (PCA) was performed on TMM-normalised log-transformed counts of all predicted lncRNAs and all coding genes across all samples using the PCATools package. Modules’ global expression was assessed by calculating eigengenes, defined as the first principal component that summarises the expression of all genes (coding or non-coding) within each module. Spearman’s rank correlation was used to analyse the relationship between eigengenes and age, with ages exceeding seven years labelled as seven. Inter-module correlations were evaluated using the plotEigengeneNetworks, with the “signed=TRUE” option.

Step 4: Exploration of lncRNAs-mRNA co-expression network and candidates selection

The strength and number of the correlations between annotated genes and lncRNAs were quantified from the network table exported from the initial adjacency matrix, using an adjacency threshold of 0.05, through the exportNetworkToCytoscape function. Candidates selected for a focused overview were ranked based on having the highest number of co-expressed genes and/or the highest expression levels. Immune gene lists were extracted from the Human collection C7 from GSEA MSigDB, to assess the proximity between predicted lncRNAs and genes involved in immunity in humans.

To assess the sequence conservation of predicted lncRNAs loci, Phastcons scores were generated from the “15

mammals” Multiz alignment MAF file available on the Senckenberg genome browser [59]. For accessibility reasons these scores were processed with the PAST tool [60] using the 470 MultiZ phylogenetic model available on genome browser [61], for which only the corresponding 15 species were used for scoring.

The genomic intervals and exonic structures of the selected lncRNAs were visualized using BAM and GFF3 files with the Gviz packages [62]. Selected lncRNAs candidates, along with the top 20 most correlated coding genes (based on adjacency scores), were extracted for further analysis. Their expression levels, represented as Z-scored TMM-normalised values, were visualised alongside relevant metadata using the ComplexHeatmap package [63].

Results

Characterisation of the lncRNA set

We employed the chromosome-level assembly of *M. myotis* published by the Bat1K consortium to identify and quantify protein-coding genes and lncRNAs in bats. To validate our quantification method using this new annotation, we compared the expression of 5,545 protein-coding genes in blood samples, which were also predicted by Huang et al. 2019 using the *Myotis lucifugus* genome (MyoLuc2.0) as reference. A strong correlation between the two studies was observed ($R=0.94$, $P<2.2e^{-16}$, Spearman’s correlation test), confirming the reliability of the quantification part of our pipeline (Fig. 2A).

We annotated a total of 9,236 lncRNA genes, including 3,761 intergenic lncRNAs (2,207 monoexonic and 1,554 polyexonic) and 5,475 overlapping antisense lncRNAs (3,502 monoexonic and 1,973 polyexonic), all expressed in at least one tissue in the dataset. This represents a 65.4% increase in the total number of annotated expressed genes ($n=14,115$), highlighting the expanded scope and depth of our lncRNA predictions (Fig. 2B).

The predicted lncRNAs were generally shorter than annotated coding genes, with an average length of 1,853nt compared to 2,086nt for coding genes. However, the length distribution varied significantly depending on the lncRNA type and number of exons (Fig. 2C). Monoexonic lncRNAs exhibited a binomial distribution with

two distinct peaks, representing short and long forms. Surprisingly, polyexonic lncRNAs were the shortest across all groups, with their distribution peaked around the minimum length threshold of 200nt, while monoexonic lncRNAs were the longest, with a median length surpassing that of coding genes (Fig. 2C). Although all predicted lncRNA types had a lower median GC content than coding genes, polyexonic lncRNAs from both groups showed higher median GC percentages (48.7 for intergenic and 50.5% for antisense) compared to their monoexonic counterparts (43.9% and 44.5%, respectively) (Fig. 2D).

The predicted lncRNAs exhibited tissue specificity, with the brain having the highest number of tissue-specific lncRNAs (1,386), followed by the kidney (904), blood (731), Liver (333) and Heart (252). A small subset of lncRNAs were shared amongst all tissues. Notably, a marked contrast between solid tissues and blood: the 4 solid tissues collectively expressed a high proportion of their lncRNAs (2,651), while most of the lncRNAs in blood were tissue-specific (Fig. 2E). As expected, predicted lncRNAs showed lower expression levels than annotated coding genes across all tissues. Blood exhibited lower overall expression levels compared to other tissue types, but with higher maximum expression levels (Fig. 2F). Interestingly, lncRNAs in solid tissues showed slightly higher expression than their intergenic counterparts.

Identification of predicted lncRNAs

A BLAST search was performed to determine the identity and homology of the lncRNAs by comparing them with human lncRNAs from the NONCODEv6 database. Both intergenic and overlapping antisense lncRNAs showed similar numbers of significant hits, with 477 and 466 matches, respectively [Sup file 3]. Several lncRNAs were identified as orthologs of well-known and highly conserved lncRNAs, including *XIST*, *NEAT1*, *MALAT1*, *GASS* and *HOTAIR*, among others [Sup file 4].

Despite the high content of viral sequences in the *M. myotis* genome, only a small number of hits ($n=8$) were found in the RefSeq database of viral genomes or corresponded to endoviral loci within the *M. myotis* genome

(See figure on next page.)

Fig. 2 The landscape of predicted protein-coding and lncRNA across tissues. **A)** Spearman correlation of gene expression between the current study and Huang et al. 2019 estimated from 100 blood *M. myotis* samples. 5,545 common genes identified in both studies were used in the analysis. **B)** The distribution of expressed protein-coding and lncRNAs identified in this study. lncRNAs were subcategorised into monoexonic lncRNAs, monoexonic lncRNAs, polyexonic lncRNAs, and polyexonic lncRNAs. **C)** Distribution of the length of expressed coding and lncRNAs across tissues. **D)** GC content distribution of coding and lncRNAs across tissues. **E)** Upset plot showing the number of expressed lncRNAs that were shared across tissues or tissue-specific. **F)** The distribution of coding and lncRNA gene expression across tissues. The raw expression counts were TMM-normalised

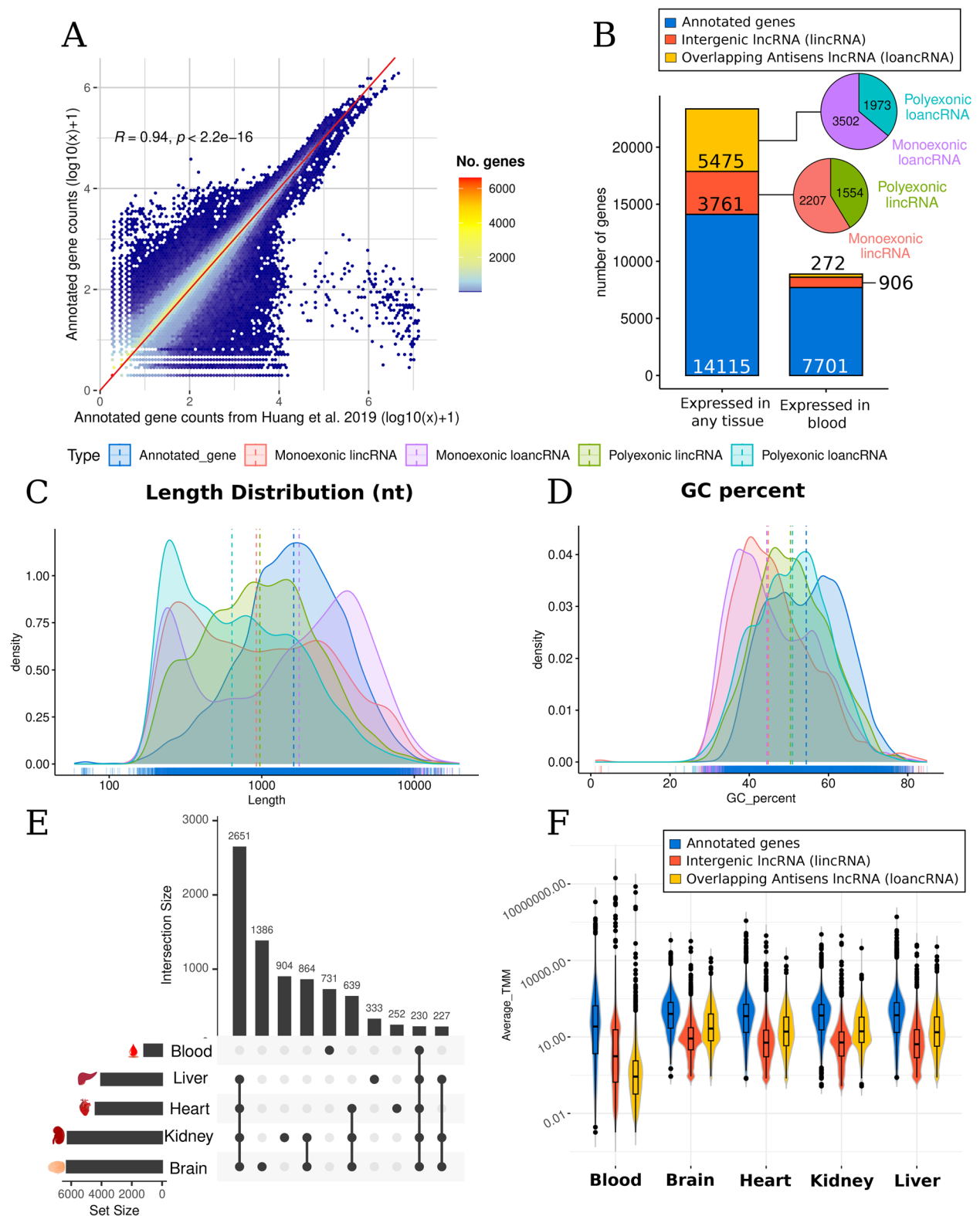


Fig. 2 (See legend on previous page.)

($n=6$). Among these, two hits matched an endogenous virus originally found in *Desmodus rotundus*, the common vampire bat. Five of the eight viral hits were against retroviruses, including the reticuloendotheliosis virus (avian retrovirus), Woolly monkey sarcoma virus, bovine herpesvirus 6, and Friend murine leukemia virus. All lncRNAs overlapping endogenous viral sequences were associated with gammaretroviral types [Sup file 5].

