IMMUNOGENOMICS

Regulatory evolution of innate immunity through co-option of endogenous retroviruses

Edward B. Chuong, Nels C. Elde,*† Cédric Feschotte*†

Endogenous retroviruses (ERVs) are abundant in mammalian genomes and contain sequences modulating transcription. The impact of ERV propagation on the evolution of gene regulation remains poorly understood. We found that ERVs have shaped the evolution of a transcriptional network underlying the interferon (IFN) response, a major branch of innate immunity, and that lineage-specific ERVs have dispersed numerous IFN-inducible enhancers independently in diverse mammalian genomes. CRISPR-Cas9 deletion of a subset of these ERV elements in the human genome impaired expression of adjacent IFN-induced genes and revealed their involvement in the regulation of essential immune functions, including activation of the AIM2 inflammasome. Although these regulatory sequences likely arose in ancient viruses, they now constitute a dynamic reservoir of IFN-inducible enhancers fueling genetic innovation in mammalian immune defenses.

*Corresponding author. E-mail: nelde@genetics.utah.edu (N.C.E.); cedric@genetics.utah.edu (C.F.) †These authors contributed equally to this work.

Changes in gene regulatory networks underline many biological adaptations, but the mechanisms promoting their emergence are not well understood. Transposable elements (TEs), including endogenous retroviruses (ERVs), have been proposed to facilitate regulatory network evolution because they contain regulatory elements that can amplify in number and/or move throughout the genome (1–3). Genomic studies support this model (4), revealing that a substantial fraction of TE-derived noncoding sequences evolve under selective constraint (3, 5), are frequently bound by transcription factors (6–10), and often exhibit cell type–specific chromatin states consistent with regulatory activity (11, 12). These observations implicate TEs as a potential source of lineage-specific cis-elements capable of rewiring regulatory networks, but the adaptive consequences of this process for specific physiological functions remain largely unexplored.

We investigated the evolution of gene regulatory networks induced by the proinflammatory cytokine interferon-γ (IFNG). Interferons are proinflammatory signaling molecules that are released upon infection to promote transcription of innate immunity factors, collectively defined as IFN-stimulated genes (ISGs) (13). ISGs are regulated by cis-regulatory elements that are bound by IFN interferon regulatory factor) and STAT (signal transducer and activator of transcription) transcription factors upon activation of IFN signaling pathways (13). Although innate immune signaling pathways are conserved among mammals, the transcriptional outputs of these pathways differ across species (14, 15), likely reflecting lineage-specific adaptation in response to independent host-pathogen conflicts. Thus, these pathways provide useful systems that allow us to investigate whether TE-derived regulatory elements influence biological outcomes.

To explore the influence of TEs on IFNG-inducible regulatory networks, we examined their contribution to IFRC1 and STAT1 binding sites with the use of published chromatin immunoprecipitation sequencing (ChIP-seq) data for three human cell lines treated with IFNG: K562 myeloid-derived cells, HeLa epithelial-derived cells, and primary CD14+ macrophages (16, 17). Our initial analysis revealed 27 TE families enriched within IFNG-induced binding peaks in at least one of the data sets examined (18) (table S1 and fig. S1, A and B) and included TEs previously predicted to be cis-regulatory elements (11, 19). These sequences contained evolutionarily young to ancient TE families, of which the majority (20 of 27) originated from long terminal repeat (LTR) promoter regions of ERVs (Fig. 1A). These data suggest that ERVs, which arose from ancient retroviral infections and currently constitute 8% of the human genome (20), represent a source of novel binding sites bound by IFNG-inducible transcription factors. We next investigated whether these ERVs may contribute to IFNG-inducible regulation of adjacent cellular genes. ERV bound by STAT1 and/or IRF1 in CD14+ macrophages were strongly enriched near ISGs (binomial test, P = 1.4 × 10−30; Fig. 1B and fig. S2), determined from a matched RNA-seq data set (table S2) (18, 20). A complementary approach using the genomic regions enrichment of annotations tool (GREAT) (20) revealed enrichment of CD14+ STAT1-bound and/or IRF1-bound ERVs near genes annotated with immune functions (fig. S3, A and B). These findings suggest a potentially widespread role for ERVs in the regulation of the human IFNG response.

