Investigation of Innate Immune Responses in *Eptesicus* Bat Cells via Comprehensive Analysis

林　咸亨

Lin, Hsien-Hen

Contents

List of Figure and Table

Abbreviation

DEG: differentially expressed gene

DET: differentially expressed transposable element

FDR: false discovery rate

IFN: interferon

ISG: interferon-stimulated gene

LINE: long interspersed nuclear elements

LTR: long terminal repeats

MERS: Middle East respiratory syndrome

Poly (I:C): polyinosinic-polycytidylic acid

SARS: severe acute respiratory syndrome

SINE: short interspersed nuclear elements

TE: transposable element

TLR: Toll-like receptors

Abstract

Bats, the order Chiroptera, have been reported to harbor zoonotic viruses without obvious symptoms, acting as reservoirs. It is thus hypothesized that bats possess distinctive immune systems in response to viral infection. Following the hypothesis, bats of the genus *Eptesicus* have been revealed to show unique immune responses, such as a novel binding motif of repressor identified in *Eptesicus fuscus* that suppress inflammation. However, the responses of the immune system in *Eptesicus* bats remain largely unknown. Here, I applied comprehensive analysis to illustrate the expression profiles in two *Eptesicus* bat cell lines, Efk3B and EnK, for elucidating the innate immune responses, the first line of immunity.

The immune responses in Efk3B and EnK cells were induced by polyinosinic-polycytidylic acid [poly (I:C)] treatments. The gene expression profiles were found to be similar in two cell lines, but divergent differentially expressed genes were also identified, respectively. I further revealed that the upregulated genes were distinct between *Eptesicus* bat cells and human epithelial cells in response to poly (I:C) induction, particularly enriching in interferon γ pathways in bat cells. Besides, some upregulated genes, such as DDX60 and IFIT1 involved in antiviral functions, were shown to exist in multiple copies in *E. fuscus* genome. Moreover, the basal expression levels of several immune-related genes, including hub genes IFIT2 and IFIT3, were found to be higher in bat cells than in four human cell lines under normal conditions without induction, suggesting that bat cells might be in activated immune status. These results indicated that the innate immune responses in *Eptesicus* bat cell lines are distinguishable from those in human cell lines.

In spite of applying existing *E. fuscus* genome reference, I performed de novo assembly and identified unannotated novel transcripts that were upregulated after induction, indicating that the existing gene annotations were still incomplete, and more details have not been revealed yet. Additionally, the landscape of transposable elements (TEs) and novel microRNAs (miRNA) in bat cells were recognized, some of which were expressing and upregulated in response to poly (I:C) treatment, suggesting that the innate immune responses in *Eptesicus* bat cells influenced/were influenced by different pathways via yet known mechanisms. These data extend current *E. fuscus* genome reference,

granting novel information for further analysis.

This is the first study to reveal the gene expression profiles, including mRNAs, TEs, and miRNAs, of the innate immune responses in *Eptesicus* bat cells. Our data provide basic and innovative insights into bat innate immunity as well as represent a valuable resource for future research into bat immune systems and the biology of *Eptesicus* bats.

Chapter 1

Literature Review and Introduction

1.1 Bats are reported to act as virus reservoirs

Bats are a diverse group of mammals belonging to order Chiroptera, which can be further divided into Yinpterochiroptera, including megabats and some of microbats, and Yangochiroperata, comprising of the remaining microbats. As the second species-rich mammal, bats consist of around 20% of all named mammals and are widely distributed all over the world, living in every continent except Antarctica (Burgin, Colella, Kahn, & Upham, 2018). There are over 1300 species of bats with extraordinary diversity in characteristics such as diet, morphology and size, and therefore bat species are complex and distinguishable (Fenton & Simmons, 2015).

Bats are considered natural reservoirs of various zoonotic viruses, and their characteristics makes them more capable of transmitting and spreading viruses. Compared to other mammals with similar metabolic rate ratios and body mass, bats have relatively longer lifespans. For example, microbats have life spans more than 25 years, and the documented greatest longevity is 35 years for a little brown bat (Calisher, Childs, Field, Holmes, & Schountz, 2006). Moreover, bats are the only mammal with abilities of flight, and particular species of bats travel large distances during seasonal migrations (Holland, 2006). Taking into the wide distribution, mixed populations, as well as above characteristics, bats are able to spread viruses for farer distances and longer durations.

Over 200 viruses have been detected in bats, some of which cause severe disease in humans and other animals (Moratelli & Calisher, 2015). Coronaviruses that cause severe respiratory syndromes, including severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and recent underdetermined Coronavirus disease 2019 (COVID-19) which is still wreaking havoc, have been reported to be harbored by bats, which are believed to cause diseases via spillover events (Anthony et al., 2017; Hu et al., 2017; Zhou & Shi, 2021). The transmission routes through bat excretion of Hendra virus, a species of henipavirus spreading in horses and humans in Australia, has been revealed in a previous study (Edson et al., 2015). Filoviruses, such as Marburg virus (MARV) and Ebolavirus (EBOV) causing severe hemorrhagic fevers in humans and primates, have also been characterized in different species of bats (Goldstein et al., 2018; Schuh et al., 2017; Yang et al., 2019). Surprisingly, previous studies identified asymptomatic bats harboring these pathogenic viruses (Leroy et al., 2005; Munster et

al., 2016; Olival & Hayman, 2014). Therefore, the development of the virus-tolerant phenotype in bats and the possible underlying mechanisms preventing symptoms of infection of viruses that are deadly in other species gradually become critical research topics.

1.2 An overview of innate immunity

Innate immunity, the first line of defense and rapid responses to foreign pathogens, has been widely studied. Although the core elements and functions are conserved, various components and mechanisms involve in the innate immunity of different species of mammals, resulting in high diversity (Riera Romo, Perez-Martinez, & Castillo Ferrer, 2016). In general, the conserved mechanisms include external barriers, humoral elements, and cellular effectors with individual characteristics and variations. External barriers consist of skin and its secretion, providing mechanical protection and preventing pathogen infiltration (Matsui & Amagai, 2015); humoral elements, including lysozyme, complement systems, and various classes of cytokines, react to pathogen invasion, exhibit anti-pathogen activity, and induce further immune responses (Bowdish, Davidson, & Hancock, 2005; Dunkelberger & Song, 2010; Lin & Leonard, 2019); cellular effectors, such as natural killer cells and myeloid phagocytic cells, recognize pathogens and contribute to activation of adaptive immunity (Bendelac, Savage, & Teyton, 2007; Rosales, 2020).

Interferons (IFNs), a class of cytokines, are released when sensing pathogens and invasive elements. For instance, exogenous viral DNAs and RNAs are recognized by pattern recognition receptors, such as Toll-like receptors (TLRs) and RIG-I-like receptors, and the following signaling pathways are activated, leading to the production of IFNs (Schoggins & Rice, 2011). Three types of IFNs have been characterized in mammals (Negishi, Taniguchi, & Yanai, 2018). Type I IFNs, including IFN-α and IFN-β, are well known as the antiviral functions that are conserved in vertebrates; type II IFNs, or IFN-γ, mainly function in antiviral responses and cellular mechanisms, such as cell growth and apoptosis; type III IFNs, or IFN-λ, are reported to exhibit similar antiviral defenses as type I IFNs (Riera Romo et al., 2016). The release of IFNs further induces interferon-stimulated genes (ISGs) and promotes antiviral responses (Schoggins & Rice, 2011).

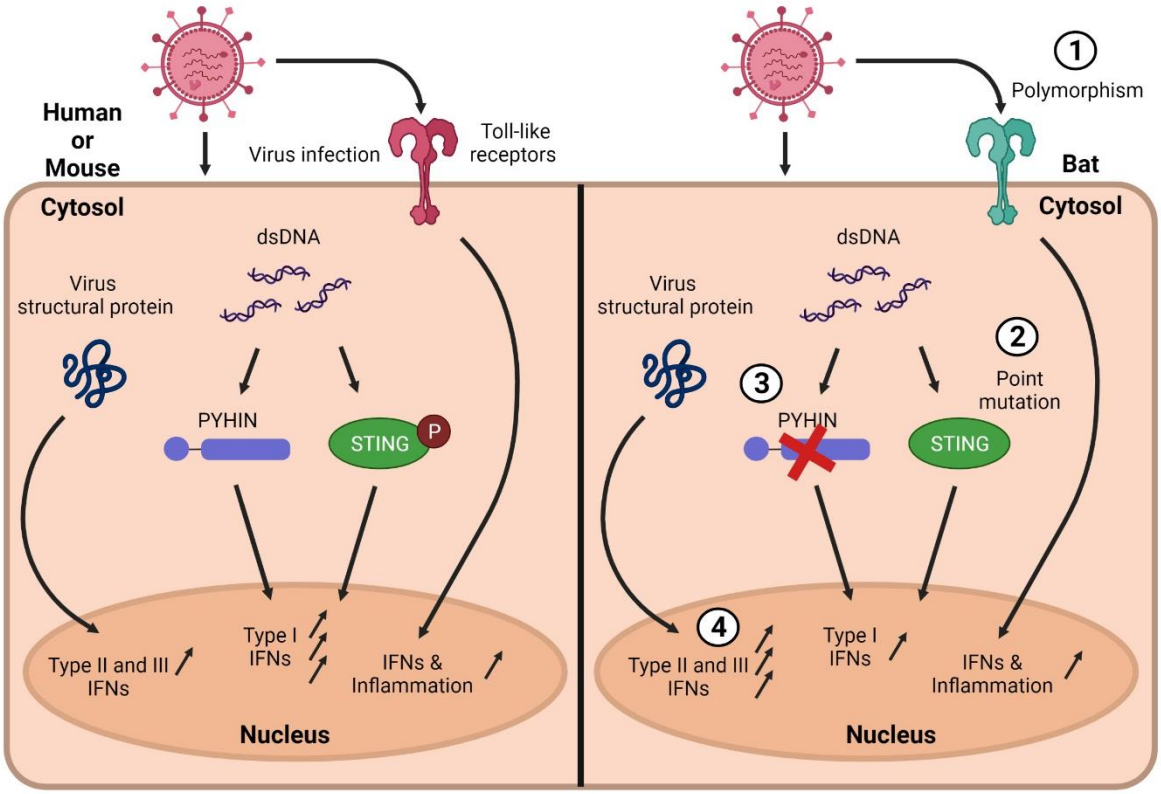
Other than the immune system itself, various cellular mechanisms and pathways also participate in or are influenced by the activation of innate immune responses. For example, microRNAs (miRNAs) have been reported to regulate innate immunity and inflammation. Previous studies have demonstrated that several transcription factors and post-transcriptional enzymes regulate the production and processing of miRNAs in response to inflammation and infection, and the produced miRNAs afterwards modulate inflammation as feedback (O'Connell, Rao, Chaudhuri, & Baltimore, 2010; Taganov, Boldin, & Baltimore, 2007). Moreover, divergent miRNA populations related to immune systems have been gradually identified in different species and cell types (Chandan, Gupta, & Sarwat, 2019; Nejad, Stunden, & Gantier, 2018; Rose et al., 2021). Another example related to the activation of innate immune responses is transposable element (TE). TEs are DNA sequences largely diverse between species. A recent study shows that different TE families are upregulated in humans and mouse cells following the induction of innate immunity (Macchietto, Langlois, & Shen, 2020). Although the consequences of TE upregulation are not fully understood yet, another research has suggested the role of TEs as enhancers, participating in the regulation of ISGs (Chuong, Elde, & Feschotte, 2016). Overall, these mechanisms elevate the complexity of innate immunity and promote uniqueness and disparity in immune responses among species.

1.3 The unique immune strategies in bats

According to the characteristics of bats that serve as reservoirs of viruses, studying the underlying mechanisms become hot topics in recent years. It is hypothesized that bats possess unique immune strategies, which have not yet been fully studied, that allow them to develop virus-tolerant phenotypes in response to viral infection, thus serving as viral reservoirs (Irving, Ahn, Goh, Anderson, & Wang, 2021).

Previous studies have shown that the basic immune responses in bats are similar to those in other mammals, including the signaling pathways and overall immune cell populations involved (Becker et al., 2019; Omatsu et al., 2008). Despite the basic immune responses, not surprisingly, it has been shown that distinct bat species demonstrate individual patterns of immunity (Fig. 1). For example, the agonist

binding sites of several Toll-like receptors in black flying fox (*Pteropus alecto*) showed unique polymorphisms, and some of the polymorphisms contained strong positive selections that reshape nucleic acid binding motifs (Cowled et al., 2011). A previous study compared the sequences of STING, a critical adaptor protein functioning in DNA sensing pathways, among several bats and non-bat mammals, and a single point mutation that caused dampened IFN activation was identified in several bat species (Xie et al., 2018). PYHIN locus, a family of proteins acting as sensors for microbial DNA and inducing type I IFN signaling pathways, were found to be missing in at least ten species of bats (Ahn, Cui, Irving, & Wang, 2016). Another research indicated that the type II and type III interferon response were stronger in bat cell lines of Egyptian rousette bat (*Rousettus aegyptiacus*) than that in human cell lines in responses to specific virus infection (Kuzmin et al., 2017). Increasing evidence indicates the existence of uncommon immune strategies in bats. Thus, it is critical to clarify these atypical features to elucidate the possible mechanisms by which bats act as virus reservoirs.



(Figure legend is in the next page.)

Figure 1. Schematic of the examples of unique immune responses in bats.

In human or mouse cells, viral elements are recognized by several components, including toll-like receptors, PYHIN protein family, and STING protein. Afterwards, a series of signaling pathways is triggered, the expression of interferons (IFN) and downstream interferon-stimulated genes (ISG) are induced, and the following immune responses are activated. By contrast, bats show various unique immune responses within these parts, including (1) the polymorphisms in toll-like receptors that underwent positive selections to recognize pathogens, (2) a point mutation in STING protein and (3) loss of PYHIN locus that cause dampened IFN activation, and (4) stronger type II and III IFN responses toward specific virus infection. The figure was created with BioRender.com.

1.4 The special immune responses in *Eptesicus* bats

Similar to other bats, bats of the genus *Eptesicus* are known to be the hosts of several zoonotic viruses, such as rhabdoviruses that cause encephalitis and flu-like symptoms (Kohl, Nitsche, & Kurth, 2021). *Eptesicus* bats in Europe are reported to host Rhabdoviruses of the genus Lyssavirus, which can cause rabies disease and lead to death of unvaccinated humans without appropriate treatments (Johnson et al., 2010).

Recently, *Eptesicus fuscus* has been reported to show distinctive immune mechanisms. In earlier research using a *E. fuscus* cell line Efk3B, a potential binding motif of repressor c-Rel was found in the promoter of TNF-α, an inflammatory cytokine that is released during acute inflammation and triggers various signaling pathways. (Banerjee, Rapin, Bollinger, & Misra, 2017). c-Rel was further confirmed to bind to the binding motif and inhibit TNF-α promoter activity, suppressing inflammatory pathology. Another study from the same research group revealed that IFN-β expression was induced after MERS coronavirus (MERS-CoV) infection in Efk3B but not in human cells, and the IRF3 signaling pathway was identified to play an important role in limiting the replication of the MERS-CoV (Banerjee, Falzarano, Rapin, Lew, & Misra, 2019). Further, a cell culture model persistently infected with MERS has been developed to study the interaction between MERS-CoV and *Eptesicus* bats (Banerjee et al., 2020). These data suggest that unique immune responses might exist in *Eptesicus* bats, and further research is required to clarify more details of the potential mechanisms.

1.5 The niche and research aim to study *Eptesicus* bat immunity

Despite the discoveries mentioned in above section, one critical niche is that the innate immune responses in *Eptesicus* bats have not yet been fully elucidate. In this study, I aimed to apply comprehensive analysis to elucidate the innate immune responses in bats of the genus *Eptesicus* (Fig. 2). I treated two *Eptesicus* bat cell lines with polyinosinic-polycytidylic acid (poly (I:C)) and analyzed the gene expression profiles. Comparing bat mRNA sequencing (mRNA-seq) data sets with human data sets, the mechanisms in response to poly (I:C) were shown to be similar in *Eptesicus* bats and humans. Besides, certain genes are uniquely upregulated only in *Eptesicus* bat cell lines. Meanwhile, I recognize several upregulated ISGs in bat cell lines existing as multi-copies. I also revealed that the basal expression levels of certain ISGs were higher in the bat cell lines. Moreover, a consequential upregulation of unannotated transcript levels was identified. In addition, I showed that the expression of several TEs was upregulated after poly (I:C) treatments. Furthermore, small RNA sequencing identified novel miRNAs, and numerous of them were upregulated in bat cells. To sum up, this study reveals the uniqueness of innate immune responses in bats of the genus *Eptesicus* and provides the basic information necessary to obtain a deeper understanding of immunity in bats.

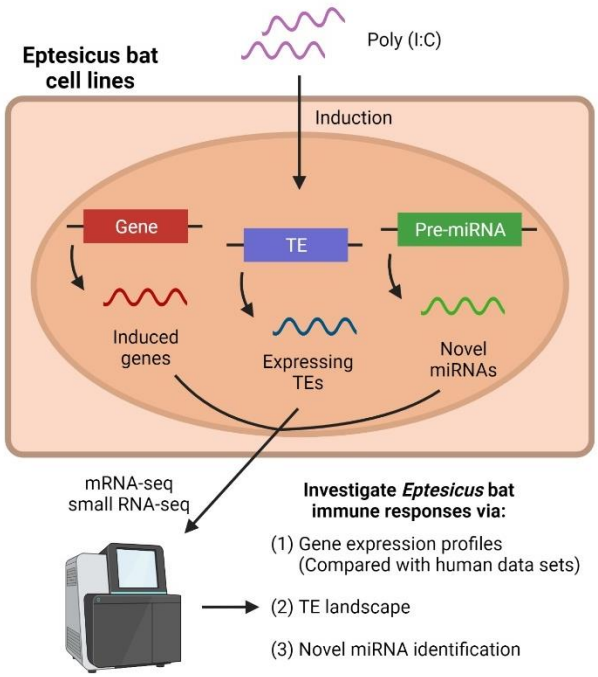


Figure 2. Schematic of studying the immune responses of *Eptesicus* bat.

I aim to investigate the immune responses in *Eptesicus* bat cell lines via gene expression profiles, TE landscape, and identifying novel miRNA. Moreover, I compared our mRNA-seq data sets with human data sets to identify the uniqueness of *Eptesicus* bat cell lines. The figure was created with BioRender.com.

Chapter 2　Result

2.1 The immune responses in Efk3B and EnK cells can be triggered by universal type I IFN and poly (I:C)

In this study, I applied two cell lines derived from *Eptesicus* bats, Efk3B and HAMOI-EnK (EnK) (Banerjee et al., 2016; Horie, Akasaka, Matsuda, Ogawa, & Imai, 2016). First, I test to trigger the innate immune responses with IFN or poly (I:C) treatments. Following the experimental design in published research, I examined the expression of IFN-β as an indicator of activated immune responses (Banerjee et al., 2017). Both cell lines were respectively treated with universal type I IFN or poly (I:C), and the level of IFN-β (IFNB1) transcripts was quantified via qRT–PCR. Consistent with the results of the published research, Efk3B cells showed increased IFN-β transcript levels over time in response to poly (I:C) treatments, which were highest after 12 hours of treatment (Fig. 3). IFN-β expression was also increased in EnK cells, and the level reached the peak after 24 hours. In contrast, IFN-β transcripts were not detected in response to treatments of universal type I IFN.

