様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　６ 年　 ９ 月　 ８ 日現在

機関番号：１４３０１
研究種目：若手研究
研究期間：2021～2023
課題番号：２１Ｋ１５０６６

研究課題名（和文）Transposable elements shape the evolution of mammalian innate immunity against pathogens

研究課題名（英文）Transposable elements shape the evolution of mammalian innate immunity against pathogens

研究代表者
ＣＨＥＮ　Ｘｕｎ（CHEN, Xun）

京都大学・高等研究院・特定助教

研究者番号：３０８８５１５８

交付決定額（研究期間全体）：（直接経費）　　3,600,000円

研究成果の概要（和文）：我々は、異なる病原体の感染に伴って発現が上昇する転移因子について検証した。これらの病原性特異的サブファミリーは、多くが霊長類進化の前段階でゲノムに挿入さた転移因子であることがわかった。さらに、若いLTR配列の多くが誤ってアノテーションされていたため、系統解析を行うことで、より高精度のアノテーションを得た。

研究成果の学術的意義や社会的意義
We revealled many pathogen specific TE subfamilies indicating their potential important functions during infection. We also found many young subfamilies were misannotated. We then proposed a new classification and annotation approach, which is key to resolve the role of TEs played during infection.

研究成果の概要（英文）：We examined transposable elements that were up-regulated upon the infection of different pathogens. We identified many LTR subfamilies that were significantly activated following Salmonella infection; we found another set of LTR subfamilies that were activated following influenza infection. These pathogen-specific subfamilies were mostly integrated before primates during evolution.
After we look into the candidate TE subfamilies, we found that many instances from these subfamilies were mis-annotated. We then performed the phylogenetic analysis to re-annotate instances from relatively young subfamilies. In the end, we validated that the new annotation that we achieved have a high sequence, epigenetics, and functional specificity validated by using the massive parallel reporter assay.

研究分野：Genomics

キーワード：Transposable element　Classification　Annotation　Epigenetics　Infection　Immunity

１．研究開始当初の背景

Microbial pathogens are first detected by the innate immune system after they breach physical and chemical barriers, and then alter adaptive immunity to the presence of infection. Transposable elements (TEs) are interspersed repetitive sequences that cover a large fraction of numerous mammalian genomes. TEs have expanded the repertoire of transcription factor binding sites in mammalian genomes and, through this process, have been co-opted in various transcriptional networks, including the immune related pathways (Bourque et al. 2018; Chuong, Elde, and Feschotte 2016). On the other hand, TE transcription, especially endogenous retroviruses, were drastically upregulated in human immune cells upon pathogen infection. Although it remains unclear how and whether TEs were upregulated consistently upon different pathogens.

TE instances were annotated as a subfamily mainly through the sequence homology between genomic sequences and curated TE consensus sequences. As we known, many TE instances were mis-annotated due to the similarity between closely related subfamilies and the evolutionary complexity within the subfamily. Although most studies on TE functionality still relied on the current problematic TE annotation files. We hypothesized that the phylogenetic relationships should be used to achieve a proper classification and annotation of TEs.

２．研究の目的

In this project, we proposed to first analyze the transcriptomic data in human macrophages before and after infection with different pathogens to identify common and pathogen-specific upregulated TEs. We also aimed to develop a novel approach to improve TE classification and annotation in the human genome and to further dissect the function of TEs upon pathogen infection.

３．研究の方法

To identify common and pathogen-specific upregulated TEs upon infection, we first downloaded the single-end RNA-seq data in human macrophages before and after infection with different pathogens, including Salmonella bacteria (SRA:SRP074274) and influenza virus (H1N1) (EBI:EGAC00001000250) (Nédélec et al. 2016; Aracena et al. 2024). We then ran TEtranscripts to quantify the TE expression level at the subfamily-level for all samples with the aligned BAM file as the input with the TE and gene annotation files. After we merged the count tables of all samples, we then used the DEseq2 tool to identify the differentially expressed subfamilies (|log2FC| >1 and log10(padj) >3) analysis before and after infection. Candidate subfamilies were kept and compared between pathogens.