MER41 is an endogenized gammaretrovirus that invaded the genome of an anthropoid primate ancestor ~45 to 60 million years ago with 7900 LTR elements, from six subfamilies (MER41A, B, C, D, E, and G), now fixed in the human genome (fig. S4A). Our analysis revealed the primate-specific MER41 family of ERVs as a source of IFNG-inducible binding sites (fig. S4B), with nearly 1000 copies in humans (N = 962) bound by
STAT1 and/or IRF1 in at least one cell type (table S3 and fig. S4C). In CD14+ macrophages, STAT1-bound MER41 elements exhibited stereotypical induction of histone H3 Lys27 (H3K27) acetylation upon IFNG stimulation, a hallmark of cis-regulatory enhancer activity (23) (Fig. 1C).

Consistent with the idea that this ERV family affects IFNG-inducible regulation, MER41B sequences were identified as enriched within STAT1 ChIP-seq peaks in IFNG-stimulated HeLa cells (19). A tandem pair of predicted STAT1 binding sites coincided with STAT1 ChIP-seq peak localization (Fig. 1D). These sites also occur in the ancestral (consensus) sequence of the MER41B subfamily (Fig. 1D) but not in the MER41A subfamily, which is characterized by a 43-base pair (bp) deletion that has eliminated these binding sites (fig. S5). MER41A sequences showed no enrichment within IFNG-inducible binding sites, despite otherwise sharing 99% sequence identity with MER41B (figs. S4B and S5). Together, these data suggest that many MER41 elements are directly bound by STAT1 upon IFNG treatment, likely owing to the presence of ancestral STAT1 binding motifs within their LTR sequences.

Next, we focused on the MER41.AIM2 ERV, which is located 220 bp upstream of the gene Absent in Melanoma 2 (AIM2), an ISG that encodes a sensor of foreign cytosolic DNA and activates an inflammatory response (24). AIM2 is IFNG-inducible in humans but is constitutively transcribed in mice (24). In humans, MER41AIM2 appears to provide the only STAT1 binding site within 50 kb of the AIM2 gene, and the element gained H3K27 acetylation upon IFNG stimulation (Fig. 2A). Therefore, the regulation of AIM2 has undergone evolutionary divergence across mammalian lineages, which in turn suggests that the transposition of MER41 upstream of AIM2 may have conferred regulation by IFN signaling in anthropoid primates.

We used the CRISPR-Cas9 system to delete the MER41.AIM2 element in HeLa cells (fig. S6) (18). Cells homozygous for the MER41AIM2 deletion (∆MER41AIM2) failed to express AIM2 upon IFNG treatment, in contrast to control cells in which AIM2 transcript levels were robustly induced by IFNG (fig. 2B). IFNG-induced AIM2 protein levels were undetectable in ∆MER41AIM2 cells (Fig. 2C), thus demonstrating that MER41AIM2 is necessary for endogenous IFNG-inducible regulation of AIM2.

We further delineated the regulatory activity of MER41AIM2 by means of luciferase reporter assays (18). MER41AIM2 was sufficient to drive IFNG-inducible reporter expression in HeLa cells, and this activity was significantly diminished by point mutations ablating the predicted STAT1 binding motifs (Fig. 2D). These binding sites are conserved across anthropoid primates (fig. S7A), and IFNG-inducible reporter activity is conserved across orthologous MER41AIM2 elements cloned from chimpanzee, rhesus macaque, and marmoset (Fig. 2D). We also confirmed that orthologs of AIM2 were all IFNG-inducible in primary fibroblasts from these species (fig. S7B). These results establish MER41AIM2 as an IFNG-inducible enhancer and suggest that it was co-opted for AIM2 regulation in an ancestor of anthropoid primates.

The binding of AIM2 to cytoplasmic double-stranded RNA from intracellular bacteria and viruses promotes the assembly of a molecular platform known as an inflammasome, which initiates pyroptotic cell death by cleaving and activating caspase-1 (25). To test whether MER41AIM2 is required for this response to infection, we infected AMER41AIM2 cells with vaccinia virus (VACV) for 24 hours and assayed secretion of the active cleaved form of caspase-1 (subunit p10) as the readout of inflammasome activity. Secreted levels of activated caspase-1 were markedly reduced in ∆MER41AIM2 cells relative to wild-type cells, and caspase-1 activation was restored by transient transfection with an AIM2 overexpression construct [pCMV-AIM2 plasmid (Fig. 2E)]. Collectively these experiments demonstrate
that MER41.AIM2 is likely a necessary element of the inflammatory response to infection.