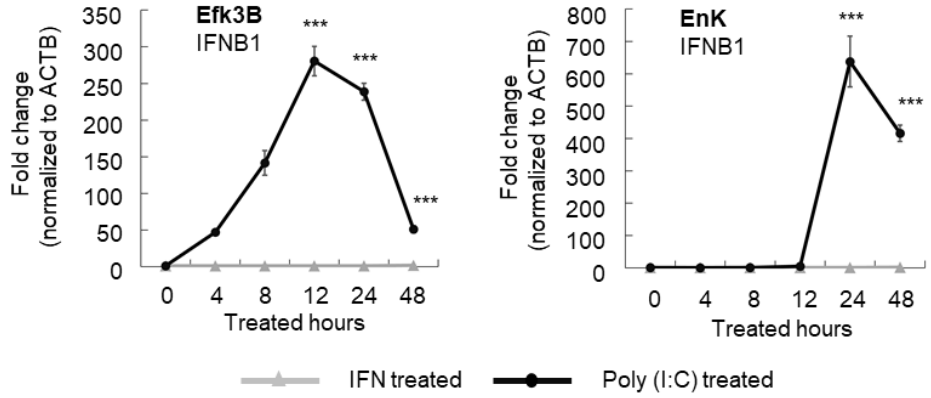
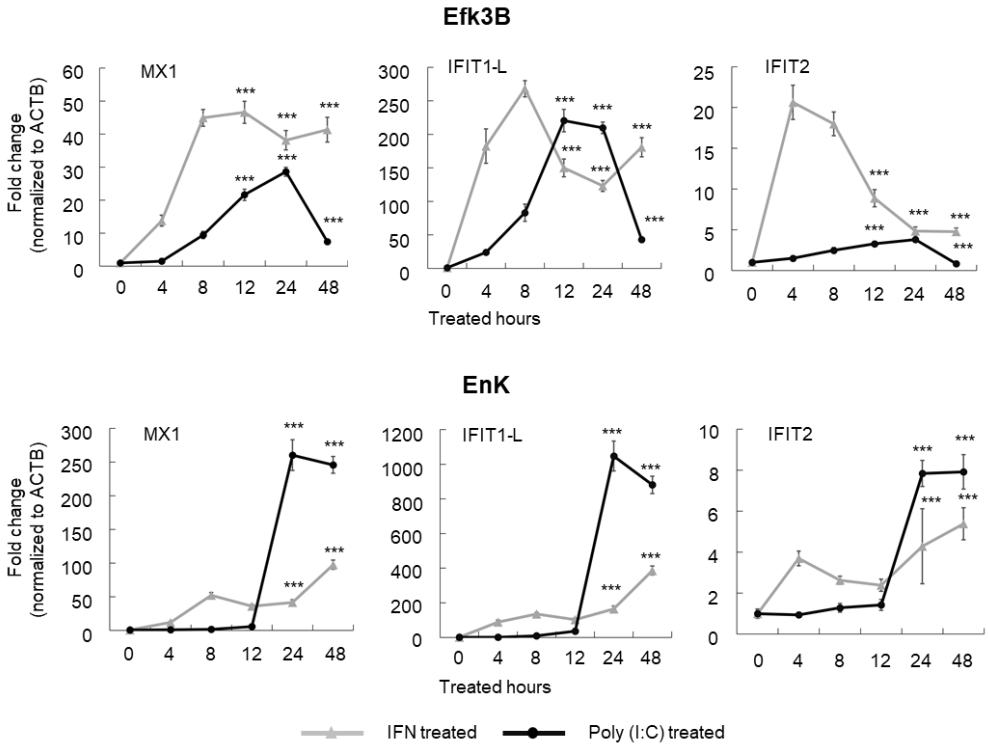


Figure 3. IFN-β transcripts were increased in Efk3B and EnK cells after poly (I:C) treatments.

The fold changes of IFNB1 transcript expression after universal type I interferon and poly (I:C) treatments in Efk3B and EnK cells. The fold-change is relative to 0 hours after normalizing with ACTB (β-actin). Data are presented as the mean ± s.e.m. of three biological replicates (n = 3). Significant differences were shown between treated samples of each time point (12, 24, and 48 hr) and untreated samples (0 hr), calculated by unpaired Student's t-test with $P < 0.001$ (***). The figures and figure legends are derived and modified from figure 1A and 1B in the original article.

Next, I further investigated the expression levels of several interferon-stimulated genes (ISGs), including MX1, IFIT1-L, and IFIT2. MX1 is a GTPase and functions in antiviral responses, which are induced by type I and type III IFNs in infected host; IFIN1 and IFIN2 are interferon-induced proteins that are rapidly induced during virus infection, recognizing viral RNAs and activating downstream signaling pathway for antiviral functions (Verhelst, Hulpiau, & Saelens, 2013; Vladimer, Gorna, & Superti-Furga, 2014). The expression levels of selected ISGs were significantly upregulated after IFN or poly (I:C) treatments overtime in both Efk3B and EnK cells (Fig. 4). In Efk3B cells after IFN treatments, MX1, IFIT1-L and IFIT2 reached the peak after 12 hours, 8 hours, and 4 hours, respectively; in contrast, MX1, IFIT1-L and IFIT2 peaked 24 hours, 12 hours, and 24 hours after poly (I:C) treatments. In EnK cells, MX1, IFIT1-L and IFIT2 reached the highest expression after 48 hours of IFN treatments; in comparison, poly (I:C) induced those ISGs to highest expression after 24 hours treatments. Although the expression of ISGs in both cell lines was significantly increased, the sensitivity or upregulated level between the two cell lines might be different, indicating that the kinetics of the immune responses were distinct between Efk3B and EnK cells. These results demonstrate that poly (I:C) and universal type I IFN can triggered the innate immune responses in both Efk3B and EnK cells.



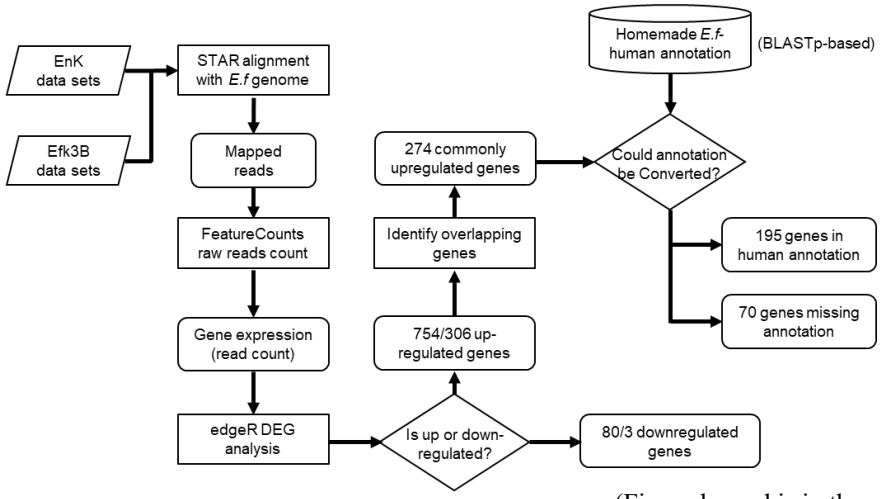
(Figure legend is in the next page.)

Figure 4. The expression of selected interferon-stimulated genes (ISGs) in Efk3B and EnK cells was induced after universal type I IFN and poly (I:C) treatments.

The fold changes in MX1, IFIT1-L, and IFIT2 expression in Efk3B universal type I interferon and poly (I:C) treatments in Efk3B and EnK cells. The fold-change is relative to 0 hours after normalizing with ACTB. Data are presented as the mean ± s.e.m. of three biological replicates (n = 3). Significant differences were shown between treated samples of each time point (12, 24, and 48 hr) and untreated samples (0 hr), calculated by unpaired Student's t-test with $P < 0.001$ (***). The figure and figure legends are derived and modified from figure 1C and 1D in the original article.

2.2 Common and unique gene expression profiles were revealed in Efk3B and EnK cells in response to poly (I:C) treatment

To comprehensively analyze the immune responses in bat cell lines, I performed mRNA sequencing for gene expression profiling. As poly (I:C) induced the expression of IFN-β, I supposed that poly (I:C) treatments affected more aspects of immune responses, which would be possible to acquire more information, and therefore decided to apply poly (I:C) treated samples to further analysis. Based on the aforementioned findings (Fig. 3 and 4), I collected Efk3B and EnK cells 12 and 24 hours after poly (I:C) treatment, respectively. Due to lack of *E. nilssonii* genome reference, I applied the *E. fuscus* genome (GCA_000308155.1 EptFus1.0) for annotation and subsequent analyses to both the Efk3B and EnK cell data sets. The pipeline and workflow shown in figure 5 were established for further analysis.



(Figure legend is in the next page.)

Figure 5. The workflow of processing Efk3B and EnK mRNA sequencing data sets.

Preprocessed reads were mapped to the *E. fuscus* genome using STAR. FeatureCounts was used for raw read counts, and R package edgeR was next used for differentially expressed gene (DEG) analysis. Up- and downregulated DEGs were identified and converted into human annotation through a homemade table for further analysis. Detailed parameters are shown in the Materials and Methods. The figure and figure legends are derived from figure 2 in the original article.

I performed differentially expressed gene (DEG) analysis to identified up- and downregulated DEGs in response to poly (I:C). The heatmaps showed that the top 500 DEGs were overall upregulated, indicating the activation of innate immune responses (Fig. 6). The genes with false discovery rate (FDR) < 0.05 and at least twofold change were isolated for further analyses. As a result, I identified a total of 409 (406 up- and 3 downregulated) and 834 (754 up- and 80 downregulated) DEGs in Efk3B and EnK cells, respectively (Fig. 7A). Among the upregulated DEGs, 274 genes were common in both Efk3B and EnK cells (Fig. 7B). These data suggest that although some upregulated DEGs are shared in Efk3B and EnK cells, uniquely upregulated DEGs are also present in each cell line.

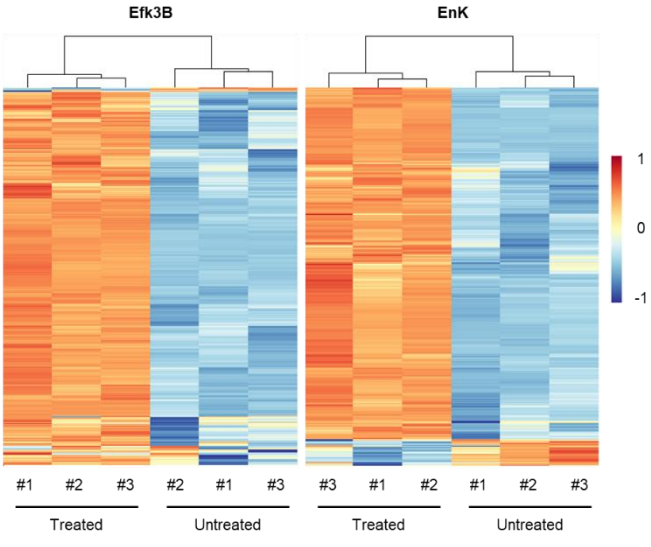


Figure 6. The expression heatmap of DEGs of Efk3B and EnK datasets.

The heatmap showing top 500 DEGs with highest varied expression between poly (I:C) treated and untreated Efk3B and EnK datasets. Scale represents normalized Z-scores generated by R pheatmap from blue-white-red. The figures and figure legends are derived and modified from supplementary figure 1 in the original article.

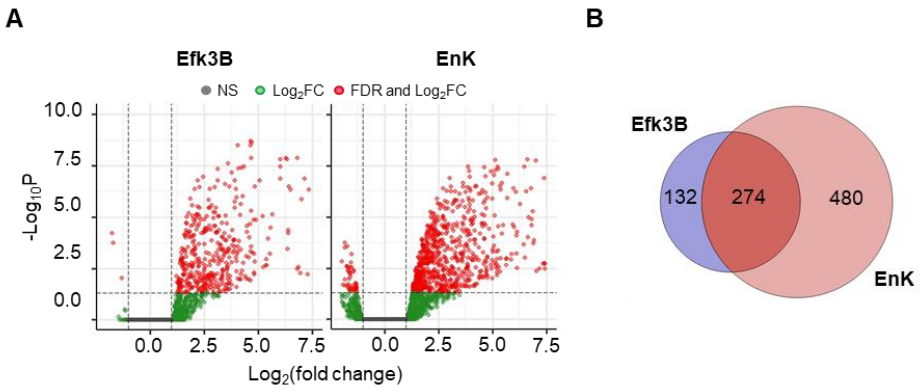


Figure 7. The DEGs in induced Efk3B and EnK cells were identified.

(A) Volcano plots of DEGs of poly (I:C) treated vs. untreated Efk3B and EnK data sets. Green dots represent genes with fold-change ≥ 2 and FDR ≥ 0.05; red dots represent genes with fold-change ≥ 2 and FDR < 0.05; gray dots represent genes with no significant difference. (B) Venn diagram of upregulated DEGs of Efk3B and EnK cells showing the common DEGs. The figure and figure legends are derived and modified from figure 3A and 3B in the original article.

Next, for further analysis and comparison, I generated a homemade conversion table and identify the orthologs of the *E. fuscus* genes in human gene annotations (details in Materials and methods). As a result, 204 in 274 *E. fuscus* annotations of common upregulated DEGs were converted into 207 human annotations, and I included these orthologous annotations for further investigations. The 204 upregulated *E. fuscus* DEGs with human annotations were isolated for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. GO analysis revealed high enrichment in biological processes related to immune responses, including “response to virus” (GO: 0009615) and “defense response to virus” (GO: 0051607) (Fig. 8A). The KEGG analysis showed comparable results, enriching in pathways involved in immune responses, such as the NOD-like receptor signaling pathway and various virus-related pathways (Fig. 8B). Combining the upregulated DEGs and enriched GO terms, the gene-concept network demonstrated that the DEGs and their corresponding functions were mainly related to antiviral responses (Fig. 8C).

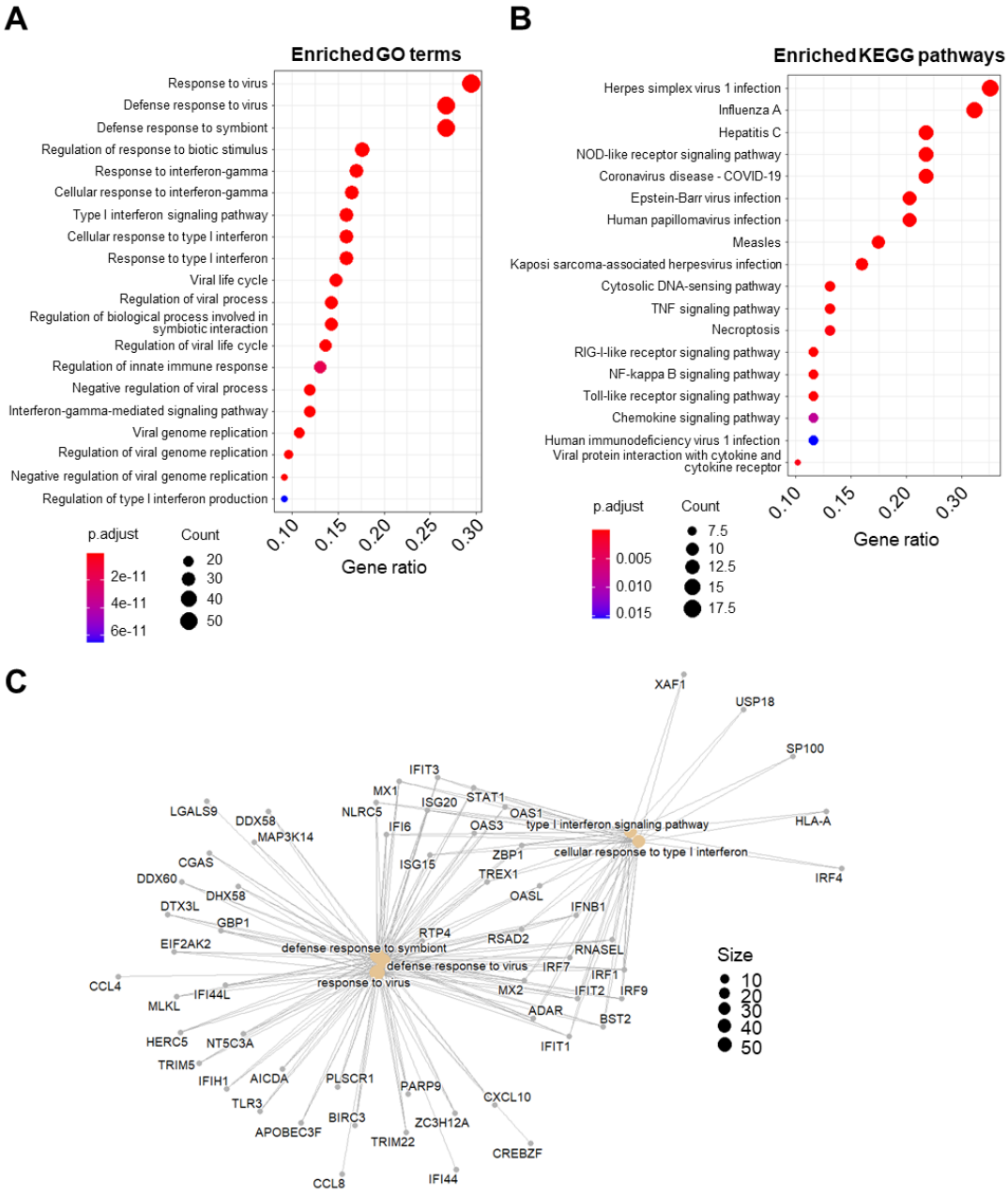


Figure 8. The common DEGs of induced Efk3B and EnK cells were mainly enriched in antiviral responses. (A) The top enriched GO term biological processes of common upregulated DEGs of Efk3B and EnK cells. The colors of dots indicate the adjusted p values. The size of the dot represents the number of genes involved in the term, and the gene ratio represents the proportion of genes occupying the whole term. (B) As per (A) for the top enriched KEGG pathways of common up-regulated DEGs of Efk3B and EnK cell. (C) The gene-concept network of linked GO terms and involved common upregulated DEGs. The dot size of GO terms represents the number of genes involved in the term. The figure and figure legends are derived and modified from figure 3C, 3D, and supplementary figure 2 in the original article.

Despite commonly upregulated DEGs, several uniquely upregulated genes were identified in EnK and Efk3B cells in response to poly (I:C) induction according to our above analysis (Fig. 7B). To investigate the outcome of this distinct expression pattern, I further applied those uniquely upregulated genes shown in figure 7B to perform GO analyses. Intriguingly, specific terms, including "response to interferon-gamma" (GO: 0034341) and "NIK/NF-kappaB signaling" (GO: 0038061), were identified in EnK cells (Fig. 9). In contrast, none was enriched using uniquely upregulate DEGs of Efk3B cells as input. These data suggest that the main mechanisms in response to exogenous double-stranded RNA (poly (I:C)) are common in Efk3B and EnK cells. In addition, uniquely upregulated DEGs indicate that divergent mechanisms might be respectively activated in Efk3B and EnK cells.