The WGCNA gene-lncRNA modules in *M. myotis* blood reveal lncRNAs linked to immunity

To identify potential regulators of immune function in *M. myotis*, we assessed the variation in the expression of the 7,701 annotated coding genes and the 1,178 predicted lncRNAs expressed in the 100 whole-blood samples from bats aged 0 to 7+ years. We first conducted a variance analysis based on blood gene expression to identify the primary factor contributing to expression variation. When examining the effect of factors such as age, sampling year, site, sequencing batch, and individual bat, we found that age explained the greatest proportion of variation. This result aligns with previous findings by Huang et al. (2019) which focused on coding genes [36]. Both types of lncRNAs showed a slightly smaller contribution of age to their expression variation compared to coding genes. Other factors, excluding residuals, had minimal impact, suggesting that other unknown factors excluding age, RNA coding attributes, and technical batch effects explain a significant part of the expression variation in bat blood within this wild population [Sup file 6].

Next, we performed a weighted gene co-expression network analysis (WGCNA) to identify modules of co-expressed annotated coding genes and lncRNAs (Fig. 3A) in *M. myotis* blood. A large proportion of the annotated genes expressed in blood, along with the majority of expressed lncRNAs in each category, were assigned to one of the identified modules: 6,097 annotated genes (79.1%), 549 lincRNAs (60.5%), and 148 loancRNAs (54.4%) (Fig. 3B). In total, 12 modules were identified and assigned unique identifiers (MEs), ordered by module size (Fig. 3C). The largest module, “ME1”, contained 2,109 genes, followed by “ME2” with 1,939 genes, and then “ME3” with 739 genes. The presence of at least one

lncRNA in each of these modules suggests the potential involvement of lncRNAs in the different functions of *M. myotis* blood cells.

To determine which co-expressed modules are involved in the immunity of *M. myotis*, we performed over-representation analysis using cell type markers and pathways databases on the annotated genes within each module (Fig. 3D–F). The first three modules were associated with pathways involved in immunity, cell cycle regulation and DNA repair. The largest module, “ME1”, exhibited significant enrichment in cell markers from almost all immune cell types, including both myeloid and lymphoid cells, excluding neutrophils. This enrichment is particularly strong for T lymphocytes. “ME3” showed enrichment for markers associated with the innate immune system, such as macrophages, neutrophils, monocytes, and natural killer cells (Fig. 3D).

ME1 appears to be strongly associated with adaptive immunity, particularly lymphocyte activation and function. The most significantly enriched pathways include “alpha-beta T-cell activation” (GO:0046631) and the “T cell receptor signaling pathway” (GO:0050852) (Fig. 3E), supported by other notable pathways like “chaperone-mediated protein folding” (GO:0061077), “viral gene expression” (GO:0019080), “B cell proliferation” (GO:0042100), and “lymphocyte apoptotic process” (GO:0070227) [Sup file 7] further supporting these observations. ME2, in contrast, is predominantly enriched for pathways linked to longevity that were previously highlighted from the same data as the pathways responsible for maintaining *M. myotis* unusual longevity [36]. Key pathways include “recombinational repair” (GO:0000725), “cell cycle checkpoint signaling” (GO:0000075), “telomere organization” (GO:0032200), and “positive regulation of autophagy” (GO:0010508). ME3 is enriched for pathways central to innate immunity and inflammatory responses, indicating a distinct role in first-line immune defense. This includes key regulators of inflammation such as tumor necrosis factor production pathways (GO:0071706), the “positive regulation of NF- κ B” (GO:0051092) (Fig. 3E), and “interferon signaling” (GO:0140888), with notable genes such as *IFNAR2* and *IFNGR2* contributing to antiviral defense.

(See figure on next page.)

Fig. 3 Weighted gene co-expression network analysis of 100 *M. myotis* blood RNA-Seq samples. **A**) Clustering dendrograms showing the expression patterns of both coding genes and lncRNAs. The first band represents the different co-expression modules initially identified, while the second band indicates merged modules that were integrated by modules with similar expression patterns. The gene clusters in white indicate that these genes cannot be assigned to any modules. **B**) Number of coding genes and lncRNAs assigned to the identified modules. **C**) The distribution of coding genes and lncRNAs in each identified module. **D**) Enrichment analysis of cell-type markers (coding genes) in each module. The numbers in circles indicate the numbers of cell-type marker genes identified in each module. **E**) Gene ontology analysis of the coding genes in each module. The numbers in circles indicate the numbers of genes in that pathway for each module

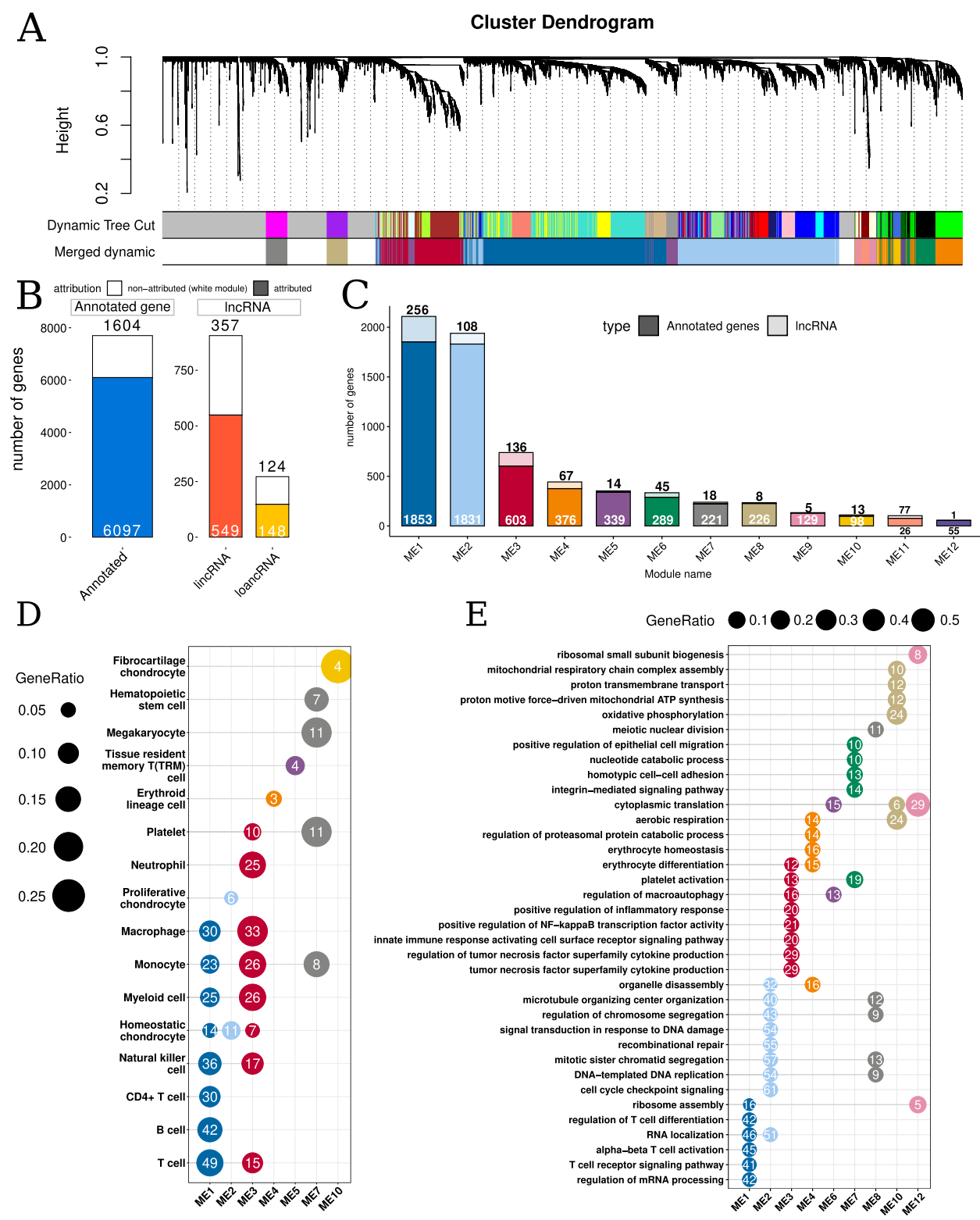


Fig. 3 (See legend on previous page.)

Additionally, ME3 is enriched for “Toll-like receptor (TLR) signaling” (GO:0071706), involving *TLR1*, 2, 4, 6, and 8, reinforcing its role in pathogen recognition and immune activation. Interestingly, the presence of “negative regulation of immune response” (GO:0050777) [Sup file 7] suggests that this module may also be involved in fine-tuning immune reactions to prevent excessive inflammation.

Overall, our results suggest that lncRNAs detected in ME1 and ME3 could be implicated in the immune response in *M. myotis* by association, with ME1 lncRNAs potentially involved in lymphocytic activity and ME3 lncRNAs potentially linked to innate immunity and inflammation. Noticeably, the percentage of lncRNAs in these immune-related modules exceeds that in ME2, suggesting that more lncRNAs are converging on immunity regulation in *M. myotis* [Sup file 8].

ME1 co-expression module reveals regulatory lncRNAs of antiviral T cells activity in *M. myotis*

To identify the most relevant immune-related lncRNAs and investigate their potential roles in immune functions, we assessed the global expression of genes from the ME1 and ME3 across the blood datasets and investigated the relationships between coding genes and lncRNAs through their co-expression network. To visualise the difference of expression of ME1 across the bat population, we mapped the expression pattern (eigengene) of ME1 as a color gradient onto the blood PCA, which revealed its expression pattern changing along the PC1 (horizontal axis). Interestingly, this overexpression was most prominent in 0 or 1 year old individuals (Fig. 4A). These younger individuals, expressing higher levels of ME1, were detected by the radio antenna system at least one year after the original sampling, showing no immediate impact of this possible immune challenge on their survival [Sup File 9]. Additionally, some top loadings of the PCA aligning with PC1 and ME1 expression patterns are

linked to cytotoxic lymphocytes, such as *SAMD3* [64], *KLRK1* [65] and *VCAN* [66] [Sup file 10].

