

RESEARCH

Open Access



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

Giulia Irene Maria Pasquesi, Conor J. Kelly, Andrea D. Ordonez and Edward B. Chuong*

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 *NLRCS5*, *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 [1, 7]. 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

*Correspondence: edward.chuong@colorado.edu

BioFrontiers Institute and Department of Molecular, Cellular & Developmental Biology, University of Colorado Boulder, 596 UCB, Boulder, CO 80309, USA



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

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.

TEs are widely speculated to have been important contributors to the evolution of bats [15–19], including bat-specific immune functions [20, 21]. 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 [15–18]. 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 (ERVs), LINEs, and SINEs. Studies in other mammalian lineages have repeatedly shown 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 [22] in ruminants [23], rodents [24], and primates [25]. TE-derived non-coding transcripts can readily form immunostimulatory double-strand RNAs or DNAs [26–28]. Finally, TEs can regulate interferon-inducible gene expression by acting as regulatory elements [29–33]. 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]. This example further supports that the recurrent independent co-option of TEs for immune functions may reflect their capacity to fuel adaptation by increasing genomic variation, especially in the context of host-pathogen coevolutionary arms races [27, 28].

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.

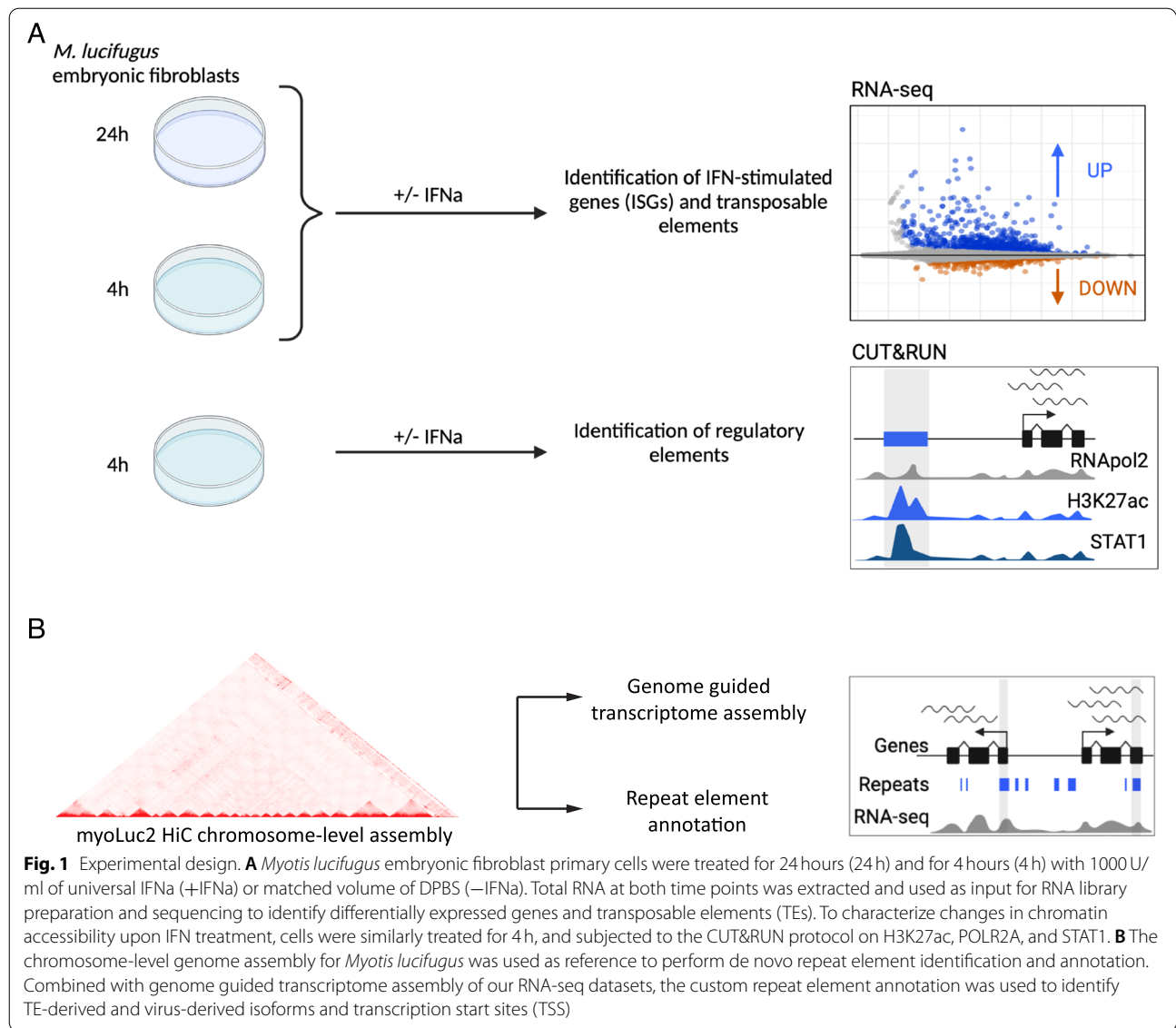
Results

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 (IFN α), and profiled the transcriptome at 0, 4 and 24h time points using RNA-seq. We confirmed cellular response to universal IFN treatment using qPCR to measure the expression of 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) [34, 35], which was the most contiguous assembly available (Scaffold N50 of ~95.5 Mb).

Prior to analyzing our functional genomic data, we performed de novo repeat identification on the myoLuc2.0_HiC assembly using RepeatModeler2 [36] and HelitronScanner [37], followed by repeat annotation using RepeatMasker [38, 39]. 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>; Fig. S2). 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 [15], 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) [40]. Compared to the previous annotation, our annotation included more species-specific Helitron and LINE1 TEs (2.3 and 1.2% more, respectively), and assigned an additional 5.3% of the genome to “Other Repeats”, a classification that groups together satellite and low complexity repeat sequences. This increase in masked repeats is congruent with the better quality of the genome assembly used in our study.

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 [42] (Fig. 2B, Fig. S3). 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. These percentages are slightly higher than what is observed in other mammals with similar TE genomic content (i.e., mouse and human, [43]). However, TE expression is highly variable between tissues and cell types for the same



(See figure on next page.)