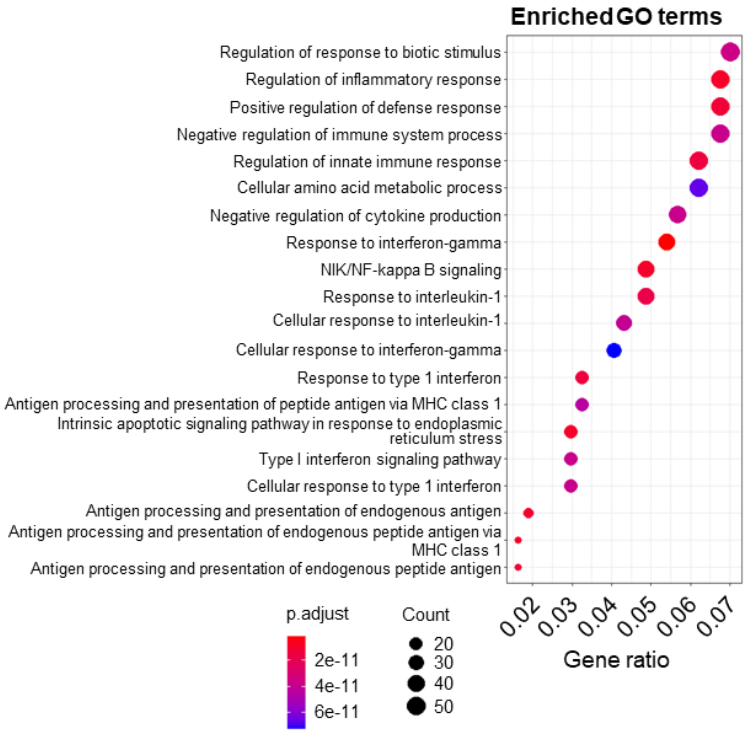


Figure 9. The GO term analysis of unique DEGs from induced EnK cells.

The top enriched GO term biological processes of unique up-regulated DEGs of EnK cells. The red the dot, the smaller the adjust p-value. The size of the dot represents the number of genes involved in the term, and the gene ratio represent the proportion of genes occupying the whole term. The figures and figure legends are derived and modified from supplementary figure 3 in the original article.

2.3 The response mechanisms to poly (I:C) differ between *Eptesicus* bat and human cell lines

To study the uniqueness of innate immune responses in bat cells, I compared our data with a published poly (I:C) treated HeLa cell mRNA-seq data set (GSE130618) to investigate the differences of gene expression profiles. The HeLa cell data set was processed using the same DEG analysis pipeline (Fig. 5). In total, 2492 DEGs (2100 up- and 392 downregulated) were identified in HeLa cells with poly (I:C) treatments (Fig. 10).

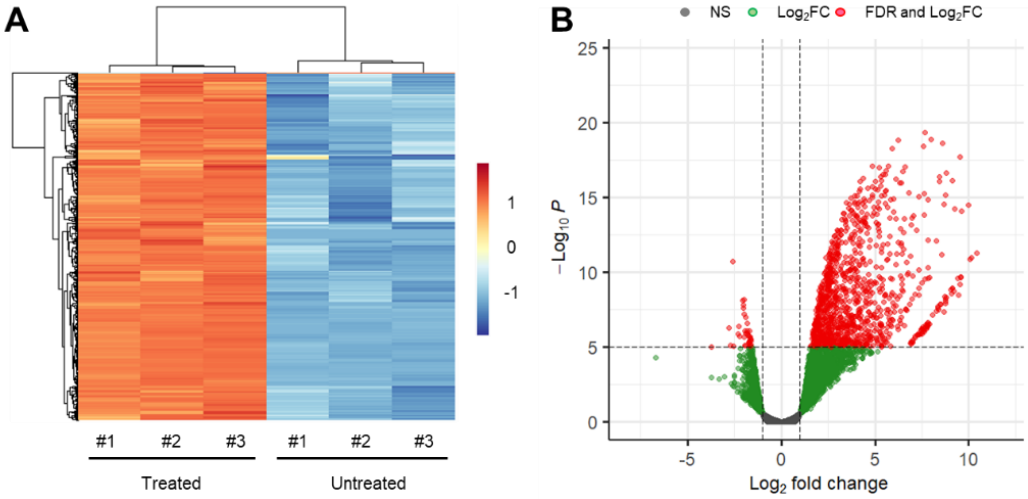


Figure 10. The DEGs of poly (I:C) treated HeLa cells were identified.

(A) The heatmap showing the top 500 DEGs with highest varied expression between poly (I:C) treated and untreated HeLa datasets. Scale represents normalized Z-scores generated by R pheatmap from blue-white-red. (B) Volcano plots of DEGs of poly (I:C) treated vs. untreated HeLa datasets. Green dots represent genes with fold-change ≧ 2; red dots represent genes with fold-change ≧ 2 and FDR < 0.05; grey dots represent genes with no significant difference. The figures and figure legends are derived and modified from supplementary figure 4 in the original article.

Considering the gene annotation of *E. fuscus* was still incomplete, I decided to be focus on the orthologs only present in both *E. fuscus* and human for the subsequent analyses. To identify shared upregulated DEGs in bat and human cells, I applied the conversion table to upregulated DEGs in Efk3B and EnK cells to obtain the corresponding human annotations. As a result, 406 and 754 upregulated

DEGs in Efk3B and EnK cells were converted into 277 and 585 human annotations, respectively. Similarly, only upregulated DEGs in HeLa cells that possessed *E. fuscus* annotations (880/2100) were remained for comparison. As the result showed, only 68 upregulated DEGs were common in the bat and human data sets (Fig. 11).

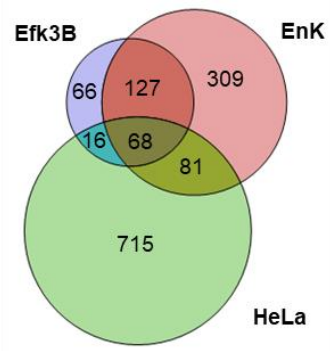


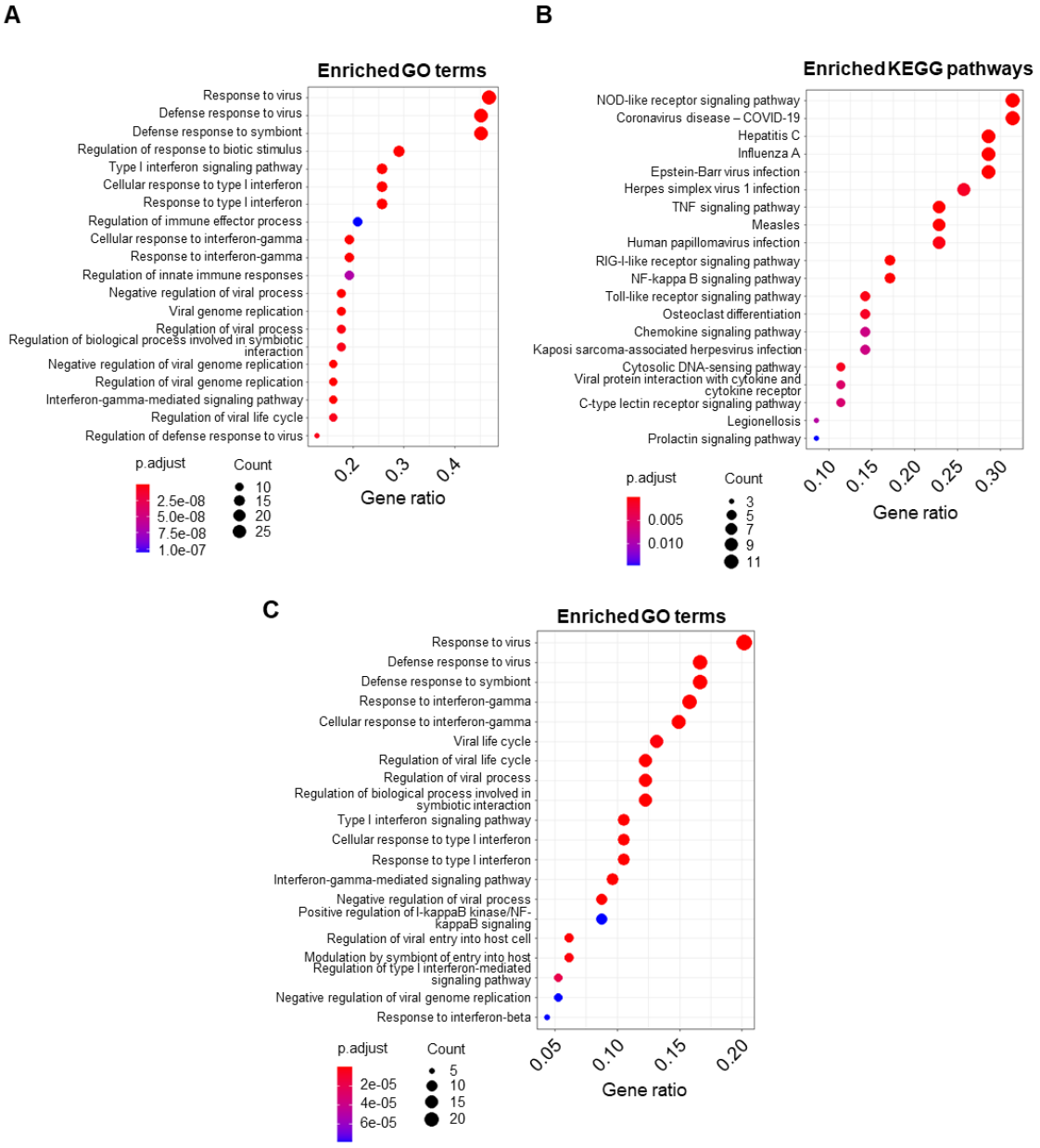
Figure 11. The common DEGs in induced *Eptesicus* bat cells and HeLa cells were identified.

Venn diagram of upregulated DEGs in Efk3B, EnK, and HeLa cells showing the common DEGs. For comparison, the annotations of *E. fuscus* were converted into human annotations. The figure and figure legends are derived and modified from figure 4A in the original article.

Subsequently, I performed GO and KEGG analyses to identified biological functions and pathways related to the upregulated DEGs. According to the results, the 68 upregulated DEGs were revealed to enrich in the terms that were similar to those of the 204 DEGs in the *Eptesicus* bat, functioning in defense responses to viruses, cellular responses to type I interferon, and virus-related pathways (Fig. 12A and 12B). The results suggest that the main mechanisms in response to poly (I:C) might be similar between *Eptesicus* bats and humans. I further analyzed the 127 uniquely upregulated DEGs in *Eptesicus* bat cells and found that GO term "the cellular response to interferon-gamma" (GO:0071346) was enriched (Fig. 12C). These data suggest that specific immune response, in particular the interferon-gamma (IFN-γ) pathway, might be more active in *Eptesicus* bat cells than in HeLa cells.

Figure 12. Different immune pathways were activated in *Eptesicus* bat cells and HeLa cells.

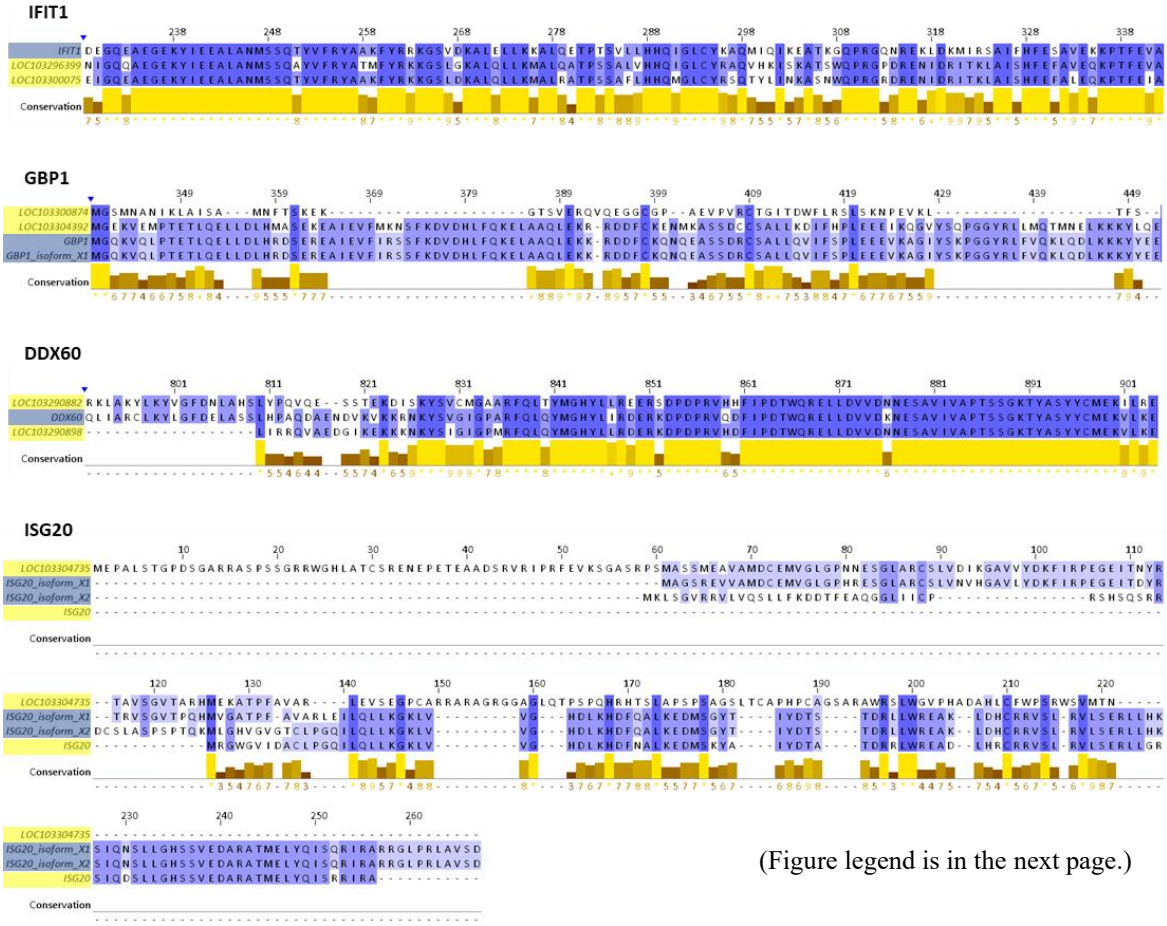
(A) The top enriched GO term biological processes of common upregulated DEGs of *Eptesicus* bat and human cells. The red the dot, the smaller the adjusted p value. The size of the dot represents the number of genes involved in the term, and the gene ratio represents the proportion of genes occupying the whole term. (B) As per (A) for the top KEGG pathways of common up-regulated DEGs of *Eptesicus* bat and human cells. (C) As per (A) for the top GO term biological processes of unique upregulated DEGs of *Eptesicus* bat cells. The figure and figure legends are derived and modified from figure 4B, 4C, and supplementary 5 in the original article. (Figure is in the next page.)



2.4 A number of upregulated ISGs were shown to exist in multi-copies

Notably, among the above 68 DEGs, there were four cases that two *E. fuscus* genes were converted into one same human gene annotation: LOC103290882 and LOC103290898 into DDX60; LOC103296399 and LOC103300075 into IFIT1; LOC103300874 and LOC103304392 into GBP1; ISG20 and LOC103304735 into ISG20. Intriguingly, I found that the nucleotide and protein sequence similarities between each copy were not always high when validating via BLASTn and BLASTp. In detail, LOC103296399/LOC103300075 (IFIT1) and LOC103300874/LOC103304392 (GBP1) were

highly similar between each copy, possessing around 90% nucleotide and protein identity; LOC103290882/LOC103290898 (DDX60) only showed 73.10% nucleotide and 49.22% protein identity; ISG20/LOC103304735 were surprisingly unable to be aligned. Parts of result of multiple sequence alignment of *E. fuscus* genes with corresponding human genes was shown in figure 13. Furthermore, LOC103290882/LOC103290898 (DDX60) closely located among *E. fuscus* genome (data not shown). This implied that these specific ISGs might duplicate and become multi-copies, which might afterward undergo mutations, insertions, or deletions throughout the evolutionary selections. However, whether those duplicated ISGs are participating in the same pathways, or they are actually playing different roles need more studies. To sum up, this data indicates that various ISGs in *Eptesicus* bats exist as multi-copies, which are simultaneously up-regulated reacting to the activation of immune responses.



(Figure legend is in the next page.)

Figure 13. Mutations, insertions, or deletions are shown in comparison of sequence similarity of *E. fuscus* multi-copies genes and corresponding human orthologs.

Parts of protein sequence similarity of *E. fuscus* IFIT1, GPB1, DDX60, and ISG20 with human orthologs. The multiple sequence alignment was performed by EMBL-EBI Clustal Omega and visualized using Jalview. The sequences are ordered by length, with the longest one at the top. The yellow boxes indicate sequences from *E. fuscus*, and the blue boxes indicate sequences from human. The darker the color of protein sequences, the higher the identity percentage. The conservation scores are calculated based on AMAS method, and star (*) means being conserved.

2.5 Unannotated transcript levels were upregulated after poly (I:C) treatment

As the gene annotation of *E. fuscus* was yet completed, I supposed that the DEG analysis was unable to identify all upregulated genes in poly (I:C) treated samples. To address this shortcoming, I applied *de novo* assembly to Efk3B data sets to obtain the transcriptomes, and I identified upregulated but yet annotated transcripts in response to poly (I:C) treatment. A total of 530 upregulated "genes" were recognized, consisting of 1577 transcripts (isoforms) (Fig. 14A). To identify and remove known transcripts, I first applied *E. fuscus* gene annotations, and 1126 transcripts (900 proteins and 226 RNAs) were aligned and identified (Fig. 14B). To further identify the existing orthologs in other species, I next aligned the remaining transcripts to 1) nonredundant protein (nr) and 2) nucleotide (nt) databases from NCBI, using BLASTx and BLASTn, respectively (details in Materials and Methods). The top hit of a transcript from the BLAST result was considered the corresponding ortholog. As the result showed, a total of 359 transcripts were aligned to gene annotations of various species (233 proteins and 126 nucleotides; Fig. 14B in appendices). No hit was identified for the remaining 92 transcripts, and thus the possible roles were remained unknown. These data reveal large numbers of upregulated transcripts without annotation, which are not included in the current *E. fuscus* genome reference yet.