As expected, ME1 expression showed a negative correlation with bat age ($R = -0.54$, $p\text{-value} = 1.9e-7$) (Fig. 4B) and demonstrated a strong inverse correlation with ME2 expression ($R = -0.8$, $p\text{-value} = 4e-27$) (Fig. 4C). In contrast, ME2 which is positively correlated with age, exhibited a reversed expression pattern along the PC1 (horizontal axis) compared to ME1 [Sup Fig. 11] suggesting an antagonistic relationship between DNA repair and T-cell/antiviral activity. ME1 expression showed a weak positive correlation with ME3 ($R = 0.29$, $p\text{-value} = 3e-3$) and a stronger positive correlation with ME5 ($R = 0.62$, $p\text{-value} = 6e-13$), a module with an as-yet-unknown function.

By assessing the relationships between coding genes and lncRNAs in ME1, we observed that lncRNAs did not uniformly co-express across the network. Instead, they were co-expressed connected in number to a limited number of central ‘hub’ coding genes involved in lymphocyte immune regulation, some of which possess antiviral properties. These ME1 “hub” genes include *ARID5B*, *CELF2*, *SEPT6*, *SATB1* (Fig. 4D).

To identify the most central lncRNAs potentially regulating bat lymphocyte immune function in ME1, we ranked candidate lncRNAs based on four criteria: expression level, number of co-expressed coding genes, orthology with human lncRNA and genomic proximity to immune genes (Fig. 4E). Most of the candidate lncRNAs were located near (within 50 kb) of an immune-related gene (94/162), suggesting a potential involvement to immune function regulation. Using expression level and number of correlated genes as criteria, we identified two standout lncRNAs candidates from ME1. The lncRNA with the highest number of co-expressed genes (number co-expressed genes = 505) and the second highest expression level (average TMM = 421) in ME1 is a poly-exonic lncRNA located between *EPHB6* and *TRBV29-1*,

(See figure on next page.)

Fig. 4 Analysis of the ME1 co-expression module. **A**) PCA of the blood expression patterns in the wild *M. myotis* samples Principal component analysis (PCA) using all the coding genes and lncRNAs characterised in the ME1 module. Color gradient represents the global expression of the ME1 module (eigengene calculated by WGCNA); **B**) Spearman's correlation analysis between the ME1 global expression (eigengene) and age of the bat in the *M. myotis* population. Only the samples with known ages and the samples over 7 years old were selected for this analysis. Samples over 7 years old (e.g. 7+) were treated as 7 years old when calculating Spearman's correlation. **C**) Spearman correlations between ME1 and the other 11 modules' global expressions. **D**) Top 10 Coding genes with the highest number of co-expressed lncRNAs in ME1; **E**) Scatter plot showing the relationship between the number of co-expressed coding genes and expression levels of each lncRNA in ME1. **F**) Genomic visualisation of predicted lncRNA: *BatLnc1*. The top session shows the genomic location of *BatLnc1* visualised using the genome browser; the below session shows the expression level and splicing variants of *BatLnc1* using an individual bat sample (MMY612) as an example, PhastCon scores indicate the conservation level of each nucleotide of *BatLnc1*. **G**) Expression heatmap of top 20 coding genes that were co-expressed with *BatLnc1* across 100 *M. myotis* blood samples. The expression counts were TMM-normalised, and further converted to z-scores. The metadata of these samples, such as age, site, year-of-capture, were color-coded (left). For age, blue dots indicate the samples with known age, while the red dots indicate a minimum of “+” age (right). Adjacency scores between *BatLnc1* and other genes are displayed (top)

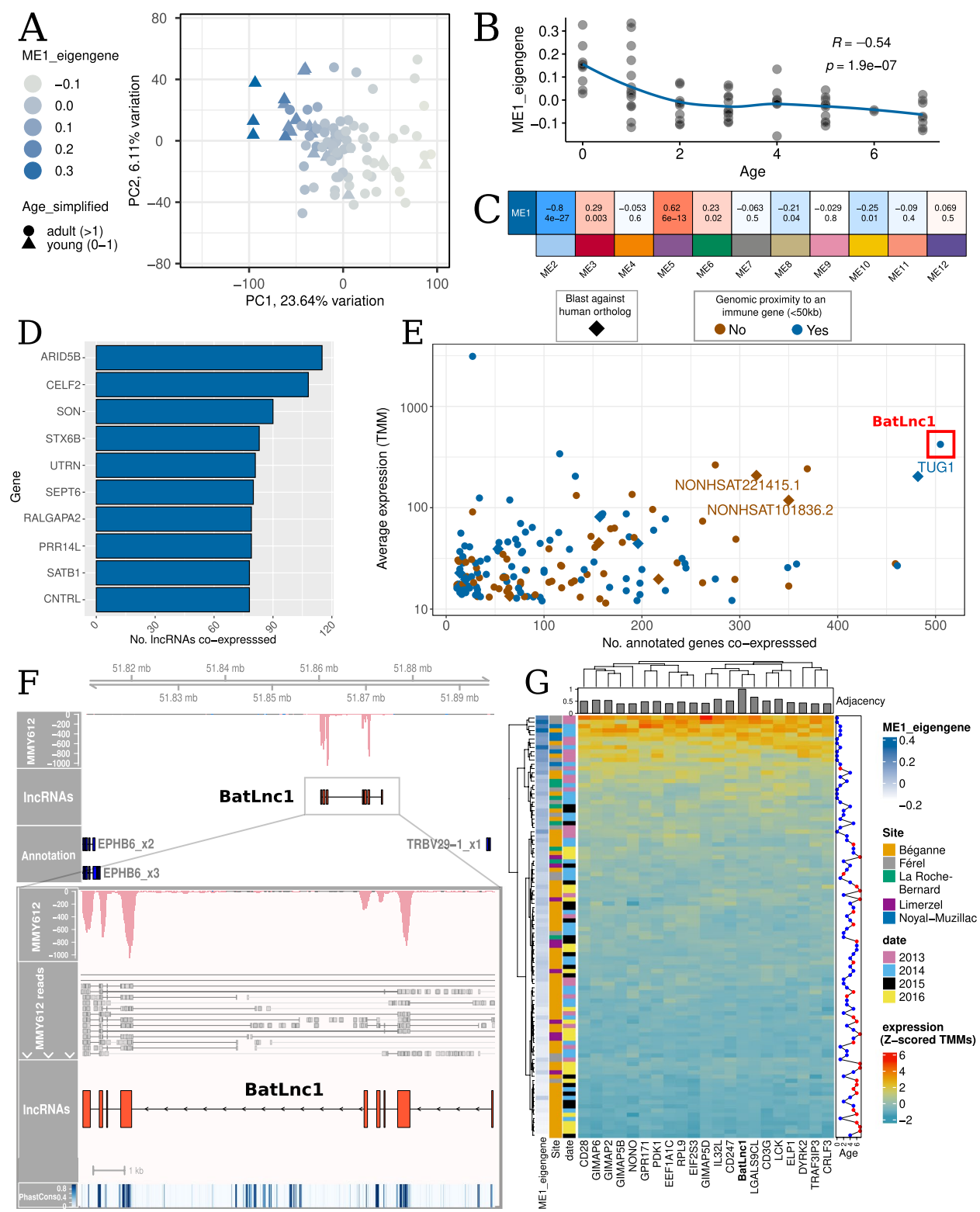


Fig. 4 (See legend on previous page.)

both of which are crucial for lymphocyte function. This lncRNA, renamed “*BatLnc1*”, does not have a known ortholog in human. However, its exons overlap genomic regions with high PFAST conservation scores compared to introns or surroundings regions (Fig. 4F), indication potential functional constraints. BLAST analysis indicates that this lncRNA may be a pseudogene derived from a TCR beta chain [Sup file 12]. The second prominent lncRNA is a conserved ortholog of the human lncRNA *TUG1* (number co-expressed genes=482, average TMM=203). Similar to its human ortholog, bat *TUG1* appears to function as a bidirectional lncRNA, positioned on the reverse strand of the promoter region of *MORC2*, an important antiviral restriction factor [Sup file 13].

To further investigate the potential functions of these two lncRNA candidates, we analysed their expression patterns alongside the top 20 most correlated annotated genes. The expression pattern of *BatLnc1* closely mirrors that of the ME1 eigengene, with enhanced expression in the blood of only a subset of young bats (aged 0–1 years). The top correlated gene, *LGALS9CL*, is known for its critical regulatory role during viral infection, and other highly correlated genes are involved in antigen-recognition T-cell receptor (TCR) activity and viral defense, such as *LCK*, *CD247*, *CD3G*, *GPR171*, *IL32L*, *NONO*, *CD28*, *DYRK2*, and *CRLF3* (Fig. 4G). A similar expression pattern is observed for the *TUG1* ortholog with different top 20 correlations. The majority of these correlated genes are HIV-related genes such as: *CCNT2*, *ZFYVE16*, *HNRNPH1*, *SF3B1*, *PRKCQ*, *DOCK10*, *ANKRD10*, *ADD3*, or antiviral genes like *UBR5* and *PTAR1* [Sup File 13].

ME3 co-expression modules reveal regulatory lncRNAs of inflammation in *M. myotis*

Using the same approach, we examined the genes and lncRNAs in the proinflammatory ME3 module. The

global expression of ME3 transposed on the PCA showed that ME3 is overexpressed only in a few individuals, particularly in one outlier, along the PC2 (vertical axis), with no apparent association with age (Fig. 5A). As for ME1, the individuals overexpressing this module were detected at least one year after, suggesting they survived their pro-inflammatory event [Sup File 9].

Additionally, some top loadings of the PCA aligning with PC2 and ME3 expression patterns [Sup file 10] are linked to sepsis like *ACVR1B* and *VNN3* [67, 68].

ME3 expression showed a weak negative correlation with age ($R=-0.28$, $p\text{-value}=1.3e-2$) (Fig. 5B), suggesting that the proinflammatory expression profile does not increase with age in *M. myotis*, consistent with previous observations [36]. ME3 also showed a weak inverse correlation with the genome protection module ME2 ($R=-0.4$, $p\text{-value}=3e-5$) and a strong positive correlation with the module ME5 ($R=0.76$, $p\text{-value}=2e-22$), respectively (Fig. 5C).