The dispersion of cis-regulatory elements propagated by the same TE family might facilitate the recruitment of multiple genes into the same regulatory network (3). We identified three additional MER41 elements within 20 kb of APOL1, IFI6, and SECTM1, which are all involved in human immunity (26–28) (Fig. 3A). As with MER41.AIM2, we used CRISPR-Cas9 to generate genomic deletions of MER41.APOL1, MER41.IFI6, and MER41.SECTM1 in HeLa cells (figs. S8 and S9). Upon treatment with IFNG, each mutant cell line exhibited significantly decreased transcript levels of the corresponding ISG relative to wild-type levels (Fig. 3B), indicating that these MER41 elements had also been co-opted as IFNG-inducible enhancers. However, in contrast to AM2, deletion of these MER41 elements did not completely abolish IFNG-induced transcript levels of these genes. This difference may be due to additional STAT1 binding sites located near these genes (Fig. 3A).

ERVs related to the primate-specific MER41 family (“MER41-like”) have been identified in

![Fig. 2. A MER41 element is essential for AIM2 inflammasome activation.](http://science.sciencemag.org/content/351/6277/1085/F2)

![Fig. 3. Multiple MER41 elements have been co-opted to regulate the IFNG response.](http://science.sciencemag.org/content/351/6277/1085/F3)
Fig. 4. IFNG-inducible ERVs are pervasive in mammalian genomes. (A) A consensus mammalian species phylogeny overlain with boxplots (median and 25th/75th percentiles) depicting the estimated age of MER41-like amplifications (18). My. million years ago; triangles depict conserved GAS motifs. (B) Luciferase reporter assays of MER41-like LTR consensus sequences from cow and dog (18). (C) Heat map of ChIP-seq signals centered on STAT1 peak summits within muroid-specific RLTR30B elements. Columns depict STAT1 ChIP-seq data from mouse bone marrow–derived macrophages (BMM) that were either untreated or treated with IFNB or IFNG. Only RLTR30B elements that are bound by STAT1 upon IFNG treatment are shown. Bottom metaprofiles represent average normalized ChIP signal across bound elements. (D) A consensus mammalian species phylogeny overlain with boxplots (median and 25th/75th percentiles) depicting the estimated age of MER41-like amplifications (18). My. million years ago; triangles depict conserved GAS motifs. (E) Luciferase reporter assay of RLTR30B consensus sequence, as in (B). [Time-calibrated phylogenies in (A) and (D) are from (34).] *P < 0.05, Student’s t test. Error bars denote SD.

most major mammalian lineages (30), raising the possibility of similar contributions to immune regulation. Further analysis, including cross-species genomic alignments, confirmed that multiple mammalian lineages were independently colonized by related MER41-like gammaretroviruses ~50 to 75 million years ago (table S4). Remarkably, we found that the tandem STAT1 binding motifs present in anthropoid MER41 are conserved in MER41-like relatives found in lemuriforms, vespertilionids, carnivores, and artiodactyls (Fig. 4A and fig. S10), which suggests that they might also have dispersed IFN-inducible enhancers in the genomes of these species. Consistent with this prediction, we found that reconstructed ancestral (consensus) sequences of MER41-like LTRs from dog and cow can drive robust IFN-inducible reporter activity in HeLa cells (Fig. 4B).

These results suggest that ERVs may have independently expanded the IFN regulatory network in multiple mammalian lineages. To further investigate this possibility, we analyzed a STAT1 ChIP-seq data set of IFNG- and IFN-β (IFNB)-stimulated primary macrophages from mouse (11), a species that lacks MER41-like elements but harbors a diverse repertoire of lineage-specific ERVs (30). Our analysis revealed a murid-specific endogenous gammaretrovirus named RLTR30B enriched for both IFNG- and IFNB-inducible STAT1 binding events (Fig. 4C and fig. S11A), which coincide with overlapping motifs corresponding to both IFNG- and IFNB-induced STAT1 binding sites located in the 5′ end of the LTR consensus sequence (Fig. 4D). Reporter assays revealed that the consensus sequence of RLTR30B also provides IFNG-inducible enhancer activity in HeLa cells (Fig. 4E). GREAT analysis also revealed significant enrichment of mouse STAT1-bound ERVs near functionally annotated immunity genes (fig. S11B).