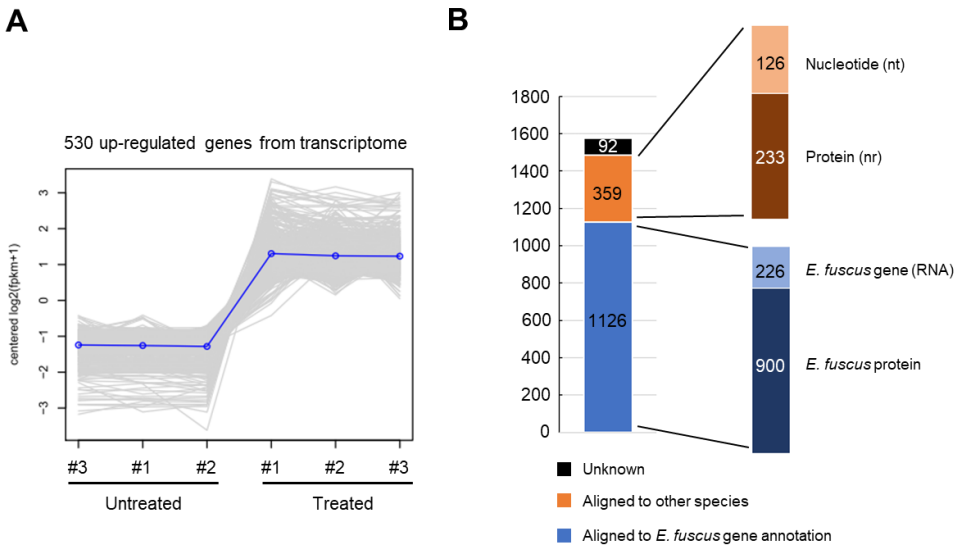


Figure 14. The up-regulated unannotated transcripts and their potential annotations were identified through alignments with existing databases.

(A) Line chart of the expression of clustered up-regulated transcripts in Efk3B de novo assembly. The transcripts with isoforms were all calculated as one gene for calculating the expression. (B) The categories and numbers of potential annotations and possible orthologs in other species of up-regulated transcripts in Efk3B datasets. The detailed steps of blast searching are described in material and methods. The figures and figure legends are derived and modified from supplementary figure 6 in the original article.

I next investigated the species possessing orthologs of unannotated *E. fuscus* transcripts. A large number of transcripts (317/359, 204 proteins and 113 nucleotides) were aligned to the gene annotations of other bat species, in particular the genus *Myotis* and genus *Pipistrellus* (Fig. 15). Moreover, only 26 of these 317 transcripts could be aligned to human gene annotations via BLASTx or BLASTn (data not shown). To conclude, these data demonstrate the upregulated but unannotated transcripts in *E. fuscus*, which might participate in the mechanisms in response to poly (I:C). Moreover, the orthologs of a large number of those unannotated transcripts exist in other bat species.

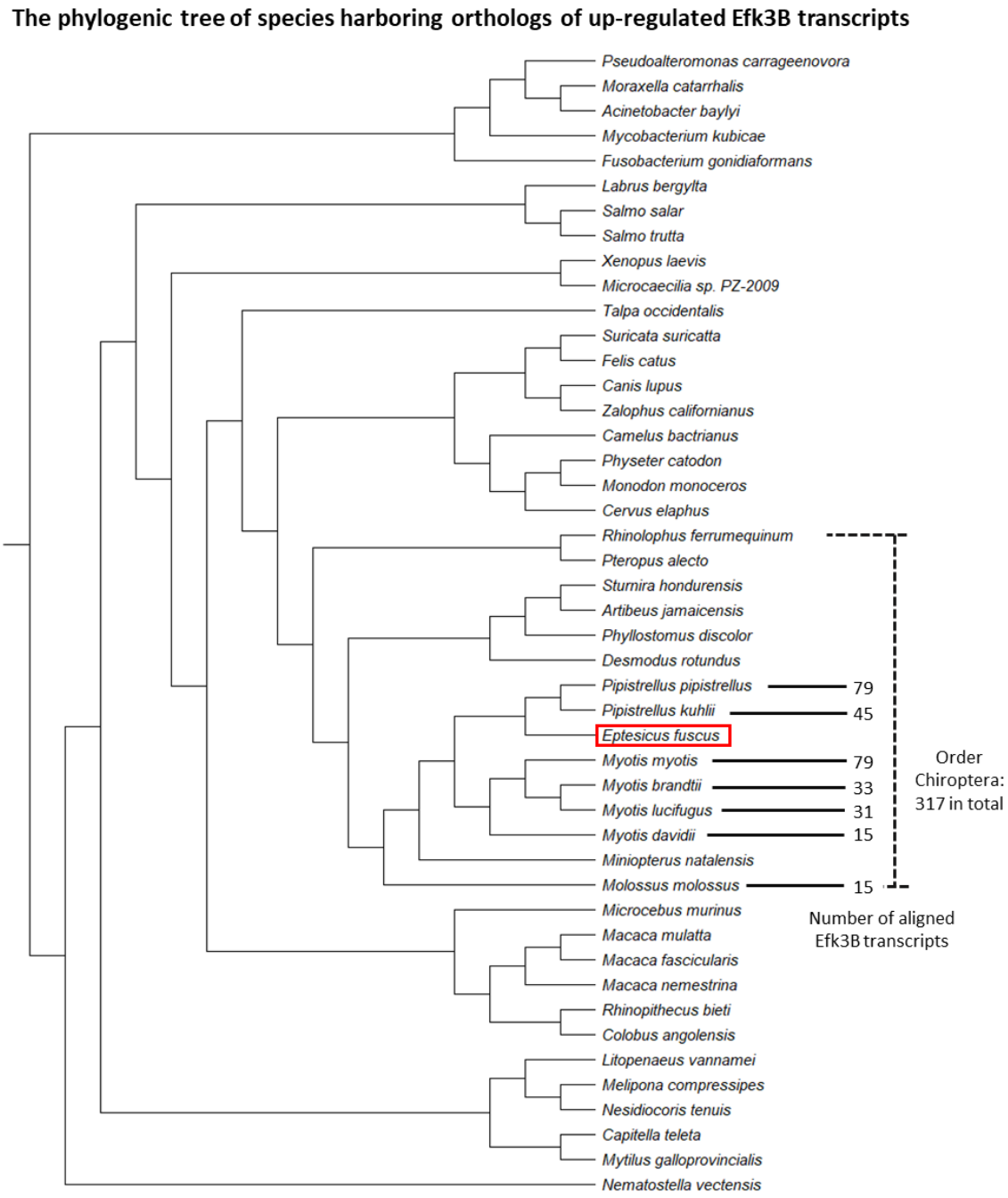


Figure 15. The phylogenic trees suggest order Chiropetera harboring most of orthologs of up-regulated unannotated Efk3B transcripts.

The species that harbored orthologs of up-regulated unannotated Efk3B transcripts were identified to generate the phylogenic tree. The number of aligned Efk3B transcripts is shown next to the species if the number is > 10. The figures and figure legends are derived and modified from supplementary figure 7 in the original article.

2.6 The basal expression level of various ISGs was higher in *Eptesicus* bat cell lines than in human cell lines

A previous study suggested that a consistently activated immune status might exist in bats, in which the basal expression of IFN-α and related ISGs in bat cell lines was shown to be higher than that in human cell lines (Zhou et al., 2016). Although the IFN-α gene is absent from or yet identified in the *E. fuscus* genome, I hypothesized that this activated immune status might also exist in other immune pathways in *Eptesicus* bats. To investigate this hypothesis, I examined ISGs expression in the Efk3B and EnK data sets without poly (I:C) treatment, following a normalized method previously proposed to obtain the basal expression level (Irving et al., 2020). I selected 12 housekeeping genes (listed in Materials and methods) that are widely conserved in mammals according to a previous study, and the expression level of each individual gene was normalized using the geometric mean of the 12 housekeeping genes (Caracausi et al., 2017). To compare the basal expression of ISGs, I incorporated different human cell lines, including the HeLa data set mentioned above and other epithelial cell lines (A549, HEK293, and HRT18) from published human cell line data sets. The gene sets consisting of ISGs were obtained from Gene Set Enrichment Analysis (GSEA) for comparison.

First, I intended to investigate the basal expression of critical and central components participating in interferon signaling pathways. The gene set “overview of interferon-mediated signaling pathway” was selected, which comprised the main genes involved in interferon signaling pathways, including various interferons, corresponding receptors, kinases, and downstream components. I found that the basal expression of selected ISGs in Efk3B and EnK cells was not obviously different from that in human cells (Fig. 16).

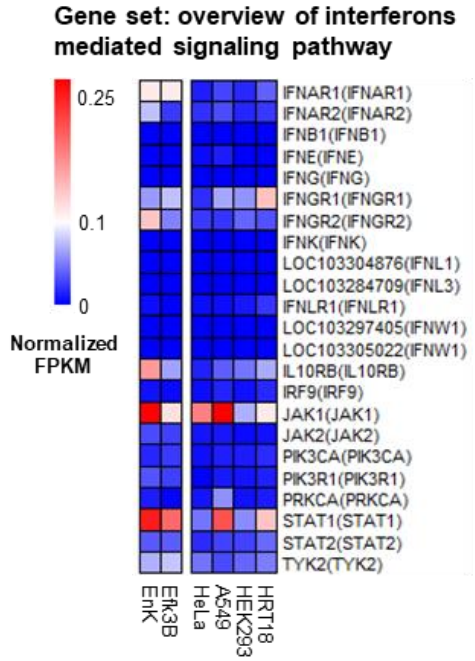


Figure 16. The basal expression levels of main regulators of interferon signaling pathways were not obviously different between *Eptesicus* bat cells and HeLa cells.

Heatmap of gene expression of the gene set "overview of interferon mediated signaling pathway" in untreated *Eptesicus* and human cell lines in the basal state. Scale represents normalized FPKM from blue-white-red. Each FPKM was normalized by dividing the geometric mean of 12 housekeeping gene FPKMs. The gene symbols outside the brackets are from *E. fuscus*, and those inside the brackets are corresponding human orthologs. The figure and figure legends are derived and modified from figure 5A in the original article.

According to the previously finding on IFN-α and related ISGs, I supposed that the components participating in type I IFN signaling pathways in *Eptesicus* bats might similarly show high basal expression. I therefore examined gene sets "IFN-β targets" and "upregulated genes in response to interferon α". As expected, several genes from these two gene sets showed higher basal expression levels in Efk3B and EnK. The expression of various hub genes of innate immunity and the antiviral response, such as DDX58 (RIG-I), IFIT2, and IFIT3, were obviously higher in bat cell lines than that in human cell lines (Fig. 17). These data indicate that specific immune pathways might be more activated in *Eptesicus* bats than in humans before induction of immune response. However, the details have not been fully clarified and require further investigation yet.

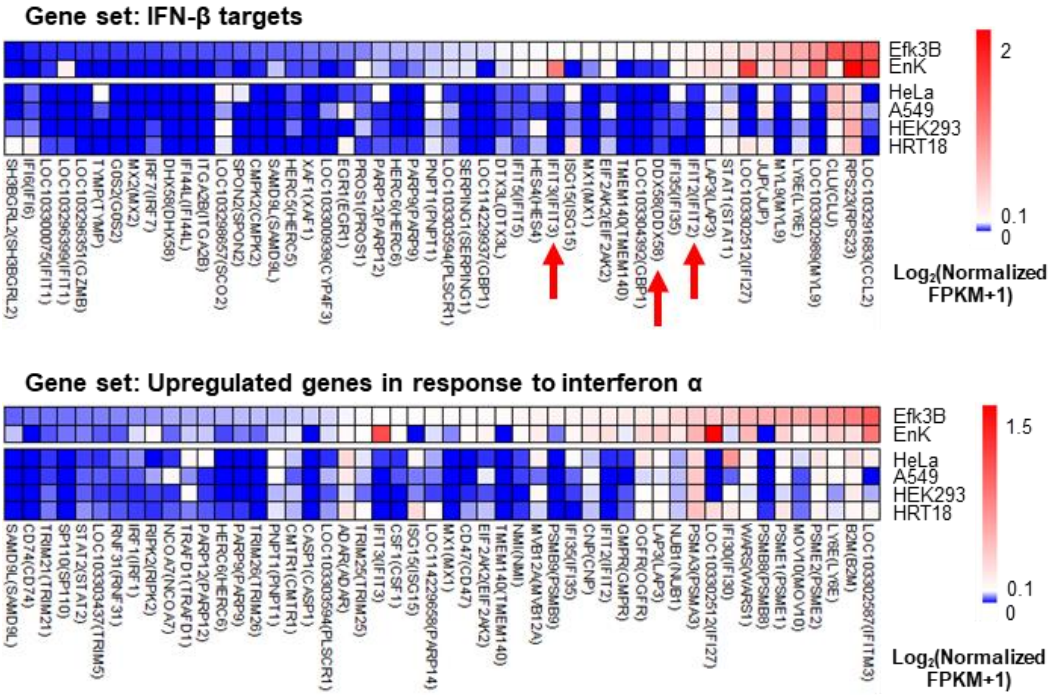


Figure 17. The basal expression levels of various interferon-stimulated genes (ISGs) before induction were higher in *Eptesicus* bat cells than in HeLa cells.

Heatmap of gene expression of the top 50 genes of the gene set "IFN-β targets" and "upregulated genes in response to interferon α" in untreated *Eptesicus* and human cell lines in the basal state, sorted by Efk3B expression. Scale represents Log_2(Normalized FPKM+1) from blue-white-red. Each FPKM was normalized by dividing the geometric mean of 12 housekeeping gene FPKMs. The gene symbols outside the brackets are from *E. fuscus*, and those inside the brackets are corresponding human orthologs. The figure and figure legends are derived and modified from figure 5B and 5C in the original article.

Due to the aforementioned discovery, I assumed that some genes might be in an activated state with high basal expression before poly (I:C) treatments. Moreover, the increased expression after induction was insufficient for the genes to be recognized as DEGs, and thus these genes were not identified and shown in previous DEG analysis (Fig. 7 and 8s). I revisited the results of the DEG analysis, and again examined the DEGs that were upregulated only in HeLa cells but not in bat cell lines. I generated a customized gene set using uniquely upregulated DEGs in HeLa cells, and the induced (treated) and basal (untreated) expression levels of Efk3B, EnK, and HeLa cells were compared in parallel (Fig. 18). Consistent with our assumption, the basal expression level of the gene set was

overall higher in Efk3B and EnK cells than in HeLa cells. Moreover, the increased level of expression after poly (I:C) treatments in bat cell lines was not as high as in HeLa cells. This finding suggests that the existing high basal expression level hindered these genes from being recognized as upregulated DEGs after poly (I:C) induction. To sum up, different from being induced after poly (I:C) treatment as in HeLa cells, specific pathways, such as pathways including IFIT2 and IFIT3, might already be in activated status under normal conditions in bat cell lines.

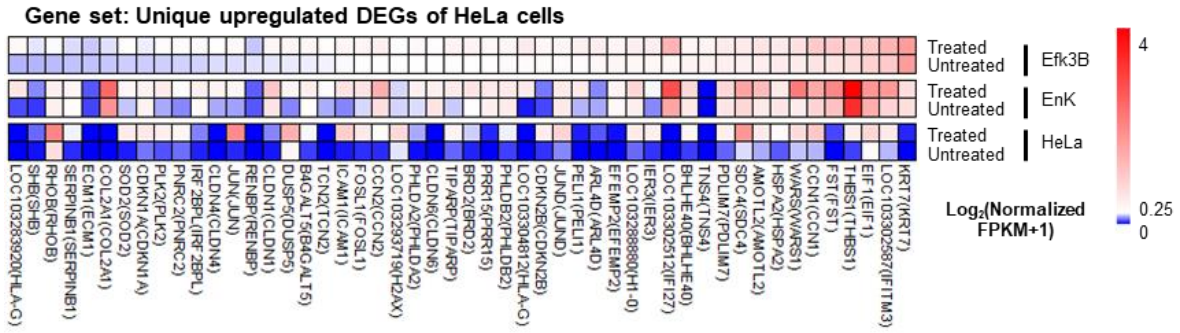


Figure 18. Several HeLa-unique upregulated DEGs showed higher basal expression level in *Eptesicus* bats cells than that in human cells.

Heatmap of the gene expression of HeLa-unique upregulated DEGs. Data from poly (I:C) treated and untreated Efk3B, EnK, and HeLa cells are shown in parallel. The order is sorted by untreated Efk3B cell expression levels, and the top 50 results are shown. Scale represents Log_2(Normalized FPKM+1) from blue-white-red. Each FPKM was normalized by dividing the geometric mean of 12 housekeeping gene FPKMs. The gene symbols outside the brackets are from *E. fuscus,* and those inside the brackets are corresponding human orthologs. The figure and figure legends are derived and modified from figure 5D in the original article.

2.7 Differentially expressed transposable elements (DETs) were identified in Efk3B and EnK cells

As a previous study has shown, the expression of several TEs is significantly increased after induction of the antiviral response (Macchietto et al., 2020). Following this finding, I investigated the landscape and expression profiles of repeat elements in response to poly (I:C) treatment in *E. fuscus* genome. Although the composition of TEs in *E. fuscus* genome have been described in previous studies,

detailed annotations are still lacking (Platt et al., 2014; Vandewege, Platt, Ray, & Hoffmann, 2016). To fulfill this gap, I applied RepeatModeler and RepeatMasker for repeat element identification to generate customized TE annotations. I removed the repeat elements that were classified as structural RNA, and the remaining repeat elements were used to construct customized TE annotations. In our identification, long interspersed nuclear elements (LINEs) and long terminal repeats (LTRs) represented 13.87% and 5.17% of *E. fuscus* genome, respectively; the proportion of short interspersed nuclear elements (SINEs) was similar, constituting 5.31% of the genome; DNA transposons (class DNA refers to class "DNA" + class "RC" in the later analysis) occupied 7.36% of the genome (Fig. 19A). Numerically, LINEs comprised 43.42% of all TEs and represented the largest TE class in *E. fuscus* genome. DNA transposons and LTRs were the second and third largest TE classes, constituting 29.35% and 14.66% of the TEs, respectively (Fig. 19B). These customized TE annotations were applied to the subsequent analysis.

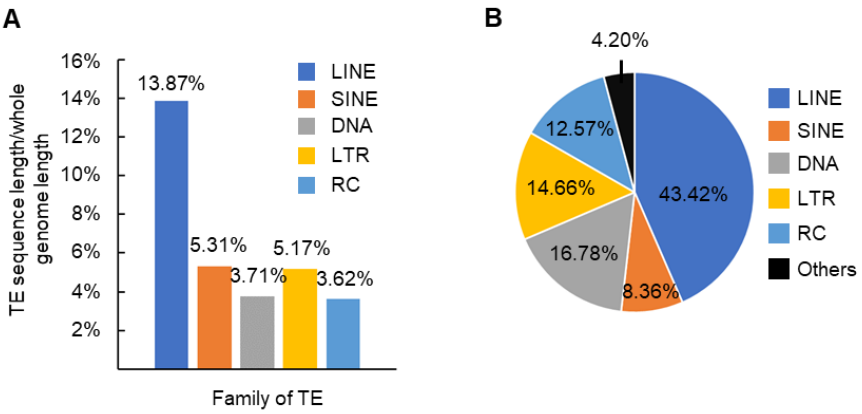


Figure 19. The transposable elements (TEs) in *E. fuscus* genome were identified and classified.

(A) Bar chart of predicted TEs identified the *E. fuscus* genome reference. The percentage represents the proportion that a class of TE occupies among the whole genomes. (B) Pie chart of the composition of predicted TEs. The percentage represents the ratio of a class of TEs to all predicted TEs. The figure and figure legends are derived and modified from figure 6A and 6B in the original article.