As in ME1, a few “hub” ME3 coding genes are connected to a substantial number of co-expressed lncRNA, indicating their central role in the network. Interestingly, the NF- κ B subunit *c-REL* possesses the largest number of co-expressed lncRNAs. Other “hub” coding genes include both negative and positive regulators of inflammation, such as *SULF2*, the hyaluronan receptor *CD44*, *SVIL*, *LYST*, and *IQGAP1* (Fig. 5D).

We observe a correlation between maximum TMM expression (to account for the few individuals overexpressing this module) and the number of genes co-expressed with the lncRNAs in ME3. A group of lncRNAs, distinguished by their exceptionally high expression, were identified as orthologs of MALAT1 (max. TMM=47,303, number co-expressed genes=226) and 5 different subparts of *NEAT1* (framed in [Fig. 5E]) coming from the same locus [Sup file 14], from which 3 parts are not reported in our analysis as conserved (*NEAT1:25* is the fragment with most co-expressed genes; max.

(See figure on next page.)

Fig. 5 Analysis of the ME3 co-expression module. **A**) PCA of the blood expression patterns in the wild *M. myotis* samples Principal component analysis (PCA) using all the coding genes and lncRNAs characterised in the ME3 module. Color gradient represents the global expression of the ME3 module (eigengene calculated by WGCNA); **B**) Spearman's correlation analysis between the ME3 global expression (eigengene) and the age of bat in the *M. myotis* population. Only the samples with known ages and the samples over 7 years old were selected for this analysis. Samples over 7 years old (e.g. 7+) were treated as 7 years old when calculating Spearman's correlation. **C**) Spearman correlations between ME3 and the other 11 modules' global expressions. **D**) Top 10 Coding genes with the highest number of co-expressed lncRNAs in ME3; **E**) Scatter plot showing the relationship between the number of co-expressed coding genes and expression levels of each lncRNA in ME3. **F**) Genomic visualisation of predicted lncRNA: *MALAT1*. The session shows the expression level and genomic location of *MALAT1* visualised using the genome browser, using an individual bat sample (MMY664) as an example, PhastCon scores indicate the conservation level of each nucleotide of *MALAT1*. **G**) Expression heatmap of top 20 coding genes that were co-expressed with *MALAT1* across 100 *M. myotis* blood samples. The expression counts were TMM-normalised, and further converted to z-scores. The metadata of these samples, such as age, site, year-of-capture, were color-coded (left). For age, blue dots indicate the samples with known age, while the red dots indicate a minimum of “+” age (right). Adjacency scores between *MALAT1* and other genes are displayed (top)

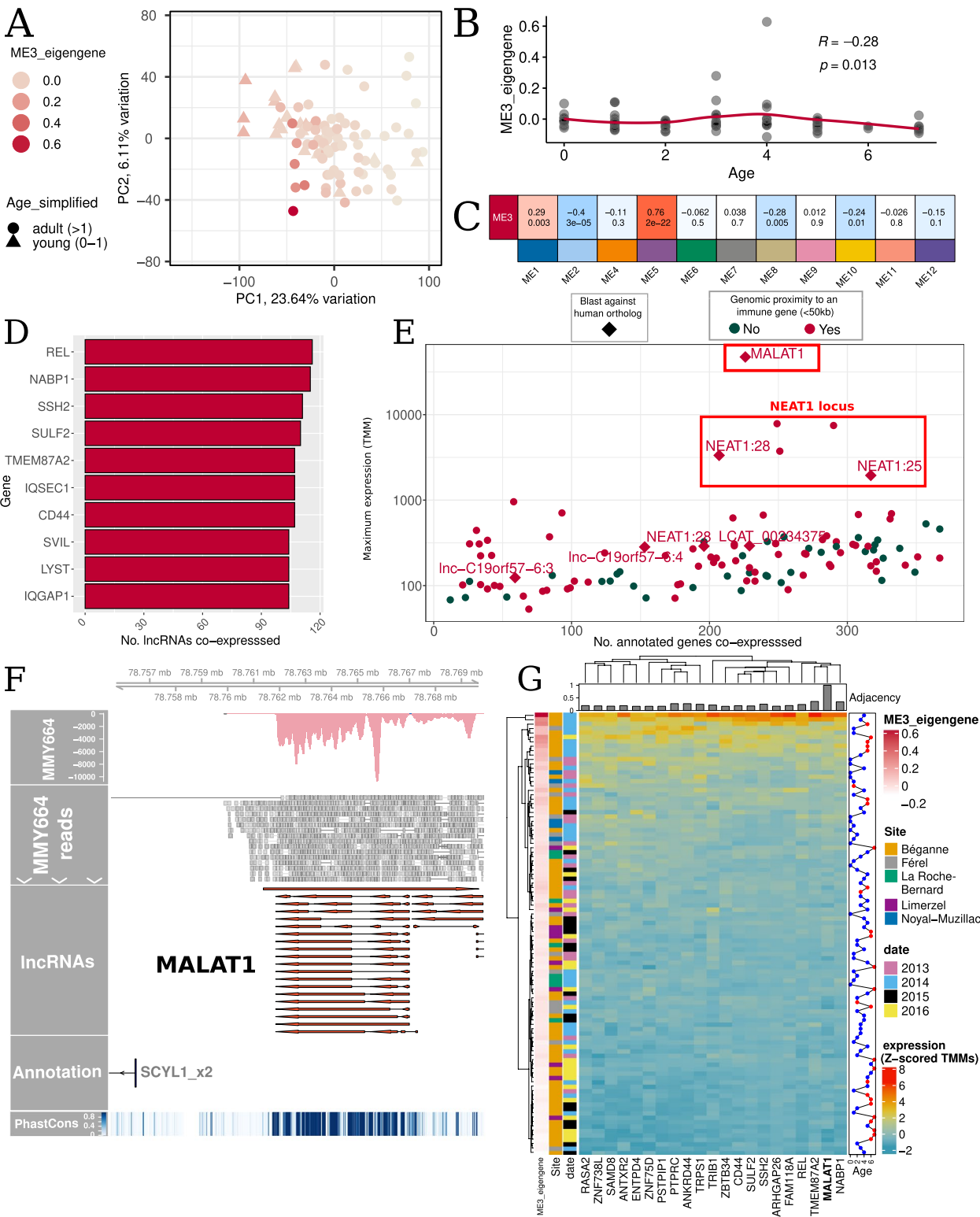


Fig. 5 (See legend on previous page.)

TMM=1,946, number co-expressed genes=317). These *NEAT1* and *MALAT1* orthologs, located between the genes *FRMD8* and *SCYL1*, share the same syntenic genomic region as in the human genome, thereby confirming their orthology (Fig. 5F).

Genes highly co-expressed with *MALAT1* in *M. myotis* blood included key inflammation-regulating genes such as *c-REL*, *TRIB1*, *CD44*, *SULF2*, along with other regulators of immunity like *NABP1*, *PTPRC* (CD45), *RASA2*, *PSTPIP1* and *ARHGAP26* (Fig. 5G). *NEAT1* showed correlations with active inflammation regulators such as *ARHGAP26*, *SULF2*, *LYST*, *KDM6B*, *CSF3R*, *LYN*, *SKAP2*, *BCL6*, *FCMR* (receptor for IgM), *INPP4A*, *LPCAT2*, *CCR1* [Sup File 14, See discussion].

Discussion

Recent studies suggest that bats' immune systems have evolved distinctly from those of other mammals, granting them remarkable tolerance and resilience to viral infections [3, 5, 7]. Despite this, the regulation of immunity by long-non coding RNA (lncRNAs) has never been explored in bats. In this study, we address for the first time this gap, by predicting and quantifying the expression of lncRNAs in bat solid tissues, and in blood at the population level, including individuals at different ages and biological states, using our newly developed pipeline of ab initio prediction and isoform validation. This approach was applied to our previously published 100 RNA-Seq data from whole blood sampled from bats of known ages sampled over six years in Brittany, France [36], and seven RNA-Seq *M. myotis* datasets from four non-blood tissues (Heart, Liver, Brain and Kidney) obtained from an euthanized bat [35].

Our first initial analysis revealed that bat's predicted lncRNAs share several similar characteristics with those of other species. These include: predominance of mono-exonic lncRNAs, smaller size, lower GC content compared to coding genes, and tissue-specific expression patterns, with a majority of specific tissue-lncRNAs in the brain. Interestingly, our results support other studies as brain is known to be enriched in tissue-specific lncRNAs in humans [69] and in brain part-specific in mouse models [70], with possible reasons among the diversity of cell types and functions concentrated in this organ. Other studies suggest that lncRNAs are important in the evolution of brain structure and function in mammals [71], and our catalogue could open the path to studying the role of lncRNAs in specific neuronal functions such as echolocation in bats. Notably, the high number of blood specific lncRNAs relative to the other solid tissues suggests the presence of lncRNAs specific to immune cells, which are abundant in blood. The identification of multiple conserved lncRNAs including the well-established

lncRNAs like *XIST*, *NEAT1*, *MALAT1* and *HOTAIR*, validates our methods and confirms the conservation and expression of some key bat lncRNAs previously reported by Mostajo et al. [26]. However, our pipeline employed a simple BLASTn search against human lncRNA databases to identify potential homologs. Given the known lack of sequence conservation in lncRNAs across evolutionary distant taxa [72], our methods likely underestimated the number of conserved lncRNAs in *M. myotis*. In addition, genomic synteny analyses would be required to confirm 1:1 orthology between human and bat lncRNAs and could provide more functional implications in the future.

One surprising finding was the scarcity of lncRNAs containing viral sequences or overlapping ERV loci across all bat tissues. An unexpected feature, given the high level of endogenous retroviral sequences in the *M. myotis* genome, the documented expression of ERV in human blood [73], and the significant expression of ERV sequences reported in *M. myotis* IPS cells [74]. This observation suggests that pervasive transcription of these ERVs may be repressed in *M. myotis* tissues.

