TITLE PAGE

Title:

Transcriptional dynamics of transposable elements in the type I IFN response in *Myotis lucifugus* cells

Authors:

Giulia Irene Maria Pasquesi[1], Conor J. Kelly[1], Andrea D. Ordonez[1] and Edward B. Chuong[1§]

§ Corresponding author:

Edward B. Chuong

edward.chuong@colorado.edu

596 UCB Boulder, CO 80309 303-735-8573

Affiliation:

1 BioFrontiers Institute and Department of Molecular, Cellular & Developmental Biology, University of Colorado Boulder

Keywords: bat immunity, transposable elements, epigenomics

Running Title: IFN-inducible transposons in *Myotis* cells

ABSTRACT

Background: Bats are a major reservoir of zoonotic viruses, and there has been growing interest in characterizing bat-specific features of innate immunity and inflammation. Recent studies have revealed bat-specific adaptations affecting interferon (IFN) signaling and IFN-stimulated genes (ISGs), but we still have a limited understanding of the genetic mechanisms that have shaped the evolution of bat immunity. Here we investigated the transcriptional and epigenetic dynamics of transposable elements (TEs) during the type I IFN response in little brown bat (*Myotis lucifugus*) primary embryonic fibroblast cells, using RNA-seq and CUT&RUN.

Results: We found multiple bat-specific TEs that undergo both locus-specific and family-level transcriptional induction in response to IFN. Our transcriptome reassembly identified multiple ISGs that have acquired novel exons from bat-specific TEs, including *NLRC5*, *SLNF5* and a previously unannotated isoform of the *IFITM2* gene. We also identified examples of TE-derived regulatory elements, but did not find strong evidence supporting genome-wide epigenetic activation of TEs in response to IFN.

Conclusion: Collectively, our study uncovers numerous TE-derived transcripts, proteins, and alternative isoforms that are induced by IFN in *Myotis lucifugus* cells, highlighting candidate loci that may contribute to bat-specific immune function.

KEYWORDS: bat immunity, transposable elements, epigenomics

BACKGROUND

Bats are increasingly recognized to be an important reservoir of zoonotic viruses, including Rabies viruses, Dengue viruses, Ebolaviruses, and Coronaviruses [1,2]. Remarkably, viral infection in bats is associated with minimal lethality and reduced inflammatory phenotypes,

which has led to extensive research aimed at uncovering bat-specific features of immunity [3–6].

Recent genomic and functional studies in bats have begun to reveal species-specific adaptations affecting innate immune responses. For example, the interferon (IFN) genes have been subject to evolutionary expansions and contractions in different bat species, and several species exhibit constitutive expression of IFNs at low levels [7] (reviewed in [1]. Bats also exhibit unique subsets of ISGs [8]. A time-course analysis comparing the type I IFN response in *Pteropus alecto* and humans revealed distinct kinetics of IFN-stimulated gene (ISG) regulation, where bats exhibit more rapid downregulation of ISGs compared to humans [9]. In addition to adaptations affecting IFN signaling, other pro-inflammatory genes are also frequently mutated or lost, including TLR genes [10], components of the inflammasome [11,12], the cGAS/STING pathway [13], and the OAS/RNASEL pathway [9,14]. These studies have begun to reveal the genetic basis for bat-specific features of immunity which could help us understand their propensity to act as viral reservoirs.

While it is clear that bats have evolved numerous unique adaptations affecting innate immune pathways, we still have a poor understanding of the genetic mechanisms responsible for these changes. Our study focuses on TEs as a potentially important yet understudied source of mutations that shape bat immunity. In studies of other mammalian lineages, there is evidence that lineage-specific TEs contribute to innate immune functions through a variety of mechanisms. For example, TE and virus-derived proteins have been repeatedly co-opted as immune proteins that often restrict viruses through dominant negative activity [15] in ruminants [16], rodents [17], and primates [18]. TE-derived non-coding transcripts can readily form immunostimulatory double-strand RNAs or DNAs [19–21]. Finally, TEs can regulate interferon-

inducible gene expression by acting as regulatory elements [22–26]. The recurrent co-option of TEs for immune functions throughout evolution may reflect their capacity to fuel adaptation by increasing genomic variation, especially in the context of host-pathogen coevolutionary arms races [27,28].

TEs are widely speculated to have been important contributors to the evolution of bats [29–33], including bat-specific immune functions [34,35]. While the genomes of all mammalian species contain numerous lineage-specific transposons, bat genomes are distinguished by recently active DNA transposons, which are extinct in most other mammalian lineages [29–32]. In addition to DNA transposons, bat genomes have been extensively shaped by other TEs typically found in other mammals, including LTR retrotransposons like endogenous retroviruses, LINEs, and SINEs. Notably, a recent report identified a *Rhinolophus*-specific LTR insertion within an exon of *OAS1,* which disrupts horseshoe bat antiviral activity to SARS-CoV2 with significant implications for the OAS1-mediated response to SARS-CoV2 in humans [14]. However, aside from this example, the potential impact of TEs on bat immunity remains largely unstudied, due to the lack of experimental and functional genomic resources available for studying bat immunology.

To conduct a comprehensive genome-wide study of TEs in bat innate immunity, we conducted transcriptomic and epigenomic profiling of the type I IFN response in *Myotis lucifugus* primary fibroblast cells. We used RNA-seq and CUT&RUN to characterize the IFN-inducible transcriptomes and regulatory elements, which allowed us to systematically examine the contribution of TEs to loci that define the bat IFN-inducible response.

RESULTS

To characterize the contribution of TEs to the IFN response in bats, we conducted transcriptomic and epigenomic profiling of the type I IFN response in *M. lucifugus* primary embryonic fibroblast cells (Fig. 1). We stimulated cells using recombinant universal IFN alpha (IFNa), and profiled the transcriptome at 0, 4 and 24h time points using RNA-seq. We confirmed cellular response to universal IFN treatment using qPCR on canonical ISGs (Fig. S1), as shown previously for *M. lucifugus* dermal fibroblasts [8]. We also profiled 0 and 4h time points using CUT&RUN to map genome-wide localization of H3K27ac, POLR2A, and STAT1. We aligned these data to a chromosome-scale HiC assembly of the little brown bat genome (myoLuc2.0_HiC) [36], which was the most contiguous assembly available (Scaffold N50 of ~95.5Mb).

Prior to analyzing our functional genomic data, we performed de-novo repeat identification on the myoLuc2.0_HiC assembly using RepeatModeler2 [37–39]) and HelitronScanner ([37–39], followed by repeat annotation using RepeatMasker [40]. We annotated 42.7% of the genome as derived from TEs, compared to 35.5% annotated in the myoLuc2 non-HiC assembly (https://www.repeatmasker.org/species/myoLuc.html). L1 LINEs represent the most abundant TE group (14.1%), followed by virus-derived elements (ERVs; 5.6%) and DNA hAT and Helitron elements (3.7% and 3.2% of the genome, respectively) (Fig. 2A). As previously identified [29], our analyses show the lineage-specific expansion of DNA transposons, first led by Helitron elements, and more recently by multiple subfamilies of hAT elements that may have been introduced by horizontal transfer (Fig. 2A) [41].

Gene and TE expression profiles upon IFN stimulation

To analyze transcriptional activity at the TE family level, we mapped RNA-seq reads to both genes and TE families using TETranscripts (Fig. 2B) [42]. On average, 6.38% of RNA-seq reads

mapped to TEs in unstimulated cells, while 7.26% of reads mapped to TEs after 4h IFN treatment, and 7.15% after 24h IFN treatment. The most abundant TE-derived transcripts we identified included L1 LINEs, DNA/hAT elements, ERVs, SINEs and L2 LINEs. We identified 45 TEs that showed significant family-level transcriptional induction at 4h (adj. *p*-val < 0.05; log2FC > 1.5), and 8 families induced at 24h according to the same cutoff thresholds (Fig. 2C). These included multiple ERV families (21), L1 LINEs (6) and DNA transposons (6) at 4h post treatment, and ERVs (5) at 24h post IFN treatment. These findings indicate that multiple bat-specific TEs show family-level transcriptional induction in response to IFN treatment, peaking at 4h and diminishing but still present at 24h post induction.

We next analyzed our RNA-seq dataset to identify IFN-stimulated genes (ISGs), as determined by DESeq2 comparing gene counts from treated to untreated samples in pairwise comparisons at 4h and at 24h post treatment (Tables S1 and S2). We first used the homology-based annotation provided by DNAzoo as a reference transcriptome. Using a cutoff of adj. *p*-val < 0.05 and log2FC > 1.5, we identified 213 upregulated transcripts (corresponding to 138 unique genes) and 1 downregulated gene at 4h, and 91 upregulated transcripts (corresponding to 66 unique genes) at 24h post IFN stimulation. Based on their expression dynamics, 4 main transcript clusters were identified (Fig. 3A): I) transcripts showing a strong response to IFN at 4h that declines at 24h; II) transcripts showing mild induction at 4h and decline at 24h; III) transcripts showing mild, stable induction; and IV) transcripts showing strong induction at 4h and rapid decline to levels similar to unstimulated cells at 24h. Both 4h and 24h post-induction ISGs were enriched for canonical ISGs and other genes involved in immune signaling (Fig. 3B). Notably, we observed induction of genes involved in DDX58/IFIH1 (RIG-I/MDA5)-mediated induction of IFNa/b at 4h, followed by induction of negative regulators of this pathway at 24h (Fig. 3B). Similarly, an enrichment for genes involved in response to cytokine stimuli was detected at 4h but not at 24h post treatment. These observations of a strong early response