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 [41] 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

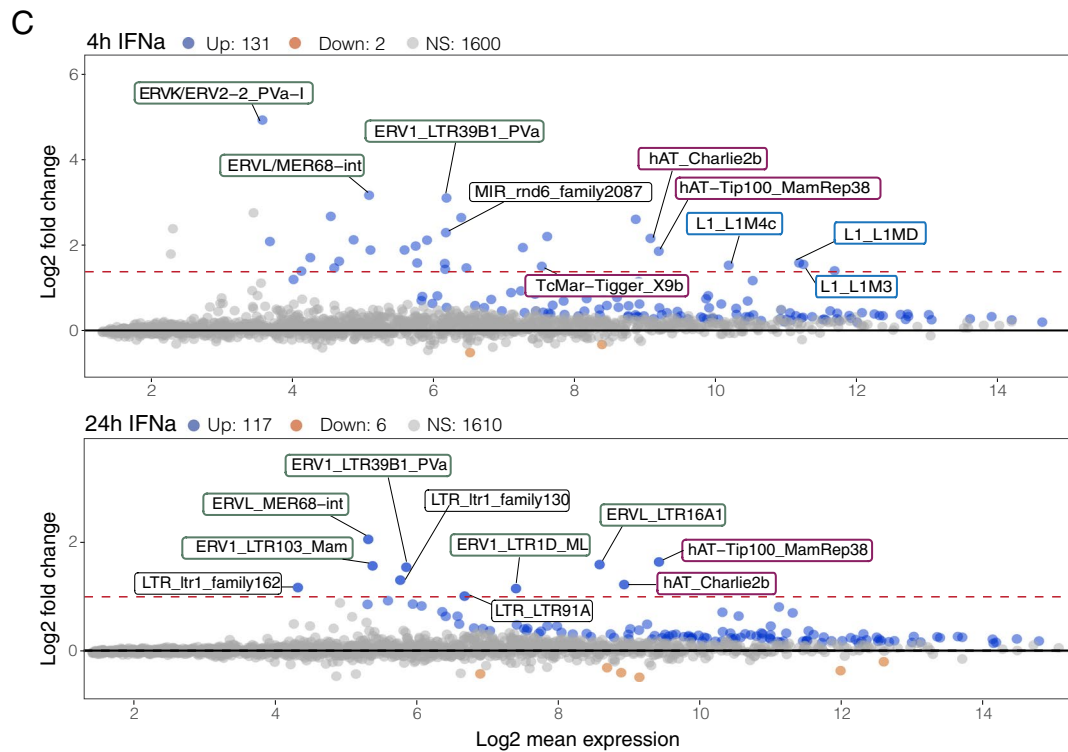
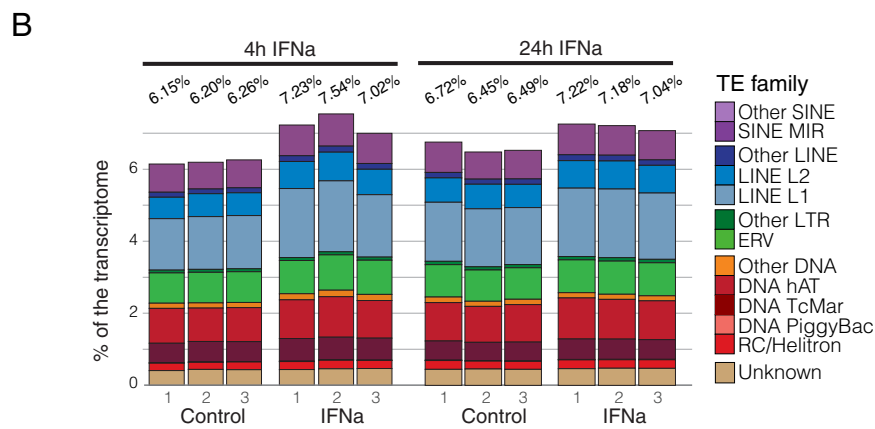
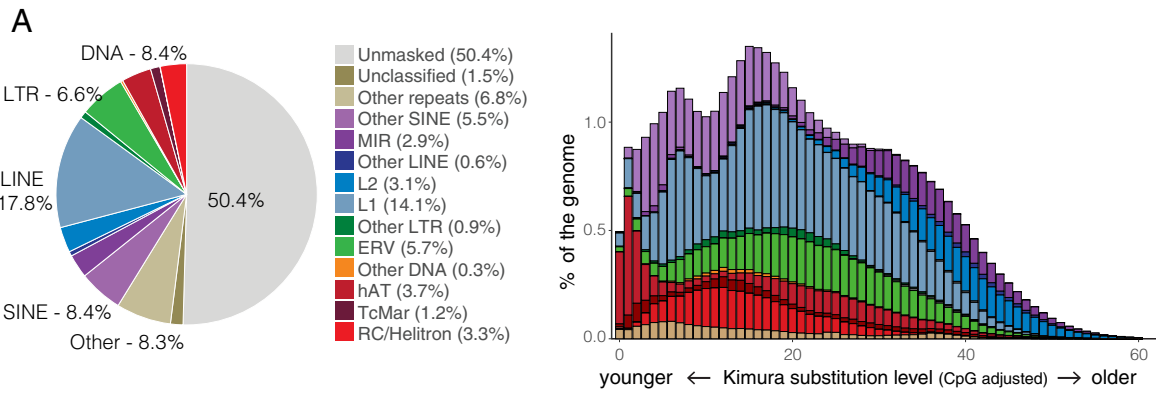


Fig. 2 (See legend on previous page.)

organism, and what we observed in primary cultured fibroblast may not reflect physiological TE transcription profiles, as TE transcription levels are affected by chromatin accessibility and TE regulatory mechanisms [43]. The most abundant TE-derived transcripts we identified included L1 LINEs, DNA/hAT elements, ERVs, SINEs and L2 LINEs. Forty-five TEs showed significant family-level transcriptional induction at 4 h (adj. p -value < 0.05; $\log_2FC > 1.5$), and 8 families induced at 24 h according to the same cutoff thresholds (Fig. 2C). These included multiple ERV families (21), L1 LINEs (6) and DNA transposons (6) at 4 h post treatment, and ERVs (5) at 24 h post IFN treatment. These findings indicate that multiple bat-specific TEs show family-level transcriptional induction in response to IFN treatment, peaking at 4 h and diminishing but still present at 24 h post induction.

As previously noticed [43], the TE transcriptional landscape recapitulates genomic TE content. When only clade specific TEs were analyzed (Fig. S3, see [Methods](#)), we found similar trends where transcriptome compositions broadly summarized the genomic landscape of TEs of similar age, with Helitrons, other SINEs and L1 LINEs being the most abundantly transcribed TEs. It is noteworthy that recent TEs (defined as TEs with a divergence from the consensus sequence less or equal 20; see [Methods](#)) make up on average 50% of the transcribed TEs, with the relationship holding true for ERVs, L1 LINEs and most of DNA elements. Since Helitrons expanded in bat genomes at the beginning of the vesper bat radiation [18], we found no changes in the transcription levels of Helitrons when analyzing only recent TEs. Similarly, SINEs are dominant in the recent transcriptome having expanded in the genome more recently. An exception to this is represented by MIR SINEs, that are mostly ancient TEs and therefore highly represented in the total TE transcriptome. Compellingly, the “other SINEs” represent almost 30% of the recent TE transcriptome (Fig. S3), but they only make up a more marginal fraction of the genome (Fig. 2). This might be indicative of recent expansion under strong purifying selection (i.e., [44]), or support the hypothesis of a stochastic transcription model of the genome [43, 45]. Additionally, we found only minimal induction of recent TEs following IFN treatment. This result is in sharp contrast with what was observed when the entire TE transcriptome was analyzed. It is

possible that the patterns here detected underlie different constraints acting specifically to restrain excessive transcription of more recent and potentially transposition-competent TEs in the context of highly regulated physiological processes.