To illustrate TE expression, I applied the TEtranscripts package, which calculated raw read counts and performed subsequent DEG analysis (Jin, Tam, Paniagua, & Hammell, 2015). A total of 606 (602 up- and 4 downregulated) and 442 (416 up- and 26 downregulated) TEs were recognized as differentially expressed TEs (DETs) in Efk3B and EnK cells, respectively (Fig. 20A). Intriguingly, although DNA transposons are just the second largest class in *E. fuscus* genome, they showed to be the largest group of upregulated elements in both Efk3B (245) and EnK (158) cells. Moreover, a large number of common upregulated TEs in Efk3B and EnK cells were identified, indicating that the induction of TEs after poly (I:C) treatments might not be random events in *Eptesicus* bat cells (Fig. 20B).

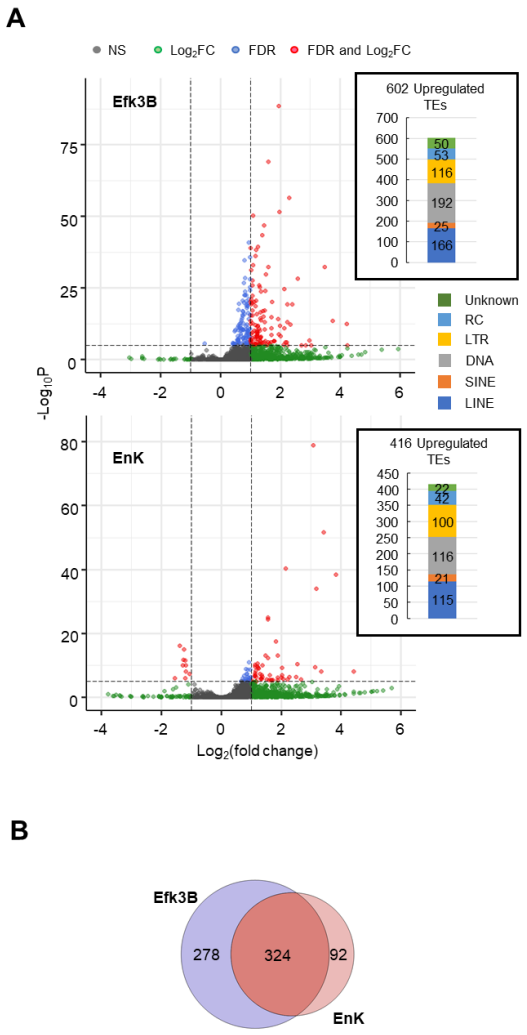


Figure 20. Groups of transposable elements (TEs) were differentially expressed in Efk3B and EnK cells after poly (I:C) treatments.

(A) Volcano plots and bar charts of differentially expressed TEs (DETs) of poly (I:C) treated vs. untreated Efk3B and EnK data sets. Green dots represent genes with fold change ≥ 2; blue dots represent genes with FDR < 0.05; red dots represent genes meeting both criteria; gray dots represent genes with no significant difference. Bar charts show the numbers of upregulated TEs of each class. (B) Venn diagram of upregulated DETs of Efk3B and EnK cells. The figure and figure legends are derived and modified from figure 6C and 6D in the original article.

2.8 The microRNA profile was influenced by poly (I:C) treatment

In previous research, specific miRNAs, which were involved in the modulation of immune responses, were reported to be induced after the activation of immune responses (Nejad et al., 2018). To investigate the profile of miRNAs in response to poly (I:C), I performed small RNA sequencing using untreated and poly (I:C) treated Efk3B cells. A flow for predicting and analyzing the expression of novel miRNAs was established (Fig. 21). I first analyzed the length distribution of whole small RNA pool. As a result, a canonical peak at approximately 22 nt was identified in both untreated and treated data sets, suggesting the existence of miRNAs (Fig. 22).

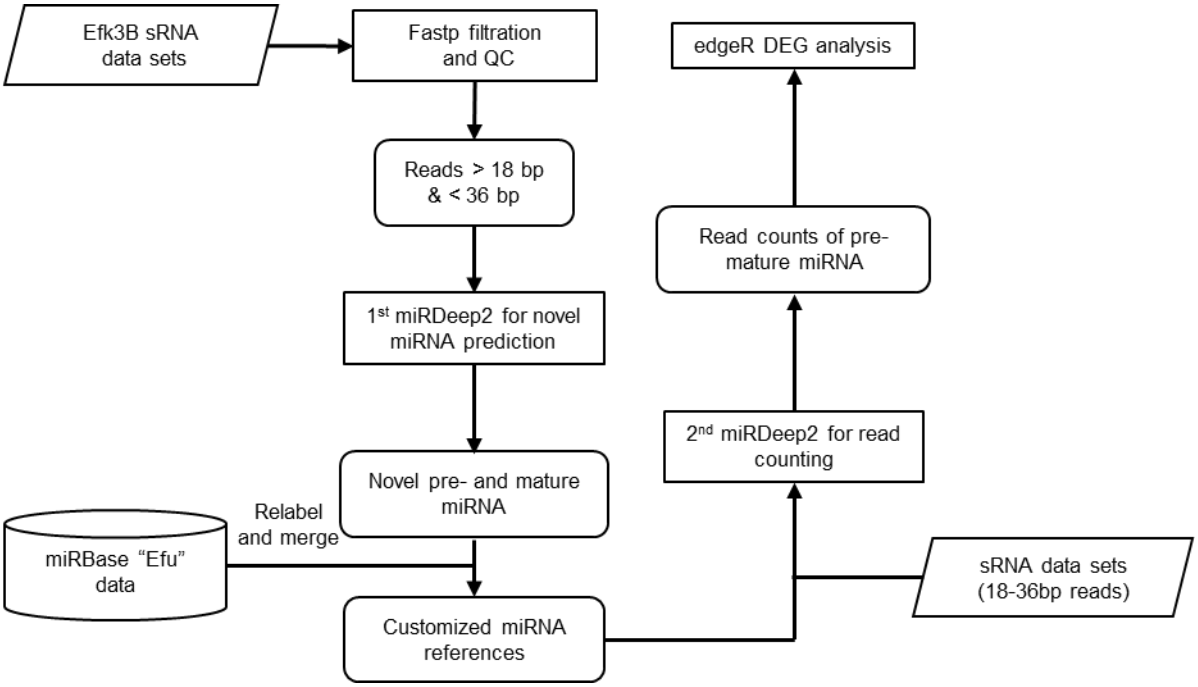


Figure 21. The workflow of processing Ef3kB and EnK small RNA-seq datasets and performing miRDeep2 novel miRNA prediction.

The raw data of small RNA-seq were filtered and only reads between 18 to 36 bp were remained. The first miRDeep2 was performed for identifying novel pre- and mature miRNA. The novel miRNAs were combined with existing miRbase "Efu" data to generate customize miRNA references, and the second miRDeep2 was performed for read counting. R Package edgeR was used for DEG analysis. The figures and figure legends are derived and modified from supplementary figure 8 in the original article.

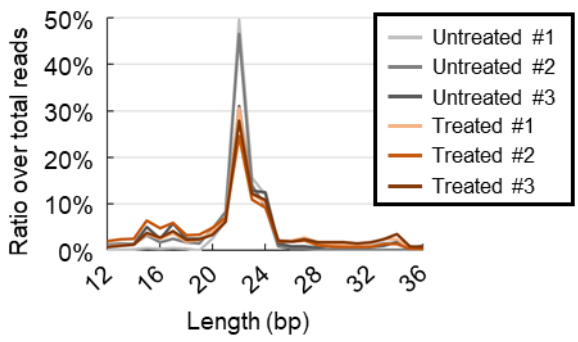
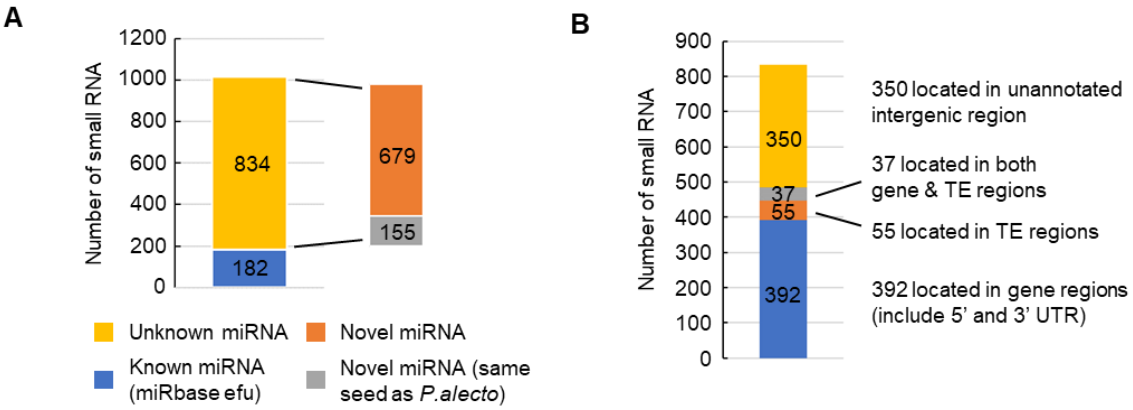


Figure 22. The length distribution of small RNA in Efk3B mainly peaked at 22 nt.

Line chart of the length distribution of Efk3B small RNA-seq data sets. The percentage represents the ratio of processed reads of each length to total processed reads. The figure and figure legends are derived and modified from figure 7A in the original article.

I next predicted novel miRNAs using miRDeep2. A total of 182 known mature miRNAs from miRBase as well as 774 novel miRNAs and 834 novel pre-miRNAs were identified in our data sets (Fig. 23A). To be noted, the seeds of 155/834 novel pre-miRNAs have been reported in *P. alecto*, indicating that orthologs of several novel *E. fuscus* miRNAs might exist in *P. alecto*. Subsequently, to elucidate the locations and potential origins of novel miRNAs, I mapped novel pre-miRNAs to *E. fuscus* genome. Approximately, half of the novel miRNAs (392/834) were aligned to genic regions, and 55 novel miRNAs were potentially generated from TE regions (Fig. 23B). Besides, 37 miRNAs simultaneously overlapped genic and TE regions, in which represented gene-TE chimeric sequences or TEs locating in gene structures. Finally, a large number of miRNAs (356/834) were remarkably produced in unannotated intergenic regions.

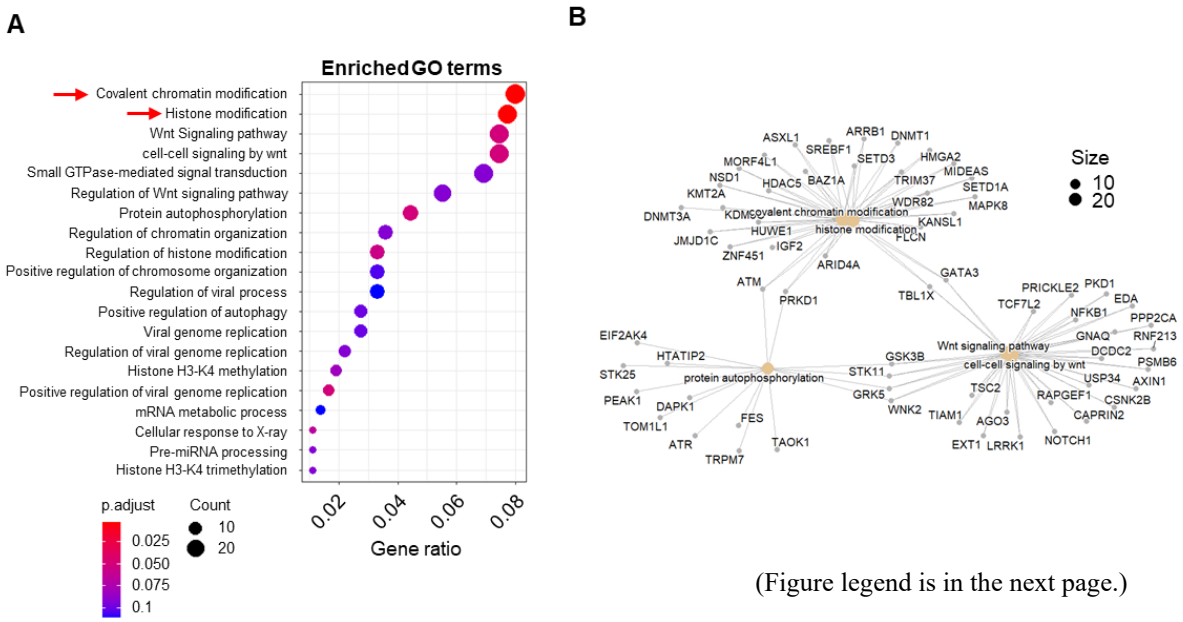


(Figure legend is in the next page.)

Figure 23. Novel miRNAs were identified, and the main locations of novel miRNAs were shown to be gene and unannotated intergenic regions.

(A) Bar chart of the numbers of known and novel miRNAs expressed in Efk3B data sets. Known miRNAs were derived from miRBase, and novel miRNAs were predicted via miRDeep2. (B) Bar chart of the numbers of miRNAs in corresponding locations among the *E. fuscus* genome in the Efk3B data sets. The figure and figure legends are derived and modified from figure 7B and 7C in the original article.

Following above data, I examined the genes that potentially produced novel pre-miRNAs. In total, 400 genes were characterized to hold 429 (392 genic + 37 genic & TE) novel pre-miRNAs that overlapped with genic regions. For subsequent analyses, the 400 genes were applied to the homemade conversion table and converted into 382 human annotations. GO term analysis of these 382 genes showed that only “covalent chromatin modification” (GO: 0016569) and “histone modification” (GO: 0016570) were significantly enriched (Fig. 24A). Furthermore, these genes were shown to participate in diverse functions and were not well grouped according to the network analysis (Fig. 24B). To sum up, these data suggest that novel miRNAs are produced from various genic regions with diverse functions, and numerous miRNAs are originated from unannotated regions that are still unclear.



(Figure legend is in the next page.)

Figure 24. Predicted novel miRNAs in gene regions was mainly originated from ISGs with diverse functions.

(A) The top enriched GO term biological processes of genes harboring novel miRNAs in Efk3B data sets. The red the dot, the smaller the adjusted p value. The size of the dot represents the number of genes involved in the term, and the gene ratio represents the proportion of genes occupying the whole term. (B) The gene-concept network of genes harboring novel miRNAs and corresponding enriched GO terms. The size of the dot represents the number of genes. The figure and figure legends are derived and modified from figure 7D and 7E in the original article.

To investigate the expression of novel miRNAs and perform DEG analysis, I constructed a customized miRNA library incorporating above predicted novel miRNAs. For this purpose, I decided to remove one untreated data set, which was considered an outliner due to its low mappability. The DEG analysis showed increased expression of 16 miRNAs (13 novel + 3 known), which were generated from 20 pre-miRNAs (13 novel + 7 known). Notably, some upregulated miRNAs were shown to be produced from several ISGs, including SLC9A2, DLG3, MX1, ZC3H12A, RNF213, and ADAR (Table 1). Coinciding with this result, MX1, ZC3H12A, RNF213, and ADAR were DEGs that were identified in previous DEG analysis in Efk3B mRNA-seq data sets, indicating that the upregulated miRNAs might be the consequences of activation of immune responses. In summary, the activation of innate immune responses after poly (I:C) treatments leads to production of groups of novel miRNAs, affecting the miRNA profile in Efk3B cells. However, whether those novel miRNAs are involved in the regulation of innate immune responses or not is still unclear.

miRNA	Pre-miRNA	Gene location (Efk3B Symbol)
efu-miR-1249	efu-mir-1249	KIAA0930
efu-miR-9215a efu-miR-9215b	efu-mir-9215a-1	SLC9A2†
	efu-mir-9215a-2	
	efu-mir-9215a-3	
	efu-mir-9215a-4	
	efu-mir-9215a-5	
	efu-mir-9215b	DLG3†
efu-novel-miR-65	efu-novel-mir-65	MX1†
efu-novel-miR-109	efu-novel-mir-109	RNF213†
efu-novel-miR-142	efu-novel-mir-142	YIPF3
efu-novel-miR-225	efu-novel-mir-225	
efu-novel-miR-250	efu-novel-mir-250	
efu-novel-miR-327	efu-novel-mir-327	LIX1L
efu-novel-miR-340	efu-novel-mir-340	
efu-novel-miR-368	efu-novel-mir-368	
efu-novel-miR-479	efu-novel-mir-479	
efu-novel-miR-555	efu-novel-mir-555	ZC3H12A†
efu-novel-miR-633	efu-novel-mir-633	GPKOW
efu-novel-miR-687	efu-novel-mir-687	
efu-novel-miR-712	efu-novel-mir-712	ADAR†

†These genes are ISGs, as defined by the Interferome database (Rusinova et al., 2013).

Table 1. Differentially expressed miRNAs, pre-miRNAs and their gene locations in Efk3B cell. The table is derived from table 1 in the original article.

Chapter 3

Discussion and Conclusion

Various bat species have been reported to show distinct characteristics of innate immune systems, such as distinguished signaling pathways and high basal expression of ISGs (Ahn et al., 2016; Xie et al., 2018; Zhou et al., 2016). In recent years, high-throughput sequencing has become gradually common, and increased data sets have been published, which facilitate the studying on not only model species but also non-model species. However, data sets from many non-model species, including bats of the genus *Eptesicus*, are still relatively lacking and limited. In this study, I applied two *Eptesicus* bat cell lines after poly (I:C) treatments and aimed to reveal the innate immune responses of those bat cells. I comprehensively analyzed the gene expression profiles in two cell lines, including mRNA, miRNA, and TE expression profiles. The expression profiles between the two bat cell lines were shown to be similar to a certain extent after poly (I:C) treatments, but several genes were uniquely upregulated in each cell line (Fig. 7 and 8). I also revealed that the upregulated genes were distinguishable between the *Eptesicus* bat cells and human epithelial cells (Fig. 11). Moreover, compared to human cells, the basal expression levels of various ISGs were higher in bat cells (Fig. 17 and 18). This report shows that innate immune responses in bats are possibly unique from those in humans, and the results and data sets provide resources for future studies, contributing to a deeper understanding of bat immunity.

In this study, I compared the expression profiles between bat cells and human HeLa cells treated by poly (I:C), elucidating the uniqueness of immune responses in *Eptesicus* bat (Fig. 11). Although common immune pathways were shared between bat and human cells, many genes were uniquely upregulated in *Eptesicus* bat cells, in particular those related to IFN-γ pathways (Fig. 12). IFN-γ is known to participate in innate immune responses and to involve in the stimulation of adaptive immune responses (Lee & Ashkar, 2018), including antiviral activity against mononegaviruses, such as the Ebola virus and Hendra virus, in mouse models and bat cell lines (Janardhana et al., 2012; Lee & Ashkar, 2018; Rhein et al., 2015). In a previous study, Kuzmin et al. also revealed that IFN-γ demonstrated the antiviral effect against filoviruses infections along with IFN-α and IFN-β in *R. aegyptiacus* cell lines, as well as the expression of bat type II IFN could only induce the innate immune responses in rousette cells but not in human cells, suggesting a host species specificity (Kuzmin et al., 2017). The activation of IFN-γ pathways in *Eptesicus* bat cell lines in this study might be supportive evidence of this species

specificity. However, surprisingly, the expression of IFN-γ transcripts (IFNG) was not detected in either untreated or poly (I:C) treated samples of *Eptesicus* bat cells (data not shown). How the IFN-γ pathways are activated in response to poly (I:C) treatment and whether these pathways practically contribute to antiviral responses in *Eptesicus* bats require further investigation. It should be noted that as I applied cell lines derived from bat or human epithelial cells to our studies, the possibility that the above characteristics might be attributed to the differences in cell type but not species cannot be excluded. Further studies are required to illustrate the roles of those uniquely upregulated genes in bat cells.