148 followed by a decline by 24h upon IFN stimulation are consistent with observations in the black
149 flying fox (*Pteropus alecto*) [9].
150
151 *TE contribution to ISG transcript structure*
152 To improve detection of potentially unannotated IFN-induced TE-derived transcripts and
153 isoforms, we conducted genome-guided transcriptome reassembly on our combined RNA-seq
154 dataset using StringTie2 [43] (Supplementary Data 1), and annotated assembled genes based
155 on homology using the SwissProt database. This yielded an expanded transcriptome with
156 68,110 transcripts (32,137 transcripts matching an annotated gene). After performing pairwise
157 differential expression analyses (Table S3) we identified 1243 IFN-inducible transcripts
158 corresponding to 836 StringTie genes (740 IFN-inducible transcripts corresponding to 449
159 StringTie genes matching Swissprot) at 4h, and 717 transcripts corresponding to 500 StringTie
160 genes (385 transcripts, 239 matching Swissprot) at 24h. Of these, 606 transcripts corresponding
161 to 392 genes (358 transcripts, 214 genes matching Swissprot) were shared between 4h and
162 24h treatment time.
163
164 We used our improved transcriptome reassembly to investigate the contribution of TEs to both
165 constitutive and IFN-inducible transcript structures. First, we identified transcripts that contained
166 exonized TEs based on the overlap of exon (>50% sequence length) and annotated TE
167 features. Considering all expressed (TPM ≥ 0.5) multi-exon transcripts, we found a subset of
168 1039 transcripts corresponding to 749 StringTie genes (648 transcripts and 470 genes with
169 homology to annotated genes in the SwissProt database) that contained at least one TE-derived
170 exon (Table S4).
171

Focusing first on transcripts reconstructed from constitutively expressed genes, we identified *EEF1A1* as an example where 271 bp (representing part of the third and fourth exons; coding exons 2 and 3) is annotated as derived from a Zisupton DNA transposon. Since Zisupton DNA elements have likely been lost during the tetrapoda radiation [44], and their presence and activity has been confirmed only in fish species among vertebrates, we further verified this finding through multiple BLAST searches (Supplementary Data 2). First, we aligned the StingTie transcript (STRG.21546.2, Scaffold 10: 12654609-12761039) to the transcript sequence deposited on the NCBI database (XM_006089332.3). After confirming the match between the two transcripts, we used the deposited mRNA sequence as a query against the Repbase transposable element database [45] and against bats and mammalian mRNA collections. All *Microchiroptera*, *Vespertilionidae* in particular, share highly homologous coding exons 3 and 4 (Supplementary Data 3), suggesting that the exonization event occurred before or during the *Microchiroptera* and *Pteropodinae* radiation. Despite the shared homology, no Zisupton matching features were found in the human or other euarchonta *EEF1A1* homologues. Therefore, our analysis successfully identified bat *EEF1A1* as an example of a gene that has likely been altered by a bat-specific TE.

We next filtered for IFN-inducible transcripts containing TE-derived exon(s), and found 44 transcripts from 34 genes (16 of which annotated by SwissProt homology) induced at 4h, and 31 transcripts from 24 genes (11 annotated) induced at 24h (Table S4). These included genes with established roles in immune responses, like *PARP9*, *SLFN5* (Fig. S2A) and a candidate novel isoform of *IFITM2* that has not been previously identified (Fig. S2B). This analysis reveals that numerous constitutively expressed and IFN-inducible genes have acquired exons from bat-specific TEs either in coding regions or UTRs (Table 1).

We next aimed to identify potential examples of co-opted TE-derived proteins (such as syncytins; [46]) which may not be masked by RepeatMasker due to their age. We used *tblastx* to query expressed StringTie transcripts against reference TE protein sequence libraries specific for retrotransposons and DNA elements (see Methods). We found a total of 7 StringTie genes with a known annotated ortholog based on the concatenated StringTie transcriptome assembly, and 156 StringTie genes based on intact ORF prediction (Table S5; Methods). By using the intact ORF prediction approach we identified 9 ISGs (i.e., *GBP1*, *DDX58* and *PARP14*) with at least one retrotransposon-derived feature, mostly from ERVK and L1 LINEs. Most of these TE-derived sequences reside in the last exon, where they provide both the stop codon and the 3'UTR, or novel coding and/or regulatory sequences. Finally we followed the same approach to identify novel protein-coding sequences matching known viral proteins that could represent domesticated viral proteins [47]. By leveraging the gEVE and a custom syncytin protein database we found a total of 3 constitutively expressed StringTie genes with homology to syncytin proteins and 9 with homology to either a pol, gag, retrotransposase or AP viral proteins (Table S6). Only one pol (RVT1)-derived gene, *UBP18*, was differentially expressed upon IFN treatment .

In addition to examining the TE contribution to protein-coding sequences, we also searched for examples of TE-derived promoters. We collapsed StringTie transcript coordinates to their transcription start site (TSS), and intersected transcript TSSs with the generated TE annotation. We identified a total of 11 transcripts with a TSS deriving from a TE that are IFN-inducible at 4h, 9 of which are shared with the 24h subset (Table S7). Most of these belong to genes known to be involved in immune function and regulation, like *NLRC5* (Fig. 4), *EIF2AK2*, *GBP1*, *MX1*, *MX2*, *PHF11*, *SAMD9* and *XAF1*, while others like *PARP14* and *CS012* are not canonical ISGs. Notably, we also observed recurrent usage of TE-derived promoters for histocompatibility loci located in Scaffolds 13 and 20, where 78 transcripts (corresponding to 21 annotated genes) at

4h and 7 transcripts (corresponding to 3 annotated genes) at 24h have TE-derived TSSs or coding exons. While the histocompatibility locus may be prone to sequence assembly artifacts related to its highly polymorphic and complex structure, our analysis suggests that TEs may have influenced the evolution of the bat histocompatibility locus, which has been proposed to underlie bat-specific immune function [48,49].

Epigenetic profiling upon IFN stimulation

Having analyzed the contribution of TEs to ISGs, we next asked how TEs contribute to inducible regulatory elements defined by H3K27ac, POLR2A, or STAT1 activity in response to IFN. We observed an increase in STAT1 signal at the predicted promoters for *IRF9* and *PSME2* in response to type I IFN (Fig. 5A). Using spike-in normalized CUT&RUN data at 0 and 4h, we used DESeq2 to define IFN-inducible regulatory elements, as performed previously [22]. Unexpectedly, we did not observe robust IFN-inducible chromatin changes that are characteristic of IFN-stimulated cells from other species [22,50]. We did not identify any H3K27ac inducible elements with an FDR < 0.05 (Fig. 5B; Table S8), and instead defined a set of 1113 elements showing an increase of H3K27ac signal with a relaxed significance threshold of unadjusted *p*-val < 0.1 (Fig. S3A). This set of IFN-inducible elements was enriched for interferon-stimulated response element (ISRE) motifs (E-value 1.43×10^{-10}) (Table S9), consistent with their activation by IFN stimulation. Thus, while our CUT&RUN analysis successfully identified some elements showing IFN-inducible activity, our analysis reveals surprisingly modest chromatin-level changes, despite robust ISG induction according to RT-qPCR from RNA taken from the same fraction of cells (Fig. S1).

Of these regions, we found that 466 out of 1113 fully overlapped at least one TE (Table S10). Additionally, we identified 766 inducible, STAT1-bound TEs that fall within 100kb of an ISG

(Table S11). This includes an LTR14_ML element that may be functioning as the promoter for the *NLRC5* locus in addition to an intronic Ves2_ML SINE element (Fig. 4). However, in contrast to previous studies in other species [22–26], we did not observe any overrepresented TE families within this set (Fig. S3B; Table S12). The only subfamilies that overlapped more than 10 IFN-inducible H3K27ac regions correspond to the ancient mammalian MIR and L2 families that predate the evolution of bats (Fig. 5C; Fig. S3C). Querying H3K27ac regions from either untreated or IFN-stimulated conditions independently, we observed only very modest enrichment of the MIR3, AmnSINE1, and LTR13C_ML families (Fig. 5D; Fig. S4A, S4B). Taken together, our analysis indicates that TEs contribute to hundreds of regulatory elements involved in IFN signaling, but in contrast to studies in other species, we did not identify enrichment of lineage-specific TE families within IFN-inducible regulatory elements. However, given that our CUT&RUN analysis revealed a relatively minimal set of inducible regulatory elements at a genome-wide level, we were limited in our ability to identify enriched TE families.

DISCUSSION

Our study characterizes the transcriptional and epigenetic dynamics of bat TEs in the IFN response in *M. lucifugus* cells. To facilitate our ability to map TEs in our functional genomic data, we conducted both RNA-seq and CUT&RUN using 150bp paired end reads, and generated an improved repeat annotation using a chromosome-scale assembly. Our analyses revealed that TEs have shaped the IFN-inducible transcriptome, but we did not find strong evidence for a global role for TEs in shaping associated epigenetic changes. Functional studies will be necessary to validate whether any of the elements identified in this study have significance for bat immunity, but given the growing number of validated examples in human and mouse, it is likely that some TEs have been co-opted for innate immune function in bats.

For our study, we generated a matched transcriptomic and epigenomic datasets profiling the type I IFN response in *M. lucifugus* primary cells. Our transcriptomic analysis of the IFN response in *M. lucifugus* embryonic fibroblasts confirms previously reported features of bat innate immunity. We found that these cells respond to IFN stimulation at the transcriptional level, with a stronger and broad induction of ISGs at earlier time points (4 hours post treatment). We also found that only a small subset of genes that were overexpressed at 4h maintain high expression levels at 24h, whereas most genes show a reduction in expression to lower levels or levels similar to those recorded in unstimulated cells. This is in agreement with gene expression profiling in *Pteropus alecto* [9]. In parallel to gene expression, we characterized expression profiles of TE-derived transcripts, and found similar trends. Total TE expression was higher at 4h post IFN treatment, with more and more diverse TE families being differentially expressed both in comparison to unstimulated and 24h cells. While the expression of these transcripts is not directly indicative of function, IFN-inducible expression of bat-specific TE families may act as a source of non-coding transcripts that can further activate innate immune pathways, akin to the "viral mimicry" pathway characterized in human cancers [19,20].

We also explored whether specific TEs may have affected the transcript structure of host genes, by screening for gene transcripts that share homology with transposable elements or viral proteins in coding regions and transcription start sites (TSS). Our genome-guided transcriptome assembly identified multiple instances of TE-derived and viral protein-derived exonization events, both as alternative (*IFITM2*) or conserved (*PARP9*) exons and TSS (*NLRC5*). Some of these transcripts represent canonical (*PARP9*, *DDX60*) or non-canonical (*PLAAT3*, *SP140L*) ISGs. These analyses provide strong evidence that TEs have been co-opted into the exons of bat ISGs, and some of these exonization events may have significant functional consequences. For example, our analysis identified multiple *Myotis*-specific TE insertion and exonization events affecting the *NLRC5* gene. NLRC5 has been identified as a key regulator of MHC class I-

dependent immune responses [51], and may be involved in the regulation of inflammasome activation and type I IFN responses [52]. Further studies are needed to validate the potential effects of these TE-derived sequences, but it is possible that *Myotis*-specific TEs have altered NLRC5 function and/or regulation.