To re-annotate the TE subfamilies, we first downloaded the TE annotation files from the Repeatmasker database, and then retrieved the sequences of TE instances using BEDtools. We then performed the phylogenetic analysis for each subfamily using MAFFT, PRANK, trimAL, IQ-TREE2 and other tools. We then clustered the TE consensus sequences (Repbase) using Cytoscape tool for the network analysis and subfamilies with similar consensus sequences were categorized into a subfamily group. After

we determined and obtained the clusters per subfamily based on the evolutionary tree and internal branch lengths, we further re-constructed the rooted tree including all cluster consensus sequences from the same subfamily group. In the end, TE instances from the clusters with amongst the top branch lengths to others were grouped as the new subfamilies. We then performed the lentiMPRA experiment to experimentally validate the new subfamilies have a higher specificity compared to the originally annotated subfamily.

4．研究成果

Using the RNA sequencing (RNA-seq) data in human macrophages during Salmonella infection, we identified 542 and 30 subfamilies that were significantly upregulated and downregulated after infection, respectively (**Figure 1A**). To examine whether the expression of these subfamilies was also changed upon the infection with other pathogens, we then performed the comparative analysis with the expression change during influenza infection (Chen et al. 2023). In the end, we identified 90, 16 and 36 subfamilies that were upregulated upon Salmonella, influenza, and both pathogens, respectively **(Figure 1B)**.

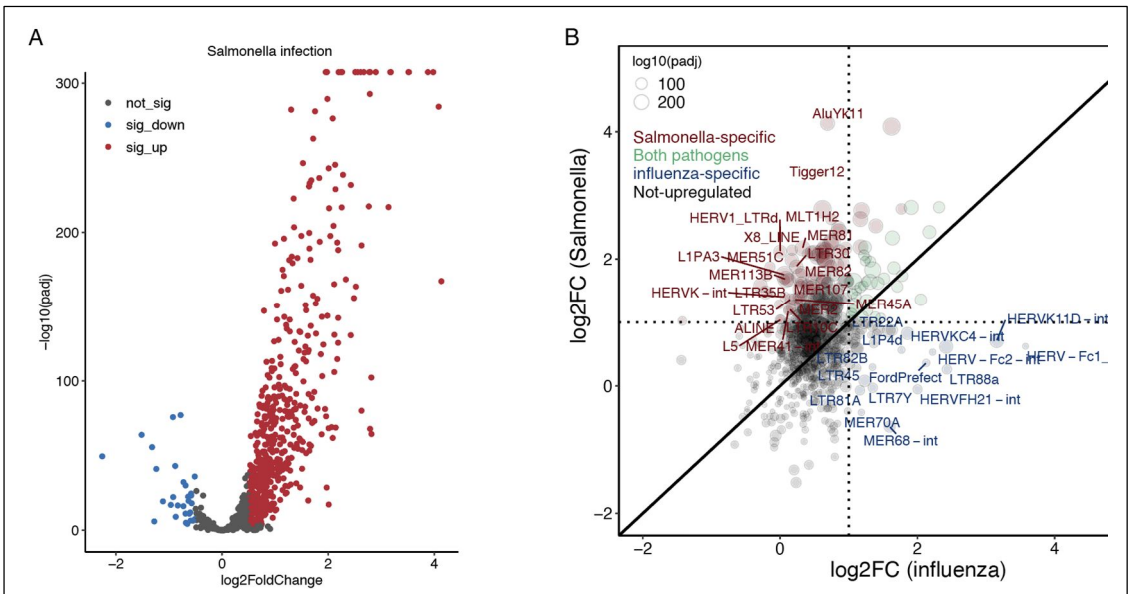

Figure 1. TE expression changes during pathogen infection. (A) TE expression changes during Salmonella infection. (B) Comparison of the upregulated subfamilies after the infection with Salmonella and influenza infection. Top candidate pathogen-specific subfamilies were labeled.