To select the lncRNAs with potential immune function from the blood sampled from our wild bat population, we were faced with two challenges. First, as bats do not typically show external signs of disease, we were unsure of their disease or immune status. However, we overcame this by performing an unsupervised analysis to identify immune co-expression gene networks changing expression across the bat population. Second, determining the exact function of a lncRNA remains challenging as most lncRNAs predicted have unknown functions and are generally poorly conserved in sequence and biological mechanism. This second challenge was addressed by constructing a co-expression network including both annotated coding genes and lncRNAs. This network allowed us to identify two major co-expression modules representing both adaptive and innate immune responses, which in turn enabled us to attribute their respective lncRNAs to these two mechanisms. The first immune module, ME1, contained 256 lncRNAs related to lymphocyte T activation and viral gene expression, while the second immune module, ME3, included 136 lncRNAs related to inflammation.

The global expression of these two modules enabled us to identify a few individuals within the wild *M. myotis* population exhibiting significant changes of immune state, plausibly linked to pathogen exposure. These individual bats were detected by antenna in the subsequent years, indicating that these immune responses and gene expression changes are typical and not indicative of terminal illness. This may also reflect a difference in the maturation of immune system between juveniles and adults. The T-cell activation

and viral related gene in ME1 overexpressed in a few young individuals may suggest an activation of antiviral immunity [75] in few susceptible, yet immunologically naive members of the colony after exposure to potential viral pathogen. The strong overexpression of ME3 observed in only a small number of individuals sampled from the same roost and year suggests an inflammatory response to a shared condition. Additionally, we observed an antagonistic relationship between the lymphocyte activity of ME1 and DNA repair pathways (ME2), while the expression of both of these modules were correlated to the age of the bats. This antagonism could reflect a trade-off between immune reaction and DNA repair pathways, as T-cells undergo multiple cycles of continuous duplications after pathogen exposure [76]. Alternatively, if the ME1 expression network is driven by a viral infection as suspected, it could indicate that viruses hijack DNA repair mechanisms to facilitate replication and immune escape [77]. This potential antagonism should be considered in the future analyses of longevity in wild bats, as shifts of immune status could impact gene expression patterns related to ageing.

Through co-expression analyses, we showed that numerous lncRNAs were co-regulated with key immune pathways in *M. myotis* blood, with many of these lncRNAs co-expressed with central immune regulators of antiviral activity or inflammation: *ARID5B*, *CELF2*, *SEPT6* and *SATB1* for the antiviral activity of ME1 and *REL* (*c-REL*), *SULF2*, *CD44*, *SVIL*, *LYST*, and *IQGAP1* for the inflammation regulation of ME3. The substantial number of lncRNAs co-expressed with *c-REL* further highlighted its role as a major regulator of bat inflammation and suggest that part of its regulatory mechanisms could involve lncRNAs. In lncRNA research, it is common practice to select a few high-priority candidates for future functional validations based on their potential for biological relevance. For this purpose, we selected four lncRNAs of high translational potential: *BatLnc1* (a newly characterised lncRNA) and *TUG1* for ME1, and *MALAT1* and *NEAT1* for ME3. These candidates were prioritised based on outstanding conservation, number of co-expressed genes in the network and/or expression level.

BatLnc1 is a polyexonic lncRNA that exhibits a high sequence conservation score, suggesting that it may be derived from part of the beta chain of the M1-specific chain of the T-cell receptor. The expression of *BatLnc1* was most strongly correlated with the Galectin9C-like (*LGALS9CL*), a copy gene of the immunomodulatory lectin GALECTIN9. GALECTIN9 is a suggested marker of diverse viral infections [78–80] with crucial roles in viral infection regulation, through mechanisms such as

limiting immunopathological damage [81–83], regulating interferon response [84–87], or by direct viral restriction [88–90]. In addition, *BatLnc1* expression is strongly correlated with other genes involved in T cell activation (*LCK*, *CD247*, *CD3G*, *GPR171*, *CD28*), and antiviral interferon signalling (*IL32L*, *NONO*, *DYRK2* and *CRLF3*) [91–94]. Based on these associations, we concluded that *BatLnc1* could be an important non-coding regulator of antiviral defense in *M. myotis*.

The second candidate selected from ME1 was the bat ortholog of *TUG1*. *TUG1* is an ubiquitous expressed lncRNA that is highly conserved in humans and mice. Initially discovered in the study of murine retina development [95], *TUG1* has recently been implicated in viral biology: its expression has been observed increasing in cases of viral infection of COVID-19 [96] or HIV [97, 98] and to have a positive or negative action on HIV physiopathological effects, replication, or reactivation [97–99]. Surprisingly, some key genes co-expressed with *TUG1* have been shown to directly or indirectly interact with HIV biology and transcription, or have been found to change expressions under HIV contamination: *CCNT2* [100], *HNRNP1* [101], *SF3B1* [102], *PRKCQ* [103], *ZFYVE16* [104], *DOCK10* and *ANKRD10* [105]. Though our current knowledge on *TUG1* does not provide enough evidence to conclude if this bat-*TUG1* acts as a repressor of retroviral activity or as a target facilitating viral degradation, our findings suggest that *TUG1* is a key player in the bat virus-host relationship. Further investigation in bats and other model organisms is required.

In ME3, we identified two highly conserved and well known ncRNAs: *MALAT1* and *NEAT1*. These two lncRNAs, which colocalised between *FRMD8* and *SCYL1*, are known for their extreme expression level and high conservation across the entire animal kingdom [106], as was also observed in our results. Their involvement is most likely pro-inflammatory, as indicated by their co-expressions with inflammation-related genes. *NEAT1* is known to activate inflammation by interacting with key regulators such as NFkB and TLR4 [107], and inflammatory components like NLRP3, NLRC4, and AIM2 [30]. *MALAT1*, on the other hand, has both pro-inflammatory [107–110] and anti-inflammatory [111, 112] roles. Interestingly, our alignments and predictions reported splice variants of *MALAT1* which are not typically observed in other species, suggesting that alternative splicing of *MALAT1* in bats may drive new functional outcomes.

In this study, we developed a comprehensive pipeline to identify and quantify lncRNAs across tissues and discovered a few key lncRNAs associated with immunity in *M. myotis* that distinguish them from other mammals. More research is required to ascertain if these immune associated lncRNAs are unique to *Myotis myotis* or shared

Table 1 Overview of selected lncRNAs and other candidates. Summary of the 4 selected lncRNAs in this study in addition to supplemental candidates with interesting characteristics that could not be studied extensively. Information on genomic localisation, number of co-expressed genes, average expression level and comments on the reason of their selection have been displayed

id	Alternative name or orthologue	Module	Reason for selection and comments	Average TMM	nb co-expressed genes	genome_contig	start (principal isoform)	stop (principal isoform)	Orientation	lncRNA_type	Length (principal isoform)	GC_percent (principal isoform)	nb_exons (principal isoform)
BatLnc_Mmyo_Intergenic_scaffold_m19_p_11G5906	<i>BatLncI</i>	ME1	Highest number of co-expressed genes	421.7794	505	scaffold_m19_p_11	51860350	51873439	-	Inter-genic lncRNA (lncRNA)	1381	54.89	9
BatLnc_Mmyo_Intergenic_scaffold_m19_p_20G211	<i>TUG1</i>	ME1	Second highest number of co-expressed genes	203.5984	482	scaffold_m19_p_20	585143	594776	-	Inter-genic lncRNA (lncRNA)	6970	43.74	6
BatLnc_Mmyo_Intergenic_scaffold_m19_p_10G9343	<i>MALAT1</i>	ME3	Very high maximum expression	14764.89	226	scaffold_m19_p_10	78761846	78767310	-	Inter-genic lncRNA (lncRNA)	4695	41.68	4
BatLnc_Mmyo_Intergenic_scaffold_m19_p_10G9375	<i>NEAT1.25</i>	ME3	Very high maximum expression	219.5041	317	scaffold_m19_p_10	78820282	78822192	-	Inter-genic lncRNA (lncRNA)	1584	58.52	3
BatLnc_Mmyo_Intergenic_scaffold_m19_p_2G8635		ME1	Close to <i>GBP6</i> (Interferon (IFN)-inducible GTPase), fourth highest number of co-expressed genes	28.0283	459	scaffold_m19_p_2	67195941	67199810	-	Inter-genic lncRNA (lncRNA)	709	46.83	4
BatLnc_Mmyo_OA.scaffold_m19_p_1G16694		ME1	Overlap <i>RORA</i> , an important gene for the function of Th4 and Th17 cells (10.1371/journal.pone.0251233, https://www.nature.com/articles/s41467-020-20385-9)	15.24445	224	scaffold_m19_p_1	176262064	176279643	+	Overlapping Antisense lncRNA (lncRNA)	3085	45.41	3
BatLnc_Mmyo_Intergenic_scaffold_m19_p_18G6167		ME3	Ex-aequo highest number of co-expressed genes in ME3	68.93299	367	scaffold_m19_p_18	32961501	32969028	-	Inter-genic lncRNA (lncRNA)	7527	46.45	1

Table 1 (continued)

id	Alternative name or orthologue	Module	Reason for selection and comments	Average TMM	nb co-expressed genes	genome_contig	start (principal isoform)	stop (principal isoform)	Orientation	lncRNA_type	Length (principal isoform)	GC_percent (principal isoform)	nb_exons (principal isoform)
BatLnc_Mmyo_OA.scaffold_m19_p_8G519		ME3	Ex-aequo highest number of co-expressed genes in ME3, overlap <i>HECW2</i>	21.4605	367	scaffold_m19_p_8	6154728	6156175	+	Overlapping Antisense lncRNA (lnc-cRNA)	927	53.83	2
BatLnc_Mmyo_Intergenic.scaffold_m19_p_7G3451		ME3	Fragment of an low antisense expression coming from the promoter of (interferon gamma receptor 1)	33.14992	248	scaffold_m19_p_7	25080813	25086933	+	Intergenic lncRNA (lnc-cRNA)	6120	42.79	1

across all bats. A recent study [113] comparing 115 mammal genomes showed extensive and excessive selection acting on immune coding genes in the ancestral lineage of bats. We could hypothesise that similar evolutionary selection pressures apply to our discovered lncRNAs given their associated role in bat immune response and therefore maybe a shared ancestral adaptation. With future genome and transcriptomic datasets from other species, our pipeline can be used for future multi-species comparative analyses of lncRNAs in bats to determine conserved and lineage specific immune bat lncRNAs of interest.