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 4 h and at 24 h 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 -value < 0.05 and $\log_2FC > 1.5$, we identified 213 upregulated transcripts (corresponding to 138 unique genes) and 1 downregulated gene at 4 h, and 91 upregulated transcripts (corresponding to 66 unique genes) at 24 h 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 4 h that declines at 24 h; II) transcripts showing mild induction at 4 h and decline at 24 h; III) transcripts showing mild, stable induction; and IV) transcripts showing strong induction at 4 h and rapid decline to levels similar to unstimulated cells at 24 h. Both 4 h and 24 h 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 IFN α /b at 4 h, followed by induction of negative regulators of this pathway at 24 h (Fig. 3B). Similarly, an enrichment for genes involved in response to cytokine stimuli was detected at 4 h but not at 24 h post treatment. These observations of a strong early response followed by a decline by 24 h upon IFN stimulation are consistent with observations in the black flying fox (*Pteropus alecto*) [9].

TE contribution to ISG transcript structure

To improve detection of potentially unannotated IFN-induced TE-derived transcripts and isoforms, we conducted genome-guided transcriptome reassembly on our combined RNA-seq dataset using StringTie2 [46] (Supplementary Data 1), and annotated assembled genes based on homology using the SwissProt database. This yielded an expanded transcriptome with 68,110 StringTie-assembled transcripts, with 32,137 transcript

(See figure on next page.)

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

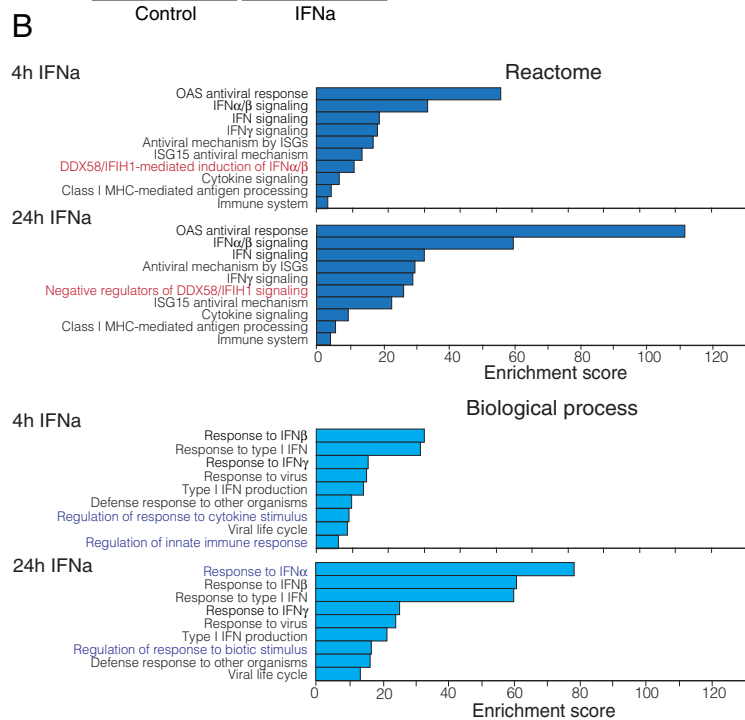
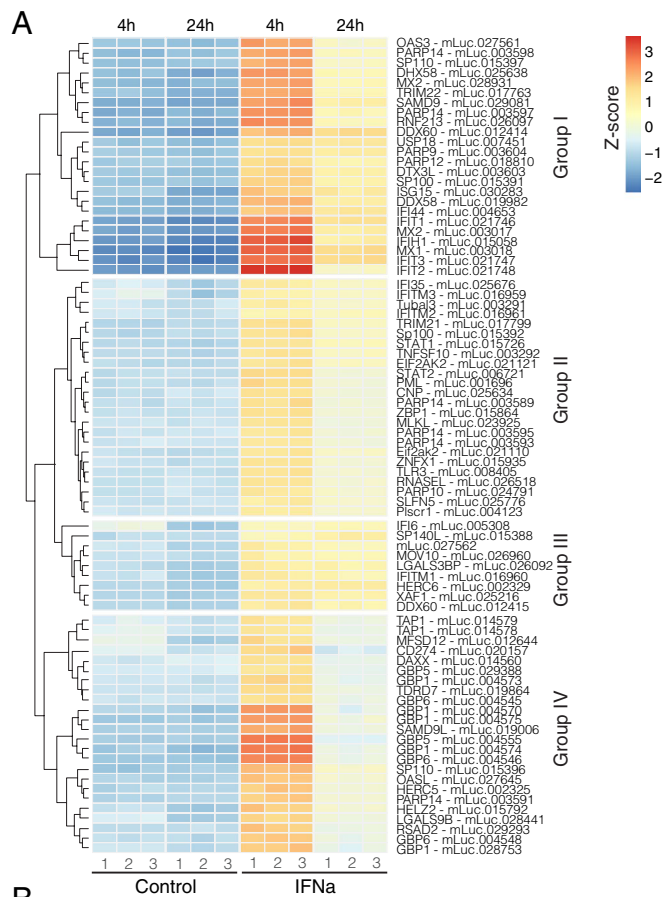


Fig. 3 (See legend on previous page.)

isoforms matching a likely protein-coding gene with an ortholog in Swissprot, and the rest most likely corresponding to long non-coding RNAs and bat-specific genes. After performing pairwise differential expression analyses (Table S3) we identified 1243 IFN-inducible transcripts corresponding to 836 StringTie genes (740 transcripts, corresponding to 449 genes which had a homologous match in Swissprot) at 4h, and 717 transcripts corresponding to 500 StringTie genes (385 transcripts, corresponding to 239 genes matching Swissprot) at 24h. Of these, 606 transcripts corresponding to 392 genes (358 transcripts, 214 genes matching Swissprot) were shared between 4h and 24h treatment time.

We used our improved transcriptome reassembly to investigate the contribution of TEs to both constitutive and IFN-inducible transcript structures. First, we identified transcripts that contained exonized TEs based on the overlap of exon (>50% sequence length) and annotated TE features. Considering all expressed (TPM \geq 0.5) multi-exon transcripts, we found a subset of 1039 transcripts corresponding to 749 StringTie genes (648 transcripts and 470 genes matching Swissprot) that contained at least one TE-derived exon (Table S4).

