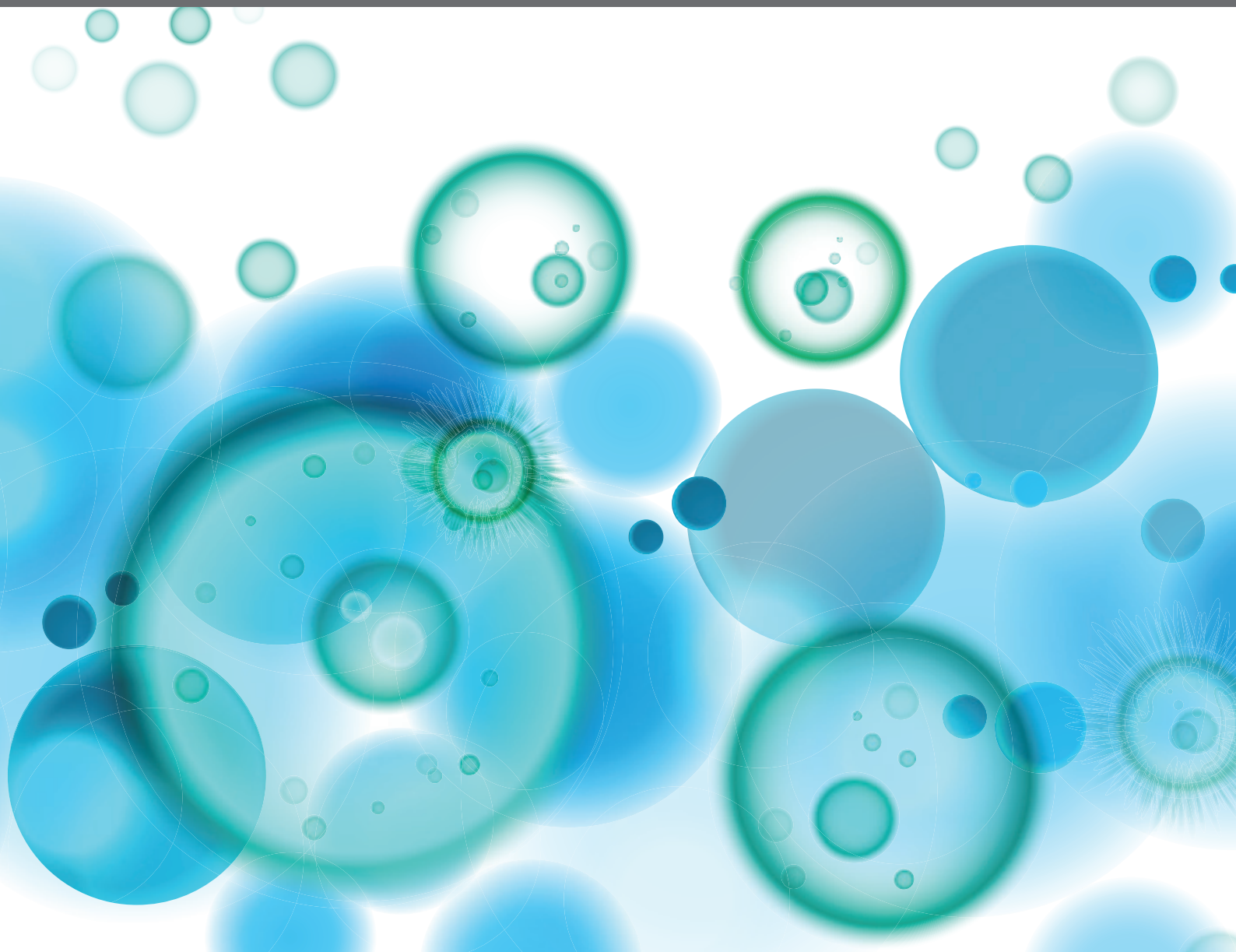


THE PAST AND THE FUTURE OF HUMAN IMMUNITY UNDER VIRAL EVOLUTIONARY PRESSURE

EDITED BY: Gkikas Magiorkinis and Tara P. Hurst

PUBLISHED IN: Frontiers in Immunology and Frontiers in Microbiology





frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence.

The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714

ISBN 978-2-88963-229-9

DOI 10.3389/978-2-88963-229-9

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

THE PAST AND THE FUTURE OF HUMAN IMMUNITY UNDER VIRAL EVOLUTIONARY PRESSURE

Topic Editors:

Gkikas Magiorkinis, National and Kapodistrian University of Athens, Greece

Tara P. Hurst, Birmingham City University, United Kingdom



Image: Gkikas Magiorkinis

"Black and Red Queen evolutionary dynamics" by

Gkikas Magiorkinis is licensed under CC-BY

There is a long-standing evolutionary battle between viruses and their hosts that continues to be waged. The evidence of this conflict can be found on both sides, with the human immune system being responsive to new viral challenges and viruses having developed often sophisticated countermeasures. The "arms race" between viruses and hosts can be thought as an example of the "Red Queen" race, an evolutionary hypothesis inspired from the dialogue of Alice with the Red Queen in Lewis Carroll's "Through the Looking-Glass". At the same time, viruses have a minimal genomic content as they have evolved to hitchhike biological machinery of their hosts (or other co-infecting viruses). The minimalistic viral genome could be thought as the result of a "Black Queen" evolution, a theory inspired from the card game Heart, where the winner is the one with the fewest points at the end.

The effects of this arms race are evident in the evolution of the human immune system. This system is capable of responding to diverse viral challenges, utilizing both the ancient innate immune system and the more recently evolved adaptive immune system of jawed vertebrates. It is now well-known that the two systems are linked, with innate immunity hypothesized to have provided raw material for the emergence of the adaptive immune response. The adaptive immune response comprises several protein families (including B and T cell receptors, MHC and KIR proteins, for example) that are encoded by complex and variable genomic regions. This complexity enables for responsive genetic changes to occur in immune cells,

such as the ability of genomic hypervariable regions in B cells to recombine in order to produce more specific antibodies. Indeed, the human immune system is thought to be continually evolving via various mechanisms such as changes in the genes encoding immune receptors and the regulatory sequences that control their expression. For example, there is some evidence that exogenous viral infections can alter the expression of endogenous retroviruses, some of which contribute to the immune response.

Viral countermeasures can include encoding decoy receptors for the signalling molecules of the immune response, altering the gene expression of adaptive immune cells during chronic infection or using host enzymes to facilitate viral immune escape. As the articles herein show, the immune system continues to be challenged by viral infections and these challenges continue to shape how the immune system combats pathogens, thus viruses and human immunity are continuously part of “Red and Black Queen” evolutionary dynamics.

We had the pleasure of working with Jonas Blomberg as a reviewer during the course of the Research Topic and his untimely passing was a great loss. Prof. Blomberg made significant contributions, including to the nomenclature of endogenous retroviruses (ERVs), the evolution and characterization of specific human ERV (HERV) and the contribution of ERVs to diseases such as cancer. It is with great respect for his contributions to the ERV field that we dedicate this eBook to his memory.