Together, our findings reveal IFN-inducible enhancers introduced and amplified by ERVs in many mammalian genomes. On occasion, these elements have been co-opted to regulate host genes encoding immunity factors. Although we have shown that ERVs play a functional role regulating innate immune pathways in human HeLa cells, further studies will be necessary to extend our findings to primary hematopoietic cells and other species such as mouse. We speculate that the prevalence of IFN-inducible enhancers in the LTRs of these ancient retroviruses is not coincidental, but may reflect former viral adaptations to exploit immune signaling pathways promoting viral transcription and replication (32). Indeed, several extant viruses, including HIV, possess IFN-inducible cis-regulatory elements (33). It would be ironic if viral molecular adaptations had been evolutionarily recycled to fuel innovation and turnover of the host immune repertoire. Regardless of how these sequences originated, our study illuminates how selfish genetic elements have contributed raw material that has been repurposed for cellular innovation.

REFERENCES AND NOTES

18. See supplementary materials on Science Online.
Expression homeostasis during DNA replication

Yoav Voichék,* Raz Bar-Ziv,* Naama Barkai†

Genome replication introduces a stepwise increase in the DNA template available for transcription. Genes replicated early in S phase experience this increase before late-replicating genes, raising the question of how expression levels are affected by DNA replication. We show that in budding yeast, messenger RNA (mRNA) synthesis rate is buffered against changes in gene dosage during S phase. This expression homeostasis depends on acetylation of H3 on its internal K56 site by Rtt109/Asf1. Deleting these factors, mutating H3K56 or up-regulating its deacetylation, increases gene expression in S phase in proportion to gene replication timing. Therefore, H3K56 acetylation on newly deposited histones reduces transcription efficiency from replicated DNA, complementing its role in guarding genome stability. Our study provides molecular insight into the mechanism maintaining expression homeostasis during DNA replication.

The synthesis of mRNA depends on protein factors binding to the DNA template. During the cell cycle, DNA dosage increases at discrete times in S phase, whereas cell volume increases continuously, introducing considerable temporal variations in DNA concentration. How these variations in DNA level affect mRNA synthesis was examined in classical studies (1). In bacteria, mRNA production follows gene dosage, so that the expression of each gene increases rapidly after its replication (2–4). By contrast, experiments in eukaryotic cells, ranging from yeast to mammals (5–7), indicate a limited dependency of gene expression on DNA dosage, prompting the hypothesis that transcription of newly replicated DNA is transiently repressed (8).

We extended previous studies, which measured total mRNA synthesis (9, 10), or focused on individual genes (11, 12) by directly comparing the expression of early versus late-replicating genes during S phase. If replicated loci produce more mRNA than unreplicated ones, then expression of genes that replicate early should increase relative to the expression of late-replicating genes during S phase (fig. S1A). In contrast, we find that the relative expression of early- versus late-replicating genes remained relatively constant in budding yeast, progressing synchronously through S phase after release from α-factor or hydroxyurea (HU) arrest and did not correlate with DNA replication timing (Fig. 1, A and B, and figs. S1 and S2). Further, the synthesis rates of early-replicating genes increased by only ~20% relative to late-replicating genes, significantly less than the ~70% increase in DNA content (Fig. 1A). We also examined cells arrested in the beginning of S phase after 3-hour treatment with HU. Despite the stable increase in DNA content of early-replicated genes, their expression increased by a mere 5% relative to that of late nonreplicated genes, suggesting that buffering under this S phase–arrest condition is even stronger than in cycling cells (Fig. 1C and fig. S8D). Taken together, our results are consistent with previous studies showing that during S phase, DNA dosage has a limited influence on mRNA synthesis rates.

In contrast to mRNA levels, the binding of RNA polymerase II to DNA did correlate with DNA content in HU-arrested cells and after release into S phase. Still, the increase in PolII binding to replicated genes (30%) was lower than expected by the increase in DNA content (Fig. 2A and fig. S3). In HU-arrested cells, early-replicated genes were depleted of elongating PolII (Fig. 2B). However, this difference was specific to HU arrest and disappeared upon release, before the completion of replication. Therefore, reduced PolII binding to replicated DNA may partially account for the buffering of gene expression, with additional differences in elongation capacity that increases buffering in HU-arrested cells.