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. Zisupton DNA transposons are not abundant in the *M. lucifugus* genome (our masking recovered only one family with 174 copies). It is possible that they were horizontally transferred into bat genomes as has happened in multiple fish species [47]. Age analysis of Zisupton genomic copies suggests that they were likely recently introduced (Kimura 2 distance from the consensus sequence of \sim 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. We further verified this finding through multiple BLAST searches (Supplementary Data 2). First, we aligned the *EEF1A1* StringTie transcript (STRG.21546.2, Scaffold 10: 12654609-12,761,039) 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 [48] 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. S4A) and a candidate novel isoform of *IFITM2* that has not been previously identified (Fig. S4B). 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 [49];) which may not be masked by RepeatMasker due to their age or low copy number. 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 [50]. 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 (RV1)-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 *NLRCS5* (Fig. 4), *EIF2AK2*, *GBP1*, *MX1*, *MX2*, *PHF11*, *SAMD9* and *XAF1*, while others like *PARP14* and *CS012* are

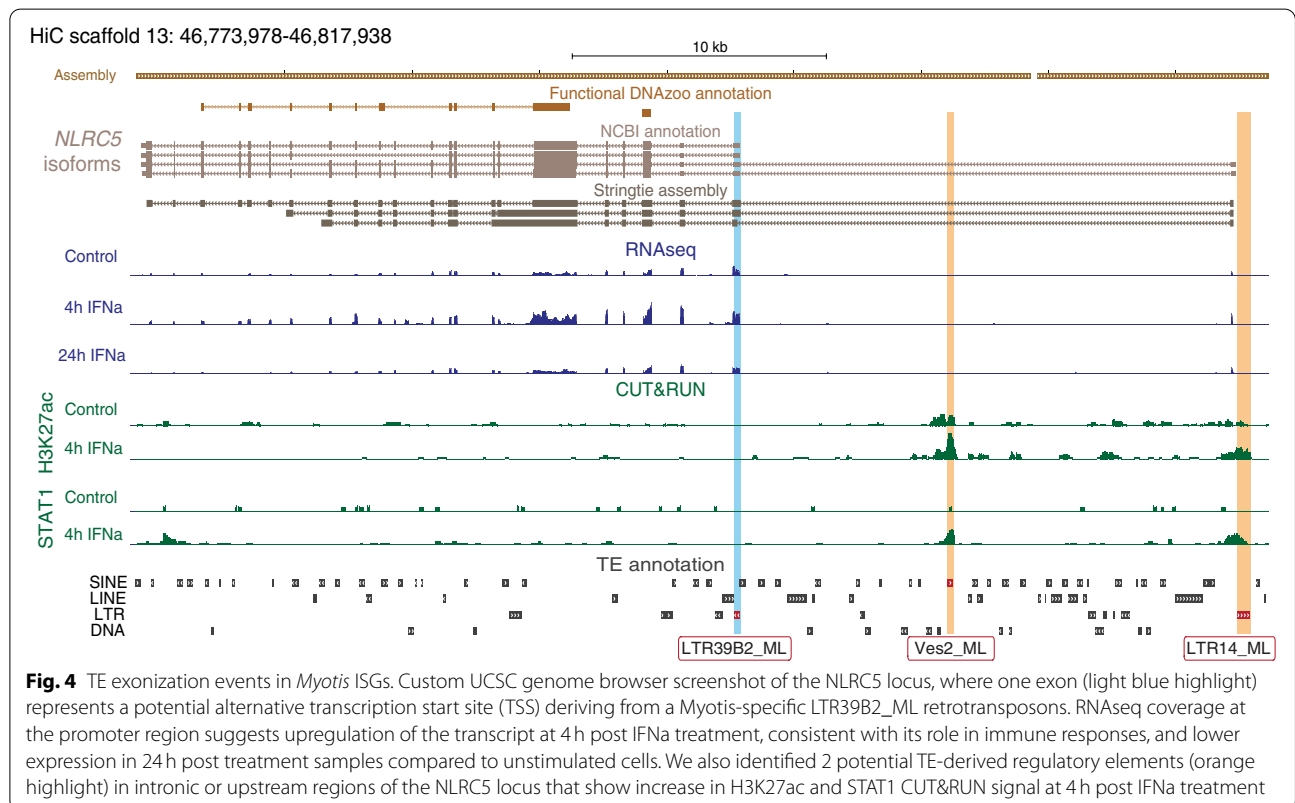
Table 1 List of candidate genes with TE-derived exons

| Gene | TE ID and Location | | |
|--------|--------------------|--------------------------------------|-------------------------------------------|
| | TE | Location | Type of exon |
| PARP9 | L1_Canid | Exon | constitutive |
| GBP1 | L1M5 | 3'UTR | StringTie alternative; ncbi constitutive |
| DDX60 | MER92B | 3'UTR | constitutive |
| NLRCS | Multiple (3 TEs) | Alternative TSS, regulatory elements | constitutive |
| CS012 | LTR10A_ML | 3'UTR | StringTie alternative; ncbi constitutive |
| XAF1 | L1MD | 3'UTR | constitutive |
| SLFN5 | Ves | E2 (5'UTR) | constitutive - truncated in some isoforms |
| IFITM2 | LTR1D_ML | 3'UTR | StringTie constitutive (missing in ncbi) |
| PLAT3 | Ves2_ML | Last Exon + 3'UTR | constitutive |
| PLAT3 | MLT1M | E2 | constitutive |
| SP140L | L1-2a_EF | E7 - 3'UTR | alternative |
| SP140L | LTR18C_ML | E6 - 3'UTR | alternative |

Transposable element (TE) exonization events were predicted by intersecting coordinates of annotated TEs and de-novo transcriptome assembly of RNA-seq data for *Myotis lucifugus*. Each candidate was inspected in our custom UCSC genome browser track (https://genome.ucsc.edu/s/GiuliaPasquesi/myoLuc2_HiC) and selected if it showed multiple lines of support (RNA expression, chromatin profile, additional gene/isoform annotation)

non-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 4 h and 7 transcripts (corresponding to 3 annotated genes) at 24 h 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 [51, 52].



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 [29]. Unexpectedly, we did not observe robust IFN-inducible chromatin changes that are characteristic of IFN-stimulated cells from other species [29, 30]. We did not identify any H3K27ac inducible elements with a false discovery rate (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-value* < 0.1 (Fig. S5A). 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 *NLRCS* locus in addition to an intronic Ves2_ML SINE element (Fig. 4). However, in contrast to previous studies in other species [29–33], we did not observe any overrepresented TE families within this set (Fig. S5B; 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. S5C). 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. S6). Taken together, our analysis indicates that TEs have contributed

hundreds of regulatory elements involved in IFN signaling, but in contrast to studies in other species, we did not identify genome-wide 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 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 (4h 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

(See figure on next page.)