Citation: Magiorkinis, G., Hurst, T. P., eds. (2019). The Past and the Future of Human Immunity Under Viral Evolutionary Pressure. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88963-229-9

Table of Contents

- 06 Editorial: The Past and the Future of Human Immunity Under Viral Evolutionary Pressure**
Tara Patricia Hurst and Gkikas Magiorkinis
- 09 Inflation vs. Exhaustion of Antiviral CD8+ T-Cell Populations in Persistent Infections: Two Sides of the Same Coin?**
Emanuele Marchi, Lian Ni Lee and Paul Klenerman
- 20 RNA Structure—A Neglected Puppet Master for the Evolution of Virus and Host Immunity**
Redmond P. Smyth, Matteo Negroni, Andrew M. Lever, Johnson Mak and Julia C. Kenyon
- 27 Influenza Virus—Host Co-evolution. A Predator-Prey Relationship?**
Konstantinos Voskarides, Eirini Christaki and Georgios K. Nikolopoulos
- 34 APOBEC3G Regulation of the Evolutionary Race Between Adaptive Immunity and Viral Immune Escape is Deeply Imprinted in the HIV Genome**
Faezeh Borzooee, Krista D. Joris, Michael D. Grant and Mani Larijani
- 54 Human Endogenous Retroviruses are Ancient Acquired Elements Still Shaping Innate Immune Responses**
Nicole Grandi and Enzo Tramontano
- 70 The Exaptation of HERV-H: Evolutionary Analyses Reveal the Genomic Features of Highly Transcribed Elements**
Patrick Gemmell, Jotun Hein and Aris Katzourakis
- 80 Related Endogenous Retrovirus-K Elements Harbor Distinct Protease Active Site Motifs**
Matthew G. Turnbull and Renée N. Douville
- 98 Transcriptional Modulation of Human Endogenous Retroviruses in Primary CD4+ T Cells Following Vorinostat Treatment**
Cory H. White, Nadejda Beliakova-Bethell, Steven M. Lada, Michael S. Breen, Tara P. Hurst, Celsa A. Spina, Douglas D. Richman, John Frater, Gkikas Magiorkinis and Christopher H. Woelk
- 108 Variable Baseline Papio cynocephalus Endogenous Retrovirus (PcEV) Expression is Upregulated in Acutely SIV-Infected Macaques and Correlated to STAT1 Expression in the Spleen**
Emmanuel Atangana Maze, Claire Ham, Jack Kelly, Lindsay Ussher, Neil Almond, Greg J. Towers, Neil Berry and Robert Belshaw
- 123 Pockets of HIV Non-infection Within Highly-Infected Risk Networks in Athens, Greece**
Leslie D. Williams, Evangelia-Georgia Kostaki, Eirini Pavlitina, Dimitrios Paraskevis, Angelos Hatzakis, John Schneider, Pavlo Smyrnov, Andria Hadjickou, Georgios K. Nikolopoulos, Mina Psychogiou and Samuel R. Friedman
- 131 Immunomodulatory Function of HBeAg Related to Short-Sighted Evolution, Transmissibility, and Clinical Manifestation of Hepatitis B Virus**
Anna Kramvis, Evangelia-Georgia Kostaki, Angelos Hatzakis and Dimitrios Paraskevis

- 143** *Hepatitis B Virus Adaptation to the CD8+ T Cell Response: Consequences for Host and Pathogen*
Sheila F. Lumley, Anna L. McNaughton, Paul Klenerman, Katrina A. Lythgoe and Philippa C. Matthews
- 157** *Impact of Interferon- α Receptor-1 Promoter Polymorphisms on the Transcriptome of the Hepatitis B Virus-Associated Hepatocellular Carcinoma*
Timokratis Karamitros, George Papatheodoridis, Dimitrios Paraskevis, Angelos Hatzakis, Jean L. Mbisa, Urania Georgopoulou, Paul Klenerman and Gkikas Magiorkinis
- 168** *Evolution of Two Major Zika Virus Lineages: Implications for Pathology, Immune Response, and Vaccine Development*
Jacob T. Beaver, Nadia Lelutiu, Rumi Habib and Ioanna Skountzou



Editorial: The Past and the Future of Human Immunity Under Viral Evolutionary Pressure

Tara Patricia Hurst^{1*} and Gkikas Magiorkinis^{2*}

¹ Department of Life Sciences, School of Health Sciences, Birmingham City University, Birmingham, United Kingdom,

² Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece

Keywords: human viruses, endogenous retrovirus, immunity, evolution, disease

Editorial on the Research Topic

The Past and the Future of Human Immunity Under Viral Evolutionary Pressure

“Nothing in biology makes sense except in the light of evolution” was argued by Theodosius Dobzhansky in 1973. After more than 100 years since Charles Darwin’s *Origin of Species*, it was increasingly realized that the theory of evolution can help us disentangle biological mechanisms. Here we asked contributors to see their antiviral research interests “through the looking glass” of evolution and to consider how evolutionary theory may help us understand human antiviral immunity. Thus, our special topic explored the scope of the evolution of human antiviral immunity, whether this included short-term effects or long-term impact. We specifically asked contributors to provide thought-provoking manuscripts that could fall within the scope of co-evolution of viruses and the human immune response. They submitted articles discussing the ongoing dynamics of co-existing with ancient endogenous retroviruses, as well as the pressures exerted by extant exogenous viruses such as influenza and human immunodeficiency virus (HIV-1). Our understanding of antiviral immunity was broadened by papers exploring mechanisms such as T cell function and RNA structural components.

In a Hypothesis and Theory article, Marchi et al. explore the effects of persistent infection on T cell responses by analyzing gene expression profiles and gene networks in publicly-available data sets. Importantly, they identified T-box 21 (Tbx21)-mediated gene expression as a hallmark of inflation but not exhaustion of memory CD8⁺ T cells. This finding has potential to shape our understanding of the genes that define memory T-cell populations. Further, this shows how different viral infections can influence either exhaustion or inflation. This highlights the adaptability of the immune response to viral infection but also how infection could alter gene expression to shift T cell phenotypes.

In one of two mini-reviews in the topic, Smyth et al. discuss the contribution of RNA structure to antiviral immunity. RNA is much more than an intermediary between DNA and protein; it can adopt complex three-dimensional structures that may be altered by single nucleotide mutations, with consequent effects on its functions. Thus, since RNA structures are critical to the replication of many viruses, they are also substrates upon which evolution can act. This can include non-coding RNAs produced during replication, as well as genomic RNA and transcripts. Viral RNA is sensed by different pattern recognition receptors (PRRs) of the innate immune response, such as Toll-like receptor 3 (TLR3) and retinoic acid inducible gene I (RIG-I). There is thus pressure on viral RNA to evade immune detection such as by nucleotide and codon usage bias (adenosine by HIV-1), the presence of specific secondary structures that restrict transcription to low levels, or the maintenance of single-stranded RNA regions to avoid PRR detection. In turn, viral RNA structures have influenced the human immune system, such as the evolution of the human leukocyte antigen (HLA) locus.

OPEN ACCESS

Edited and reviewed by:

Ian Marriott,
University of North Carolina at
Charlotte, United States

*Correspondence:

Tara Patricia Hurst
tara.hurst@bcu.ac.uk
Gkikas Magiorkinis
gkikasmag@gmail.com

Specialty section:

This article was submitted to
Microbial Immunology,
a section of the journal
Frontiers in Immunology

Received: 16 August 2019

Accepted: 17 September 2019

Published: 02 October 2019

Citation:

Hurst TP and Magiorkinis G (2019)
Editorial: The Past and the Future of
Human Immunity Under Viral
Evolutionary Pressure.
Front. Immunol. 10:2340.
doi: 10.3389/fimmu.2019.02340

The co-evolution of influenza virus and host immunity is discussed in the second mini-review (Voskarides et al.). The authors argue that this is an example of antagonistic evolution, with influenza virus a predator and the human host as prey. This conflict manifests in the genetic variability of Influenza viruses, a result of the absence of proof-reading by the RNA polymerase, that continually challenges the host immune system. In response, the human host has genetic major histocompatibility complex (MHC) diversity which facilitates T cell-mediated immunity to diverse pathogens.

APOBEC3G (apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3G, A3G) is a cellular cytidine deaminase that mutates retroviral genomes, thereby being a key host cell mechanism to debilitate viruses that embed in the genome. In an original article herein, the authors suggest that the action of A3G extends to adaptive immunity, suggesting that HIV-1 cytotoxic T cell (CTL) escape mutants are generated through A3G activity (Borzoee et al.). The authors use a bioinformatics approach to simulate A3G mutations and show that HIV-1 mutants are biased toward cytotoxic T cell (CTL) escape mutants. These results point toward an evolutionary adaptation of the HIV-1 genome that essentially “hijacks” human antiretroviral activity to counteract T-cell immunity.