We also identified instances of TE-derived constitutively expressed genes. We verified through multiple BLAST and sequence alignments that the ~100 amino acids of the EEF1A1 protein of *Microchiroptera* and likely *Pteropodinae* bats derived from the exonization of a Zisupton DNA transposons. Although Zisupton DNA transposons are not abundant in the *M. lucifugus* genome (we recovered only one family present, with 174 copies), it is possible that they were horizontally transferred into bat genomes as has happened in multiple fish species [44]. Age analysis of Zisupton genomic copies suggests that they were likely recently introduced (divergence from consensus ~18), but were able to expand only for a short period of time (peak of divergence from the consensus sequence at 4-9), when other DNA elements have been active. This example uncovered by our analysis highlights the possibility that TEs have shaped other aspects of bat biology in addition to genes involved in immune function.

While RNA-seq profiling has been applied by an increasing number of studies to profile bat immunity at the transcriptomic level, no study to date has characterized bat immunity at the epigenomic level. Unexpectedly, while our RNA-seq analysis of *M. lucifugus* cells coincided with strong transcriptional response to IFN treatment, we observed relatively modest chromatin changes based on CUT&RUN epigenomic profiling. This observation contrasts with robust chromatin changes typically observed in IFN-treated cells from other mammalian species [22,53,54]. As a result, we did not observe strong evidence supporting family-level TE regulatory induction as observed previously in other species such as human [50] and cow [22], partly due to the lack of clearly inducible elements at a genome-wide level.

There are multiple potential explanations for our observation of a relatively modest IFN-inducible epigenetic response in bat cells. First, our study is one of the first to conduct CUT&RUN on bat cells, and the antibodies used in this study have not been fully validated in *M. lucifugus*. However, while the antibody used for STAT1 may exhibit poor recognition of bat ortholog, histones and POLR2A are highly conserved and expected to be targeted effectively by standard antibodies. Second, our study involved stimulating derived embryonic fibroblast cell culture with recombinant universal type I IFN. While these conditions nonetheless showed strong transcriptional induction of ISGs, it is possible that chromatin dynamics are different during endogenous activation of IFN responses *in vivo.*

Finally, our observations may reflect a unique attribute of bat immunity, consistent with the idea that some bat species exhibit constitutive IFN expression [7,55]. Although the type I IFN locus of *M. lucifugus* is poorly characterized, we were able to annotate at least 11 uncharacterized genes that likely reflect the expansion by gene duplication of the IFNw cluster (Fig. S5) [56], whereas no IFNa genes were identified. Of the 11 IFNw paralogues, 5 showed evidence of constitutive low expression in unstimulated cells, and induction at 4h post treatment (Fig. S5). These observations suggest a scenario where *Myotis* epigenomes are “primed” due to constitutive expression of IFN, and may be capable of driving robust transcriptional activation without exhibiting epigenetic changes typically associated with inducible chromatin activity, such as increased H3K27ac or POLR2A levels.

CONCLUSIONS

Our study provides a first systematic investigation of the contribution by TEs to the bat type I IFN response. We uncover numerous examples of TE-derived transcripts, alternative exons,

and regulatory elements that shape the genomic response to IFN in *M. lucifugus*. Our study suggests that TEs in other bat lineages such as *Pteropus* and *Rhinolophus* may also shape IFN-inducible transcriptomes, which may motivate functional studies to determine their biological significance in the context of bat immunity. Our findings lend additional support for a widespread role for TE co-option in shaping the evolution of species-specific immune responses.

METHODS

Transposable Element identification and analysis

Myotis lucifugus genomic repeat elements were annotated according to homology-based and *de novo* identification approaches. Although repeat elements have been extensively annotated in bat species including *Myotis lucifugus*, we performed *de novo* TE identification using RepeatModeler2 (included in dfam-tetools-1.1) [37] and HelitronScanner v1.1 [38] to match the newly released, highly contiguous chromosome-level assembly of the little brown bat genome (myoLuc2_HiC) [36,57,58]. Given their highly repetitive nature, TE loci are hard to assemble and often incomplete; therefore genome assemblies that rely on long read assembly strategies (i.e., HiC) [36,59] are better suited for capturing full length elements over contiguous chromosome-level scaffolds.

Briefly, we performed de novo TE identification using RepeatModeler and HelitronScanner, then combined the two libraries as a single little brown bat *de novo* library that was used for homology-based TE annotation using RepeatMasker v4.1.0 [40]. To maximize element identification we followed a custom multi-step mapping strategy [60] using multiple libraries as reference for the masking process in the following order: (i) bat specific repeats included in the Repbase library provided with RepeatMasker; (ii) a bat specific library provided by Dr. Cosby;

(iii) our *de novo* little brown bat library; (iv) the entire tetrapoda Repbase library provided with RepeatMasker [45].

Cell lines and treatment

Myotis lucifugus primary embryonic fibroblast cells were a gift from Mario Capecchi. Cells were grown at 37°C and 5% CO_2 and passaged in DMEM (ThermoFisher #10566016) supplemented with 10% FBS, 5% MEM nonessential amino acids, 100 U/mL Penicillin-Streptomycin, and 1 mM sodium pyruvate. Cells were seeded into six-well plates at an optimized density of 2×10^5 cells per well in 2ml of culturing media (or 2×10^6 cells per 15 cm dish for CUT&RUN). The following day (or 48h for CUT&RUN) cells were treated with 1000U/ml of Universal Type I IFNa resuspended in DPBS (PBL Assay Science #11200) in 2ml of culturing media; control cells were treated with equivalent volume of DPBS in 2ml of media. At four hours and 24 hours post treatment cells were harvested for RNA extraction (four hours for CUT&RUN).

RNA isolation and library preparation for RNA-seq

Following media removal, cells were washed with 1ml of DPBS and detached by adding 400ul of 0.25% trypsin per well. Following a 10 minutes incubation at 37°C, trypsin was neutralized with 1.6ml of culturing media. Cell suspensions were transferred into 1.7ml tubes and pelleted by centrifugation at 300xg for 5 minutes. Cells were then lysed in 300ul of RNA lysis buffer (Zymo Research #R1060-1-50), and stored at -80°C until RNA extraction was performed using the Quick-RNA MiniPrep kit (Zymo Research #R1054), following manufacturer's instructions.

Total RNA samples for each time point and condition were prepared in three biological replicates as described above. A NanoDrop One spectrophotometer (Thermo Fisher Scientific) was used to determine RNA concentration and quality; all samples passed quality assessment.

PolyA enrichment and library preparation was performed using the KAPA mRNA HyperPrep Kit (Kapa Biosystems #8098115702) according to the manufacturer's protocols. Briefly, 500 ng of RNA was used as input, and single-index adapters (Kapa Biosystems #08005699001) were added at a final concentration of 10 nM. Purified, adapter-ligated library was amplified for a total of 11 cycles following the manufacturer's protocol. The final libraries were pooled and sequenced on an Illumina NovaSeq 6000 (University of Colorado Genomics Core) as 150bp paired-end reads.

CUT&RUN sample and library preparation

CUT&RUN pulldowns were generated using a protocol from [61,62]. All buffers were prepared according to the "High Ca^{2+}/Low Salt" section of the protocol using 0.04% digitonin (EMD Millipore #300410). 5×10^5 viable cells were used for each pulldown. The following antibodies were used: rabbit anti-mouse IgG (1:100, Abcam #ab46540), rabbit anti-H3K27me3 (1:100, Cell Signaling #9733), rabbit anti-H3K27ac (1:100, Millipore #MABE647), rabbit anti-pRPB1-Ser5 (1:50, Cell Signaling #135235S), rabbit anti-STAT1 (1:100, Cohesion #3322), rabbit anti-pSTAT1-Ser727 (1:100, Active Motif #39634). pAG-MNase (prepared as in [61,62]) was added to each sample following primary antibody incubation at a final concentration of 700 ng/mL. Chromatin digestion, release, and extraction was carried out according to [61,62]. Yeast spike-in DNA (gift from Steven Henikoff) was added to the quenching ("1× STOP") buffer for a final concentration of 100 pg/mL. Pulldown success was determined by Qubit dsDNA High Sensitivity and TapeStation 4200 HSD5000 before proceeding with library preparation.

Libraries were generated using a modified protocol for use with the KAPA HyperPrep Kit. Briefly, the full volume of each pulldown (50 uL) was used to generate libraries according to the manufacturer's protocol with the following modifications. Freshly diluted 0.200 uM single-index

adapters (Kapa Biosystems #08005699001) were added to each library at a low concentration (9 nM) to minimize adapter dimer formation. Adapter-ligated libraries underwent a double-sided 0.8X/1.0X cleanup with KAPA Pure Beads (Kapa Biosystems #07983280001). Purified, adapter-ligated libraries were amplified using the following PCR cycling conditions: 45 s at 98°C, 15x(15 s at 98°C, 10 s at 60°C), 60 s at 72°C. Amplified libraries underwent a double-sided 0.8X/1.0X cleanup. The final libraries were quantified using Qubit dsDNA High Sensitivity and TapeStation 4200 HSD5000. Libraries were pooled and sequenced on an Illumina NovaSeq 6000 (Novogene) as 150bp paired-end reads.

Paired RT-qPCR

5×10^5 viable cells from the same CUT&RUN populations (untreated and 4h IFN) were used to extract RNA for RT-qPCR analysis to confirm induction of IFN-inducible genes prior to CUT&RUN library preparation. Cells were lysed in 300ul of RNA lysis buffer (Zymo Research, #R1060-1-50). Prepared lysates were stored at -80°C until RNA extraction was performed using the Quick-RNA MiniPrep kit (Zymo Research #R1054), following the manufacturer's instructions.