Fig. 5 Epigenomic profiling of untreated and IFN α -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. IFN α -inducible (*p-value* < 0.10, log₂FC > 0) STAT1 peaks are highlighted orange. **B** MA plot of IFN α -inducible (unadjusted *p-value* < 0.10, log₂FC > 0, blue) and IFN α -repressed (unadjusted *p-value* < 0.10, log₂FC < 0, orange) H3K27ac regions from H3K27ac CUT&RUN. Regions with an unadjusted *p-value* > 0.10 are shown in grey. Log₂ fold change values were shrunk using the apeglm function v1.8.0 [41]. **C** Heatmaps showing normalized CUT&RUN signal (signal per million reads) over IFN α -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

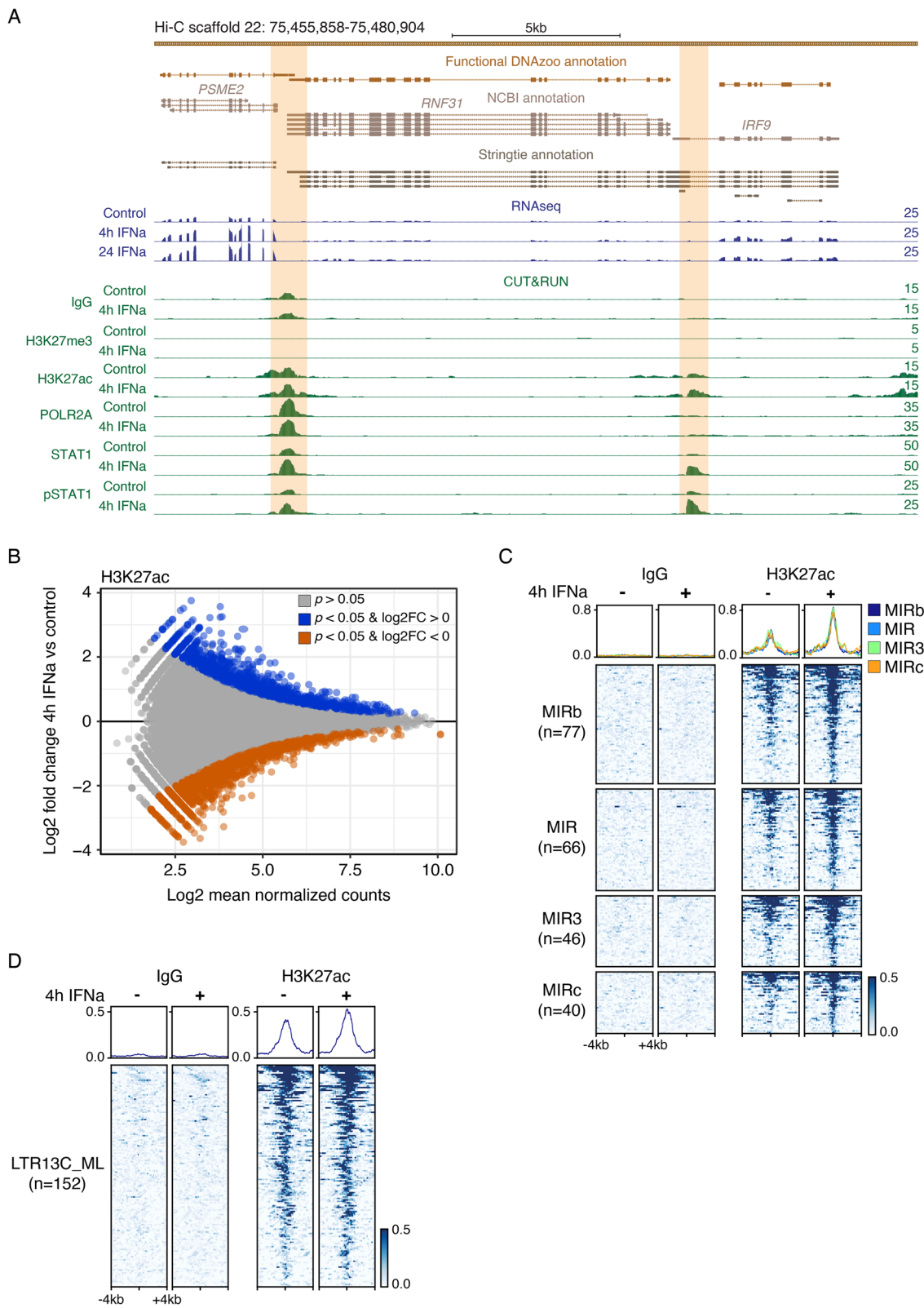


Fig. 5 (See legend on previous page.)

both in comparison to unstimulated and 24 h 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 [26, 27]. Indirect support for the hypothesis of a potential involvement of TEs in innate immunity comes from the observation that whereas the total TE transcriptome follows in composition the TE genomic content, this is not true when it comes to differentially expressed TE families. Since our analyses cannot discriminate TE expression at the level of individual loci, it is possible that only few members of a given TE family are overexpressed, but the family as a whole is not. Altogether, it is likely that IFN-responsive sequence-specific features (i.e., IFN-responsive promoters) exist in *Myotis*.

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 [53], and may be involved in the regulation of inflammasome activation and type I IFN responses [54]. 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. 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 [29, 55, 56]. As a result, we did not observe strong evidence supporting family-level TE regulatory induction as observed previously in other species such as human [30] and cow [29], 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, 57]. 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. S7) [58], whereas no IFNa genes were identified. Of the 11 IFNw paralogues, 5 showed evidence of constitutive low expression in unstimulated cells, and induction at 4 h post treatment (Fig. S7). 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. We performed de novo TE identification using RepeatModeler2 (included in dfamtools-1.1) [36] and HelitronScanner v1.1 [37] to match the newly released, highly contiguous chromosome-level assembly of the little brown bat genome (myoLuc2_HiC) [34, 35, 59, 60]. Genome assemblies that rely on long read assembly strategies (i.e., HiC [34, 61]) 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 [39]. To maximize element identification we followed a custom multi-step mapping strategy [62] 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 [19]; (iii) our de novo little brown bat library; (iv) the entire tetrapoda Repbase library provided with RepeatMasker [48].

Cell lines and treatment

Myotis lucifugus primary embryonic fibroblast cells were a gift from Mario Capocchi (University of Utah). Cells were grown at 37°C and 5% CO₂ 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 2 ml of culturing media (or 2×10^6 cells per 15 cm dish for CUT&RUN). The following day (or 48 h for CUT&RUN) cells were treated with 1000 U/ml of Universal Type I IFN α resuspended in DPBS (PBL Assay Science #11200) in 2 ml of culturing media; control cells were treated with equivalent volume of DPBS in 2 ml of media. At 4 and 24 h post treatment cells were harvested for RNA extraction (4 h for CUT&RUN).

RNA isolation and library preparation for RNA-seq

Following media removal, cells were washed with 1 ml of DPBS and detached by adding 400 μ l of 0.25% trypsin per well. Following a 10 minutes incubation at 37°C, trypsin was neutralized with 1.6 ml of culturing media. Cell suspensions were transferred into 1.7 ml tubes and pelleted

by centrifugation at 300 \times g for 5 minutes. Cells were then lysed in 300 μ l 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 150 bp paired-end reads.