A cluster of papers within the topic discuss the role of endogenous retroviruses (ERV) in the immune response. This includes an up-to-date review on the shaping of the immune response by human ERVs (HERVs) (Grandi and Tramontano). Further, four original research articles explore diverse aspects of ERVs and immunity. Firstly, ERV sequences within the genome are a potential source of genetic diversity that are known to be co-opted by the host. This has been shown most recently to include HERV-H sequences contributing to stem cell identity. Gemmell et al. explore the nature of HERV-H loci and their transcription further, showing positive correlations between particular long-terminal repeats (LTRs) and cell types (type II LTRs and embryonic cells).

Secondly, Turnbull and Douville analyzed HERV-K protease sequences within the human genome. They show that transcription of HERV-K proteases varies according to the functional motifs, with two motifs (DTGAD, DTGVD) being more abundant among all HERV-K integrations, but also more active, compared to the less common motifs, in diseases such as amyotrophic lateral sclerosis and breast cancer. These results are important not only for understanding the interactions among HERV-K integrations evolutionarily, but also to appreciate the diversity of enzymatic targets when designing anti-HERV based treatments.

Thirdly, the effect of the histone deacetylase (HDAC) inhibitor vorinostat on HERV expression in CD4+ T cells was analyzed using RNA-Seq (White et al.). Vorinostat is one latency reversing agent being examined in HIV-1 shock and kill strategies. The study found over 2000 HERV elements were modulated by vorinostat, notably the downregulation of ERVL and upregulation of HERV-9 elements, the latter of which was confirmed by digital droplet PCR. This study identified three HERV-9 members (LTR12 elements) that were upregulated by vorinostat in a dose-dependent manner. The relevance of this

finding lies in the transcriptional regulation of certain genes by LTR12 elements, including the pro-apoptotic genes TP3 and TNFRSF10B (White et al.). In agreement with our previous work using RT-qPCR, members of the HERV-K (HML-2), HERV-W and HERV-FRD families were not found to be upregulated by vorinostat (1). Notably the study by White et al. unlike that of Hurst et al., used uninfected, primary CD4+ T cells to avoid confounding effects of HIV-1 infection.

Finally, ERVs can be upregulated in response to acute exogenous viral infections and this may contribute to the antiviral response. In the paper by Maze et al. the expression of *Papio cynocephalus* Endogenous Retrovirus (PcEV) was found to be upregulated in macaque plasma and tissues (peripheral blood mononuclear cells (PBMCs) and spleen) in response to simian immunodeficiency virus (SIV) infection. Further, upregulation of the interferon (IFN)-stimulated gene (ISG), STAT1, an important marker of the early antiviral response, correlated with the increased PcEV RNA levels (Maze et al.). The authors posit a role for ERVs in the activation of the innate immune response.

The effect of immunity-virus interface at the population level was examined by Williams et al. where they present the perplexing phenomenon of a “pocket” of uninfected individuals in a connected network of recently HIV-1-infected individuals. This raises the question of how a cluster of connected HIV-1 exposed individuals could remain uninfected within an HIV-1 outbreak. The authors discuss potential explanations including the possibility that long-term infections act as a “firewall” in spreading the infections to the full-population. The latter could be explained by the existence of an epidemic equilibrium of co-existence where human-HIV co-evolution would be acting in the absence of treatment.

The evolutionary gameplay underlying the pathogenesis of hepatitis B virus's e antigen (HBVeAg) is reviewed and explored by Kramvis et al. HBVeAg, as the authors explain, is a tolerogen, an antigen that increases the tolerance of HBV infection by the human host. There is significant variability in HBVeAg seroconversion and transmission efficiency among genotypes, which the authors suggest can be explained as a result of short-sighted evolution. The virus-host interplay of HBV is further explored by Lumley et al. on the effects of CD8+ T cell responses. The authors review the evidence that CD8+ T cells play an important role in the chronicity of HBV infection as well as the disease outcomes such as cirrhosis and hepatocellular carcinoma (Lumley et al.). The role of interferon- α receptor-1 (IFNAR1) promoter polymorphisms on HBV infection are explored by Karamitros et al. with respect to the development of hepatocellular carcinoma. The authors show that the variable tandem repeat [VNTR: -77(GT)n] has a significant impact in the transcription profile of IFNAR1 and this may be relevant in cancer pathophysiology.

Finally, Beaver et al., have reviewed evidence of how Zika virus (ZIKV) evolution can explain the observed patterns of pathogenesis. The authors review evidence showing that the immune responses against African isolates differ from those observed against Asian ZIKV lineages, and discuss how

this could explain observed pathogenic differences among ZIKV strains.

From ancient germline infections to modern epidemics, outbreaks or emerging infections, our contributors explored the evolutionary interplay of human immunity and viruses at the molecular, individual, or population level. They offered thought-provoking hypotheses for further exploration, reviewed publications, and provided novel results in the field of human antiviral immunity field. This shows the

power of integrating evolutionary theory with research of antiviral immunity and we hope that our special topic will further promote evolutionary thinking across fields of human pathogen research.

AUTHOR CONTRIBUTIONS

TH and GM wrote the draft and edited the final version of the manuscript.

REFERENCES

1. Hurst T, Pace M, Katzourakis A, Phillips R, Klennerman P, Frater J, et al. Human endogenous retrovirus (HERV) expression is not induced by treatment with the histone deacetylase (HDAC) inhibitors in cellular models of HIV-1 latency. *Retrovirology*. (2016) 13:10. doi: 10.1186/s12977-016-0242-4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Hurst and Magiorkinis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Inflation vs. Exhaustion of Antiviral CD8+ T-Cell Populations in Persistent Infections: Two Sides of the Same Coin?

Emanuele Marchi^{1,2*}, Lian Ni Lee^{1,2} and Paul Klenerman^{1,2}

¹ Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, ² Translational Gastroenterology Unit, John Radcliffe Hospital, Oxford, United Kingdom

OPEN ACCESS

Edited by:

Gkikas Magiorkinis,
National and Kapodistrian University
of Athens, Greece

Reviewed by:

Ian Humphreys,
Cardiff University, United Kingdom
Anastasia Samsonova,
Saint Petersburg State University,
Russia

*Correspondence:

Emanuele Marchi
emanuele.marchi@ndm.ox.ac.uk

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 02 August 2018

Accepted: 23 January 2019

Published: 06 March 2019

Citation:

Marchi E, Lee LN and Klenerman P
(2019) Inflation vs. Exhaustion of
Antiviral CD8+ T-Cell Populations in
Persistent Infections: Two Sides of the
Same Coin? *Front. Immunol.* 10:197.
doi: 10.3389/fimmu.2019.00197

Persistent virus infection can drive CD8+ T-cell responses which are markedly divergent in terms of frequency, phenotype, function, and distribution. On the one hand viruses such as Lymphocytic Choriomeningitis Virus (LCMV) Clone 13 can drive T-cell “exhaustion”, associated with upregulation of checkpoint molecules, loss of effector functions, and diminished control of viral replication. On the other, low-level persistence of viruses such as Cytomegalovirus and Adenoviral vaccines can drive memory “inflation,” associated with sustained populations of CD8+ T-cells over time, with maintained effector functions and a distinct phenotype. Underpinning these divergent memory pools are distinct transcriptional patterns—we aimed to compare these to explore the regulation of CD8+ T-cell memory against persistent viruses at the level of molecular networks and address whether dysregulation of specific modules may account for the phenotype observed. By exploring in parallel and also merging existing datasets derived from different investigators we attempted to develop a combined model of inflation vs. exhaustion and investigate the gene expression networks that are shared in these memory pools. In such comparisons, co-ordination of a critical module of genes driven by Tbx21 is markedly different between the two memory types. These exploratory data highlight both the molecular similarities as well as the differences between inflation and exhaustion and we hypothesize that co-ordinated regulation of a key genetic module may underpin the markedly different resultant functions and phenotypes *in vivo*—an idea which could be tested directly in future experiments.