Total RNA samples for each time point and condition were prepared in three biological replicates as described above. A NanoDrop One spectrophotometer (Thermo Fisher Scientific) was used to determine RNA concentration and quality; all samples passed quality assessment. RNA expression levels for *CTCF*, *STAT1*, and *IFIH1* were quantified using the Luna Universal One-Step RT-qPCR Kit (New England Biolabs #E3005L) according to the manufacturer's instructions. In brief, for each reaction 25 ng of RNA was combined with 5ul 2× Luna Universal One-Step Reaction Mix, 0.5ul 20× Luna WarmStart RT Enzyme Mix, 0.4ul 10uM forward primer, and 0.4ul 10uM reverse primer. Reactions were amplified using a CFX384 Touch Real-Time

PCR Detection System (Bio-Rad) with the following PCR cycling conditions: 10 min at 55°C, 1 min at 95°C, 40x(10 s at 95°C, 30 s at 60°C). On-target amplification was assessed by melt curve analysis. Two biological replicates were included for each treatment condition, and each biological replicate was run in technical duplicate. Statistical significance was assessed using a two-tailed paired Student's t-test with a threshold of *p*-val < 0.05.

Transcriptome analyses

Paired-end 150bp read length FASTQ files were quality and adapter trimmed using BBDuk v.38.05 [63]; quality check was performed using FastQC v0.11.8 [64] and inspected through MultiQC v1.7 [65]. Filtered FASTQ files were then mapped to the myoLuc2_HiC genome using a 2-pass approach in STAR v2.7.3a [66]. STAR was run following default parameters and allowing for multi-mapping reads (with options '*–outAnchorMultimapNmax 100 –winAnchorMultimapNmax 100 –outFilterMultimapNmax 100*'), a requisite for the inclusion of TE-mapping reads in the output files. The annotation file available on DNAzoo, here referred to as "functional annotation" (www.dropbox.com/sh/xt300ht42mihjov/AADoENW7RTvR3jTh1a8qUOmRa) was used as reference for the mapping process. For the second pass of mapping we filtered out novel junctions that mapped to the mitochondrial genome, HiC_scaffold_93 based on the most likely alignment hit to the reference mitochondrial genome performed using LASTZ v1.02.00 [67]. Resulting alignment files in sorted *.bam* format were then provided as input for TE and gene expression quantification in TETranscripts v2.1.4 [42] using the same gene annotation and our custom TE annotation derived from RepeatMasker. Pairwise differential expression analyses at 4h and 24h post IFN treatment were performed in DESeq2 v1.32 [68]. Functional enrichment analyses of differentially expressed genes (adj. *p*-val < 0.05, log2FC > 1.5) were performed using the WebGestalt web tool [69].

475

476 *Genome guided transcriptome assembly and analysis*

477 Short-read RNA-seq alignment files generated by running STAR (see previous paragraph)
478 were merged, sorted and indexed using SAMtools v1.10 [70], and the resulting *.bam* file was
479 used as input for genome guided transcriptome assembly in StringTie v1.3.3b [43]. StringTie
480 was run following default parameters, except that the minimum number of spliced reads
481 required to align across a junction was increased from 1 to 5 using option '*-j 5*'. The resulting
482 StringTie gtf output (Supplementary Data 1) was then converted to FASTA format using the
483 *gffread* utility using options '*-M, -F, and -Z*' included in Cufflinks v.2.2.1 [71]. We followed an
484 homology-based approach to annotate assembled StringTie genes and isoforms. StringTie
485 gene sequences were queried against the SwissProt database [72] through the *blastx* search
486 algorithm in BLAST 2.12.0+ [73] using options '*-max_target_seqs 1 -evalue 10*'. Matches were
487 then filtered for shared sequence identity equal to or greater than 50%.

488

489 To find transcripts with coding regions that intersect TEs, we applied BEDTools v2.28.0 [74] to
490 filter for events where at least 50% of the sequence of one exon derived from a TE with option '-
491 *f 0.5*'. Briefly, exon coordinates were extracted from the StringTie annotation and intersected
492 with our custom TE annotation for *Myotis lucifugus*. To narrow down the list of candidate
493 StringTie transcripts and limit redundant of false positive matches, the output was filtered for
494 multi-exon transcripts with a transcripts per million (TPM) value equal to or higher than 0.5. TPM
495 values were quantified at the isoform level by RSEM v1.3.0 [75]. Filtered StringTie transcript
496 candidates were cross-referenced against the results of differential expression analysis in
497 DESeq2 v1.32 [68] (Table S3) to identify TE exonization events in ISGs. Candidates ISGs were
498 verified by visually inspecting candidates against the DNAzoo and NCBI annotations for *Myotis*
499 *lucifugus* and RNA-seq coverage tracks. To identify TEs that might be contributing to alternative
500 transcriptional start sites (TSS), we used a custom python script to extract the TSS coordinates

from the StringTie annotation and intersected this collapsed file with the TE annotation file as previously described.

Finally, we queried StringTie-derived assembled transcripts against databases of DNA transposons and retrotransposons extracted from the Repbase repository of TE reference sequences [45]. *tblastx* was used according to previously specified parameters and resulting hits were further filtered for alignments greater than or equal to 300bp with a sequence identity greater than or equal to 90%. To identify any transcripts with intact protein-coding sequences of viral origin, we ran *blastx* against i) the gEVE repository of retroviral proteins [76] and ii) a custom database of syncytin proteins collected from the ncbi repository [77]. The output of gEVE blast was filtered for alignments greater than or equal to 200bp with a shared sequence identity greater than or equal to 50%. The output of syncytin blast was filtered for alignments greater than 100bp with shared sequence identity of 50% and above. The same blast analysis for TE sequences and protein databases was carried out on identified open reading frames (ORFs) larger than 50aa found by running the function *usearch* [78] (*usearch -fastx_findorfs -orfstyle 7 -mincodons 16)* on the StringTie assembled transcripts file.

CUT&RUN analysis

Adapters and low quality reads were trimmed using BBDuk v38.05 [63] using options *'ktrim=r k=34 mink=11 hdist=1 tpe tbo qtrim=r trimq=10*'. Trimmed reads were aligned to the myoLuc2_HiC assembly using Bowtie 2 v2.2.9 [79] with options '*--local –very-sensitive-local –no-unal –no-mixed –no-discordant -I 10 -X 700*', and only uniquely mapping reads with a minimum MAPQ of 10 were retained. Fragments aligning to the mitochondrial genome were removed. Trimmed reads independently aligned to the *S. cerevisiae* assembly (GCF_000146045.2) using Bowtie 2 v2.2.9 [79] with options *'--local –very-sensitive-local –no-*

526 *unal –no-mixed –no-discordant –no-overlap –no-dovetail -I 10 -X 700*'. myoLuc2_HiC read
527 depth was normalized according to the number of fragments aligned to the *S. cerevisiae*
528 assembly for each sample, and normalized bigWigs corresponding to read coverage per 1
529 million normalized reads were generated using deepTools v3.0.1 [80,81] for heatmap
530 visualization.
531
532 Peak calling was performed using complete and size subsetted alignment files with MACS2
533 v2.1.1 [82] in a two-step process where separate sets of peaks were called with 1) single-end
534 options '*--format BAM --shift=-75 --extsize=150*' and 2) paired-end option '*--format BAMPE*'. For
535 both modes only peaks with an unadjusted *p*-val < 0.01 were retained. Peaks from each mode
536 were subsequently merged. IgG peaks were subtracted from each pulldown peak set to
537 minimize background. Only the top 20,000 peaks by descending MACS2 peak score were
538 retained for further analysis.
539
540 To identify IFN-inducible CUT&RUN peaks, the top 20,000 peaks for all samples for a particular
541 pulldown (across all replicates, untreated and IFN-stimulated) were concatenated into a single
542 list, and aligned fragments from each individual sample were counted for all peaks using
543 BEDTools v2.28.0 [74]. IFN-inducible peaks were called using DESeq2 v1.26.0 [68], however
544 we were unable to identify peaks that were significantly upregulated in response to IFN with an
545 FDR < 0.10. We therefore took a more relaxed approach, retaining all peaks with an unadjusted
546 *p*-val < 0.10 and log2FC > 0. log2FC values were shrunken using the apeglm function v1.8.0
547 [83] for visualization. Motif analysis was performed using XSTREME v5.4.1 [84] with options '--
548 *minw 6 --maxw 20 --streme-motifs 20 --align center*' querying against the JASPAR CORE 2018
549 vertebrates database [85].
550

To assess the contributions of TEs in regulating the IFN response, we intersected IFN-inducible H3K27ac peaks with all annotated TEs, requiring that all reported TEs are fully contained within the peak. We further characterized the regulatory landscape by identifying STAT1-marked TEs as a function of distance to the nearest ISG (FDR<0.05, log2FC>1.5) transcriptional start site using BEDTools v2.28.0 [74]. To assess family-level enrichment, GIGGLE v0.6.3 [86] was used to create a database of TEs in the myoLuc2_HiC genome using the custom TE database as described above. IFN-inducible H3K27ac peaks were then queried against the TE database. Results were ranked by descending Giggle enrichment score, and enriched TE families were identified according to the odds ratio, Fisher's two-tailed *p*-val, and number of overlaps. The TE heatmaps were prepared by selecting elements within various families that overlapped either IFN-inducible H3K27ac regions or any H3K27ac regions from untreated or IFN conditions. Signal from *S. cerevisiae* spike-in, CPM normalized bigwigs was plotted as heatmaps using deepTools v3.0.1 [80,81].

565 **REFERENCES**

566 1. Gorbunova V, Seluanov A, Kennedy BK. The World Goes Bats: Living Longer and Tolerating
567 Viruses. Cell Metab. 2020;32:31–43.

568 2. Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, et al. A comparison
569 of bats and rodents as reservoirs of zoonotic viruses: are bats special? Proc Biol Sci.
570 2013;280:20122753.

571 3. Schuh AJ, Amman BR, Sealy TK, Spengler JR, Nichol ST, Towner JS. Egyptian rousette bats
572 maintain long-term protective immunity against Marburg virus infection despite diminished
573 antibody levels. Sci Rep. 2017;7:8763.

574 4. Halpin K, Hyatt AD, Fogarty R, Middleton D, Bingham J, Epstein JH, et al. Pteropid bats are
575 confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus
576 transmission. Am J Trop Med Hyg. 2011;85:946–51.

577 5. Swanepoel R, Leman PA, Burt FJ, Zachariades NA, Braack LE, Ksiazek TG, et al.
578 Experimental inoculation of plants and animals with Ebola virus. Emerg Infect Dis. 1996;2:321–
579 5.