As with all in silico characterisations, our analyses cannot fully address the functional potential of these lncRNAs [114], and in vitro validations are ultimately required to confirm whether these lncRNAs perform the predicted function especially in the context of bats. As the fragmented nature of the *NEAT1* prediction revealed, predicting exact isoforms from short-read RNA-Seq remains a challenge, and exact splice variants events in our candidates such as *NEAT1* and *MALAT1*, would need to be further explored by ribo-depleted long-read sequencing. In addition, the lack of complete knowledge about the disease status of these bats means that we cannot fully identify the pathogens responsible for these immune-expression changes observed in this subset of individuals. More focused pathogen detection and further research on immune markers in our *M. myotis* colony are required to clarify the immune responses in wild bats. While we focused on four key lncRNAs, there are others among the hundreds of potential candidates that could also have immune-related functions (Table 1).

Conclusions

In summary, we have identified and characterised a novel landscape of lncRNA in one of the long-lived bat species *M. myotis*. This has extended the transcriptome annotation for this new emerging model species for immune and longevity studies. We have characterised these lncRNA in form and correlated them with traits of interest, including immunity. We have provided both a novel pipeline to identify lncRNA in bats and new lncRNAs of interest for future functional assays and ultimately therapeutic potential, shedding light on the regulation of bats' unusual immune response.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11485-1>.

Supplemental File 1. Metadata used for the RNA-seq dataset used for analysis, including sampling ID ("MMY" number), RNA-seq reference, reference of miRNA-Seq data (if available on SRA database), individual transponder ID, Colony and Year of sampling, RNA-Seq batch and

comments. Supplemental File 2. Analysis of the topology of the Whole blood RNA-Seq WGCNA network for different soft-threshold powers. Left panel indicates the impact of soft-power on the Mean scale-free topology fit index, Right panel indicates the impact of soft power on the Mean connectivity. Supplemental File 3. Analysis of the topology of the Whole blood RNA-Seq WGCNA network for different soft-threshold powers. Left panel indicates the impact of soft-power on the Mean scale-free topology fit index, Right panel indicates the impact of soft power on the Mean connectivity. Supplemental File 4. Significant blastn hits against human NONCODE v6 database (e-value < 0.001, query coverage >60%). Supplemental File 5. Significant blast hits against viral RefSeq viral genomes database (e-value < 0.001, query coverage >60%), or lncRNAs overlapping ERV regions annotated in *M. myotis* genome browser (more than 50% covered). Supplemental File 6. Analysis of variance of coding genes, lncRNAs and lincRNAs in whole blood cohort, explainable by the different metadata available in Supplemental File 1. Residual variance represents the contribution from uncharacterized variables. Supplemental File 7. ClusterProfiler results of Gene Ontology analysis enriched from coding genes of different co-expression modules in blood RNA-Seq. Supplemental File 8. Percentage of lncRNA present in each mRNA-lncRNA co-expression module. Supplemental File 9. Table representing results of passive antenna data crossed with level of expression of immune and DNA repair co-expression modules (eigengene), date of last detection of the individuals studied in the whole-blood cohorts, relative to their year of sampling, bold text indicate latest year of detection based on capture data. Supplemental File 10. PCA of the *M. myotis* blood RNA-Seq based on coding genes and lncRNAs patterns with the most contributing coding gene loadings shown (top 10 by PC). Supplemental File 11. Patterns of expression (Eigengene) of the DNA repair co-expression module ME2. Left: PCA of the blood expression patterns (coding genes + lncRNAs) in the wild *M. myotis* population, with expression of ME2. Right: Spearman correlation between ME2 eigengene and age of bat in *M. myotis* population. Bottom: Inter-module spearman correlations with ME2 based on eigengenes. Supplemental File 12. Online Blastn result from the sequence of *BatLnc1* on "Core nucleotide Database", top 10 hits by bit score. Supplemental File 13. Left: Genomic visualisation of lncRNA prediction of *TUG1* on *Myotis* genome. Right: Zscored TMM expression of *TUG1* orthologue associated with the top 20 most correlated coding genes expression in the *M. myotis* population blood. Supplemental File 14. Left: Genomic visualisation of lncRNA prediction of *NEAT1* on *Myotis* genome. The framed fragments (*NEAT1*:25) have been used for co-expression heatmap. Right: Zscored TMM expression of the *NEAT1*:25 orthologue associated with the top 20 most correlated coding genes expression in the *M. myotis* population blood.

Acknowledgements

Not applicable.

Authors' contributions

ET, SR, ZH, SC designed the study and wrote the manuscript. SR and SC coded the analysis pipeline, analysed the public RNAseq data, and generated the figures. FT organised the annual *M. myotis* mark-recapture studies, generated passive antenna detection data, and provided inputs on bat's immunology. WH, GH advised on pipeline and study design and shared their expertise on lncRNAs and bioinformatics. All authors contributed to the text, read and approved the final manuscript.

Funding

This research was supported by a Science Foundation Ireland (SFI) Future Frontiers Award 19/FFP/6790, a SFI grant 18/CRT/6214 and a European Research Council Synergy grant (no. 101118919) awarded to Emma C. Teeling.

Data availability

The RNAseq datasets analysed during the current study are already in public access in the SRA public database. *M. myotis* RNAseq from solid tissues (heart, liver, kidney, brain) are available under the PRJNA572574 accession code and the blood RNAseq dataset from Huang et al. 2019 is available under the PRJNA503704 accession code. The reference sequences used for the *M. myotis* genome and transcriptome are available at the following link <https://bds>.

mpi-cbg.de/hillerlab/Bat1KPilotProject/. *M. myotis* mitogenome and its gene annotation are available in Genbank under the KT901455.1 accession code. The code and custom files used for lncRNA prediction, along with the custom references and the lncRNA annotation will be available in the following github repository : <https://github.com/UCDBatLabBioinformatics/BatLnc>.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹School of Biology and Environmental Science, University College Dublin, Belfield, Dublin, Ireland. ²School of Biodiversity, One Health & Veterinary Medicine, College of Medical, Veterinary, and Life Sciences, University of Glasgow, Glasgow, UK. ³Earlham Institute, Norwich, UK. ⁴School of Biological Sciences, University of East Anglia, Norwich, UK.

Received: 30 January 2025 Accepted: 13 March 2025

Published online: 05 April 2025

References

- Teeling EC, Springer MS, Madsen O, Bates P, O'Brien SJ, Murphy WJ. A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science*. 2005;307(5709):580–4.
- Austad SN. Methusalem's Zoo: how nature provides us with clues for extending human health span. *J Comp Pathol*. 2010;142(Suppl 1):S10–21.
- Irving AT, Ahn M, Goh G, Anderson DE, Wang LF. Lessons from the host defences of bats, a unique viral reservoir. *Nature*. 2021;589(7842):363–70.
- Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev*. 2006;19(3):531–45.
- Banerjee A, Baker ML, Kulcsar K, Misra V, Plowright R, Mossman K. Novel insights into immune systems of bats. *Front Immunol*. 2020;24:11.
- Letko M, Seifert SN, Olival KJ, Plowright RK, Munster VJ. Bat-borne virus diversity, spillover and emergence. *Nat Rev Microbiol*. 2020;18(8):461–71.
- Wang LF, Gamage AM, Chan WOY, Hiller M, Teeling EC. Decoding bat immunity: the need for a coordinated research approach. *Nat Rev Immunol*. 2021;21(5):269–71.
- Ahn M, Anderson DE, Zhang Q, Tan CW, Lim BL, Luko K, et al. Dampened NLRP3-mediated inflammation in bats and implications for a special viral reservoir host. *Nat Microbiol*. 2019;4(5):789–99.
- Xie J, Li Y, Shen X, Goh G, Zhu Y, Cui J, et al. Dampened STING-dependent interferon activation in bats. *Cell Host Microbe*. 2018;23(3):297–301.e4.
- Ahn M, Cui J, Irving AT, Wang LF. Unique loss of the PYHIN gene family in bats amongst mammals: implications for inflammasome sensing. *Sci Rep*. 2016;6(1):21722.
- Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, et al. Comparative analysis of bat genomes provides insight into the evolution of flight and immunity. *Science*. 2013;339(6118):456–60.
- Banerjee A, Rapin N, Bollinger T, Misra V. Lack of inflammatory gene expression in bats: a unique role for a transcription repressor. *Sci Rep*. 2017;7(1):2232.
- Glennon NB, Jabado O, Lo MK, Shaw ML. Transcriptome profiling of the virus-induced innate immune response in *Pteropus vampyrus* and its attenuation by Nipah virus interferon antagonist functions. *J Virol*. 2015;89(15):7550–66.
- Hölzer M, Schoen A, Wulle J, Müller MA, Drosten C, Marz M, et al. Transcriptomes of cells from the microbat *myotis daubentonii*. *Science*. 2019;19:647–61.
- Shaw AE, Hughes J, Gu Q, Behdenna A, Singer JB, Dennis T, et al. Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol*. 2017;15(12):e2004086.
- Zhou P, Tachedjian M, Wynne JW, Boyd V, Cui J, Smith I, et al. Contraction of the type I IFN locus and unusual constitutive expression of IFN- α in bats. *Proc Natl Acad Sci U S A*. 2016;113(10):2696–701.
- Hayward JA, Tachedjian M, Cui J, Cheng AZ, Johnson A, Baker ML, et al. Differential evolution of antiretroviral restriction factors in pteropid bats as revealed by APOBEC3 gene complexity. *Mol Biol Evol*. 2018;35(7):1626–37.
- Fuchs J, Hölzer M, Schilling M, Patzina C, Schoen A, Hoenen T, et al. Evolution and antiviral specificities of interferon-induced Mx proteins of bats against Ebola, influenza, and other RNA viruses. *J Virol*. 2017;91(15):e00361–e417.
- Fernandes AP, Águeda-Pinto A, Pinheiro A, Rebelo H, Esteves PJ. Evolution of TRIM5 and TRIM22 in bats reveals a complex duplication process. *Viruses*. 2022;14(2):345.
- Hayward JA, Tachedjian M, Johnson A, Irving AT, Gordon TB, Cui J, et al. Unique evolution of antiviral tetherin in bats. *J Virol*. 2022;96(20):e01152–e1222.
- Kacprzyk J, Hughes G, Pálsson-McDermott E, Quinn S, Puechmaille S, O'Neill L, et al. A potent anti-inflammatory response in bat macrophages may be linked to extended longevity and viral tolerance. *Acta Chiropterologica*. 2017;19:219–28.
- Mattick JS, Amaral PP, Carninci P, Carpenter S, Chang HY, Chen LL, et al. Long non-coding RNAs: definitions, functions, challenges and recommendations. *Nat Rev Mol Cell Biol*. 2023;24(6):430–47.
- Gloss BS, Dinger ME. The specificity of long noncoding RNA expression. *Biochim Biophys Acta BBA - Gene Regul Mech*. 2016;1859(1):16–22.
- Nguyen Q, Carninci P. Expression specificity of disease-associated lncRNAs: toward personalized medicine. In: Morris KV, editor. Long non-coding RNAs in human disease. Cham: Springer International Publishing; 2016. p. 237–58.
- Riquier S, Mathieu M, Bessière C, Boureux A, Ruffe F, Lemaître JM, et al. Long non-coding RNA exploration for mesenchymal stem cell characterisation. *BMC Genomics*. 2021;22(1):412.
- Mostajo NF, Lataretu M, Krautwurst S, Mock F, Desirò D, Lamkiewicz K, et al. A comprehensive annotation and differential expression analysis of short and long non-coding RNAs in 16 bat genomes. *NAR Genomics Bioinforma*. 2020;2(1):lqz006.
- Peltier DC, Roberts A, Reddy P. LNCing RNA to immunity. *Trends Immunol*. 2022;43(6):478–95.
- Suarez B, Prats-Mari L, Unfried JP, Fortes P. LncRNAs in the type I interferon antiviral response. *Int J Mol Sci*. 2020;21(17):6447.
- Walther K, Schulte LN. The role of lncRNAs in innate immunity and inflammation. *RNA Biol*. 2021;18(5):587–603.
- Zhang P, Cao L, Zhou R, Yang X, Wu M. The lncRNA Neat1 promotes activation of inflammasomes in macrophages. *Nat Commun*. 2019;10(1):1495.
- Yan R, Liang X, Hu J. MALAT1 promotes colonic epithelial cell apoptosis and pyroptosis by sponging miR-22-3p to enhance NLRP3 expression. *PeerJ*. 2024;18(12):e18449.
- Yu SY, Dong B, Tang L, Zhou SH. LncRNA MALAT1 sponges miR-133 to promote NLRP3 inflammasome expression in ischemia-reperfusion injured heart. *Int J Cardiol*. 2018;254:50.
- Nishitsuji H, Ujino S, Yoshio S, Sugiyama M, Mizokami M, Kanto T, et al. Long noncoding RNA #32 contributes to antiviral responses by controlling interferon-stimulated gene expression. *Proc Natl Acad Sci U S A*. 2016;113(37):10388–93.
- Hu J, Zhang L, Zheng X, Wang G, Chen X, Hu Z, et al. Long noncoding RNA #61 exerts a broad anti-influenza virus effect by its long arm rings. *Antiviral Res*. 2023;121(5):105637.
- Jebb D, Huang Z, Pippel M, Hughes GM, Lavrichenko K, Devanna P, et al. Six reference-quality genomes reveal evolution of bat adaptations. *Nature*. 2020;583(7817):578–84.
- Huang Z, Whelan CV, Foley NM, Jebb D, Touzalin F, Petit EJ, et al. Longitudinal comparative transcriptomics reveals unique mechanisms underlying extended healthspan in bats. *Nat Ecol Evol*. 2019;3(7):1110–20.