Keywords: exhaustion, inflation, bioinformatics, LCMV (lymphocytic choriomeningitis virus), CMV (cytomegalovirus)

INTRODUCTION

CD8+ T cell responses play a critical role in control of many virus infections. In the case of Lymphocytic Choriomeningitis (LCMV) it has been extensively modeled in the mouse and the dynamics, specificity, and function of antiviral CD8+ T cell responses are well-understood. One feature of this infection is the development of CD8+ T cell exhaustion, a feature first described in the pre-tetramer era as loss of function and finally deletion in the presence of persisting viruses such as DOCILE strains (1), and subsequently investigated at a molecular level using LCMV Clone 13, which shows similar features (2). Mapping the transcriptional underpinning of CD8+ T cell

exhaustion was crucial in defining its mechanisms—key of which include expression of checkpoint molecules such as PD-1 (3). These discoveries have been very influential in understanding both antiviral and also anti-cancer responses and have driven the development of new checkpoint blockade therapies.

In contrast, persistent infection with cytomegaloviruses (CMVs)—human and murine CMV (MCMV)—is linked with the development of memory “inflation” (4). This is marked by the late expansion and maintenance of a number of CD8+ T cell pools directed at a subset of peptides (5). Their phenotype lacks the expression of checkpoint molecules, rather showing acquisition of markers of cellular differentiation over time. Importantly, in contrast to the exhausted phenotype, these cells retain strong effector functions. The transcriptional underpinning of this—and of the related model of memory inflation driven by adenoviral vectors (6)—has also been explored, and this is also highly distinct from the development of conventional “central” long-lived memory cells (7).

Both models are associated with viral persistence—high level viremia in the case of LCMV, and very low level local reactivation in the case of MCMV. How is the development of these two apparently very distinct forms of CD8+ T cell memory driven? To address this we aimed to compare the gene expression profiles and gene networks in these different settings. Such datasets are complex to generate and require highly reproducible and well-established models, coupled with adequate T cell numbers specific for individual epitopes. Thus, currently such data are valuable and remain an important resource to explore. Typically, even with newer techniques such as RNA-Seq, such studies only focus on immune responses to a single pathogen rather than comparing diverse pathogens.

To achieve the comparison we sought, distinct datasets generated from different platforms by two different laboratories were merged, and an integrated model created and subjected to validation. We explored whether two data sets generated from comparable experimental designs were sufficiently suitable to be merged, despite differences in array platforms, mouse suppliers and viruses, and whether this could provide some new insights into the relationship between these different T cell populations.

As outlined in **Figure 1A**, data sets comparing memory Inflation against conventional memory (7) and CD8 T cell Exhaustion vs. conventional memory (8) studies include comparable or analogous samples in which CD8 transcriptomics from early and later stages of viral infections were referred to Naïve cells. In our model, samples from these experiments were expected to be assimilated to three broad categories, the *common reference* samples (Naïve CD8 T cells, the most comparable expression profiles), the *intermediate phenotype* (CD8 T cells from not Inflating or Exhausted samples) and an *Extreme phenotype* (CD8 T cells Inflating or Exhausted, hence expected to diverge significantly from the rest of samples and hypothetically from each other. The reliability of integration could be evaluated from the expected distribution of expression profiles in exploratory analysis (i.e. principal component analysis, PCA) and further data mining could explore the validity of the model and hypotheses generated (**Figure 1A**).

Via such an integrative and comparative bioinformatics analysis of expression profiles from inflation and exhaustion

murine models, we describe similarities and dissimilarities of the phenomena at a transcriptomic level. We observe some predicted and also some unexpected features, with some possible practical implications for future experimental design. The results presented here explore the use of such a workflow as one approach to integrate valuable existing datasets from different platforms. While data from such *in silico* comparisons cannot reach the quality and accuracy of those derived from a single unified *in vivo* experiment, there is an opportunity that existing, publicly available data can be studied further in order to address questions not originally anticipated at the time datasets were generated, and to help develop new ideas for the field.

RESULTS AND DISCUSSION

PCA of Expression Data From *Inflation* and *Exhaustion* Models Reveals Comparable Events Between the Phenomena

We first addressed the overall transcriptomic similarities between memory inflation and exhaustion by re-exploring a previously generated dataset, GSE73314 (7). This dataset was derived from mice infected with MCMV, tracking one inflationary response, M38, and one conventional memory response (M45) at both early (acute, d7) and late (memory, d50) timepoints.

Antigen-specific T cell populations were FACS-sorted following tetramer staining and gene expression analyzed using microarray. In parallel inflationary and conventional responses to beta-galactosidase expressed in a replication-deficient human Adenovirus serotype 5 construct (HuAd5-lacZ) were studied following HuAd5-lacZ immunization. Responses to the inflationary epitope βgal_{96} (referred to as D8V in this paper) and the conventional epitope βgal_{497} (referred to as I8V in this paper) were analyzed again at the peak acute (d21) and late memory (d100) time points. The analysis of this dataset has been previously described, including principal components analysis based on all informative genes or subsets of transcription factors, and importantly the close relationships between these expression profiles and those of human “inflationary” populations derived from studies of CMV were confirmed (7). Data from **Figure 2A** from the paper are reproduced here [previously published as Figure S4 from (7)] which depicts the first 3 principal components of these data prior to addition of the integrated exhaustion dataset. A clustering of data derived from the acute timepoint for the relevant tetramers [D8V d21 (blue), I8V d21 (yellow), M38 d7 (magenta), and M45 d7 (pink)] is observed. However, with reference to the naive (green) dataset, the inflationary populations at the late timepoints [D8V d100 (red), M38 d50 (turquoise)] sit slightly further segregated than the acute samples, while the conventional memory pools at the late timepoints (M45 d50) have shifted in the opposite direction, with some re-expression of profiles (M45 d50) linked to resting naive cells. I8V at d100 are slightly divergent from M45 at d50, possibly reflecting subtle intrinsic differences (e.g., tissue tropism, antigen levels during infection) between the two infection models, especially as MCMV is a replicating virus with periodic episodes of viral