580 6. Munster VJ, Adney DR, van Doremalen N, Brown VR, Miazgowicz KL, Milne-Price S, et al.
581 Replication and shedding of MERS-CoV in Jamaican fruit bats (Artibeus jamaicensis). Sci Rep.
582 2016;6:21878.

583 7. Zhou P, Tachedjian M, Wynne JW, Boyd V, Cui J, Smith I, et al. Contraction of the type I IFN
584 locus and unusual constitutive expression of IFN-α in bats. Proc Natl Acad Sci U S A.
585 2016;113:2696–701.

586 8. Shaw AE, Hughes J, Gu Q, Behdenna A, Singer JB, Dennis T, et al. Fundamental properties
587 of the mammalian innate immune system revealed by multispecies comparison of type I
588 interferon responses. PLoS Biol. 2017;15:e2004086.

589 9. De La Cruz-Rivera PC, Kanchwala M, Liang H, Kumar A, Wang L-F, Xing C, et al. The IFN
590 Response in Bats Displays Distinctive IFN-Stimulated Gene Expression Kinetics with Atypical
591 RNASEL Induction. J Immunol. 2018;200:209–17.

592 10. Escalera-Zamudio M, Zepeda-Mendoza ML, Loza-Rubio E, Rojas-Anaya E, Méndez-Ojeda
593 ML, Arias CF, et al. The evolution of bat nucleic acid-sensing Toll-like receptors. Mol Ecol.
594 2015;24:5899–909.

595 11. Ahn M, Cui J, Irving AT, Wang L-F. Unique Loss of the PYHIN Gene Family in Bats
596 Amongst Mammals: Implications for Inflammasome Sensing. Sci Rep. 2016;6:21722.

597 12. Ahn M, Anderson DE, Zhang Q, Tan CW, Lim BL, Luko K, et al. Dampened NLRP3-
598 mediated inflammation in bats and implications for a special viral reservoir host. Nat Microbiol.
599 2019;4:789–99.

600 13. Xie J, Li Y, Shen X, Goh G, Zhu Y, Cui J, et al. Dampened STING-Dependent Interferon
601 Activation in Bats. Cell Host Microbe. 2018;23:297–301.e4.

602 14. Wickenhagen A, Sugrue E, Lytras S, Kuchi S, Noerenberg M, Turnbull ML, et al. A
603 prenylated dsRNA sensor protects against severe COVID-19. Science. 2021;374:eabj3624.

604 15. Frank JA, Feschotte C. Co-option of endogenous viral sequences for host cell function. Curr
605 Opin Virol. 2017;25:81–9.

606 16. Mura M, Murcia P, Caporale M, Spencer TE, Nagashima K, Rein A, et al. Late viral
607 interference induced by transdominant Gag of an endogenous retrovirus. Proc Natl Acad Sci U
608 S A. 2004;101:11117–22.

609 17. Young GR, Yap MW, Michaux JR, Steppan SJ, Stoye JP. Evolutionary journey of the
610 retroviral restriction gene Fv1. Proc Natl Acad Sci U S A. 2018;115:10130–5.

611 18. Blanco-Melo D, Gifford RJ, Bieniasz PD. Co-option of an endogenous retrovirus envelope
612 for host defense in hominid ancestors. Elife [Internet]. 2017;6. Available from:
613 http://dx.doi.org/10.7554/eLife.22519

614 19. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA
615 Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous
616 Retroviruses. Cell. 2015;162:974–86.

617 20. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-
618 Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous
619 Transcripts. Cell. 2015;162:961–73.

620 21. Schmidt N, Domingues P, Golebiowski F, Patzina C, Tatham MH, Hay RT, et al. An
621 influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to
622 stimulate host antiviral immunity. Proc Natl Acad Sci U S A. 2019;116:17399–408.

623 22. Kelly CJ, Chitko-McKown C, Chuong EB. Ruminant-specific retrotransposons shape
624 regulatory evolution of bovine immunity [Internet]. bioRxiv. 2021 [cited 2022 Feb 25]. p.
625 2021.10.01.462810. Available from: https://www.biorxiv.org/content/10.1101/2021.10.01.462810

626 23. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-
627 option of endogenous retroviruses. Science. American Association for the Advancement of
628 Science; 2016;351:1083–7.

629 24. Srinivasachar Badarinarayan S, Shcherbakova I, Langer S, Koepke L, Preising A, Hotter D,
630 et al. HIV-1 infection activates endogenous retroviral promoters regulating antiviral gene
631 expression. Nucleic Acids Res. 2020;48:10890–908.

632 25. van de Lagemaat LN, Landry J-R, Mager DL, Medstrand P. Transposable elements in
633 mammals promote regulatory variation and diversification of genes with specialized functions.
634 Trends Genet. 2003;19:530–6.

635 26. Bogdan L, Barreiro L, Bourque G. Transposable elements have contributed human
636 regulatory regions that are activated upon bacterial infection. Philos Trans R Soc Lond B Biol
637 Sci. 2020;375:20190332.

638 27. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms
639 races. Annu Rev Genet. Annual Reviews; 2012;46:677–700.

640 28. Katzourakis A, Aswad A. Evolution: Endogenous Viruses Provide Shortcuts in Antiviral
641 Immunity. Curr Biol. Elsevier; 2016;26:R427–9.

642 29. Ray DA, Feschotte C, Pagan HJT, Smith JD, Pritham EJ, Arensburger P, et al. Multiple
643 waves of recent DNA transposon activity in the bat, Myotis lucifugus. Genome Res.
644 2008;18:717–28.

645 30. Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, et al. A Helitron
646 transposon reconstructed from bats reveals a novel mechanism of genome shuffling in
647 eukaryotes. Nat Commun. 2016;7:10716.

648 31. Mitra R, Li X, Kapusta A, Mayhew D, Mitra RD, Feschotte C, et al. Functional
649 characterization of piggyBat from the bat Myotis lucifugus unveils an active mammalian DNA
650 transposon. Proc Natl Acad Sci U S A. 2013;110:234–9.

651 32. Pritham EJ, Feschotte C. Massive amplification of rolling-circle transposons in the lineage of
652 the bat Myotis lucifugus. Proc Natl Acad Sci U S A. 2007;104:1895–900.

653 33. Cosby RL, Judd J, Zhang R, Zhong A, Garry N, Pritham EJ, et al. Recurrent evolution of
654 vertebrate transcription factors by transposase capture. Science [Internet]. 2021;371. Available
655 from: http://dx.doi.org/10.1126/science.abc6405

656 34. Skirmuntt EC, Escalera-Zamudio M, Teeling EC, Smith A, Katzourakis A. The Potential Role
657 of Endogenous Viral Elements in the Evolution of Bats as Reservoirs for Zoonotic Viruses.
658 Annual Reviews; 2020 [cited 2022 Feb 9]; Available from:
659 https://www.annualreviews.org/doi/abs/10.1146/annurev-virology-092818-015613

660 35. Skirmuntt EC, Katzourakis A. The evolution of endogenous retroviral envelope genes in bats
661 and their potential contribution to host biology. Virus Res. 2019;270:197645.

662 36. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo
663 assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds.
664 Science. 2017;356:92–5.

665 37. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for
666 automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A.
667 2020;117:9451–7.

668 38. Xiong W, He L, Lai J, Dooner HK, Du C. HelitronScanner uncovers a large overlooked
669 cache of Helitron transposons in many plant genomes. Proc Natl Acad Sci U S A.
670 2014;111:10263–8.

671 39. Tempel S. Using and understanding RepeatMasker. Methods Mol Biol. 2012;859:29–51.

672 40. RepeatMasker Home Page [Internet]. [cited 2022 Apr 12]. Available from:
673 http://www.repeatmasker.org

674 41. Ray DA, Pagan HJT, Thompson ML, Stevens RD. Bats with hATs: evidence for recent DNA
675 transposon activity in genus Myotis. Mol Biol Evol. 2007;24:632–9.

676 42. Jin Y, Tam OH, Paniagua E, Hammell M. TEtranscripts: a package for including
677 transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics.
678 2015;31:3593–9.

679 43. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome
680 assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 2019;20:278.

681 44. Böhne A, Zhou Q, Darras A, Schmidt C, Schartl M, Galiana-Arnoux D, et al. Zisupton—A
682 Novel Superfamily of DNA Transposable Elements Recently Active in Fish. Mol Biol Evol.
683 Oxford Academic; 2011;29:631–45.

684 45. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in
685 eukaryotic genomes. Mob DNA. 2015;6:11.

686 46. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, et al. Syncytin is a captive retroviral
687 envelope protein involved in human placental morphogenesis. Nature. 2000;403:785–9.

688 47. Horie M, Kobayashi Y, Honda T, Fujino K, Akasaka T, Kohl C, et al. An RNA-dependent
689 RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus. Sci
690 Rep. 2016;6:25873.

691 48. Zhang G, Cowled C, Shi Z, Huang Z, Bishop-Lilly KA, Fang X, et al. Comparative analysis of
692 bat genomes provides insight into the evolution of flight and immunity. Science. 2013;339:456–
693 60.

694 49. Pavlovich SS, Lovett SP, Koroleva G, Guito JC, Arnold CE, Nagle ER, et al. The Egyptian
695 Rousette Genome Reveals Unexpected Features of Bat Antiviral Immunity. Cell.
696 2018;173:1098–110.e18.

697 50. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-
698 option of endogenous retroviruses. Science. 2016;351:1083–7.

699 51. Kobayashi KS, van den Elsen PJ. NLRC5: a key regulator of MHC class I-dependent
700 immune responses. Nat Rev Immunol. 2012;12:813–20.

701 52. Benkő S, Kovács EG, Hezel F, Kufer TA. NLRC5 Functions beyond MHC I Regulation-What
702 Do We Know So Far? Front Immunol. 2017;8:150.

703 53. Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, et al. Latent enhancers
704 activated by stimulation in differentiated cells. Cell. Elsevier; 2013;152:157–71.

705 54. Qiao Y, Giannopoulou EG, Chan CH, Park S-H, Gong S, Chen J, et al. Synergistic
706 activation of inflammatory cytokine genes by interferon-γ-induced chromatin remodeling and toll-
707 like receptor signaling. Immunity. 2013;39:454–69.

708 55. Bondet V, Le Baut M, Le Poder S, Lécu A, Petit T, Wedlarski R, et al. Constitutive IFNα
709 Protein Production in Bats. Front Immunol. 2021;12:735866.