When we look at the annotation of LTR subfamilies we previously identified, we found that many sequences may be mis-annotated based on the distribution of divergence rates relative to the consensus sequences. We developed a new approach based on the phylogenetic analysis to re-annotate the sequences from the young LTR subfamilies (**Figure 2A**). To validate it, we picked MER11A/B/C subfamilies as examples and the sequences from these subfamilies were classified into four reconstituted MER11 new subfamilies, including MER11_G1/G2/G3/G4. We next functionally validated our new approach by looking at the regulatory potential and motifs in MER11_G1/G2/G3/G4 new subfamilies using MPRA experiment (**Figure 2B**). After that, we applied it to a total of 53 originally annotated subfamilies and successfully grouped them into 75 new subfamilies, which showed an increased transcription factor binding specificity.

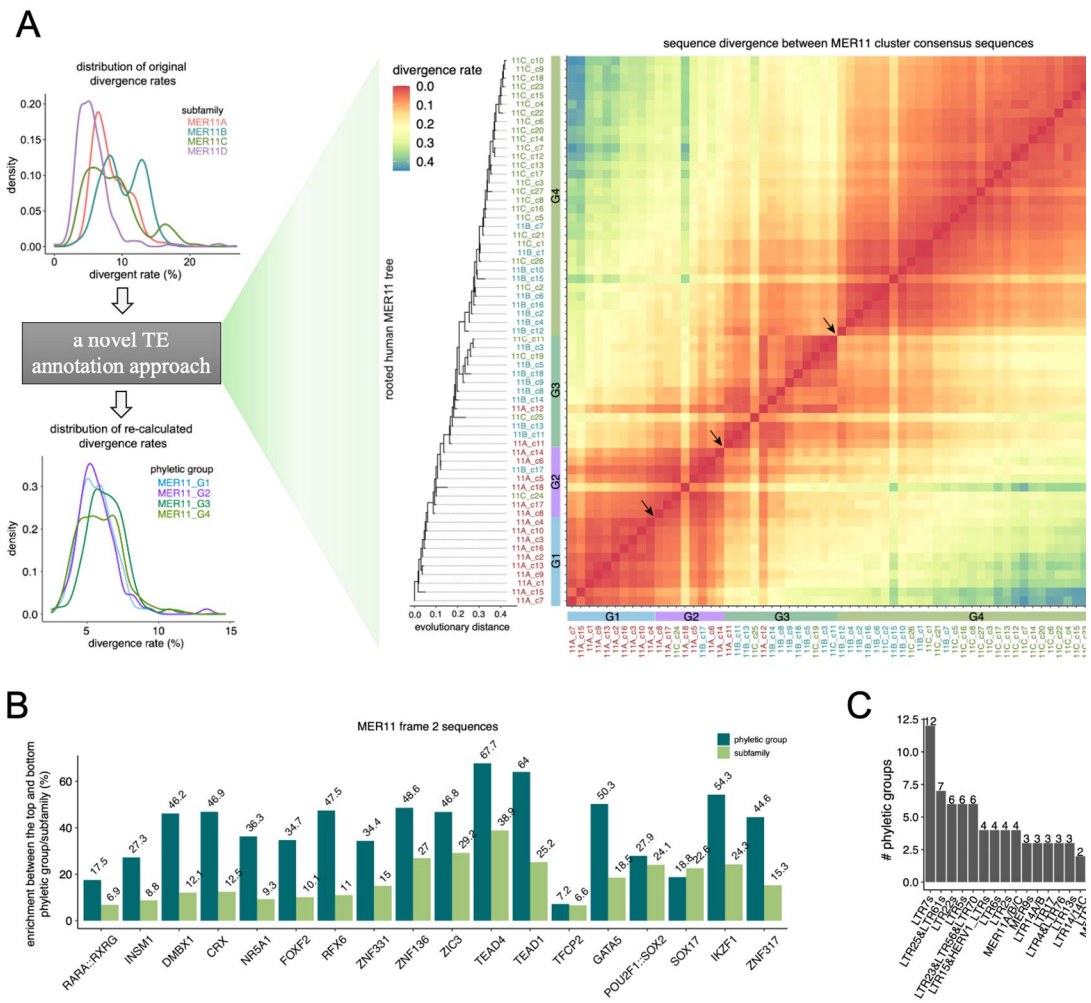

Figure 2. A novel TE classification and annotation approach. (A) The novel annotation approach for re-annotating TE instances. MER11A/B/C subfamilies are shown as examples. Left panel shows the divergence rates relative to each consensus sequence before and after the re-annotation. Right panel shows the reconstructed rooted tree of all cluster consensus sequences from the MER11A/B/C subfamily group. (B) The novel approach has a significantly higher specificity of functional motifs for the new subfamilies (phyletic groups) compared to the original subfamilies using the MPRA technology. Enrichment was computed as the proportion of instances containing each motif in the top new subfamily, the one with the highest proportion, minus the bottom new subfamily, the one with the lowest proportion. Motif enrichment was computed similarly between the top and the bottom original subfamily. (C) Numbers of newly annotated subfamilies per subfamily group. Sequences from a total of 53 subfamilies were re-annotated into 75 new subfamilies using our novel approach.

References