CUT&RUN sample and library preparation

CUT&RUN pulldowns were generated using a protocol from [63, 64]. All buffers were prepared according to the "High Ca²⁺/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 [63, 64]) 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 [63, 64]. Yeast spike-in DNA (gift from Steven Henikoff) was added to the quenching ("1 \times 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 μ l) was used to generate libraries according to the manufacturer's protocol with the following modifications. Freshly diluted 0.200 μ M 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 150 bp paired-end reads.

Paired RT-qPCR

5×10^5 viable cells from the same CUT&RUN populations (untreated and 4 h 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 QuickRNA 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 $2\times$ Luna Universal One-Step Reaction Mix, 0.5ul $20\times$ 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 (BioRad) with the following PCR cycling conditions: 10 min at 55°C , 1 min at 95°C , 40x(10s at 95°C , 30s 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 $p\text{-value} < 0.05$.

Transcriptome analyses

Paired-end 150bp read length FASTQ files were quality and adapter trimmed using BBDuk v.38.05 [65]; quality check was performed using FastQC v0.11.8 [66] and inspected through MultiQC v1.7 [67]. Filtered FASTQ files were then mapped to the myoLuc2_HiC genome using a 2-pass approach in STAR v2.7.3a [68]. 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 DNAAZoo, 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 [69]. 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. To analyze clade-specific TE expression we filtered our custom TE annotation file to include only TEs with a Kimura2D ≤ 20 . Pairwise differential expression analyses at 4h and 24h post IFN treatment were performed in DESeq2 v1.32 [70]. Functional enrichment analyses of differentially expressed genes (adj. $p\text{-value} < 0.05$, $\log_2\text{FC} > 1.5$) were performed using the WebGestalt web tool [71].

Genome guided transcriptome assembly and analysis

Short-read RNA-seq alignment files generated by running STAR (see previous paragraph) were merged, sorted and indexed using SAMtools v1.10 [72], and the resulting .bam file was used as input for genome guided transcriptome assembly in StringTie v1.3.3b [46]. StringTie was run following default parameters, except that the minimum number of spliced reads required to align across a junction was increased from 1 to 5 using option '*-j 5*'. The resulting StringTie gtf output (Supplementary Data 1) was then converted to FASTA format using the *gffread* utility using options '*-M, -E, and -Z*' included in Cufflinks v2.2.1 [73]. We followed an homology-based approach to annotate assembled StringTie genes and isoforms. StringTie gene sequences were queried against the SwissProt database [74] through the *blastx* search algorithm in BLAST 2.12.0+ [75] using options '*-max_target_seqs 1 -evalue 10*'. Matches were then filtered for shared sequence identity equal to or greater than 50%.

To find transcripts with coding regions that intersect TEs, we applied BEDTools v2.28.0 [76] to filter for events where at least 50% of the sequence of one exon derived from a TE with option '*-f 0.5*'. Briefly, exon coordinates were extracted from the StringTie annotation and intersected with our custom TE annotation for *Myotis lucifugus*. To narrow down the list of candidate StringTie transcripts and limit redundant of false positive matches, the output was filtered for multi-exon transcripts with a transcripts per million (TPM) value equal to or higher than 0.5. TPM values were quantified at the isoform level

by RSEM v1.3.0 [77]. Filtered StringTie transcript candidates were cross-referenced against the results of differential expression analysis in DESeq2 v1.32 [70] (Table S3) to identify TE exonization events in ISGs. Candidates ISGs were verified by visually inspecting candidates against the DNazoo and NCBI annotations for *Myotis lucifugus* and RNA-seq coverage tracks. To identify TEs that might be contributing to alternative 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 [48]. *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 [78] and ii) a custom database of syncytin proteins collected from the ncbi repository [79]. 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* [80] (*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 [65] 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 [81] 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 [81] with options '*--local -very-sensitive-local -no-unal -no-mixed -no-discordant -no-overlap -no-dovetail -I 10 -X 700*'. myoLuc2_HiC read depth was normalized according to the number of fragments aligned to the *S. cerevisiae* assembly for each sample, and normalized

bigWigs corresponding to read coverage per 1 million normalized reads were generated using deepTools v3.0.1 [82, 83] for heatmap visualization.

Peak calling was performed using complete and size subsetted alignment files with MACS2 v2.1.1 [84] in a two-step process where separate sets of peaks were called with 1) single-end options '*--format BAM --shift=-75 --extsize=150*' and 2) paired-end option '*--format BAMPE*'. For both modes only peaks with an unadjusted *p-value* < 0.01 were retained. Peaks from each mode were subsequently merged. IgG peaks were subtracted from each pulldown peak set to minimize background. Only the top 20,000 peaks by descending MACS2 peak score were retained for further analysis.

To identify IFN-inducible CUT&RUN peaks, the top 20,000 peaks for all samples for a particular pulldown (across all replicates, untreated and IFN-stimulated) were concatenated into a single list, and aligned fragments from each individual sample were counted for all peaks using BEDTools v2.28.0 [76]. IFN-inducible peaks were called using DESeq2 v1.26.0 [70], however we were unable to identify peaks that were significantly upregulated in response to IFN with an FDR < 0.10. We therefore took a more relaxed approach, retaining all peaks with an unadjusted *p-value* < 0.10 and log₂FC > 0. log₂FC values were shrunken using the *apeglm* function v1.8.0 [41] for visualization. Motif analysis was performed using XSTREME v5.4.1 [85] with options '*--minw 6 --maxw 20 --streme-motifs 20 --align center*' querying against the JASPAR CORE 2018 vertebrates database [86].

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 [76]. To assess family-level enrichment, GIGGLE v0.6.3 [87] 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 Gigggle enrichment score, and enriched TE families were identified according to the odds ratio, Fisher's two-tailed *p-value*, 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 [82, 83].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13100-022-00277-z>.

Additional file 1: Fig. S1. *Myotis lucifugus* embryonic fibroblast responsiveness to IFN α stimulation. **Fig. S2.** *Myotis lucifugus* genomic repeat element composition. **Fig. S3.** *Myotis lucifugus* total and recent TE transcriptome composition. **Fig. S4.** TE exonization events in *Myotis* ISGs. **Fig. S5.** A) Heatmaps showing normalized CUT&RUN signal (signal per million reads) over IFN α -inducible H3K27ac (unadjusted p -val < 0.10, log₂FC > 0, $n = 1113$). B) Volcano plot visualizing family-level TE enrichment for IFN α -inducible (unadjusted p -val < 0.10, log₂FC > 0) H3K27ac regions. TE families with a Fisher's two-tailed p -val < 0.05 were defined as enriched (blue) or depleted (green) according to the reported odds ratio. Nonsignificant families are shown in grey. C) Heatmaps showing normalized CUT&RUN signal (signal per million reads) over IFN α -inducible (unadjusted p -val < 0.10, log₂FC > 0) H3K27ac-marked L2 families. **Fig. S6.** Heatmaps showing normalized CUT&RUN signal (spike-in signal per million reads) over H3K27ac-marked A) MIR3 and B) AmnSINE1 families. **Fig. S7.** Reconstruction of the type I IFN locus.