710 56. Kepler TB, Sample C, Hudak K, Roach J, Haines A, Walsh A, et al. Chiropteran types I and
711 II interferon genes inferred from genome sequencing traces by a statistical gene-family
712 assembler. BMC Genomics. 2010;11:444.

713 57. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution
714 map of human evolutionary constraint using 29 mammals. Nature. 2011;478:476–82.

715 58. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, et al. The
716 Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with
717 chromosome-length scaffolds for under $1000 [Internet]. bioRxiv. 2018 [cited 2022 Apr 18]. p.
718 254797. Available from: https://www.biorxiv.org/content/10.1101/254797v1

719 59. Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive
720 technique to capture the conformation of genomes. Methods. 2012;58:268–76.

721 60. Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, et al. Squamate
722 reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals.
723 Nat Commun. 2018;9:2774.

724 61. Janssens D, Henikoff S. CUT&RUN: Targeted in situ genome-wide profiling with high
725 efficiency for low cell numbers v3 [Internet]. protocols.io. ZappyLab, Inc.; 2019. Available from:
726 https://www.protocols.io/view/cut-amp-run-targeted-in-situ-genome-wide-profiling-zcpf2vn

727 62. Meers MP, Bryson TD, Henikoff JG, Henikoff S. Improved CUT&RUN chromatin profiling
728 tools. Elife [Internet]. 2019;8. Available from: http://dx.doi.org/10.7554/eLife.46314

729 63. BBMap [Internet]. SourceForge. [cited 2022 Apr 18]. Available from:
730 https://www.sourceforge.net/projects/bbmap/

731 64. Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data
732 [Internet]. [cited 2022 Apr 18]. Available from:
733 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

734 65. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for
735 multiple tools and samples in a single report. Bioinformatics. 2016;32:3047–8.

736 66. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast
737 universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

738 67. CCGB: Miller Lab, LASTZ [Internet]. [cited 2022 Feb 20]. Available from:
739 https://www.bx.psu.edu/~rsharris/lastz/

740 68. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-
741 seq data with DESeq2. Genome Biol. 2014;15:550.

742 69. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit
743 with revamped UIs and APIs. Nucleic Acids Res. 2019;47:W199–205.

744 70. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
745 Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.

746 71. Pertea G, Pertea M. GFF Utilities: GffRead and GffCompare. F1000Res [Internet]. 2020;9.
747 Available from: http://dx.doi.org/10.12688/f1000research.23297.2

748 72. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement
749 TrEMBL in 2000. Nucleic Acids Res. 2000;28:45–8.

750 73. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+:
751 architecture and applications. BMC Bioinformatics. 2009;10:421.

752 74. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features.
753 Bioinformatics. 2010;26:841–2.

754 75. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or
755 without a reference genome. BMC Bioinformatics. 2011;12:323.

76. Nakagawa S, Takahashi MU. gEVE: a genome-based endogenous viral element database provides comprehensive viral protein-coding sequences in mammalian genomes. Database [Internet]. 2016;2016. Available from: http://dx.doi.org/10.1093/database/baw087

77. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res. 2022;50:D20–6.

78. Edgar RC. Search and clustering orders of magnitude faster than BLAST. Bioinformatics. 2010;26:2460–1.

79. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.

80. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 2014;42:W187–91.

81. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44:W160–5.

82. Liu T. Use Model-Based Analysis of ChIP-Seq (MACS) to Analyze Short Reads Generated by Sequencing Protein–DNA Interactions in Embryonic Stem Cells. In: Kidder BL, editor. Stem Cell Transcriptional Networks: Methods and Protocols. New York, NY: Springer New York; 2014. p. 81–95.

83. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35:2084–92.

777 84. Grant CE, Bailey TL. XSTREME: Comprehensive motif analysis of biological sequence
778 datasets [Internet]. bioRxiv. 2021 [cited 2021 Sep 24]. p. 2021.09.02.458722. Available from:
779 https://www.biorxiv.org/content/10.1101/2021.09.02.458722v1

780 85. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al.
781 JASPAR 2020: update of the open-access database of transcription factor binding profiles.
782 Nucleic Acids Res. 2020;48:D87–92.

783 86. Layer RM, Pedersen BS, DiSera T, Marth GT, Gertz J, Quinlan AR. GIGGLE: a search
784 engine for large-scale integrated genome analysis. Nat Methods. 2018;15:123–6.

785

DECLARATIONS

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Newly generated RNAseq and CUT&RUN raw files have been deposited under the GEO SuperSeries accession GSE200833. Processed data, including TE annotation, can be visualized as a UCSC genome browser custom track here: https://genome.ucsc.edu/s/GiuliaPasquesi/myoLuc2_HiC.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by xxx.

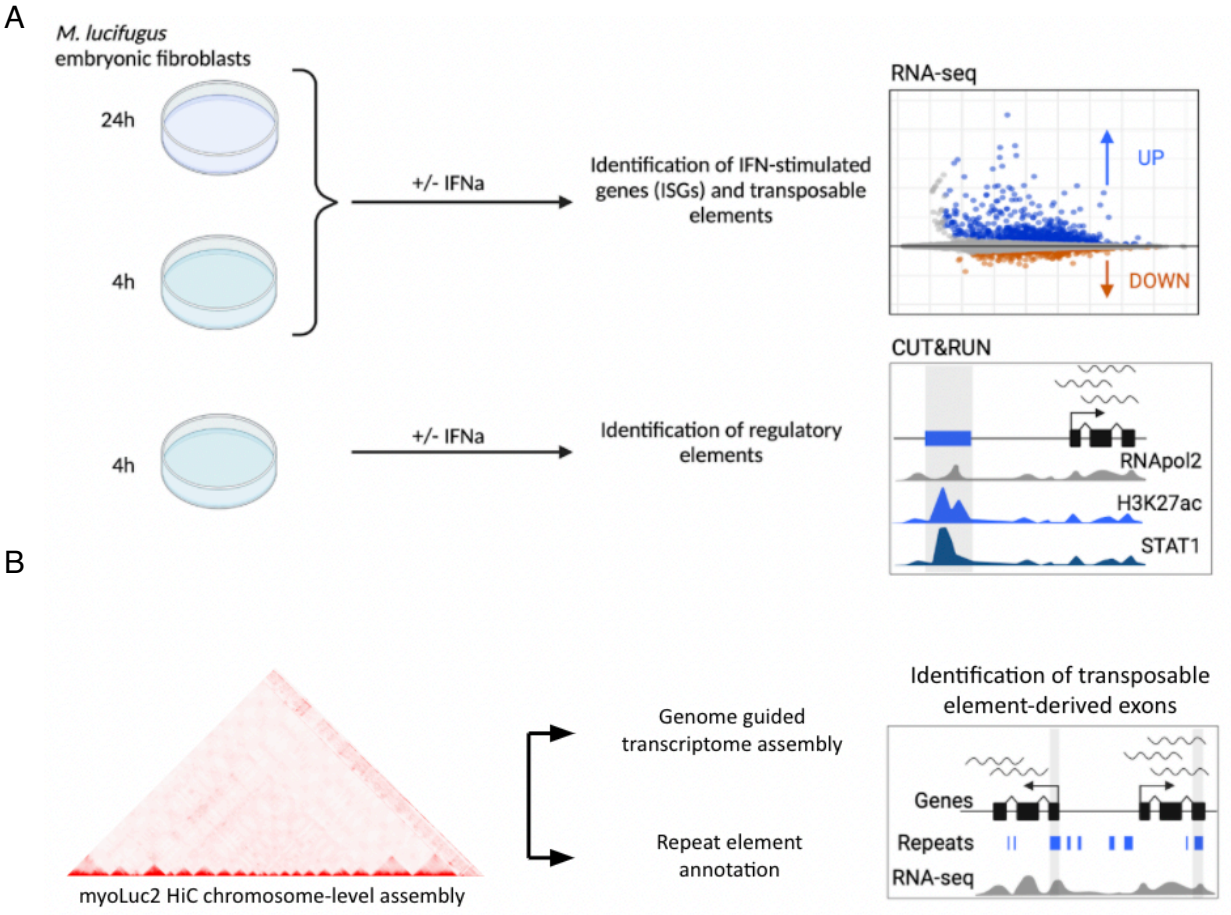
Authors' contributions

EC and GP designed the study. GP and CK performed experiments. GP, CK, AO and EC analyzed data and interpreted results. GP and CK created and edited figures and tables. EC, GP and CK wrote the manuscript with input from all co-authors. All authors gave final approval for publication.

Acknowledgements

Unpublished genome assemblies are used with permission from the DNA Zoo Consortium (dnazoo.org). *Myotis lucifugus* primary embryonic fibroblast cells were a kind gift of Mario Capecchi (University of Utah). We thank the BioFrontiers Computing core for technical support during this study.