Higher basal expression was shown in some immune-related genes, and it might be one of the unique characteristics of bat immune strategies. Although IFN-β was not constitutively expressed in bat cells according to our data, I indeed identified higher expression levels of putative ISGs in bat cells than in human cells (Fig. 1, 17, and 18). It is thus possible that the high basal expression of ISGs involved in type I IFN pathways might be a critical factor in response to viral infection in bats, which in the end lead to the virus-tolerating phenotype. The detailed mechanisms of regulation and potential outcome of the high basal expression require further studies.

I also observed different patterns in gene expression between Efk3B and EnK cells. A total of 409 and 834 genes were identified as DEGs in Efk3B and EnK cells, respectively, and 274 upregulated genes were shared in both cell lines (Fig. 7B). Gene ontology analyses showed that general biological processes related to immune responses were commonly induced, and additional pathways might be uniquely activated in EnK cells (Fig. 8 and 9). However, as mentioned above, although both Efk3B and EnK cells were derived from the kidney and are supposedly epithelial cells, it is also not clear if the differences were attributable to the distinctness between species and/or cell types. Further studies are required to reveal the biological roles of these differences.

Our data showed that noncoding RNAs (ncRNAs) are induced after poly (I:C) treatments, suggesting that ncRNAs might be involved in the innate immunity of bat cells. Among 274 upregulated DEGs shared in Efk3B and EnK cells in response to poly (I:C), 70 genes are termed as uncharacterized ncRNAs in NCBI. In previous studies, ncRNAs have been shown to be involved in various aspects of immunity. For example, in mouse embryonic fibroblasts, ncRNAs were shown to inhibit NF-κB DNA

binding activity and further downregulate the inflammatory response; specific ncRNAs were suggested to react to IL-1β activation and subsequently modulate inflammatory activity in human epithelial cells and fibroblasts (Hadjicharalambous & Lindsay, 2019). However, the potential roles of ncRNAs in bats are still relatively unknown. Our data provide clues to investigating the functions of these upregulated ncRNAs in *Eptesicus* bat cells.

I preliminarily revealed the expression profiles of miRNAs in response to poly (I:C) in *Eptesicus* bat cell lines (Fig. 23 and 24). Previous studies have suggested that miRNAs participate in the modulation of innate immunity (Nguyen et al., 2018). As mentioned, some miRNAs were shown to regulate immune responses in feedback loops or directly target viral RNAs, inhibiting viral replication. Although the potential roles of miRNAs in bat immunity are still unclear, as the induced expression patterns have shown in this study, it is possible that some of the novel miRNAs play roles in the regulation of innate immune responses in *Eptesicus* bat cells. Further study is necessary to clarify the possible functions of newly-identified novel miRNAs.

Similar to miRNAs, TEs are also suggested to participate in immune responses. However, the TE expression profiles in response to induction of immune responses have not yet been illustrated. I indeed detected obvious upregulation of TEs, many of which were common in Efk3B and EnK cells (Fig. 20). The recent study indicates that TE expression is upregulated during virus infection, which further activates innate immune responses in human and mouse cells (Macchietto et al., 2020). Other earlier studies suggested that the upregulation of TE expression results in IFN activation in cancer cells and pluripotent cells (Grow et al., 2015; Mu, Ahmad, & Hur, 2016; Roulois et al., 2015). Nevertheless, the roles of TEs in bat immunity is relatively unknown. Whether these upregulated TEs are involved in the regulation of immune responses or are just concomitant with activated immune responses requires further investigation.

Our assembly data represent a valuable resource for gene annotations of the genomes of *Eptesicus* bats. Although the annotations of the *E. fuscus* genome are provided in NCBI database, it is still incomplete, and many annotations were not yet experimentally validated. Here, I performed *de novo* assembly using Efk3B data sets and identified 359 unannotated but upregulated transcripts after poly

(I:C) treatments (Fig. 14). Intriguingly, among the 359 unannotated transcripts, orthologs of 42 transcripts were not identified in other species of bats, suggesting the discovery of novel transcripts specifically expressing in *Eptesicus* bats. Furthermore, I identified novel miRNAs as described above (Fig. 14 and 15). Therefore, our data will be useful for more precise annotation of the *E. fuscus* genome, contributing to future studies on *E. fuscus*.

I found several up-regulated ISGs existing as multi-copies in *Eptesicus* bats, which might also point to potential unique immune responses (Fig. 7 and 13). These multi-copies ISGs are simultaneously upregulated in response to poly (I:C) treatments, suggesting their concurrent reactions to the immune responses. Previous data suggests that ISGs shared by various species are more likely to exist as multiple copies, and the ISGs could have different copy number in different species (Shaw et al., 2017). The impact of copy number variations has been studied in cancers, and the results suggest that the copy number is highly correlated with gene expression (Myhre et al., 2013; Shao et al., 2019). Another report reveals that copy number variation might be related to autoimmune diseases (Yim, Jung, Chung, & Chung, 2015). However, the exact role of these multi-copies ISGs in bat cells is still unclear. Whether these multi-copies ISGs make influence on the innate immune response, playing similar or distinct roles in *Eptesicus* bat, required further research.

Genome references of *E. nilssonii* are still lacking, and it is thus difficult to analyze the data obtained from EnK cells. In this study, I analyzed the data sets obtained from EnK basing on the *E. fuscus* genome. This strategy simplifies the processes and provides preliminary results. However, it indeed has limitations, such as expectedly higher mismatches during mRNA-seq mapping. These mismatches could further affect the results of DEG analysis and interrupt GO term analysis. Thus, unique patterns in *E. nilssonii* might not have been identified due to above limitation. Additionally, as miRDeep2 prediction requires the genome sequence, I could not perform miRDeep2 analysis using EnK data sets. Another limitation is that the gene expression profiles are definitely influenced by the cell types and culture conditions. To our knowledge, the published HeLa data set, using epithelial cell line and poly (I:C) treatment, was the one most similar to our samples when this study was undergone. However, Efk3B and EnK cells are kidney-derived cells, different from cervical cancer HeLa cells.

Furthermore, these cell lines were grown in different culture conditions. It is necessary to include more cell types, culture conditions, and treatments for further analysis in future studies. Therefore, as an increasing number of high-throughput data studies and relatively complete genome references become available in the future, a deeper analysis could be performed to reveal additional information.

In conclusion, this study provides supportive and novel perspectives into the research of bat immunity via studying *Eptesicus* bats, contributing to the clarification of bat immune functions. Further understanding the mechanisms by which bats respond to viral infection is critical for developing innovative methods and therapies for defending against pathogenic diseases in humans.

Chapter 4

Materials and methods

4.1 Cell culture

Efk3B cells (*Eptesicus fuscus* kidney epithelial cells; Kerafast, Boston, Massachusetts, USA) (Banerjee et al., 2016) were cultured in Dulbecco's Minimal Essential Medium (DMEM; Thermo Fisher Scientific, Waltham, Massachusetts, USA) containing 10% fetal bovine serum (Thermo Fisher Scientific), penicillin/streptomycin (Nacalai Tesque, Kyoto, Japan), and 1% GlutaMax (Thermo Fisher Scientific). EnK cells (*Eptesicus nilssonii* kidney-derived cells) (Horie et al., 2016) were grown in low glucose DMEM (Nacalai Tesque) with 10% fetal bovine serum (Thermo Fisher Scientific) and penicillin/streptomycin (Nacalai Tesque). Cells were maintained in a 37 °C humidified incubator with 5% CO_2.

4.2 Polyinosinic-polycytidylic acid (Poly (I:C)) and universal type I interferon (IFN) treatment

Efk3B and EnK cells were seeded in 24-well plates at concentrations of 7.5×10^4 and 2.5×10^4 cells/well for poly (I:C) and universal type I interferon treatment. For poly (I:C) treatments, a working concentration of 4 μg/mL low molecular weight poly (I:C) (Invitrogen, Waltham, Massachusetts, USA) was transfected into cells using 2 μL TransIT-X2 (Mirus Bio, Boston, Massachusetts, USA; for EnK cells) or PEI MAX (Polysciences, Warrington, Pennsylvania, USA; for Efk3B cells) mixed with 70 μL Opti-MEM™ I Reduced Serum Medium (Thermo Fisher Scientific). For IFN treatments, universal type I interferon (PBL Assay Science, Piscataway, New Jersey, USA) was diluted with culture medium to a working concentration of 1000 units/mL. Cells were harvested after 4, 8, 12, 24, and 48 hours of treatment and were used for RNA extraction. Untreated cells at 0 hours were used as control groups.

4.3 RNA extraction, reverse transcription, and qRT–PCR

NucleoSpin® RNA Plus (TaKaRa, Otsu, Japan) was used to extract total RNA following the manufacturer's instructions. A total of 200 ng of the extracted RNA was reverse transcribed into cDNA with a Verso cDNA Synthesis Kit (Thermo Fisher Scientific) with the oligo dT primer included in the

kit as per the manufacturer's directions for a 1-hour reverse transcription reaction at 50 °C.

To quantify the expression level of selected interferon-stimulated genes (ISGs), qRT–PCR was performed with Luna Universal qPCR Master Mix (New England Biolabs, Ipswich, Massachusetts, USA) using a CFX96 Touch Real-Time PCR Detection System (BioRad, Hercules, California, USA). Samples containing the above cDNA were prepared following the manufacturer's instructions. As the *E. nilssonii* genome sequence was not yet available, primers for qRT–PCR were designed based on the NCBI *E. fuscus* genome (GCA_000308155.1 EptFus1.0) as described below. After selecting the primer-targeting sites, I amplified the fragments containing the sites through conventional PCR using Efk3B and EnK cDNA samples. Through Sanger sequencing and comparison of amplicons from Efk3B and EnK samples, I designed common primer sets for the qRT–PCR assays. If necessary, additional primers were designed according to the nucleotide differences between the two cell lines. The qRT–PCR primer sequences are listed as below:

Gene	Primer sequence (5'-Sequence-3')
ACTB	Forward: CCTGGGCATGGAATCCTGTGGC
	Reverse: TCCTTCTGCATCCTGTCGGCGA
IFNB1	Forward: GCGTGCTCCGATTCCGACAGAG
	Reverse: TCCTTCTGGACCTGCTGTGGCT
MX1	Forward: TGGAGGCGCTGTCTGGTGTCTC
	Reverse: CTCAGTTTCAGCACCAGGGGGC
IFIT1-L	Forward: ACCAAAACACTGCAAATTGCTGCC
	Reverse (for Efk3B): GCTGTTCCAGCCTAGCTTGGAT
	Reverse (for EnK): GGTGTTCCAGCCGAGGTTGGAT
IFIT2	Forward: AATCCAGCAGCAGCACCCTGAC
	Reverse: TTCGGAGAGTCTGCCCATGCCA

The qPCR conditions were as follows: 10-minute initial denaturation step at 95 °C; two-step cycling for 35 cycles at 95 °C/5 sec, 60 °C/30 sec with absorbance readings between each cycle; and a final step of 95 °C/10 sec to generate the dissociation curve with absorbance readings for the dissociation curve acquired at every 0.5 °C from 65–95 °C. Two housekeeping genes (ACTB and GAPDH) were tested, and as there was no variation in normalized Cq and relative fold changes, I chose ACTB for normalization.

4.4 mRNA sequencing

The extracted total RNA was qualified using High Sensitivity RNA ScreenTape (Agilent, Santa Clara, California, USA) and 4200 TapeStation system (Agilent) and confirmed that all RIN^e values ranged from 8.7-9.9. To prepare strand-specific paired-end libraries, the extracted total RNA was first processed via a Dynabeads™ mRNA Purification Kit (Thermo Fisher Scientific) to enrich poly-A tailed mRNA. The enriched samples were quantified via Qubit® RNA HS Assay Kits (Life Technology, Carlsbad, California, US), and 1.5-15.2 ng of RNA samples were used for library preparation. The NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina® (New England Biolabs) was used for library preparation according to the manufacturer's instructions. In detail, the condition for fragmentation was 10 minutes at 94 °C; the condition of size selection was set for isolating approximately 400 bp inserts by using Sera-Mag™ (Cytiva, Tokyo, Japan), and the cycle number for PCR enrichment ranged from 12-15 cycles depending on the concentration of the RNA input. The constructed libraries were diluted 20 times and qualified using High Sensitivity D5000 ScreenTape (Agilent) and the 4200 TapeStation system. If necessary, the libraries were subjected to dual size selection with Sera-Mag™ to adjust the final library size, ranging between 450-600 bp. Finally, the libraries were quantified using Qubit® dsDNA HS Assay Kits (Life Technology) to calculate the final concentration of the samples for mRNA sequencing. Sequencing was performed using the NextSeq 500/550 Mid Output Kit v2 (300 Cycles) on the paired-end Illumina NextSeq platform (NextSeq 500, 2 × 150 bp; Illumina, San Diego, California, USA).

4.5 Preprocessing for mRNA sequencing

Raw reads were processed by fastp (v0.20.1) for adapter trimming and quality control with the following parameters: -l 35 -y -3 -W 3 -M 20 -x --detect_adapter_for_pe (Chen, Zhou, Chen, & Gu, 2018). The published HeLa cell data set (GSE130618), HEK293 cell, A549 cell, and HRT18 cell data sets (GSE165900) was downloaded from NCBI and processed with the same procedures as above.

4.6 Annotation conversion and differentially expressed gene analysis

To convert the *E. fuscus* annotation into human annotation, I generated a conversion based on BLASTp (v2.10.0+; Camacho et al., 2009). Protein annotations from *E. fuscus* (NCBI: txid 29078; GCA_000308155.1 EptFus1.0) and humans (NCBI: txid9606; GRCh38.p13) were used. The database for BLASTp was made using human protein annotation with default parameters. To simplify the output and interpretation, I performed BLASTp using the parameters -max_target_seqs 1 -outfmt 6 -evalue 1e-5 to identify the most similar orthologs between *E. fuscus* and humans. The output table contained NCBI IDs, which were further converted into gene symbols for further analysis.

As an *E. nilssonii* genome reference was unavailable, the analysis of both the Efk3B and EnK data sets was based on the *E. fuscus* genome. To perform differentially expressed gene (DEG) analysis, the processed raw reads of the EnK and Efk3B data sets were mapped to the *E. fuscus* genome, and those of the HeLa data set were mapped to the human genome (GRCh38.p13) using STAR (v 2.7.8a) with the following parameters: --outMultimapperOrder Random --outSAMmultNmax 1 (Dobin et al., 2013). SAMtools (v1.10) was used to transform the STAR results from the sam file to the bam file (H. Li et al., 2009). Afterward, raw read counts of the genes were calculated based on *E. fuscus* and human gene annotations via featureCounts (v2.0.1) with the following parameters: -p -M -O -s 2 -t exon (Liao, Smyth, & Shi, 2014). Raw read counts were next used for DEG analysis in R (v4.1.0) using the edgeR (v3.34.0) package. The threshold for identifying DEGs was set as $|\log2(\text{fold change})| \geqq 1$ and a false discovery rate (FDR) < 0.05. The visualization of heatmaps and volcano plots was performed using the pheatmap (v1.0.12) and EnhancedVolcano (v1.10.0) packages (Blighe, Rana, & Lewis, 2021; Kolde, 2019). After identifying DEGs, Gene Ontology analysis, KEGG analysis, and gene-concept network analysis were performed and visualized using the enrichplot (v1.12.2) package (Yu, 2021). Venn diagrams were generated using Venn Diagram Plotter (https://omics.pnl.gov/software/Venn-diagram-plotter).

4.7 De novo assembly and transcriptome analysis

To generate overall assembly using Efk3B data sets, I first merged all untreated and poly (I:C)

treated raw reads processed by mRNA-seq. Trinity (v2.11.0) was used for performing *de novo* assembly with the strand-specific parameter --SS_lib_type RF (Grabherr et al., 2011). The downstream quantification and analysis were performed following the instructions in the Trinity manual and using built-in scripts (Haas et al., 2013). In detail, the raw reads were aligned to *de novo* assembled transcripts using bowtie2, and the expression of transcripts was estimated by RSEM (Langmead & Salzberg, 2012; B. Li & Dewey, 2011). To identify DEGs, the expression of transcripts was counted at the gene level but not at the isoform level to simplify the process. The DEG analysis was performed using edgeR, and the threshold for identifying DEGs was set as $|\log2(\text{fold change})| \geq 2$ with a false discovery rate (FDR) < 0.001 to reduce the candidate numbers.

To identify the possible annotations of differentially expressed transcripts, I performed alignments in the following order: the transcripts were first aligned to *E. fuscus* protein annotations via BLASTx; the remaining transcripts were next aligned to *E. fuscus* mRNA annotations via BLASTn; the remaining transcripts were then aligned to the NCBI nonredundant protein database (nr) via BLASTx, and the remaining transcripts were finally aligned to the NCBI nonredundant nucleotide database (nt) via BLASTn. The transcripts without any annotation after these four-step alignments were characterized as unknown. The phylogenetic trees were generated using TimeTree with corresponding species names, and the graph was generated and modified by MEGA (Version 11.0.8) (Kumar, Stecher, Suleski, & Hedges, 2017).

4.8 Normalization for analysis of basal expression level

For the normalization of gene expression to analyze and compare the basal expression level, a normalized method was applied as previously described (Irving et al., 2020). A total of 12 housekeeping genes (gene symbols in *E. fuscus*: TBCB, CIAO2B, NOP10, GAPDH, RRAGA, LOC103289292, ACTG1, RPS27, ACTB, MIF, RPL8, and RPS18) were selected for normalization according to a previous study (Caracausi et al., 2017). To represent gene expression, “fragments per kilobase of exon per million reads mapped” (FPKM) for each gene were calculated using the raw read counts obtained above. For normalization, the FPKM of each gene was individually divided by the geometric mean of

FPKM of the selected housekeeping genes. The normalized FPKM was defined as the basal expression of each gene.

The gene sets for basal expression analysis were selected from Molecular Signatures Databases (Mootha et al., 2003; Subramanian et al., 2005; MSigDB, UC San Diego and Broad Institute). The gene sets WP_OVERVIEW_OF_INTERFERONSMEDIATED_SIGNALING_PATHWAY (M39785), HECKER_IFNB1_TARGETS (M3010), and HALLMARK_INTERFERON_ALPHA_RES-PONSE (M5911) were used for basal expression analysis.

4.9 Construction of customized transposable element annotations and TE analysis

To manually annotate repeat elements, a customized library was first generated from the *E. fuscus* genome using RepeatModeler (v2.0.2a), which identified and modeled the *de novo* transposable element (TE) family (Flynn et al., 2020). In addition, the LTR discovery pipeline (parameter: -LTRStruct) was included for more comprehensive identification. This customized library was combined with the "primary *Eptesicus* library", which was included in the RepeatMasker (v4.1.2) package (Smit, Hubley, & Green, 2013-2015). RepeatMasker was subsequently used to predict and annotate repeat elements in the *E. fuscus* genome based on the combined library. After annotation, the repeat elements classified as structural RNA, simple repeats, and low complexity repeats were removed. The remaining elements were used for customized TE annotations.