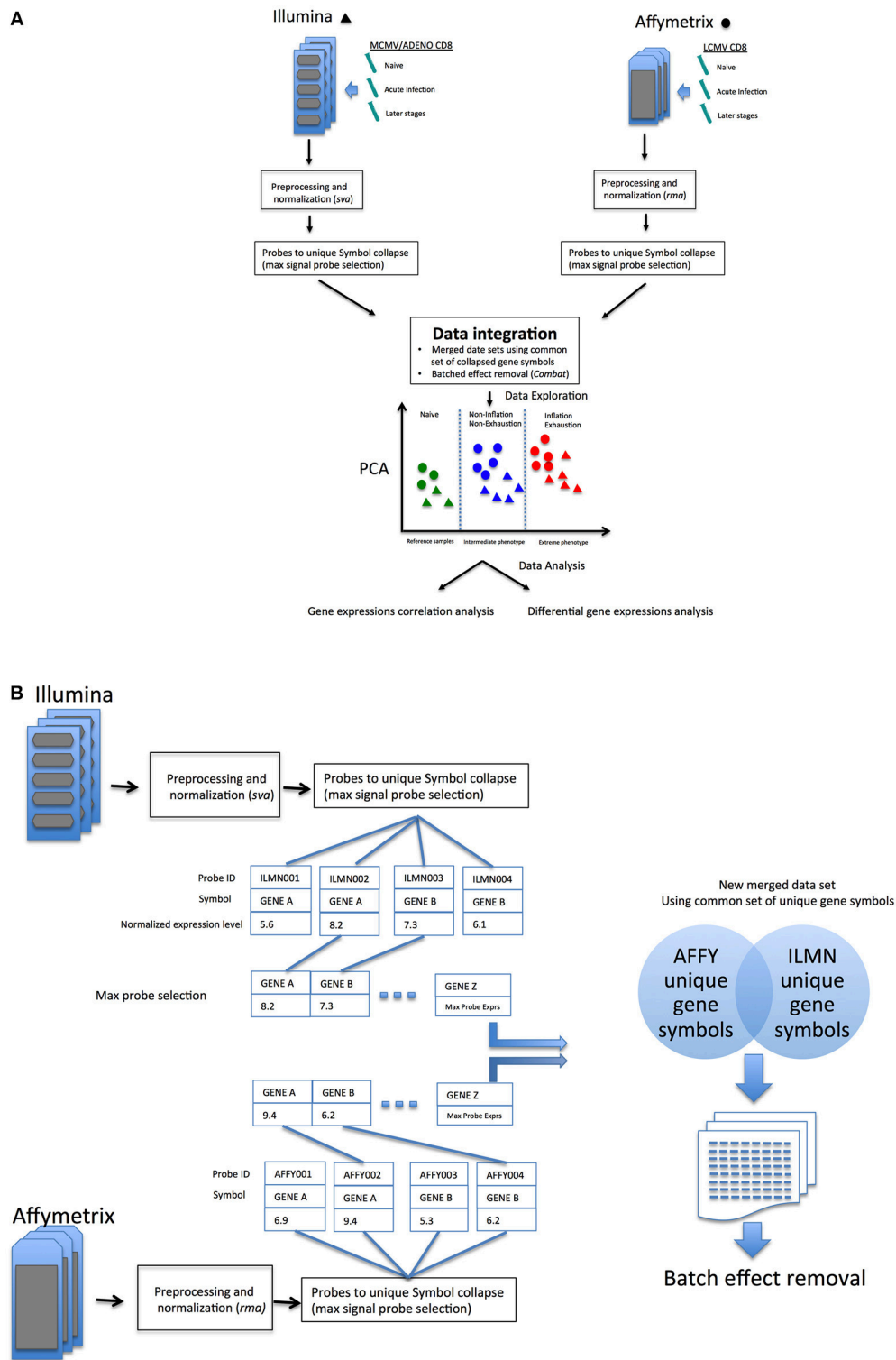
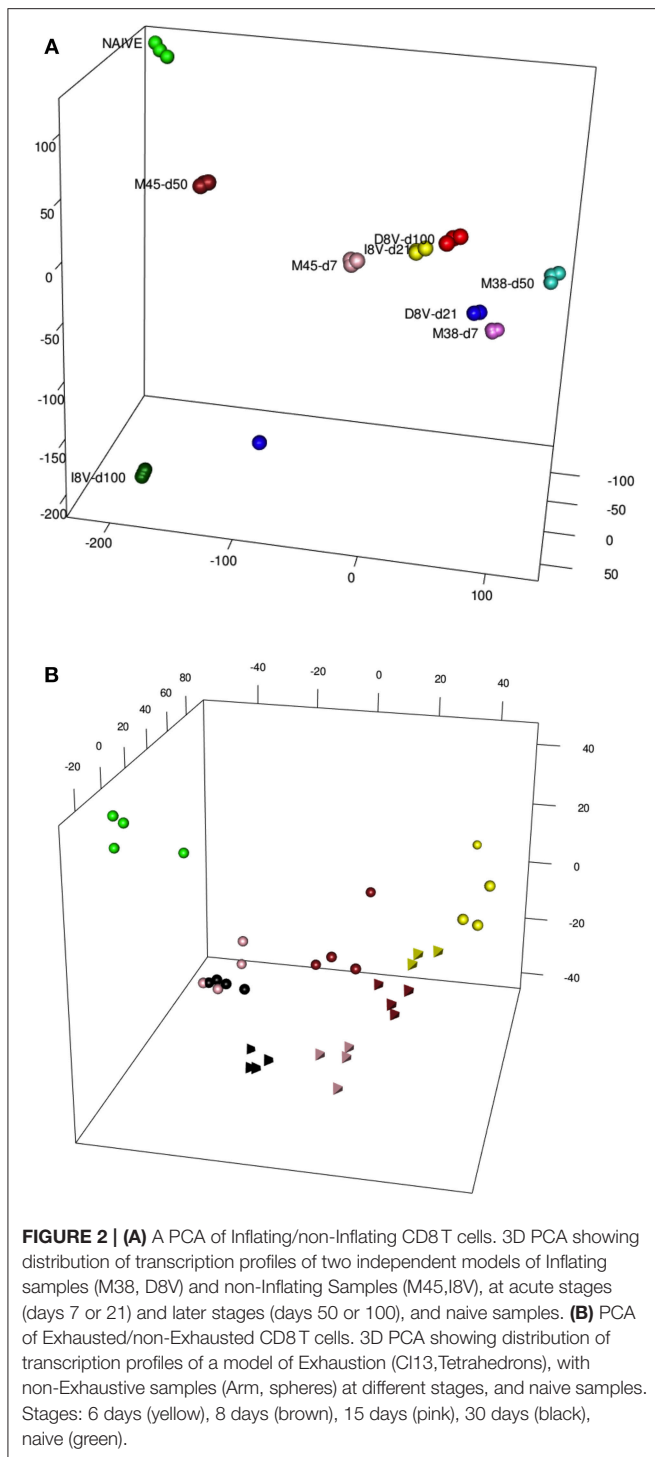


FIGURE 1 | (A) Overview of the data integration process and analysis. **(B)** Overview of probes-to-gene collapse method and data sets merge.

reactivation while AdHu5 is a non-replicating vector which does not reactivate. Nonetheless the expression of both groups of conventional memory cells at the early phase of infection

(M45 d7 and I8V d21) show reduced divergence, suggestive of a conserved response between acute and conventional cells at the molecular level.



The same type of descriptive analysis was next performed on the exhaustion data set, GSE41867 (8). The data points were derived from acutely-resolved lymphocytic choriomeningitis (LCMV) infection (Arm) or chronic LCMV infection (Arm C113) groups at days 6, 8, 15, and 30 post-infection and were generated on the Affymetrix platform. In this study, gene-expression profiles of exhausted CD8+ T-cells from mice

infected with chronic LCMV (Arm C113) were compared with functional CD8+ cells, from mice infected with the Armstrong strain of the virus. The analysis revealed a comparable layout of sample distribution (**Figure 2B**), with groups' relative distances denoting a good concordance with the experimental outcomes. Dysfunctional clone-13 specific CD8+ T-cells (tetrahedron) (15 and 30 days, pink and black, respectively), appear to drift away at late time points and do not cluster around conventional CD8+ T-cells (spheres, pink, and black). In contrast, as was previously observed in the inflation model, the gene expression profiles of cells at the early timepoints [day 6 (pink) and 8 (brown) post-infection] from acutely resolving (spheres) and chronic LCMV (tetrahedron) are seen to cluster together.

Data set integration, when it is possible and successfully achieved, should allow more direct comparisons between samples generated in independent experiments. An alternative perspective of the differences or similarities between samples, without redesigning an entirely new experiment (which could become quite large, complex and very costly) could allow researchers to investigate new hypotheses and improve experimental design. In an attempt to improve the comparative analysis of the two models, the two data sets were merged (schematic in **Figure 1A**), following a pipeline aimed to reduce batch effects from multiple sources (see methods) as much as possible, which would confound accurate detection of gene expression signals (Schematic in **Figure 1B**).

Figure 3 shows a 3D PCA of the two merged data sets [from Bolinger et al. (7), (B) and Doering et al. (8) (D)], generated with different microarray platform (Illumina and Affymetrix), plotting together Inflating and Exhausting samples (in red) with the respective functional counterparts (in blue). Projections of the first three principal components capture most of the variability (>50%) of the total set of common expressed genes between the two platforms (~14,000). The results are also shown as a dendrogram in **Figure S1A**. In order to check the robustness of the results, 5 outliers were removed and the merged dataset reprocessed (without extra normalization after ComBat processing) and the most variable genes were selected, with genes filtered by variance (interquartile range (IQR) > 0.5, $n = 1,660$) (9). As shown in **Figures S1A–C**, the sample distribution is in good agreement with the original 3D PCA (**Figure 3**).

The appropriateness of the data integration is highlighted by the fact that naive samples [in green, spheres: Affymetrix, (8); tetrahedron: Illumina, (7)], which are the most comparable samples between the two data sets cluster in close proximity to each other. Interestingly, late timepoint Inflation and Exhausted samples (in red), other than having the tendency to cluster relatively close to each other, diverge at most from naive and non-Inflation and non-Exhausted samples (in blue), which instead occupy an intermediate position in the overall plot (**Figure 3**). In accordance with their extremely differentiated phenotype, Inflating samples are placed at the furthest distance from naive samples, immediately followed by the exhausted cells.