37. Touzalin F, Petit EJ, Cam E, Stagier C, Teeling EC, Puechmaile SJ. Mark loss can strongly bias estimates of demographic rates in multi-state models: a case study with simulated and empirical datasets. *Peer Community J*. 2023;3:3.
38. Huang Z, Gallot A, Lao NT, Puechmaile SJ, Foley NM, Jebb D, et al. A nonlethal sampling method to obtain, generate and assemble whole blood transcriptomes from small, wild mammals. *Mol Ecol Resour*. 2016;16(1):150–62.
39. Jebb D, Foley NM, Puechmaile SJ, Teeling EC. The complete mitochondrial genome of the Greater Mouse-Eared bat, *Myotis myotis* (Chiroptera: Vespertilionidae). *Mitochondrial DNA Part DNA Mapp Seq Anal*. 2017;28(3):347–9.
40. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17(1):10–2.
41. Babraham Bioinformatics - FastQC a quality control tool for high throughput sequence data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Accessed 19 Dec 2024.
42. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8.
43. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinforma Oxf Engl*. 2013;29(1):15–21.
44. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5.
45. Mapleson D, Venturini L, Kaithakottil G, Swarbrick D. Efficient and accurate detection of splice junctions from RNA-seq with Portcullis. *GigaScience*. 2018;7(12):giy131.
46. Venturini L, Caim S, Kaithakottil GG, Mapleson DL, Swarbrick D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *GigaScience*. 2018;7(8):giy093.
47. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
48. Kang YJ, Yang DC, Kong L, Hou M, Meng YQ, Wei L, et al. CPC2: a fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Res*. 2017;45(W1):W12–6.
49. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res*. 2013;41(6):e74.
50. Li A, Zhang J, Zhou Z. PLEK: a tool for predicting long non-coding RNAs and messenger RNAs based on an improved k-mer scheme. *BMC Bioinformatics*. 2014;15(1):311.
51. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923–30.
52. Tarazona S, Furió-Tarí P, Turrà D, Pietro AD, Nueda MJ, Ferrer A, et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res*. 2015;43(21):e140.
53. Zhao L, Wang J, Li Y, Song T, Wu Y, Fang S, et al. NONCODEV6: an updated database dedicated to long non-coding RNA annotation in both animals and plants. *Nucleic Acids Res*. 2021;49(D1):D165–71.
54. Brister JR, Ako-adjei D, Bao Y, Blinkova O. NCBI Viral Genomes Resource. *Nucleic Acids Res*. 2015;43(Database issue):D571–7.
55. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10(1):421.
56. Hoffman GE, Schadt EE. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*. 2016;17(1):483.
57. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559.
58. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation*. 2021;2(3):100141.
59. Senckenberg Genome Browser. <https://genome.senckenberg.de/>. Accessed 27 Jan 2025.
60. Ramani R, Krumholz K, Huang YF, Siepel A. PhastWeb: a web interface for evolutionary conservation scoring of multiple sequence alignments using phastCons and phyloP. *Bioinformatics*. 2019;35(13):2320–2.
61. Hiller Lab 470 Mammals - 470 mammalian genomes aligned with Multiz by Michael Hiller's Group. <https://genome.ucsc.edu/cgi-bin/hgTracks?db=hg38&g=cons470way>. Accessed 27 Jan 2025.
62. Hahne F, Ivanek R. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol Clifton NJ*. 2016;1418:335–51.
63. Gu Z. Complex heatmap visualization. *iMeta*. 2022;1(3):e43.
64. Park BV, Freeman ZT, Ghasemzadeh A, Chattergoon MA, Rutebemberwa A, Steigner J, et al. TGF- β 1-mediated Smad3 enhances PD-1 expression on antigen-specific T cells in cancer. *Cancer Discov*. 2016;6(12):1366–81.
65. Wensveen FM, Jelenčić V, Polić B. NKG2D: a master regulator of immune cell responsiveness. *Front Immunol*. 2018;8:9.
66. Ghorbani S, Jelinek E, Jain R, Buehner B, Li C, Lozinski BM, et al. Versican promotes T helper 17 cytotoxic inflammation and impedes oligodendrocyte precursor cell remyelination. *Nat Commun*. 2022;13(1):2445.
67. Guan W, Xu J, Shi Y, Wang X, Gu S, Xie L. VNN1 as a potential biomarker for sepsis diagnosis and its implications in immune infiltration and tumor prognosis. *Front Med*. 2023;10:1236484.
68. Preechanukul A, Yimthin T, Tandhavanant S, Brummaier T, Chomkatekaew C, Das S, et al. Abundance of ACVR1B transcript is elevated during septic conditions: Perspectives obtained from a hands-on reductionist investigation. *Front Immunol*. 2023;14:1072732.
69. Derrien T, Johnson R, Bussotti G, Tanzer A, Djebali S, Tilgner H, Guernec G, Martin D, Merkel A, Knowles DG, Lagarde J. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res*. 2012;22(9):1775–89.
70. Mercer TR, Dinger ME, Sunkin SM, Mehler MF, Mattick JS. Specific expression of long noncoding RNAs in the mouse brain. *Proc Natl Acad Sci*. 2008;105(2):716–21.
71. Zimmer-Bensch G. Emerging roles of long non-coding RNAs as drivers of brain evolution. *Cells*. 2019;8(11):1399.
72. Johnsson P, Lipovich L, Grandér D, Morris KV. Evolutionary conservation of long noncoding RNAs: sequence, structure, function. *Biochim Biophys Acta*. 2014;1840(3):1063–71.
73. Balestrieri E, Pica F, Matteucci C, Zenobi R, Sorrentino R, Argaw-Denboba A, et al. Transcriptional activity of human endogenous retroviruses in human peripheral blood mononuclear cells. *BioMed Res Int*. 2015;2015(1): 164529.
74. Déjosez M, Marin A, Hughes GM, Morales AE, Godoy-Parejo C, Gray JL, et al. Bat pluripotent stem cells reveal unusual entanglement between host and viruses. *Cell*. 2023;186(5):957–974.e28.
75. Jagadeesh A, Prathyusha AMVN, Sheela GM, Bramhachari PV. T cells in viral infections: the Myriad flavours of antiviral immunity. *Dyn Immune Act Viral Dis*. 2019;5:139–48.
76. Moro-García MA, Mayo JC, Sainz RM, Alonso-Arias R. Influence of inflammation in the process of T lymphocyte differentiation: proliferative, metabolic, and oxidative changes. *Front Immunol*. 2018;1:9.
77. Saladino N, Salamango DJ. Wielding a double-edged sword: viruses exploit host DNA repair systems to facilitate replication while bypassing immune activation. *Front Virol*. 2024;4:1410258.
78. Bozorgmehr N, Mashhour S, Perez Rosero E, Xu L, Shahbaz S, Sligl W, et al. Galectin-9, a player in cytokine release syndrome and a surrogate diagnostic biomarker in SARS-CoV-2 infection. *mBio*. 2021;12(3). <https://doi.org/10.1128/mbio.00384-21>.
79. Reddy PB, Sehrawat S, Suryawanshi A, Rajasagi NK, Mulik S, Hirashima M, et al. Influence of galectin-9/Tim-3 interaction on herpes simplex virus-1 latency. *J Immunol Baltim Md 1950*. 2011;187(11):5745–55.
80. Tandon R, Chew GM, Byron MM, Borrow P, Niki T, Hirashima M, et al. Galectin-9 is rapidly released during acute HIV-1 infection and remains sustained at high levels despite viral suppression even in elite controllers. *AIDS Res Hum Retroviruses*. 2014;30(7):654–64.
81. Hu CC, Jeng WJ, Chen YC, Fang JH, Huang CH, Teng W, et al. Memory regulatory T cells increase only in inflammatory phase of chronic hepatitis B infection and related to Galectin-9/Tim-3 interaction. *Sci Rep*. 2017;7(1):15280.
82. Sehrawat S, Suryawanshi A, Hirashima M, Rouse BT. Role of Tim-3/Galectin-9 inhibitory interaction in viral-induced immunopathology: shifting the balance toward regulators. *J Immunol*. 2009;182(5):3191–201.
83. Shim JA, Park S, Lee ES, Niki T, Hirashima M, Sohn S. Galectin-9 ameliorates herpes simplex virus-induced inflammation through apoptosis. *Immunobiology*. 2012;217(6):657–66.