TEtranscripts (v2.2.1) with multiple-mode mapping was used to obtain the raw read counts of each TE (Jin et al., 2015). To improve the accuracy of counting, *E. fuscus* gene annotation was included. DEG analysis was subsequently performed by TEtranscripts using DESeq2 (Love, Huber, & Anders, 2014). The method and tools used for visualizing differentially expressed TEs (DETs) were the same as those used for the mRNA-seq DEG analysis described above.

4.10 Small RNA sequencing, novel microRNA prediction, and data

analysis

Direct-zol RNA Kits (Zymo Research, Irvine, California, USA) were used to extract total RNA, including small RNA. Extracted RNA was qualified using High Sensitivity RNA ScreenTape and 4200 TapeStation system (Agilent) to confirm that all RIN^e values ranged from 7.6-8.4. Gel size selection was performed to filter out small RNAs for further library preparation. TruSeq Small RNA Library Preparation Kits (Illumina) were used to prepare libraries according to the manufacturer's instructions. Sequencing was performed on the Illumina HiSeq platform (HiSeq 2500 system, 1 × 50 bp). Library preparation and sequencing were performed by Macrogen Japan Corp.

To trim adapters, filter selected sizes (12-36 bp), and perform quality control, raw reads were processed by fastp (v0.20.1) with the following parameters: -l 12 --length_limit 36 -y -3 -W 3 -M 20 -x -a TGGAATTCTCGGGTGCCAAGG. The length distribution of sRNA processes was calculated by a homemade bash script. The package miRDeep2 (v2.1.0.3) was used to predict novel precursors (pre-miRNAs) as well as novel microRNAs (miRNAs) and calculate raw read counts of each pre-miRNA/miRNA (Friedlander, Mackowiak, Li, Chen, & Rajewsky, 2012). To improve the accuracy of novel miRNA prediction, mature miRNAs and hairpins of *E. fuscus* and mature miRNAs of *Pteropus alecto* from miRBase were included (Release 22.1; Kozomara, Birgaoanu, & Griffiths-Jones, 2019). The script "mapper.pl" was first executed with the following parameters: -d -e -h -i -j -l 18 -m -v; the script "miRDeep2.pl" was next executed using the output of mapper.pl and the databases from miRBase. To identify the localization of novel pre-miRNAs in the *E. fuscus* genome, bedtools intersect (v2.30.0) was used with the following parameters: -s -f 1 -u (Quinlan & Hall, 2010). To define the features among the *E. fuscus* genome, *E. fuscus* gene annotations from NCBI were defined as "gene regions", customized transposable element annotations described above were defined as "TE regions", overlapping sections of genes and TEs were defined as "gene & TE regions", and the remaining unannotated regions were termed "unannotated regions". A miRNA was assigned to the regions only when it fully overlapped with the corresponding features. DEG analysis for novel miRNAs was performed using raw read counts obtained from miRDeep2. The methods and tools for mRNA-seq DEG analysis described above were again used for analysis and visualization.

4.11 Data availability

The accession number for the sequencing data reported in this paper is in the DDBJ DRA database: DRA012731.

4.12 Statistics

The qPCR data are shown as the mean and standard error of the mean (s.e.m.) of multiple biological replicates. Unpaired Student's t-tests were used to test significance. Data were considered significant if *P < 0.05, **P < 0.01, and ***P < 0.001.

Reference

Ahn, M., Cui, J., Irving, A. T., & Wang, L. F. (2016). Unique Loss of the PYHIN Gene Family in Bats Amongst Mammals: Implications for Inflammasome Sensing. *Sci Rep, 6*, 21722. doi:10.1038/srep21722

Anthony, S. J., Gilardi, K., Menachery, V. D., Goldstein, T., Ssebide, B., Mbabazi, R., . . . Mazet, J. A. (2017). Further Evidence for Bats as the Evolutionary Source of Middle East Respiratory Syndrome Coronavirus. *mBio, 8*(2). doi:10.1128/mBio.00373-17

Banerjee, A., Falzarano, D., Rapin, N., Lew, J., & Misra, V. (2019). Interferon Regulatory Factor 3-Mediated Signaling Limits Middle-East Respiratory Syndrome (MERS) Coronavirus Propagation in Cells from an Insectivorous Bat. *Viruses, 11*(2). doi:10.3390/v11020152

Banerjee, A., Rapin, N., Bollinger, T., & Misra, V. (2017). Lack of inflammatory gene expression in bats: a unique role for a transcription repressor. *Sci Rep, 7*(1), 2232. doi:10.1038/s41598-017-01513-w

Banerjee, A., Rapin, N., Miller, M., Griebel, P., Zhou, Y., Munster, V., & Misra, V. (2016). Generation and Characterization of Eptesicus fuscus (Big brown bat) kidney cell lines immortalized using the Myotis polyomavirus large T-antigen. *J Virol Methods, 237*, 166-173. doi:10.1016/j.jviromet.2016.09.008

Banerjee, A., Subudhi, S., Rapin, N., Lew, J., Jain, R., Falzarano, D., & Misra, V. (2020). Selection of viral variants during persistent infection of insectivorous bat cells with Middle East respiratory syndrome coronavirus. *Sci Rep, 10*(1), 7257. doi:10.1038/s41598-020-64264-1

Becker, D. J., Nachtmann, C., Argibay, H. D., Botto, G., Escalera-Zamudio, M., Carrera, J. E., . . . Streicker, D. G. (2019). Leukocyte Profiles Reflect Geographic Range Limits in a Widespread Neotropical Bat. *Integr Comp Biol, 59*(5), 1176-1189. doi:10.1093/icb/icz007

Bendelac, A., Savage, P. B., & Teyton, L. (2007). The biology of NKT cells. *Annu Rev Immunol, 25*, 297-336. doi:10.1146/annurev.immunol.25.022106.141711

Blighe, K., Rana, S., & Lewis, M. (2021). EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. (Version 1.10.0) [R package]. Retrieved from

https://github.com/kevinblighe/EnhancedVolcano

Bowdish, D. M., Davidson, D. J., & Hancock, R. E. (2005). A re-evaluation of the role of host defence peptides in mammalian immunity. *Curr Protein Pept Sci, 6*(1), 35-51. doi:10.2174/1389203053027494

Burgin, C. J., Colella, J. P., Kahn, P. L., & Upham, N. S. (2018). How many species of mammals are there? *Journal of Mammalogy, 99*(1), 1-14. doi:10.1093/jmammal/gyx147

Calisher, C. H., Childs, J. E., Field, H. E., Holmes, K. V., & Schountz, T. (2006). Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev, 19*(3), 531-545. doi:10.1128/CMR.00017-06

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics, 10*, 421. doi:10.1186/1471-2105-10-421

Caracausi, M., Piovesan, A., Antonaros, F., Strippoli, P., Vitale, L., & Pelleri, M. C. (2017). Systematic identification of human housekeeping genes possibly useful as references in gene expression studies. *Mol Med Rep, 16*(3), 2397-2410. doi:10.3892/mmr.2017.6944

Chandan, K., Gupta, M., & Sarwat, M. (2019). Role of Host and Pathogen-Derived MicroRNAs in Immune Regulation During Infectious and Inflammatory Diseases. *Front Immunol, 10*, 3081. doi:10.3389/fimmu.2019.03081

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics, 34*(17), i884-i890. doi:10.1093/bioinformatics/bty560

Chuong, E. B., Elde, N. C., & Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science, 351*(6277), 1083-1087. doi:10.1126/science.aad5497

Cowled, C., Baker, M., Tachedjian, M., Zhou, P., Bulach, D., & Wang, L. F. (2011). Molecular characterisation of Toll-like receptors in the black flying fox Pteropus alecto. *Dev Comp Immunol, 35*(1), 7-18. doi:10.1016/j.dci.2010.07.006

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., . . . Gingeras, T. R. (2013).

STAR: ultrafast universal RNA-seq aligner. *Bioinformatics, 29*(1), 15-21. doi:10.1093/bioinformatics/bts635

Dunkelberger, J. R., & Song, W. C. (2010). Complement and its role in innate and adaptive immune responses. *Cell Res, 20*(1), 34-50. doi:10.1038/cr.2009.139

Edson, D., Field, H., McMichael, L., Vidgen, M., Goldspink, L., Broos, A., . . . Smith, C. (2015). Routes of Hendra Virus Excretion in Naturally-Infected Flying-Foxes: Implications for Viral Transmission and Spillover Risk. *PLoS One, 10*(10), e0140670. doi:10.1371/journal.pone.0140670

Fenton, M. B., & Simmons, N. B. (2015). *Bats: A World of Science and Mystery*: University of Chicago Press.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A, 117*(17), 9451-9457. doi:10.1073/pnas.1921046117

Friedlander, M. R., Mackowiak, S. D., Li, N., Chen, W., & Rajewsky, N. (2012). miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res, 40*(1), 37-52. doi:10.1093/nar/gkr688

Goldstein, T., Anthony, S. J., Gbakima, A., Bird, B. H., Bangura, J., Tremeau-Bravard, A., . . . Mazet, J. A. K. (2018). The discovery of Bombali virus adds further support for bats as hosts of ebolaviruses. *Nat Microbiol, 3*(10), 1084-1089. doi:10.1038/s41564-018-0227-2

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., . . . Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol, 29*(7), 644-652. doi:10.1038/nbt.1883

Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., . . . Wysocka, J. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature, 522*(7555), 221-225. doi:10.1038/nature14308

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., . . . Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for

reference generation and analysis. *Nat Protoc, 8*(8), 1494-1512. doi:10.1038/nprot.2013.084

Hadjicharalambous, M. R., & Lindsay, M. A. (2019). Long Non-Coding RNAs and the Innate Immune Response. *Noncoding RNA, 5*(2). doi:10.3390/ncrna5020034

Holland, R. A. (2006). Orientation and navigation in bats: known unknowns or unknown unknowns? *Behavioral Ecology and Sociobiology, 61*(5), 653-660. doi:10.1007/s00265-006-0297-7

Horie, M., Akasaka, T., Matsuda, S., Ogawa, H., & Imai, K. (2016). Establishment and characterization of a cell line derived from Eptesicus nilssonii. *J Vet Med Sci, 78*(11), 1727-1729. doi:10.1292/jvms.16-0274

Hu, B., Zeng, L. P., Yang, X. L., Ge, X. Y., Zhang, W., Li, B., . . . Shi, Z. L. (2017). Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog, 13*(11), e1006698. doi:10.1371/journal.ppat.1006698

Irving, A. T., Ahn, M., Goh, G., Anderson, D. E., & Wang, L. F. (2021). Lessons from the host defences of bats, a unique viral reservoir. *Nature, 589*(7842), 363-370. doi:10.1038/s41586-020-03128-0

Irving, A. T., Zhang, Q., Kong, P. S., Luko, K., Rozario, P., Wen, M., . . . Wang, L. F. (2020). Interferon Regulatory Factors IRF1 and IRF7 Directly Regulate Gene Expression in Bats in Response to Viral Infection. *Cell Rep, 33*(5), 108345. doi:10.1016/j.celrep.2020.108345

Janardhana, V., Tachedjian, M., Crameri, G., Cowled, C., Wang, L. F., & Baker, M. L. (2012). Cloning, expression and antiviral activity of IFNgamma from the Australian fruit bat, Pteropus alecto. *Dev Comp Immunol, 36*(3), 610-618. doi:10.1016/j.dci.2011.11.001

Jin, Y., Tam, O. H., Paniagua, E., & Hammell, M. (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq data sets. *Bioinformatics, 31*(22), 3593-3599. doi:10.1093/bioinformatics/btv422

Johnson, N., Vos, A., Freuling, C., Tordo, N., Fooks, A. R., & Muller, T. (2010). Human rabies due to lyssavirus infection of bat origin. *Vet Microbiol, 142*(3-4), 151-159. doi:10.1016/j.vetmic.2010.02.001

Kohl, C., Nitsche, A., & Kurth, A. (2021). Update on Potentially Zoonotic Viruses of European Bats.

Vaccines (Basel), 9(7). doi:10.3390/vaccines9070690

Kolde, R. (2019). pheatmap: Pretty Heatmaps. (Version 1.0.12) [R package]. Retrieved from https://CRAN.R-project.org/package=pheatmap

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Res, 47*(D1), D155-D162. doi:10.1093/nar/gky1141

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol, 34*(7), 1812-1819. doi:10.1093/molbev/msx116

Kuzmin, I. V., Schwarz, T. M., Ilinykh, P. A., Jordan, I., Ksiazek, T. G., Sachidanandam, R., . . . Bukreyev, A. (2017). Innate Immune Responses of Bat and Human Cells to Filoviruses: Commonalities and Distinctions. *J Virol, 91*(8). doi:10.1128/JVI.02471-16

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods, 9*(4), 357-359. doi:10.1038/nmeth.1923

Lee, A. J., & Ashkar, A. A. (2018). The Dual Nature of Type I and Type II Interferons. *Front Immunol, 9*, 2061. doi:10.3389/fimmu.2018.02061

Leroy, E. M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassanin, A., Yaba, P., . . . Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. *Nature, 438*(7068), 575-576. doi:10.1038/438575a

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics, 12*, 323. doi:10.1186/1471-2105-12-323

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics, 30*(7), 923-930. doi:10.1093/bioinformatics/btt656

Lin, J. X., & Leonard, W. J. (2019). Fine-Tuning Cytokine Signals. *Annu Rev Immunol, 37*, 295-324.

doi:10.1146/annurev-immunol-042718-041447

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol, 15*(12), 550. doi:10.1186/s13059-014-0550-8

Macchietto, M. G., Langlois, R. A., & Shen, S. S. (2020). Virus-induced transposable element expression up-regulation in human and mouse host cells. *Life Sci Alliance, 3*(2). doi:10.26508/lsa.201900536

Matsui, T., & Amagai, M. (2015). Dissecting the formation, structure and barrier function of the stratum corneum. *Int Immunol, 27*(6), 269-280. doi:10.1093/intimm/dxv013

Mootha, V. K., Lindgren, C. M., Eriksson, K. F., Subramanian, A., Sihag, S., Lehar, J., . . . Groop, L. C. (2003). PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet, 34*(3), 267-273. doi:10.1038/ng1180

Moratelli, R., & Calisher, C. H. (2015). Bats and zoonotic viruses: can we confidently link bats with emerging deadly viruses? *Mem Inst Oswaldo Cruz, 110*(1), 1-22. doi:10.1590/0074-02760150048

Mu, X., Ahmad, S., & Hur, S. (2016). Endogenous Retroelements and the Host Innate Immune Sensors. *Adv Immunol, 132*, 47-69. doi:10.1016/bs.ai.2016.07.001

Munster, V. J., Adney, D. R., van Doremalen, N., Brown, V. R., Miazgowicz, K. L., Milne-Price, S., . . . Bowen, R. A. (2016). Replication and shedding of MERS-CoV in Jamaican fruit bats (Artibeus jamaicensis). *Sci Rep, 6*, 21878. doi:10.1038/srep21878

Myhre, S., Lingjaerde, O. C., Hennessy, B. T., Aure, M. R., Carey, M. S., Alsner, J., . . . Sorlie, T. (2013). Influence of DNA copy number and mRNA levels on the expression of breast cancer related proteins. *Mol Oncol, 7*(3), 704-718. doi:10.1016/j.molonc.2013.02.018

Negishi, H., Taniguchi, T., & Yanai, H. (2018). The Interferon (IFN) Class of Cytokines and the IFN Regulatory Factor (IRF) Transcription Factor Family. *Cold Spring Harb Perspect Biol, 10*(11). doi:10.1101/cshperspect.a028423

Nejad, C., Stunden, H. J., & Gantier, M. P. (2018). A guide to miRNAs in inflammation and innate immune responses. *FEBS J, 285*(20), 3695-3716. doi:10.1111/febs.14482

Nguyen, T. H., Liu, X., Su, Z. Z., Hsu, A. C., Foster, P. S., & Yang, M. (2018). Potential Role of MicroRNAs in the Regulation of Antiviral Responses to Influenza Infection. *Front Immunol, 9*, 1541. doi:10.3389/fimmu.2018.01541

O'Connell, R. M., Rao, D. S., Chaudhuri, A. A., & Baltimore, D. (2010). Physiological and pathological roles for microRNAs in the immune system. *Nat Rev Immunol, 10*(2), 111-122. doi:10.1038/nri2708

Olival, K. J., & Hayman, D. T. (2014). Filoviruses in bats: current knowledge and future directions. *Viruses, 6*(4), 1759-1788. doi:10.3390/v6041759

Omatsu, T., Bak, E. J., Ishii, Y., Kyuwa, S., Tohya, Y., Akashi, H., & Yoshikawa, Y. (2008). Induction and sequencing of Rousette bat interferon alpha and beta genes. *Vet Immunol Immunopathol, 124*(1-2), 169-176. doi:10.1016/j.vetimm.2008.03.004

Platt, R. N., 2nd, Vandewege, M. W., Kern, C., Schmidt, C. J., Hoffmann, F. G., & Ray, D. A. (2014). Large numbers of novel miRNAs originate from DNA transposons and are coincident with a large species radiation in bats. *Mol Biol Evol, 31*(6), 1536-1545. doi:10.1093/molbev/msu112

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics, 26*(6), 841-842. doi:10.1093/bioinformatics/btq033

Rhein, B. A., Powers, L. S., Rogers, K., Anantpadma, M., Singh, B. K., Sakurai, Y., . . . Maury, W. (2015). Interferon-gamma Inhibits Ebola Virus Infection. *PLoS Pathog, 11*(11), e1005263. doi:10.1371/journal.ppat.1005263

Riera Romo, M., Perez-Martinez, D., & Castillo Ferrer, C. (2016). Innate immunity in vertebrates: an overview. *Immunology, 148*(2), 125-139. doi:10.1111/imm.12597

Rosales, C. (2020). Neutrophils at the crossroads of innate and adaptive immunity. *J Leukoc Biol, 108*(1), 377-396. doi:10.1002/JLB.4MIR0220-574RR

Rose, S. A., Wroblewska, A., Dhainaut, M., Yoshida, H., Shaffer, J. M., Bektesevic, A., . . . Immunological Genome, C. (2021). A microRNA expression and regulatory element activity atlas of the mouse immune system. *Nat Immunol, 22*(7), 914-927. doi:10.1038/s41590-021-00944-y

Roulois, D., Loo Yau, H., Singhania, R., Wang, Y., Danesh, A., Shen, S. Y., . . . De Carvalho, D. D. (2015). DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell, 162*(5), 961-973. doi:10.1016/j.cell.2015.07.056

Rusinova, I., Forster, S., Yu, S., Kannan, A., Masse, M., Cumming, H., . . . Hertzog, P. J. (2013). Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res, 41*(Database issue), D1040-1046. doi:10.1093/nar/gks1215

Schoggins, J. W., & Rice, C. M. (2011). Interferon-stimulated genes and their antiviral effector functions. *Curr Opin Virol, 1*(6), 519-525. doi:10.1016/j.coviro.2011.10.008