817 FIGURES



818

819 **Fig. 1 - Experimental design. A**) *Myotis lucifugus* embryonic fibroblast primary cells were treated for 24
820 hours (24h) and for 4 hours (4h) with 1000U/ml of universal IFNa (+IFNa) or matched volume of DPBS (-
821 IFNa). Total RNA at both time points was extracted and used as input for RNA library preparation and
822 sequencing to identify differentially expressed genes and transposable elements (TEs). To characterize
823 changes in chromatin accessibility upon IFN treatment, cells were similarly treated for 4h, and subjected
824 to the CUT&RUN protocol on H3K27ac, POLR2A, and STAT1. **B**) The chromosome-level genome
825 assembly for *Myotis lucifugus* was used as reference to perform *de novo* repeat element identification
826 and annotation. Combined with genome guided transcriptome assembly of our RNA-seq datasets, the
827 custom repeat element annotation was used to identify TE-derived and virus-derived isoforms and
828 transcription start sites (TSS).
829

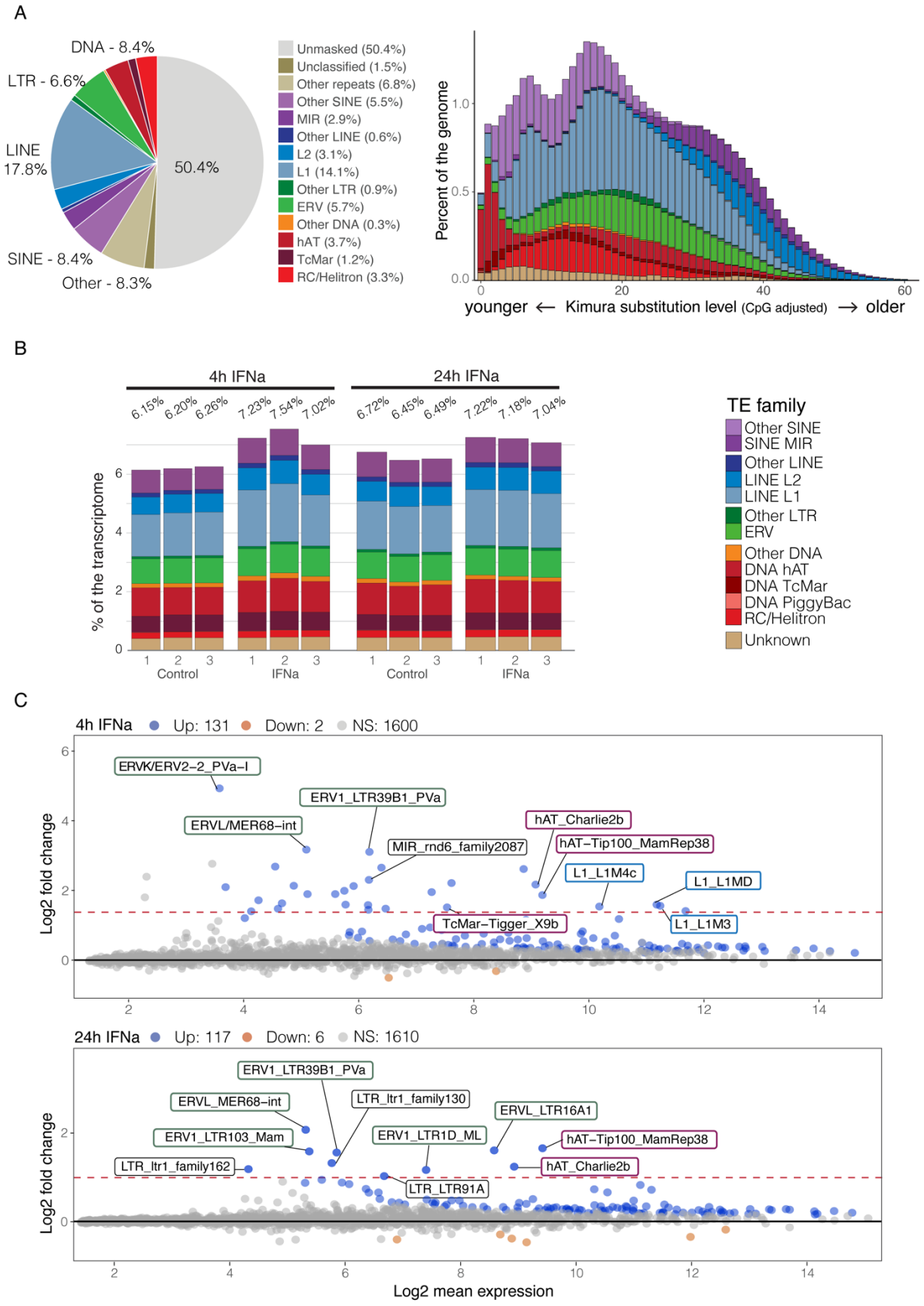
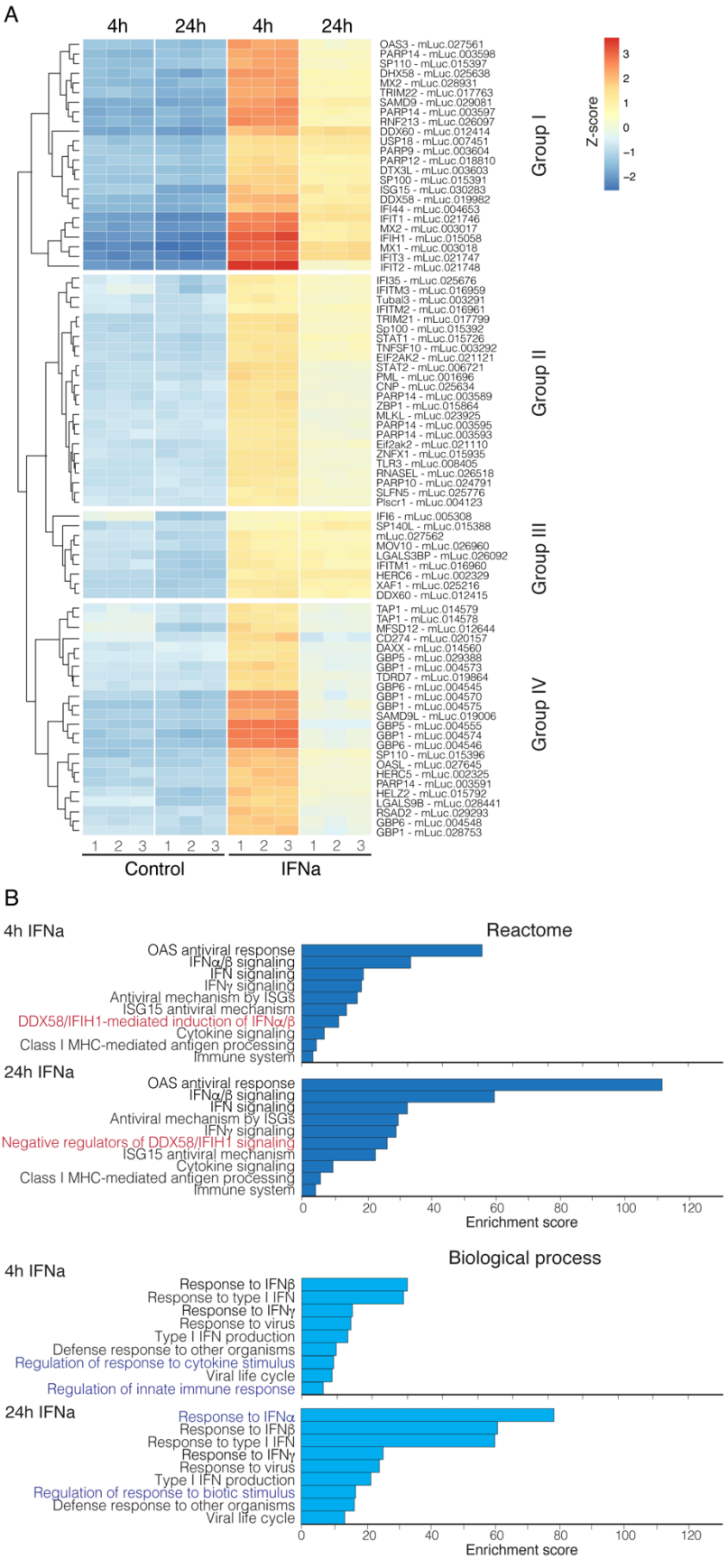


Fig. 2 - Repeat element composition, evolutionary dynamics and transcriptional profiles. Repeat elements were annotated by combining *de novo* identification and homology-based searches. **A**) Pie chart shows the relative abundance of main TE families (42.7% total) as percentage of the genome; histogram shows the composition as percentage of the genome of major TE superfamilies as a function of the divergence (Kimura2Distance) from the reference consensus sequence of each TE. Given the correlation between divergence from the consensus and time of transposition, the lower the K2D (left) and the younger a TE is. As previously identified in Myotis and other bat species, we recovered recent expansion and activity of multiple DNA elements, in particular Helitrons and more recently hAT elements. **B**) Histograms show expression levels of major TE families as a fraction (%) of normalized read counts from RNA sequencing data of IFNa-stimulated and unstimulated cells at 4h and 24h post treatment. **C**) MAplots of apeglm [83] transformed data show differentially expressed TEs at 4h (top) and 24h (bottom) post IFN treatment. For the 4h time point, only the top three TEs per family with the highest Log2 fold change were labelled. For both time points, only TEs that met a threshold of adjusted p-value < 0.05 and log2 fold change > 1.5 (accepted fdr = 0.05) were labelled. Counts of upregulated (blue dots) and downregulated (orange dots) TEs are based on adjusted p-value (<0.05; accepted fdr = 0.05) only. Among induced TEs at 4h (45 total based on our cutoffs) we recovered for the most part ERV retrotransposons (green outline), DNA hAT transposons (red outline) and L1 LINEs (light blue outline). At 24h post treatment we only found 10 families that met the filtering criteria, for the majority ERV shared with the 4h dataset.



851

852 **Fig. 3 - Gene expression profiles and enrichment analyses. A**) Heatmap shows the 140 genes with
853 highest expression variance across samples in vsd transformed data (clustering method: “euclidean”).
854 Embryonic fibroblast cells presented four different profiles of gene induction dynamics through time: i)
855 genes with rapid induction and decline at 24h (Group I); ii) genes with mild induction at 4h and low decline
856 at 24h or with stable induction (Groups II and III); and iii) genes with rapid induction at 4h and rapid
857 decline (Group IV). **B**) Bar graphs show the results of functional overrepresentation analysis (ORA) on
858 differentially expressed genes at 4h and 24h post IFNa treatment. Although most terms are shared
859 between the two time points, we found the DDX58/IFIH1 pathway to be active at 4h post IFNa stimulation,
860 whereas it is subject to negative regulation at 24h. Similarly, the response to cytokine stimuli is present at
861 4h post treatment, but not at 24h.
862