Batch effects may still play a role in this type of data analysis, and this was investigated by *pvca* (10). Prior to ComBat processing, “platform” factor (Affymetrix, Illumina) was assessed as the major source of batch effect; following batch effect removal

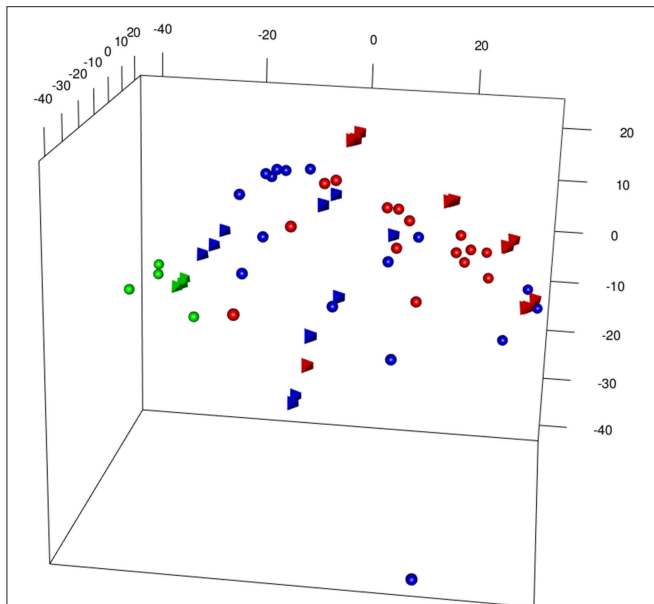


FIGURE 3 | 3DPCA of merged samples from Inflating and Exhausted models' data sets following batch effects removal between different microarray platforms (spheres: Affymetrix, Exhaustion study, tetrahedron: Illumina, Inflation study). The plot is showing overall distribution of Naive (green), Non-Inflating and Non-exhausting (blue), and Inflating and Exhausting (red) samples using the total common set of expressed genes (~14,000).

the “State” factor (Naive, Resolving, Inflating_Exhausting, as batch describing three simplified categories of samples: reference sample, intermediate phenotype, and extreme phenotype) represent the major source of variability (Figure S7). This initial, descriptive statistical approach suggests that immune responses demonstrating Inflation and Exhaustion share some common features at molecular level—despite the divergence that is observed at phenotypic level.

It is reasonable to postulate the existence of a common set of genes behind these immune responses; a pathway, in which the behavior and the intrinsic dynamics of its components could determine the fate of T-cells toward either Inflation or Exhaustion.

Weighted Gene Co-expression Network Analysis of Inflating Samples

In order to test the hypothesis of a shared set of genes behind opposite immune responses, yet highly related at the origin, we performed Weighted Gene Co-Expression Network Analysis [WGCNA, (11)] on the Inflation data set only (i.e., without any potential artifacts generated through merging), prior filtering the genes by variance ($n = 2,231$), with the intention of detecting a module of genes characteristic of the phenomenon.

The dendrogram in Figure 4A shows gene hierarchical clustering highlighting modules of genes with high interconnectivity based on TOM similarity (Topological Overlap Measure, a robust measure of network proximity). The second largest detected module (blue) was found enriched with immune relevant genes, after checking for the most statistically

represented GO category (*GOenrichmentAnalysis* in WGCNA package, Tables S1A,B). Reactome pathway enrichment analysis (12) confirmed the overrepresentation of immune relevant pathways within blue module (Figure S2A). Genes such as *Tbx21* and *Eomes*, known to have pivotal roles in T-cell differentiation, are contained in the blue module, along with other genes, such as *E2f2*, involved in cell proliferation and consistent with the Inflation phenotype.

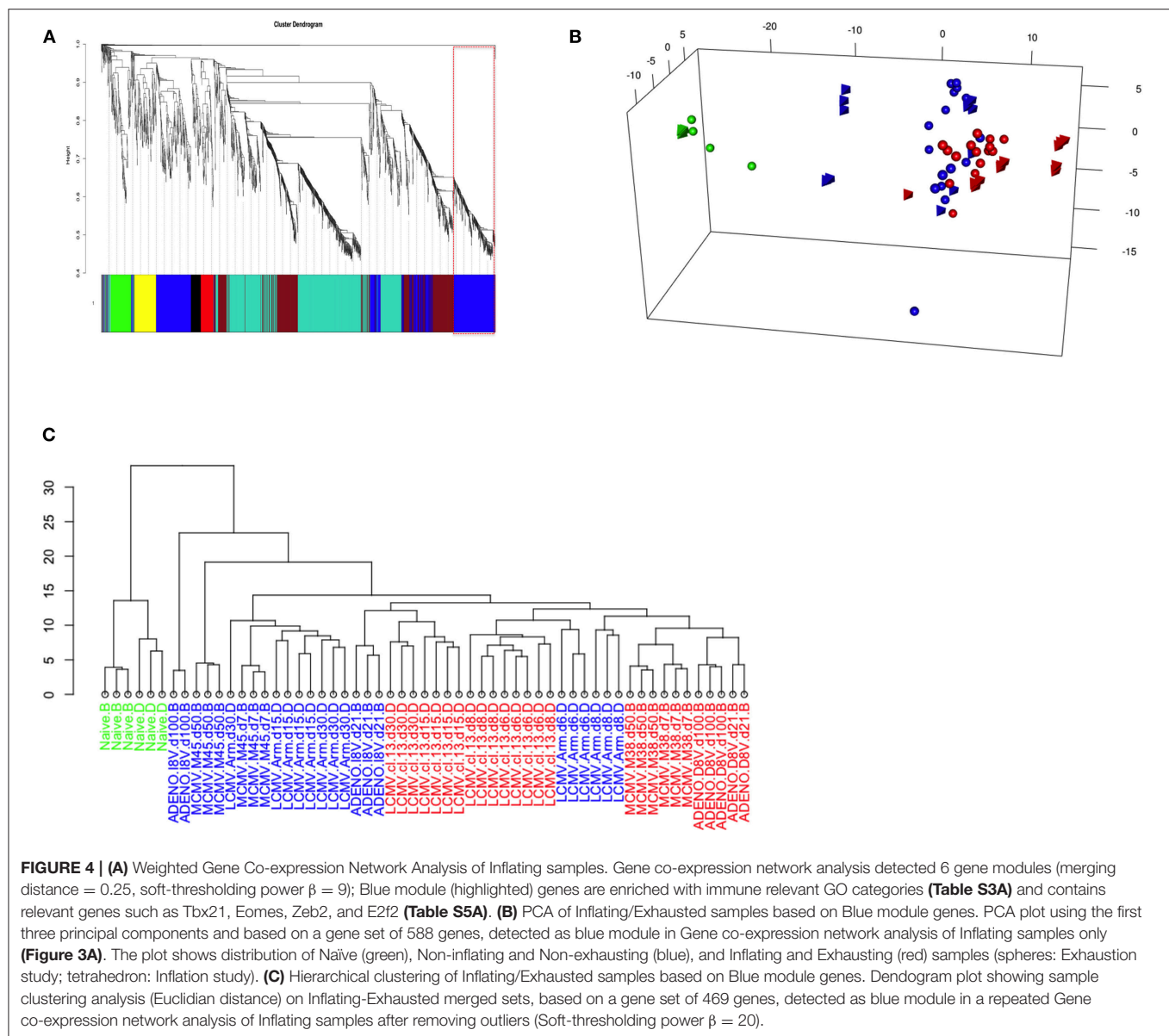
We therefore repeated PCA on the whole integrated data set, using the 588 genes assigned to the blue module, to test if sample distribution in respect to these genes was unaltered, improved or changed (Figure 4B); it was observed, in fact, that the sample layout was conserved: independent naive samples (in green, spheres, and tetrahedrons), that could function as calibrating samples, were clustered even closer, while inflating samples (red, tetrahedrons) were separated further compared to the PCA plot using the whole gene set (Figure 3). In general, the rest of Inflating and Exhausted samples appear to cluster tighter in the context of this subset of genes (blue module genes).