Schuh, A. J., Amman, B. R., Sealy, T. K., Spengler, J. R., Nichol, S. T., & Towner, J. S. (2017). Egyptian rousette bats maintain long-term protective immunity against Marburg virus infection despite diminished antibody levels. *Sci Rep, 7*(1), 8763. doi:10.1038/s41598-017-07824-2

Shao, X., Lv, N., Liao, J., Long, J., Xue, R., Ai, N., . . . Fan, X. (2019). Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med Genet, 20*(1), 175. doi:10.1186/s12881-019-0909-5

Shaw, A. E., Hughes, J., Gu, Q., Behdenna, A., Singer, J. B., Dennis, T., . . . Palmarini, M. (2017). Fundamental properties of the mammalian innate immune system revealed by multispecies comparison of type I interferon responses. *PLoS Biol, 15*(12), e2004086. doi:10.1371/journal.pbio.2004086

Smit, A., Hubley, R., & Green, P. (2013-2015). RepeatMasker Open-4.0 (Version v4.1.2). Retrieved from http://www.repeatmasker.org

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., . . . Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A, 102*(43), 15545-15550. doi:10.1073/pnas.0506580102

Taganov, K. D., Boldin, M. P., & Baltimore, D. (2007). MicroRNAs and immunity: tiny players in a big field. *Immunity, 26*(2), 133-137. doi:10.1016/j.immuni.2007.02.005

Vandewege, M. W., Platt, R. N., 2nd, Ray, D. A., & Hoffmann, F. G. (2016). Transposable Element

Targeting by piRNAs in Laurasiatherians with Distinct Transposable Element Histories. *Genome Biol Evol, 8*(5), 1327-1337. doi:10.1093/gbe/evw078

Verhelst, J., Hulpiau, P., & Saelens, X. (2013). Mx proteins: antiviral gatekeepers that restrain the uninvited. *Microbiol Mol Biol Rev, 77*(4), 551-566. doi:10.1128/MMBR.00024-13

Vladimer, G. I., Gorna, M. W., & Superti-Furga, G. (2014). IFITs: Emerging Roles as Key Anti-Viral Proteins. *Front Immunol, 5*, 94. doi:10.3389/fimmu.2014.00094

Xie, J., Li, Y., Shen, X., Goh, G., Zhu, Y., Cui, J., . . . Zhou, P. (2018). Dampened STING-Dependent Interferon Activation in Bats. *Cell Host Microbe, 23*(3), 297-301 e294. doi:10.1016/j.chom.2018.01.006

Yang, X. L., Tan, C. W., Anderson, D. E., Jiang, R. D., Li, B., Zhang, W., . . . Shi, Z. L. (2019). Characterization of a filovirus (Mengla virus) from Rousettus bats in China. *Nat Microbiol, 4*(3), 390-395. doi:10.1038/s41564-018-0328-y

Yim, S. H., Jung, S. H., Chung, B., & Chung, Y. J. (2015). Clinical implications of copy number variations in autoimmune disorders. *Korean J Intern Med, 30*(3), 294-304. doi:10.3904/kjim.2015.30.3.294

Yu, G. (2021). enrichplot: Visualization of Functional Enrichment Result. (Version 1.12.2). Retrieved from https://yulab-smu.top/biomedical-knowledge-mining-book/

Zhou, P., & Shi, Z. L. (2021). SARS-CoV-2 spillover events. *Science, 371*(6525), 120-122. doi:10.1126/science.abf6097

Zhou, P., Tachedjian, M., Wynne, J. W., Boyd, V., Cui, J., Smith, I., . . . Baker, M. L. (2016). Contraction of the type I IFN locus and unusual constitutive expression of IFN-alpha in bats. *Proc Natl Acad Sci U S A, 113*(10), 2696-2701. doi:10.1073/pnas.1518240113

Acknowledgements

I would like to show my great appreciation and dedicate this thesis to all the people who have provided my enormous physically and mentally supports and suggestions in this period:

Dr. Keizo Tomonaga, for his inclusiveness and kindness, accepting me to join the laboratory and guiding me to overcome different obstacles in both life and academia with lots of patience;

Dr. Masayuki Horie, for his instructions, guidance, advice, and kind understanding that support me to study, do research, and finish the paper and thesis smoothly. All members in Horie group: *Shohei Kojima*, *Garcia Bea Clarise Baluyot*, *Yahiro Mukai*, *Junna Kawasaki*, and *Koichi Kitao*, for the important discussion and comments to polish and refine my study and research;

All members in Tomonaga lab: *Makino*-sensei, *Yamashita*-san, *Tanaka*-san, *Komorizono*-san, *Kanda*-san, *Sakai*-san, *Nabeka*-san, *Iwata*-san, and former members: *Yanai*-san, *Onimaru*-san, *Watanabe*-kun, *Kim*-san, and *Kakuya*-san, for establishing the wonderful and comfortable atmosphere in lab as well as together maintaining the lab in great conditions for working;

Nishimoto-san, *Tsuji*-san, and the staff of the Student Affairs Section, for helping me with tons of inquiries and questions in school affairs and life;

Comrades for several years: 蔡孟義，張平，龔穎之，聶良育，王德瑋，張育瑄，黃怡瑄; friends coming to Japan together three years ago: 家穎，德穎，永佳，凱鴻，志剛; an important friend, 嘉敏; all friends in the crew of Granblue Fantasy that we actually never met but I know we are close, for all the mental supports during this period, which are definitely irreplaceable.

And my family, for their immeasurable, unlimited, infinite, endless, incalculable love behind me, accompanying me during all the time.

I am grateful to Dr. *Takefumi Kondo* and Ms. *Yukari Sando* for performing the NextSeq sequencing. I thank Dr. *Keiko Takemoto* (Institute for Frontier Life and Medical Sciences, Kyoto University, Japan) for technical support.

I appreciate Japan-Taiwan Exchange Association (公益財団法人日本台湾交流協会) for

providing me the scholarship, and therefore I am able to start my studying and research in Japan.

This study was supported by JSPS KAKENHI JP20H05682 (KT) and JP18K19443 (MH); MEXT KAKENHI JP16H06429 (KT), JP16K21723 (KT), JP16H0643 (KT), JP17H05821 (MH) and JP19H04833 (MH); and the Hakubi project at Kyoto University (MH).

The work, which is being presented in this thesis entitled "**Investigation of Innate Immune Responses in *Eptesicus* Bat Cells via Comprehensive Analysis**", is mainly based on the published article:

A comprehensive profiling of innate immune responses in Eptesicus bat cells

Hsien-Hen Lin, Masayuki Horie, Keizo Tomonaga

Microbiology and Immunology, 2021, in press

林咸亨 Hsien-Hen Lin, 2021/12

The acknowledgement in Mandarin (for myself) is as below.

轉眼間已經開始提筆寫致謝了。想要感謝的老師同事與朋友已在上述提及。即便感恩之情無法言盡，一再重述又似乎略顯濫情，請原諒我就此打住。最後的最後，就留段話給自己吧。

一直為研究的結果苦惱，尤其開始準備畢業審查後，心情好像回到碩士班口試結束時，那個下定決心唸博士班的自己。當時不停懷疑自己的能力不足，整個碩士班沒有太大的長進，也沒有可以特別凸顯的強項。到了現在，從寫 paper 到準備論文這段時間，回顧了這段時間的研究成果，再次咀嚼了一模一樣的心情。

然而回想剛到日本時的不安，直到現在或許稱得上安身立命，一路上說是歷盡風霜似乎也沒有如此辛苦不堪，但絕不是風平浪靜毫無波瀾。最初一句日文都說不出口，到現在雖然還是不太會說，仍自己處理了生活上大大小小的事情；自己收集資料並準備就職活動，接連的拒絕百般折磨，但最後獲得了兩間公司的職缺。同樣檢視自己的研究歷程，比起碩士班當下其實仍更獨立了些，更有想法了些，更懂得批判了些，也更有擔當了些吧。若這是種無法拔除的本能，願對自己的懷疑能在重要時刻成為推進動力，盼對自己的不滿能在抉擇當下化作指路的星。

於是迎來學生生活最後一個冬天。三年時光荏苒，所有的歡愉快樂與苦難折磨只在倏忽，卻真切深刻。也許可以有點自信地說，最後的確有稍微貼近當初想像中的樣貌吧？結束總伴隨開始，儘管前方的路渺若煙雲，肯定如京都的雪一般，值得殷殷期盼。

林咸亨 2021.12

Appendices

Coding for analysis

Adaptor trimming and quality control for sequencing raw data via fastp (bash)

```
## For paired-end mRNA-seq data sets
fastp -w 16 \
    -i <fastq_R1> -o < output fastq_R1 > -I < fastq_R2> -O <output fastq_R2> \
    -j <output json> -h <output html> \
    -l 35 -y -3 -W 3 -M 20 -x --detect_adapter_for_pe

## For single-end small RNA-seq data sets
fastp -w 16 \
    -i <fastq> -o < output fastq > \
    -j <output json> -h <output html> \
    -l 12 --length_limit 36 -y -3 -W 3 -M 20 -x -a TGGAATTCTCGGGTGCCAAGG
```

Genome mapping (bash)

```
## Construct STAR index
STAR --runThreadN 32 --sjdbOverhang 149 --runMode genomeGenerate \
    --genomeDir <output director> --genomeFastaFiles <genome fasta> --sjdbGTFfile <gtf/gff >

## STAR mapping (output one result for multiple mapped reads)
STAR --runThreadN 32 --readFilesCommand zcat --outReadsUnmapped Fastx \
    --outSAMmultNmax 1 --outMultimapperOrder Random --genomeDir <STAR index> \
    --readFilesIn <input fastq_R1> <input fastq_R2> --outFileNamePrefix <output director> \

## Transform sam to bam, extract mapped reads, sort bam file, and create bam index
samtools sort -@ 32 -O bam -o <output bam> -T <output directory> <input sam>
samtools view -@ 32 -h -b -F 4 <input bam> > <output mapped bam>
samtools sort -@ 32 -O bam -o <output sorted bam> -T <output directory> <input mapped bam>
samtools index <input sorted bam>
```

Raw read counting and fpkm calculation of mRNA-seq (bash)

```
## Obtain raw read counts using featureCounts
featureCounts -T 36 -p -M -O -s 2 -t exon -g gene_name \
        -a <annotation gtf> -o <output txt> <all input bam> 2>&1 tee <log txt>

## Calculate fpkm
# Combine biological repeats
awk 'BEGIN{
    OFS="\t"
    }{
    print $1,$2+$3+$4,$5+$6+$7,$8
    }' <input raw read counts> > <processed raw read count>

# Calculate fpkm using established perl script
# rpkm_script_beta.pl from https://github.com/decodebiology/rpkm_rnaseq_count.git
perl rpkm_script_beta.pl <processed raw read count> 2:3 4 > <output fpkm txt>
```

edgeR differentially expressed gene (DEG) analysis (R)

```
# Input samples and make matrix.
combined_table <- read.table(<processed raw read count>, header = T,
                                    col.names = c("Genes","treated_1", "treated_2", "treated_3",
```

```
                                                    "untreated_1", "untreated_2", "untreated_3",
                                                    "Length"))
rownames(combined_table) <- combined_table$Genes
combined_table <- combined_table[,c(5:7,2:4)]

# Construct DGE matrix for F-test
group <- factor(c(1, 1, 1, 2, 2, 2))
DGE_table <- DGEList(counts = combined_table, genes = rownames(combined_table), group = group)
keep <- filterByExpr(DGE_table, group = group)
DGE_table <- DGE_table[keep,,keep.lib.sizes = FALSE]
DGE_table <- calcNormFactors(DGE_table)
design <- model.matrix(~group)
DGE_table <- estimateDisp(DGE_table, design)

## Calculate DEGs
# Perform F-test
fit <- glmQLFit(DGE_table, design)
qlf <- glmQLFTest(fit, coef = 2)
topTags(qlf)

tr <- glmTreat(fit, coef = 2, lfc = 1)
topTags(tr, n = 15, p.value = 0.05)

# Make volcano MA plot
EnhancedVolcano(topTags(qlf, n = Inf, p.value = 1)$table,
                lab = rownames(topTags(qlf, n = Inf, p.value = 1)$table),
                selectLab = "none",
                xlim = c(-2, 7.5), ylim = c(0, 10),
                title = NULL, subtitle = NULL,
                pCutoff = 0.05,
                FCcutoff = 1,
                x = 'logFC',
                y = 'FDR')

# Selected DEGs with > 2 fold-changes, and p-value < 0.05.
# Seperate upregulate and downregulate genes.
deg <- topTags(qlf, n = Inf, p.value = 0.05)$table
up <- row.names(deg[deg$logFC > 0,])
down <- row.names(deg[deg$logFC < 0,])
```

Generate homemade conversion table via BLASTp (bash and R)

```
## Perform blasp using bash
makeblastdb -in <human protein fasta> -parse_seqids \
      -blastdb_version 5 -title <name of database> -out <human protein blasp database> -dbtype prot
blastp -query <E. fuscus protein fasta> \
      -db <human protein blasp database> -max_target_seqs 1 \
      -outfmt 6 -evalue 1e-5 -num_threads 32 > <output outfm6 file>

## Input the outfm6 file into R and transform RefSeq ID into gene symbol
# Input outfmt6 file
blastp <- read.table("<output outfm6 file>")
blastp <- blastp[1:2]
colnames(blastp) <- c("Ef_protein","Homo_protein")
blastp <- unique(blastp)
```

```
# Input curated annotation, generated from NCBI annotation table
# Information includes “Accession”, “Start”, “Stop”, “Gene Symbol”, “NCBI Protein ID”
Ef_annotation <- read.table("<E.fuscus curated annotation txt>", header = TRUE)
Ef_annotation_curated <- unique(Ef_annotation[, 4:5])
Homo_annotation <- read.table("<human curated annotation txt>", header = TRUE)
Homo_annotation_curated <- unique(Homo_annotation[, 4:5])

# Generate final conversion table
temp <- merge(blastp, Homo_annotation_curated, by.x = "Homo_protein", by.y = "Protein")
colnames(temp) <- c("Homo_protein", "Ef_protein", "Homo_Locus")
temp2 <- merge(temp, Ef_annotation_curated, by.x = "Ef_protein", by.y = "Protein")
colnames(temp2) <- c("Ef_protein", "Homo_protein", "Homo_Locus", "Ef_Locus")
Homo_Ef_annotation <- unique(test2[, 3:4])
```

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis (R)

```
# GO analysis via enrichGO
# cnetplot() used for network analysis (not shown)
ego <- enrichGO(gene = <sorted E. fuscus target gene with human annotation>,
                universe = <whole E. fuscus gene with human annotation>,
                keyType = "SYMBOL",
                OrgDb = org.Hs.eg.db,
                ont = "BP",
                pAdjustMethod = "BH",
                pvalueCutoff  = 0.01,
                qvalueCutoff  = 0.05,
                readable      = FALSE)

# KEGG analysis via enrichKEGG
kk <- enrichKEGG(gene = gene.df$ENTREZID,
                 organism = 'hsa',
                 pvalueCutoff = 0.05)
```

De novo assembly using Trinity (bash)

```
Trinity --seqType fq --SS_lib_type RF --max_memory 400G --CPU 36 \
        --left <input fastq_R1> --right <input fastq_R2> --output <output directory>
```

Identify differentially expressed transcripts from de novo assembly (bash)

```
## Follow RSEM flow from Trinity instruction and use built-in scripts.
# Reference preparation.
align_and_estimate_abundance.pl --transcripts <transcripts fasta> \
    --est_method RSEM --aln_method bowtie2 --trinity_mode \
    --SS_lib_type RF --thread_count 32 --prep_reference

# Map mRNA-seq raw data to transcripts via bowtie2 and perform raw read count via RSEM.
align_and_estimate_abundance.pl \
    --transcripts <transcripts fasta> \
    --seqType fq --SS_lib_type RF \
    --samples_file <list containing groups and paths of input mRNA-seq raw fasta> \
    --output_dir <output directory> \
    --est_method RSEM --aln_method bowtie2 --trinity_mode --thread_count 32

# Generate matrix for edgeR DEGs analysis.
abundance_estimates_to_matrix.pl --est_method RSEM \
        --gene_trans_map <gene_trans_map file generated above> \
        --name_sample_by_basedir \
```

```
        --quant_files <list containing paths of all RSEM.isoforms.results generated above>

# Check transcripts TPM by accumulation.
<TrinityPath>/util/misc/count_matrix_features_given_MIN_TPM_threshold.pl \
        RSEM.gene.TPM.not_cross_norm |
        tee RSEM.genes.TPM.not_cross_norm.counts_by_min_TPM

# Perform edgeR DEGs analysis.
run_DE_analysis.pl --matrix RSEM.gene.counts.matrix --method edgeR \
        --samples_file <list containing groups and paths of input mRNA-seq raw fasta> \
        --reference_sample untreated --output <output directory>

# Extract and cluster DEGs, use -P 1e-3 and -C 2, a stricter threshold.
analyze_diff_expr.pl --matrix RSEM.gene.TMM.EXPR.matrix -P 1e-3 -C 2

# Partition genes into expression clusters.
define_clusters_by_cutting_tree.pl -R diffExpr.P1e-3_C2.matrix.RData --Ptree 60
```

Identify repeat elements among genomes using RepeatModeler and RepeatMasker (bash)

```
## Run RepeatModeler to generate repeat libraries
<RepeatModelerPath>/BuildDatabase -name EptFus <genome fasta>
<RepeatModelerPath>/RepeatModeler -database EptFus -pa 40 -LTRStruct

## Run RepeatMasker to identify repeat elements.
# Generate customized libraries using existing libraries and own RepeatModeler result.
<RepeatMaskerPath>/famdb.py -i <RepeatMaskerPath>/Libraries/RepeatMaskerLib.h5 \
        families --format fasta_name --ancestors --descendants Eptesicus > Eptesicus-families.fa
cat Eptesicus-families.fa EptFus-families.fa > combined_EptFus.fa

# Perform RepeatMasker
<RepeatMaskerPath>/RepeatMasker -pa 40 -gff -lib combined_EptFus.fa <input genome fasta>
```

Raw read counting and DEG analysis for transposable element using TEtranscripts (bash)

```
TEtranscripts -t <treatment bam> -c <control bam> --GTF <gene gtf> --TE <TE gtf> \
        --strand reverse --sortByPos --project <output name> --mode uniq
```

miRDeep2 Prediction (bash)

```
mapper.pl <input config txt> -d -e -h -i -j -l 18 -m -v -o 32 \
        -k TGGAATTCTCGGGTGCCAAGG -p <genome bowtie index> \
        -s <reads_collapsed.fa> -t <reads_vs_genome.arf>

miRDeep2.pl <reads_collapsed.fa> <genome fasta> <reads_vs_genome.arf>\
        <known mature miRNA fasta> <known mature miRNA fasta from related species> \
        <known hairpin fasta> 2> <log txt>
```

Alignment of miRNA and genome regions using intersectBed (bash)

```
intersectBed -s -F 1 -wa \
        -a <genome region including gene, TE, etc., in bed format> -b <miRNA bed file> |
        sort -u -k 4,4 > <output bed file>
```