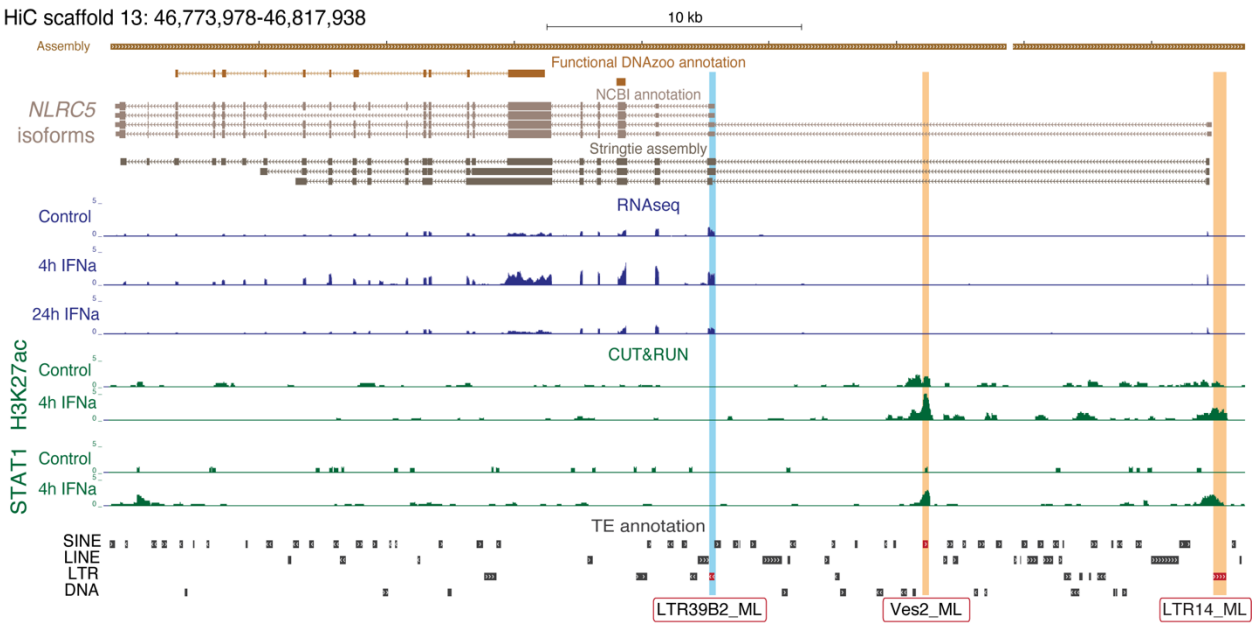
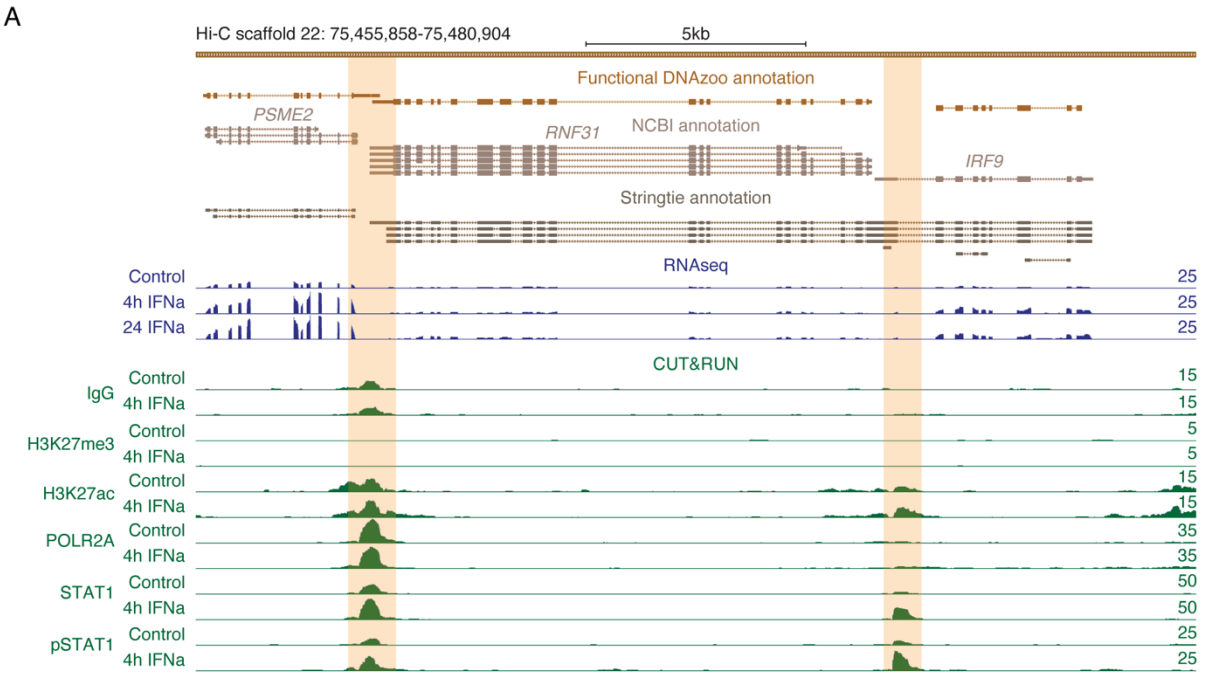
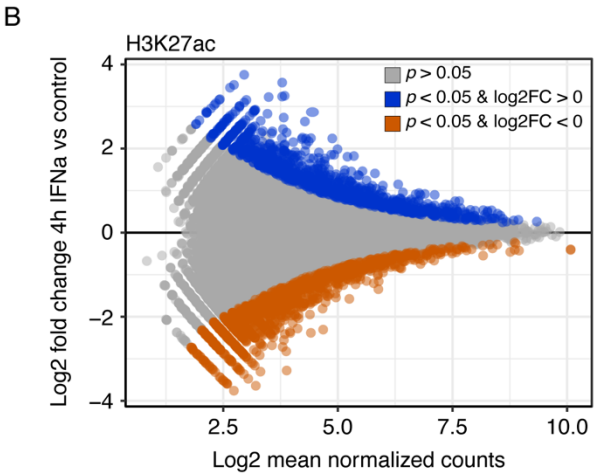
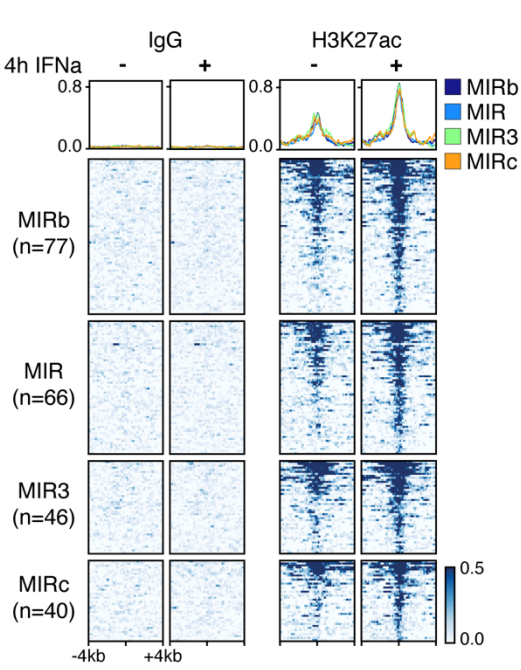
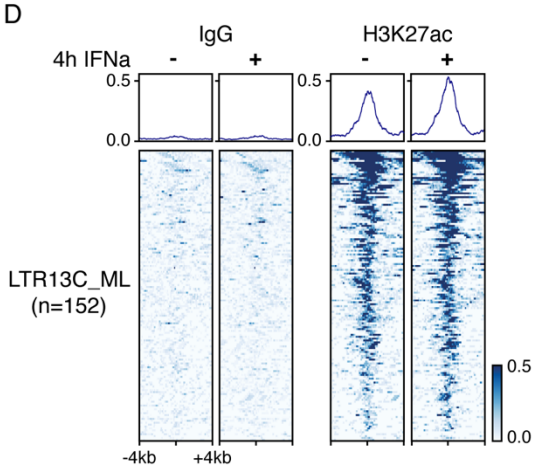


Fig. 4 - TE exonization events in *Myotis* ISGs. Custom UCSC genome browser screenshot of the NLRC5 locus, where one exon (light blue highlight) represents a potential alternative transcription start site (TSS) deriving from a Myotis-specific LTR39B2_ML retrotransposons. RNAseq coverage at the promoter region suggests upregulation of the transcript at 4h post IFNa treatment, consistent with its role in immune responses, and lower expression in 24h post treatment samples compared to unstimulated cells. We also identified 2 potential TE-derived regulatory elements (orange highlight) in intronic or upstream regions of the NLRC5 locus that show increase in H3K27ac and STAT1 CUT&RUN signal at 4h post IFNa treatment.









875

Fig. 5 - Epigenomic profiling of untreated and IFNa-stimulated embryonic fibroblast primary cells. **A)** Genome browser view of the *IRF9* and *PSME2* loci. RNA-seq and CUT&RUN tracks are normalized per million reads. Signal track maxima are indicated on the right of each track. IFNa-inducible (*p*-val < 0.10, log2FC > 0) STAT1 peaks are highlighted orange. **B)** MA plot of IFNa-inducible (unadjusted *p*-val < 0.10, log2FC > 0, blue) and IFNa-repressed (unadjusted *p*-val < 0.10, log2FC < 0, orange) H3K27ac regions from H3K27ac CUT&RUN. Regions with an unadjusted *p*-val > 0.10 are shown in grey. Log2 fold change values were shrunken using the apeglm function v1.8.0 [83] .**C)** Heatmaps showing normalized CUT&RUN signal (signal per million reads) over IFNa-inducible (unadjusted *p*-val < 0.10), H3K27ac-marked MIR families. **D)** Heatmaps showing normalized CUT&RUN signal (signal per million reads) over H3K27ac-marked LTR13C_ML families.

887 TABLES
888

	TE ID and Location	
Gene	**TE**	**Location**
PARP9	L1_Canid	Exon
GBP1	L1M5	3'UTR
DDX60	MER92B	3'UTR
NLRC5	Multiple (3 TEs)	Alternative TSS, regulatory elements
CS012	LTR10A_ML	3'UTR
XAF1	L1MD	3'UTR
SLFN5	Ves	E2 (5'UTR)
IFITM2	LTR1D_ML	3'UTR
PLAT3	Ves2_ML	Last Exon + 3'UTR
PLAT3	MLT1M	E2
SP140L	L1-2a_EF	E7 - 3'UTR
SP140L	LTR18C_ML	E6 - 3'UTR

889
890 **Table 1 - List of candidate genes with TE-derived exons.**
891 Transposable element (TE) exonization events were predicted by intersecting coordinates of annotated
892 TEs and de-novo transcriptome assembly of RNA-seq data for *Myotis lucifugus*. Each candidate was
893 inspected in our custom UCSC genome browser track
894 (https://genome.ucsc.edu/s/GiuliaPasquesi/myoLuc2_HiC) and selected if it showed multiple lines of
895 support (RNA expression, chromatin profile, additional gene/isoform annotation).
896
897