84. Brooks AK, Lawson MA, Rytych JL, Yu KC, Janda TM, Steelman AJ, et al. Immunomodulatory factors galectin-9 and interferon-gamma synergize to induce expression of rate-limiting enzymes of the kynurenine pathway in the mouse hippocampus. *Front Immunol.* 2016;7:422.
85. van den Hoogen LL, van der Heijden EHM, Hillen MR, Mertens JS, Fritsch-Stork RDE, Radstake TRDJ, et al. Galectin-9 reflects the interferon signature and correlates with disease activity in systemic autoimmune diseases. Response to: 'Biomarkers: to be or not to be' by Yavuz and Rönnblom. *Ann Rheum Dis.* 2020;79(1):e9–e9.
86. Imaizumi T, Kumagai M, Sasaki N, Kurotaki H, Mori F, Seki M, et al. Interferon- γ stimulates the expression of galectin-9 in cultured human endothelial cells. *J Leukoc Biol.* 2002;72(3):486–91.
87. Park WS, Jung WK, Park SK, Heo KW, Kang MS, Choi YH, et al. Expression of galectin-9 by IFN- γ stimulated human nasal polyp fibroblasts through MAPK, PI3K, and JAK/STAT signaling pathways. *Biochem Biophys Res Commun.* 2011;411(2):259–64.
88. Carreca AP, Gaetani M, Busà R, Francipane MG, Gulotta MR, Perricone U, et al. Galectin-9 and interferon-gamma are released by natural killer cells upon activation with interferon-alpha and orchestrate the suppression of hepatitis C virus infection. *Viruses.* 2022;14(7):1538.
89. Machala EA, Avdic S, Stern L, Zajonc DM, Benedict CA, Blyth E, et al. Restriction of human cytomegalovirus infection by Galectin-9. *J Virol.* 2019;93(3). <https://doi.org/10.1128/jvi.01746-18>.
90. Miyakawa K, Nishi M, Ogawa M, Matsunaga S, Sugiyama M, Nishitsuji H, et al. Galectin-9 restricts hepatitis B virus replication via p62/SQSTM1-mediated selective autophagy of viral core proteins. *Nat Commun.* 2022;13:531.
91. Lahaye X, Gentili M, Silvín A, Conrad C, Picard L, Jouve M, et al. NONO detects the nuclear HIV capsid to promote cGAS-mediated innate immune activation. *Cell.* 2018;175(2):488–501.e22.
92. Ribeiro-Dias F, Saar Gomes R, de Lima Silva LL, dos Santos JC, Joosten LAB. Interleukin 32: a novel player in the control of infectious diseases. *J Leukoc Biol.* 2017;101(1):39–52.
93. An T, Li S, Pan W, Tien P, Zhong B, Shu HB, et al. DYRK2 negatively regulates type I interferon induction by promoting TBK1 degradation via Ser527 phosphorylation. *PLoS Pathog.* 2015;11(9):e1005179.
94. Yan X, Zheng W, Geng S, Zhou M, Xu T. Cytokine receptor-like factor 3 negatively regulates antiviral immunity by promoting the degradation of TBK1 in Teleost fish. *J Virol.* 2023;97(1):e01792–22.
95. Young TL, Matsuda T, Cepko CL. The noncoding RNA taurine upregulated gene 1 is required for differentiation of the murine retina. *Curr Biol.* 2005;15(6):501–12.
96. Tayel SI, El-Masry EA, Abdelal GA, Shehab-Eldeen S, Essa A, Muharram NM. Interplay of lncRNAs NEAT1 and TUG1 in incidence of cytokine storm in appraisal of COVID-19 infection. *Int J Biol Sci.* 2022;18(13):4901.
97. Mishra T, Phillips S, Zhao Y, Wilms B, He C, Wu L. Epitranscriptomic m6A modifications during reactivation of HIV-1 latency in CD4+ T cells. *mBio.* 2024;15(11):e02214–24.
98. Pillai PP, Kannan M, Sil S, Singh S, Thangaraj A, Chivero ET, et al. Involvement of lncRNA TUG1 in HIV-1 Tat-Induced Astrocyte Senescence. *Int J Mol Sci.* 2023;24(5):4330.
99. Li J, Chen C, Ma X, Geng G, Liu B, Zhang Y, et al. Long noncoding RNA NRON contributes to HIV-1 latency by specifically inducing tat protein degradation. *Nat Commun.* 2016;13(7):11730.
100. Fong YW, Zhou Q. Relief of two built-in autoinhibitory mechanisms in P-TEFb is required for assembly of a multicomponent transcription elongation complex at the human immunodeficiency virus type 1 promoter. *Mol Cell Biol.* 2000;20(16):5897–907.
101. Kutluay SB, Emery A, Penumutthu SR, Townsend D, Tenneti K, Madison MK, et al. Genome-wide analysis of heterogeneous nuclear ribonucleoprotein (hnRNP) binding to HIV-1 RNA reveals a key role for hnRNP H1 in alternative viral mRNA splicing. *J Virol.* 2019;93(21):e01048–e1119.
102. Kyei GB, Meng S, Ramani R, Niu A, Lagisetty C, Webb TR, et al. Splicing factor 3B subunit 1 interacts with HIV tat and plays a role in viral transcription and reactivation from latency. *mBio.* 2018;9(6):18.
103. Bermejo M, López-Huertas MR, Hedgpeth J, Mateos E, Rodríguez-Mora S, Maleno MJ, et al. Analysis of protein kinase C θ inhibitors for the control of HIV-1 replication in human CD4+ T cells reveals an effect on retrotranscription in addition to viral transcription. *Biochem Pharmacol.* 2015;94(4):241–56.
104. Voss JG, Dobra A, Morse C, Kovacs JA, Danner RL, Munson PJ, et al. Fatigue-related gene networks identified in CD14+ cells isolated from HIV-infected patients—Part I: research findings. *Biol Res Nurs.* 2013;15(2):137–51.
105. Petralia MC, Nicoletti F, Tancheva L, Kalfin R, Fagone P, Mangano K. Gene co-expression network modular analysis reveals altered immune mechanisms in HIV-HAND. *Brain Sci.* 2022;12(10):1378.
106. Weghorst F, Torres Marcén M, Faridi G, Lee YCG, Cramer KS. Deep conservation and unexpected evolutionary history of neighboring lncRNAs MALAT1 and NEAT1. *J Mol Evol.* 2024;92(1):30–41.
107. Pan Y, Wang T, Zhao Z, Wei W, Yang X, Wang X, et al. Novel insights into the emerging role of Neat1 and its effects downstream in the regulation of inflammation. *J Inflamm Res.* 2022;15:557–71.
108. Cai LJ, Tu L, Huang XM, Huang J, Qiu N, Xie GH, et al. lncRNA MALAT1 facilitates inflammasome activation via epigenetic suppression of Nrf2 in Parkinson's disease. *Mol Brain.* 2020;13(1):130.
109. Yang H, Liang N, Wang M, Fei Y, Sun J, Li Z, et al. Long noncoding RNA MALAT-1 is a novel inflammatory regulator in human systemic lupus erythematosus. *Oncotarget.* 2017;8(44):77400–6.
110. Zhou Q, Run Q, Li CY, Xiong XY, Wu XL. lncRNA MALAT1 promotes STAT3-mediated endothelial inflammation by counteracting the function of miR-590. *Cytogenet Genome Res.* 2020;160(10):565–78.
111. Masoumi F, Ghorbani S, Talebi F, Branton WG, Rajaei S, Power C, et al. Malat1 long noncoding RNA regulates inflammation and leukocyte differentiation in experimental autoimmune encephalomyelitis. *J Neuroimmunol.* 2019;15(328):50–9.
112. Zhao G, Su Z, Song D, Mao Y, Mao X. The long noncoding RNA MALAT1 regulates the lipopolysaccharide-induced inflammatory response through its interaction with NF- κ B. *FEBS Lett.* 2016;590(17):2884–95.
113. Morales AE, Dong Y, Brown T, et al. Bat genomes illuminate adaptations to viral tolerance and disease resistance. *Nature.* 2025;638:449–58.
114. Ponting CP, Haerty W. Genome-wide analysis of human long noncoding RNAs: a provocative review. *Annu Rev Genomics Hum Genet.* 2022;23:153–72.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.