It is important to note that the same analysis performed using other gene modules, e.g. the turquoise module, can disrupt the PCA sample layout we observe using all genes (or blue module genes) and it generates a less informative PCA plot or hierarchical clustering dendrogram. The clustering analysis using genes from the Turquoise module (not enriched in immune relevant GO terms, Tables S1A,B or Reactome pathways, Figure S2B) produced a sample layout (Figure S3) that is noticeably different from the layout produced when employing blue module genes. Module detection, clustering analysis, GO terms and Reactome pathway enrichment (Figures S2A,B) were consistent after removing outliers and changing parameters to fit best a scale free topology (more appropriate soft-thresholding power, $\beta = 20$). The equivalent approach of hierarchical sample clustering analysis, employing again blue module genes, shows clearly that Inflating and Exhausted samples cluster together (in red), along with recently activated samples at acute phases from both data sets (Figure 4C).

Graphical Representation of Gene Networks Obtained in Inflating and Exhausted Data Sets

A separate WGCNA was executed on Exhaustion data set, matching the criteria utilized for the Inflation network inference, and similarly a module containing *Tbx21* was detected with a gene composition significantly enriched in immunological pathways. Indeed, there is a considerable gene overlap between the two modules that is even more evident within the context of transcription factor genes where the overlap reaches the 41% of the features (Figure 5A).

To have a further insight in the two related *Tbx21* modules a graphical representation of the nodes connectivity was produced. A circular topology was imposed in both gene modules to make the networks comparable and observe whether the gene hubs (nodes with the highest connectivity) were conserved or changed. Edges between genes reflect their direct correlations and

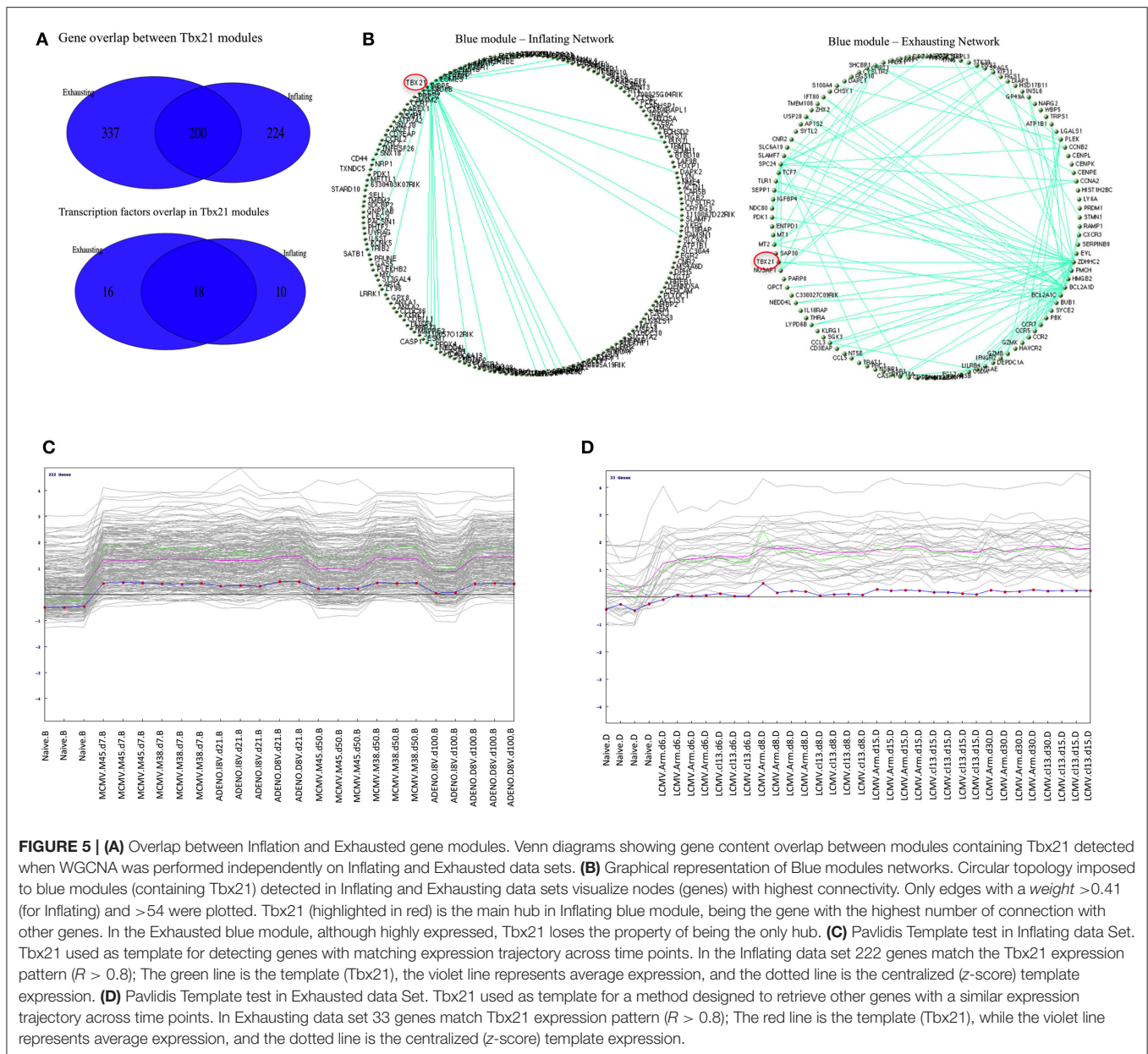


a similar correlation with “third party” genes, measured in the TOM matrix.

We discovered that in the module network generated using the Inflation data set, Tbx21 appeared to be a dominant hub, connecting with the vast majority of genes in the module, emphasized clearly when only edges with a *weight* > 0.41 were plotted (**Figure 5B**). Tbx21 has been previously described to be pivotal in controlling CD8 T cell activation (13) and this finding would extend the role of Tbx21 in maintaining functionality and effector status of the T cells in a subset of long-term memory populations.

In contrast, in the corresponding graphical representation of the analogous module in the *Exhaustion* data set, Tbx21 does not appear anymore to be the unique hub, losing its central role. Even lowering the cut off (*weight* < 0.54) of plotted edges

shows evidently that other genes are more likely to function as hubs (**Figure 5B**). The **Tables S2A,C** matrices represent the TOMs (topological overlay measures) used as input to visualize blue (Tbx21) and turquoise modules networks in Cytoscape and imposing a circular topology (**Figure 5B**). In order to corroborate this finding, we analyzed the data with an alternative approach, employing a method that searches for genes following a similar trajectory to Tbx21 across the time points of the two experiments (**Figures 5C,D**). We applied the Pavlidis template matching test (14, 15) using Tbx21 as template, equally for Inflating and Exhausted modules. Consistent with network analysis we observe that in Exhaustion data there are only 33 genes that match significantly the expression trajectory of Tbx21 (**Figure 5D**), while in *Inflating* data module over 200 genes display expression patterns strongly matching the one of Tbx21 (**Figure 5C**).



In order to further explore the hypothesis that Tbx21 is the main master regulator driving the inflation and directly interacting with the genes in its module, we analyzed ChIP-Seq data publicly available (16) where antibodies against Tbx21 were used to analyze transcription factor-DNA binding in Cytotoxic T Lymphocytes from mice infected with LCMV (P14 mice). A Tbx21 binding profile was obtained running the *macs2* peak caller algorithm (17) on wild type sample (SRR2075567, Tbx21 ChIP-seq on effector P14 CD8+ T cells) and using an input control (SRR2075584) ($q < 0.01$). Following gene annotation of peaks region [Homer bioinformatics tools, homer.ucsd.edu/homer, (18)], 383 out of 588 genes in Tbx21 module obtained with Inflating samples were found with binding sites for Tbx21 (in proximity, <1 kbp); 414 genes if we include

peaks within a distance of 2 kbp. A Fisher test was performed to demonstrate that Tbx21 module was significantly enriched with peaks: highly significant peaks (*macs2 peak score* > 50) mapping to blue module (63/588) were significantly more enriched ($p = 0.0003$) than turquoise module peaks (44/805).