We hypothesized that chromatin regulators may suppress transcription from replicated DNA. To identify such factors, we examined a published data set describing how individual deletions of 165 chromatin-associated factors affect the genome-wide expression profile (13). Deleting a factor that limits transcription from replicated DNA will increase gene expression in proportion to the time at which the gene is replicated in S phase, so that early-replicated genes will increase in expression more than genes replicated late. We therefore searched for mutants in which gene expression levels were (on average) negatively correlated with gene replication timing (Fig. 3A). Of the three mutants showing the strongest correlation between gene expression and replication timing, two were involved in H3 acetylation: the acetyltransferase Rtt109 and its histone chaperone cofactor Asf1 (14–16). A similar effect was detected in expression data from fission yeast deleted of the Asf1 paralogue (fig. S4A) (17, 18). The third candidate, Tos4, is a less-characterized putative transcription factor (19). All three genes increase in expression during G1, just before DNA replication (20).

The correlation between gene expression and replication timing were highly significant in all three mutants, but the difference in expression between early- and late-replicated genes was small. This is expected, as measurements were taken in asynchronous cultures in which only a minority of cells are in S phase. To amplify the difference in mRNA levels between early- and late-replicated genes, we profiled gene expression in all three mutants synchronized by HU (Fig. 3B). In both Rtt109 and Δasf1 cells (but not Δtos4), expression levels correlated with DNA content, with genes that were already replicated showing 24 to 28% increase in expression relative to the nonreplicated genes. We next measured the mRNA levels and synthesis rates in Δrtt109 cells progressing synchronously through S phase. Indeed, expression and synthesis rates of early-replicated genes increased transiently during mid–S phase relative to late-replicating genes (Fig. 3C and fig. S5). Therefore, Rtt109 is required for buffering mRNA synthesis during DNA replication (Fig. 3C, red line).

A similar loss of buffering was observed for Δasf1 (Fig. 3D) and also for Δtos4 cells (fig. S7). The latter is particularly notable because Δtos4 did not abrogate buffering in HU-arrested conditions, consistent with an additional buffering mechanism acting upon HU arrest. To examine whether the three candidates act through the same pathway, we measured gene expression in the double deletions Δrtt109 Δasf1 and Δrtt109Δtos4. The increased expression of early-replicated genes was similar to the single deletions for the two pairs, suggesting that the three genes function

ACKNOWLEDGMENTS

Accession numbers for the published data sets analyzed in this study are available in the supplementary materials. We thank all members of the Elde and Feschotte labs for insightful discussions. We thank A. Kapusta, A. Lewis, D. Downhour, J. Carleton, and K. Cone for technical assistance, and D. Hancks and J. F. McCormick for their critical input. Supported by a Pew Charitable Trusts award and NIH grants GM085454 and GM14534 (N.C.E.) and by NIH grants GM112972 and GM109290 (C.F.). E.B.C. is a Howard Hughes Medical Institute postdoctoral fellow of the Jane Coffin Childs Fund. N.C.E. is a Pew Scholar in the Biomedical Sciences and Mario R. Capecchi Endowed Chair in Genetics. The authors declare no financial conflicts of interest.

SUPPORTING MATERIALS

www.sciencemag.org/content/351/6277/1087/suppl/DC1

Materials and Methods
Tables S1 to S6
Fig. S1 to S11
References (25–49)
30 September 2015; accepted 2 February 2016
10.1126/science.aad5497
Regulatory evolution of innate immunity through co-option of endogenous retroviruses
Edward B. Chuong, Nels C. Elde and Cédric Feschotte

Science 351 (6277), 1083-1087.
DOI: 10.1126/science.aad5497

Regulatory use of endogenous retroviruses
Mammalian genomes contain many endogenous retroviruses (ERVs), which have a range of evolutionary ages. The propagation and maintenance of these genetic elements have been attributed to their ability to contribute to gene regulation. Chuong et al. demonstrate that some ERV families are enriched in regulatory elements, so that they act as independently evolved enhancers for immune genes in both humans and mice (see the Perspective by Lynch). The analysis revealed a primate-specific element that orchestrates the transcriptional response to interferons. Selection can therefore act on selfish genetic elements to generate novel gene networks.

Science, this issue p. 1083 see also p. 1029