Aracena, Katherine A., Yen-Lung Lin, Kaixuan Luo, Alain Pacis, Saideep Gona, Zepeng Mu, Vania Yotova, et al. 2024. “Epigenetic Variation Impacts Individual Differences in the Transcriptional Response to Influenza Infection.” *Nature Genetics*, February, 1–12.

Bourque, Guillaume, Kathleen H. Burns, Mary Gehring, Vera Gorbunova, Andrei Seluanov, Molly Hammell, Michaël Imbeault, et al. 2018. “Ten Things You Should Know about Transposable Elements.” *Genome Biology* 19 (1): 199.

Chen, Xun, Alain Pacis, Katherine A. Aracena, Saideep Gona, Tony Kwan, Cristian Groza, Yen Lung Lin, et al. 2023. “Transposable Elements Are Associated with

the Variable Response to Influenza Infection." *Cell Genomics* 3 (5): 100292.

Chuong, Edward B., Nels C. Elde, and Cédric Feschotte. 2016. "Regulatory Evolution of Innate Immunity through Co-Option of Endogenous Retroviruses." *Science* 351 (6277): 1083–87.

Nédélec, Yohann, Joaquín Sanz, Golshid Baharian, Zachary A. Szpiech, Alain Pacis, Anne Dumaine, Jean-Christophe Grenier, et al. 2016. "Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens." *Cell* 167 (3): 657-669.e21.

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　1件／うちオープンアクセス　1件）

１．著者名 Chen Xun、Zhang Zicong、Yan Yizhi、Goubert Clement、Bourque Guillaume、Inoue Fumitaka	４．巻 0
２．論文標題 Cryptic endogenous retrovirus subfamilies in the primate lineage	５．発行年 2023年
３．雑誌名 BioRxiv	６．最初と最後の頁 0
掲載論文のDOI（デジタルオブジェクト識別子） 10.1101/2023.12.07.570592	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計3件（うち招待講演　0件／うち国際学会　3件）

１．発表者名 Xun Chen
２．発表標題 Transposable elements contribute to global chromatin remodeling in the human response to influenza infection
３．学会等名 44th The Molecular Biology Society of Japan（国際学会）
４．発表年 2021年

１．発表者名 Xun Chen
２．発表標題 Using MPRA to understand the functional cores of transposable element enhancer activity and evolution in primates
３．学会等名 International Human Epigenome Consortium（国際学会）
４．発表年 2022年

１．発表者名 Xun Chen
２．発表標題 Transposable elements contribute to epigenetic changes in the human response to influenza infection
３．学会等名 44th The Molecular Biology Society of Japan（国際学会）
４．発表年 2021年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関