Additional file 2: Table S1. RNAseq 4hrs Contrast IFNsUN.

Additional file 3: Table S2. RNAseq 24hrs Contrast IFNsUN.

Additional file 4: Supplementary Data 1. Annotation file in gtf format for the custom StringTie transcriptome assembly.

Additional file 5: Table S3. StringTie Contrast IFNsUN.

Additional file 6: Table S4. StringTie assembly TE-exon intersection.

Additional file 7: Supplementary Data 2. Zisupton-EEF1A1 alignment.

Additional file 8: Supplementary Data 3. EEF1A1 sequence alignment across bat species.

Additional file 9: Table S5. StringTie transcriptome vs TE reference sequences tblastx analysis.

Additional file 10: Table S6. StringTie transcriptome vs viral proteins tblastx analysis.

Additional file 11: Table S7. TE-Transcription Start Site overlap.

Additional file 12: Table S8. M. lucifugus embryonic fibroblast H3K27ac CUT&RUN DESeq2 results.

Additional file 13: Table S9. M. lucifugus embryonic fibroblast H3K27ac CUT&RUN motif analysis.

Additional file 14: Table S10. TE-overlapping IFN α -inducible H3K27ac peaks by family.

Additional file 15: Table S11. Distance from TE-overlapping STAT1 peaks to nearest ISG (log₂FC > 1.5) TSS.

Additional file 16: Table S12. Family-level TE enrichment using GIGGLE.

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. We thank the University of Colorado Genomics Shared Resource and BioFrontiers Computing core for technical support during this study.

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.

Funding

This study was supported by the NIH (1R35GM128822), the Alfred P. Sloan Foundation, the David and Lucile Packard Foundation, and the Boettcher foundation.

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.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 18 April 2022 Accepted: 9 August 2022

Published online: 06 September 2022

References

- Gorbunova V, Seluanov A, Kennedy BK. The world goes bats: living longer and tolerating viruses. *Cell Metab.* 2020;32:31–43.
- Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, et al. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? *Proc Biol Sci.* 2013;280:20122753.
- Schuh AJ, Amman BR, Sealy TK, Spengler JR, Nichol ST, Towner JS. Egyptian rousette bats maintain long-term protective immunity against Marburg virus infection despite diminished antibody levels. *Sci Rep.* 2017;7:8763.
- Halpin K, Hyatt AD, Fogarty R, Middleton D, Bingham J, Epstein JH, et al. Pteropid bats are confirmed as the reservoir hosts of henipaviruses: a comprehensive experimental study of virus transmission. *Am J Trop Med Hyg.* 2011;85:946–51.
- Swanepoel R, Leman PA, Burt FJ, Zachariades NA, Braack LE, Ksiazek TG, et al. Experimental inoculation of plants and animals with Ebola virus. *Emerg Infect Dis.* 1996;2:321–5.
- Munster VJ, Adney DR, van Doremalen N, Brown VR, Miazgowiec KL, Milne-Price S, et al. Replication and shedding of MERS-CoV in Jamaican fruit bats (*Artibeus jamaicensis*). *Sci Rep.* 2016;6:21878.
- 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;113:2696–701.
- 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:e2004086.
- De La Cruz-Rivera PC, Kanchwala M, Liang H, Kumar A, Wang L-F, Xing C, et al. The IFN response in bats displays distinctive IFN-stimulated gene expression kinetics with atypical RNASEL induction. *J Immunol.* 2018;200:209–17.
- Escalera-Zamudio M, Zepeda-Mendoza ML, Loza-Rubio E, Rojas-Anaya E, Méndez-Ojeda ML, Arias CF, et al. The evolution of bat nucleic acid-sensing toll-like receptors. *Mol Ecol.* 2015;24:5899–909.
- Ahn M, Cui J, Irving AT, Wang L-F. Unique loss of the PYHIN gene family in bats amongst mammals: implications for Inflammasome sensing. *Sci Rep.* 2016;6:21722.
- 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;4:789–99.
- 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;23:297–301.e4.
- Wickenhagen A, Sugrue E, Lytras S, Kuchi S, Noerenberg M, Turnbull ML, et al. A prenylated dsRNA sensor protects against severe COVID-19. *Science.* 2021;374:eabj3624.
- Ray DA, Feschotte C, Pagan HJT, Smith JD, Pritham EJ, Arensburger P, et al. Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Res.* 2008;18:717–28.

16. Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, et al. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nat Commun*. 2016;7:10716.
17. Mitra R, Li X, Kapusta A, Mayhew D, Mitra RD, Feschotte C, et al. Functional characterization of piggyBat from the bat *Myotis lucifugus* unveils an active mammalian DNA transposon. *Proc Natl Acad Sci U S A*. 2013;110:234–9.
18. Pritham EJ, Feschotte C. Massive amplification of rolling-circle transposons in the lineage of the bat *Myotis lucifugus*. *Proc Natl Acad Sci U S A*. 2007;104:1895–900.
19. Cosby RL, Judd J, Zhang R, Zhong A, Garry N, Pritham EJ, et al. Recurrent evolution of vertebrate transcription factors by transposase capture. *Science*. 2021;371. <https://doi.org/10.1126/science.abc6405>.
20. Skirmuntt EC, Escalera-Zamudio M, Teeling EC, Smith A, Katzourakis A. The potential role of endogenous viral elements in the evolution of bats as reservoirs for zoonotic viruses. *Annu Rev*. 2020; Available from: <https://www.annualreviews.org/doi/abs/10.1146/annurev-virology-092818-015613>.
21. Skirmuntt EC, Katzourakis A. The evolution of endogenous retroviral envelope genes in bats and their potential contribution to host biology. *Virus Res*. 2019;270:197645.
22. Frank JA, Feschotte C. Co-option of endogenous viral sequences for host cell function. *Curr Opin Virol*. 2017;25:81–9.
23. Mura M, Murcia P, Caporale M, Spencer TE, Nagashima K, Rein A, et al. Late viral interference induced by transdominant gag of an endogenous retrovirus. *Proc Natl Acad Sci U S A*. 2004;101:11117–22.
24. Young GR, Yap MW, Michaux JR, Steppan SJ, Stoye JP. Evolutionary journey of the retroviral restriction gene Fv1. *Proc Natl Acad Sci U S A*. 2018;115:10130–5.
25. Blanco-Melo D, Gifford RJ, Bieniasz PD. Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. *Elife*. 2017;6. <https://doi.org/10.7554/eLife.22519>.
26. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell*. 2015;162:974–86.
27. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-Demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell*. 2015;162:961–73.
28. Schmidt N, Domingues P, Golebiowski F, Patzina C, Tatham MH, Hay RT, et al. An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *Proc Natl Acad Sci U S A*. 2019;116:17399–408.
29. Kelly CJ, Chitko-McKown C, Chuong EB. Ruminant-specific retrotransposons shape regulatory evolution of bovine immunity. *bioRxiv*. 2021:2021.10.01.462810 Available from: <https://www.biorxiv.org/content/10.1101/2021.10.01.462810>. Cited 2022 Feb 25.
30. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*. 2016;351:1083–7 American Association for the Advancement of Science.
31. Srinivasachar Badarinarayan S, Shcherbakova I, Langer S, Koepke L, Preising A, Hotter D, et al. HIV-1 infection activates endogenous retroviral promoters regulating antiviral gene expression. *Nucleic Acids Res*. 2020;48:10890–908.
32. van de Lagemaat LN, Landry J-R, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet*. 2003;19:530–6.
33. Bogdan L, Barreiro L, Bourque G. Transposable elements have contributed human regulatory regions that are activated upon bacterial infection. *Philos Trans R Soc Lond Ser B Biol Sci*. 2020;375:20190332.
34. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356:92–5.
35. Myotis_Lucifugus. *DNA Zoo*. Available from: https://www.dnazoo.org/assemblies/Myotis_Lucifugus. Cited 2022 Jun 7.
36. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020;117:9451–7.
37. Xiong W, He L, Lai J, Dooner HK, Du C. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc Natl Acad Sci U S A*. 2014;111:10263–8.
38. Tempel S. Using and understanding RepeatMasker. *Methods Mol Biol*. 2012;859:29–51.
39. RepeatMasker Home Page. Available from: <http://www.repeatmasker.org>. Cited 2022 Apr 12.
40. Ray DA, Pagan HJT, Thompson ML, Stevens RD. Bats with hATs: evidence for recent DNA transposon activity in genus *Myotis*. *Mol Biol Evol*. 2007;24:632–9.
41. 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.
42. Jin Y, Tam OH, Paniagua E, Hammell M. Tetrascripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics*. 2015;31:3593–9.
43. Pasquesi GIM, Perry BW, Vandeweghe MW, Ruggiero RP, Schield DR, Casteo TA. Vertebrate lineages exhibit diverse patterns of transposable element regulation and expression across tissues. *Genome Biol Evol*. 2020;12:506–21.
44. Xue AT, Ruggiero RP, Hickerson MJ, Boissinot S. Differential effect of selection against LINE retrotransposons among vertebrates inferred from whole-genome data and demographic modeling. *Genome Biol Evol*. 2018;10:1265–81.
45. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57–74.
46. Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*. 2019;20:278.
47. Böhne A, Zhou Q, Darras A, Schmidt C, Scharl M, Galiana-Arnoux D, et al. Zisupton - a novel superfamily of DNA transposable elements recently active in fish. *Mol Biol Evol*. 2011;29:631–45 Oxford Academic.
48. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015;6:11.
49. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*. 2000;403:785–9.
50. Horie M, Kobayashi Y, Honda T, Fujino K, Akasaka T, Kohl C, et al. An RNA-dependent RNA polymerase gene in bat genomes derived from an ancient negative-strand RNA virus. *Sci Rep*. 2016;6:25873.
51. 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;339:456–60.
52. Pavlovich SS, Lovett SP, Koroleva G, Guito JC, Arnold CE, Nagle ER, et al. The Egyptian Rousette genome reveals unexpected features of bat antiviral immunity. *Cell*. 2018;173:1098–110.e18.
53. Kobayashi KS, van den Elsen PJ. NLRCS: a key regulator of MHC class I-dependent immune responses. *Nat Rev Immunol*. 2012;12:813–20.
54. Benkő S, Kovács EG, Hezel F, Kufer TA. NLRCS functions beyond MHC I regulation—what do we know so far? *Front Immunol*. 2017;8:150.
55. Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, et al. Latent enhancers activated by stimulation in differentiated cells. *Cell*. 2013;152:157–71 Elsevier.
56. Qiao Y, Giannopoulou EG, Chan CH, Park S-H, Gong S, Chen J, et al. Synergistic activation of inflammatory cytokine genes by interferon- γ -induced chromatin remodeling and toll-like receptor signaling. *Immunity*. 2013;39:454–69.
57. Bondet V, Le Baut M, Le Poder S, Lécru A, Petit T, Wedlarski R, et al. Constitutive IFN α protein production in bats. *Front Immunol*. 2021;12:735866.
58. Kepler TB, Sample C, Hudak K, Roach J, Haines A, Walsh A, et al. Chiropteran types I and II interferon genes inferred from genome sequencing traces by a statistical gene-family assembler. *BMC Genomics*. 2010;11:444.
59. Lindblad-Toh K, Garber M, Zuk O, Lin MF, Parker BJ, Washietl S, et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*. 2011;478:476–82.
60. Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostafa R, et al. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv*. 2018:254797 Available from: <https://www.biorxiv.org/content/10.1101/254797v1>. Cited 2022 Apr 18.
61. Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*. 2012;58:268–76.

62. Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, et al. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun.* 2018;9:2774.
63. Janssens D, Henikoff S. CUT&RUN: targeted in situ genome-wide profiling with high efficiency for low cell numbers v3. protocols.io: ZappyLab, Inc.; 2019. Available from: <https://www.protocols.io/view/cut-amp-run-targeted-in-situ-genome-wide-profiling-zcpf2vn>
64. Meers MP, Bryson TD, Henikoff JG, Henikoff S. Improved CUT&RUN chromatin profiling tools. *Elife.* 2019;8. <https://doi.org/10.7554/eLife.46314>.
65. BBMap. SourceForge. Available from: <https://www.sourceforge.net/projects/bbmap/>. Accessed Feb 2019
66. Babraham bioinformatics - FastQC A quality control tool for high throughput sequence data Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>. Version 0.11.08 - 2018/04/10
67. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* 2016;32:3047–8.
68. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
69. CCGB: Miller Lab, LASTZ. Available from: <https://www.bx.psu.edu/~rsharis/lastz/>. Accessed Feb 2021.
70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15:550.
71. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res.* 2019;47:W199–205.
72. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
73. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Res.* 2020;9. <https://doi.org/10.12688/f1000research.23297.2>.
74. Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* 2000;28:45–8.
75. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10:421.
76. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26:841–2.
77. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323.
78. Nakagawa S, Takahashi MU. gEVE: a genome-based endogenous viral element database provides comprehensive viral protein-coding sequences in mammalian genomes. *Database.* 2016;2016. <https://doi.org/10.1093/database/baw087>.
79. 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.
80. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* 2010;26:2460–1.
81. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods.* 2012;9:357–9.
82. 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.
83. 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.
84. 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: Springer New York; 2014. p. 81–95.
85. Grant CE, Bailey TL. XSTREME: comprehensive motif analysis of biological sequence datasets. *bioRxiv.* 2021; Available from: <https://www.biorxiv.org/content/10.1101/2021.09.02.458722v1>.
86. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 2020;48:D87–92.
87. Layer RM, Pedersen BS, DiSera T, Marth GT, Gertz J, Quinlan AR. GIGGLE: a search engine for large-scale integrated genome analysis. *Nat Methods.* 2018;15:123–6.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