We can formulate the hypothesis that in the case of an Inflationary response there is the presence of a “homogenous” genetic program, mainly coordinated by the transcription factor Tbx21. We postulated that a significant difference between Inflating and Exhausted populations could be more evident from differential correlations in expression between genes, rather than absolute changes in expression levels of the same genes between models. The strongest correlations that Tbx21 shows with its putative target genes in the Inflation data, highlight

its fully functional role as a master regulator, while in the Exhaustion data its task could be mitigated or disrupted by other coexisting transcriptional pathways. In order to elucidate the strongest level of correlation between Tbx21 and genes within its module, it could be informative to examine the expression levels across the time points of Tbx21 and one representative gene. E2f2 is a transcription factor involved in cell growth and proliferation. The relationship between Tbx21 and E2f2 was analyzed separately in the Exhaustion and Inflation datasets (**Figure S4**). In the Inflation time course, E2f2 shows an expression pattern that matches well to that of Tbx21, especially at later stages of infection (beyond the dashed line), while in the Exhaustion time series the pattern of expression is poorly correlated.

In support of this observation, inflating samples at the late timepoints are enriched with Reactome pathways involved in cell division and proliferation. The late timepoints of the inflation and exhausted samples within the integrated dataset were checked for enrichment of 674 curated Reactome gene sets [MSigDB Collections, (19) [c2.cp.reactome.v6.2.symbols.gmt](https://www.gsea-msigdb.org/gsea/msigdb/collections/c2.cp.reactome.v6.2.symbols.gmt)]. Direct comparison between inflated (M38, day 50) vs. exhausted cell populations (Arm Cl13 day 30) was analyzed by GSEA [Gene Set Enrichment Analysis, (20)] and it was noted that the top most enriched gene sets in Inflation are homogenous and belong to categories of cell division and DNA replication (**Table S3, Figure S5**). Conversely, the most significant Reactome pathways enriched in exhausted samples are quite heterogeneous (**Table S4, Figure S4B**). Repeating this analysis with inflating cells from the Adenovirus model (Day 50) vs. exhausted samples (Arm Cl13 day 30) yielded consistent results in spite of the presence of a clear outlier in the inflating samples (S11 D8V), which would have reduced the statistical power (data not shown). Taken together, these findings lend credence to the idea that a key difference between inflation and exhausted cells at the later timepoints is that regulation of the former's proliferative capacities is retained, akin to those of activated and highly proliferating effector and memory CD8 T cells after acute Arm infection (8).

Finally, we further explored the new potential insights gained from data integration using GSEA of Immunological signatures from MSigDB (4,872 gene sets, [c7.all.v6.2.symbols.gmt](https://www.gsea-msigdb.org/gsea/msigdb/all/c7.all.v6.2.symbols.gmt)). In the same comparison, between Inflation vs. Exhaustion at late timepoints (MCMV M38 Days 50 vs. LCMV Cl13 days 30); significantly enriched gene sets either in Inflation or Exhaustion populations (FDR < 0.25, based on gene sets permutations) from comparable studies were found to be biologically interpretable. Data and relevant examples consistent with experimental data (21) are shown in **Table S5** and **Figure S6**.

CONCLUSIONS

This analysis of previously published transcriptional data was aimed to address a simple but important question: what is the relationship between memory inflation and immune exhaustion? Both types of memory are distinct from conventional central memory development where populations contract following an

acute expansion. In the one case the cells remain functional (inflation) and in the other they lose function (exhaustion)—associated with distinct phenotypes. We hypothesized that dysregulation of a key module of genes might account for the phenotypic and functional differences seen between these T cell types.

Although it is possible to address this question using parallel analyses of existing data—and we have done this here—there is additional power in the merging of datasets, although great care must be taken to avoid artifacts due to batch and platform effects. Clearly a repeat set of experiments with mice treated with the different infections and vaccines in parallel, with a conserved pipeline for gene expression measurement and downstream data analysis would be ideal, but also impractical and costly given constraints on animal models and animal welfare in different settings (for example in the UK, LCMV Arm Cl13 remains a biosafety group 3 pathogen). We therefore think approaches of data integration—using a range of appropriate tools—from the increasing number of datasets publicly available will have important potential for future studies. Indeed this potential should only increase over time with convergence of sequencing approaches toward high-throughput RNA-Seq methods and accrual of datasets.

We observed two features of note from this analysis. Firstly, using a merged dataset, some transcriptional features of exhaustion and inflation appear to be shared, and cluster broadly closer to those derived from responses analyzed at an acute timepoint than to naive T cells or conventional memory. This is perhaps not a surprise since both inflationary and exhaustive memory are dependent on persistent/repetitive antigen stimulation. While the nature and/or the intensity of this antigen re-encounter may differ between the settings, in both cases, TCR triggering occurs and it is the response to this triggering which distinguishes the populations. While the populations clearly differ in expression of inhibitory receptors such as PD-1 and TIM-3, these represent only a relatively small set of genes within the total number up- and down-regulated in these populations compared to naive cells—and may overall contribute little to the PCA and hierarchical methods used to compare these subsets. This is not to say that such features are not critical and distinctive, but simply that other shared features may be worth exploration in future.

The second feature of interest relates to the relative connectivity of Tbx21 amongst genes within the same module—this analysis was initially performed using the primary non-merged datasets. All such genes in each setting (inflation or exhaustion) show a degree of correlation—however the role of Tbx21 as a central master transcription factor amongst these genes is fundamentally different in the inflationary pool. This fits with a well-demonstrated type 1 responsiveness driven by Tbx21 and associated to these populations, functional control of virus, and also with a differentiated phenotype which depends on functional interactions between Tbx21 and Zeb2 (16, 22). The lack of connectivity of Tbx21 in exhaustion has been previously described and is reproduced in this comparative analysis—however the cause of this has yet to be fully elucidated.

It is interesting to note that in the absence of Eomes, a transcription factor prominent in exhausted (8) but not in inflating T cells, memory CD8 T cells display a phenotype which strikingly resembles that of inflating memory cells, being IL-2 low and high in Granzyme B, Klrp-1, and Perforin (23). However, these memory cells were not able to proliferate properly in the recall response. This does not appear to be the case in inflating memory cells, and underscores the importance of the co-ordinated cell cycling module present in inflating cells which is not observed in exhausted cells. Furthermore, retention of their proliferative capacity would also explain their numerical superiority.

Infection with CMV drives not only memory inflation and conventional memory, but also is associated with the development of “peripheral memory” and tissue resident memory. Peripheral memory is linked to an intermediate expression of the chemokine receptor CX3CR1 (fractalkine receptor) and with the potential to proliferate *in vivo* as well as differentiate to both CX3CR1 high and low subsets, depending on the exposure to antigen (24, 25). Tissue resident cells have lost CX3CR1 expression like central memory cells, but have evolved tissue associated phenotypes such as expression of CD69 and CD103 (26). It is likely the combination of these multiple distinct memory types are responsible for long term virus control, especially the maintenance of inflationary pools being dependent on longer-lived central and peripheral memory cells. Further transcriptional analyses of peripheral and tissue-resident memory within the MCMV and adenovirus model vs. the LCMV model for example using such an integrated approach as shown here will be of value in defining their key distinguishing characteristics and how they are inter-related. Overall this will certainly contribute to our understanding of the host responses to chronic viral infection and hopefully to the development of novel vaccines.

In conclusion, we have tried to address an important immunological question by a deeper analysis of valuable experimentally derived datasets, and in doing so generate new ideas about the causes of exhaustion and the mechanisms involved in robust memory after vaccination. Such studies can be examined in parallel but the tools available to interrogate the data as one integrated group are attractive and this area will doubtless be explored further in future. The major issues include not only potential batch effects, but also fundamental differences in mouse strains used, housing and handling, all of which can lead to artifacts. For immunologic experiments, the presence of very well-defined/conserved cell populations (in this case naive and conventional memory CD8+ T cells) provides some important internal references and controls following dataset integration. Further mathematical tools should also be applied to assess the quality of the integration and avoid biases (e.g., the PVCA-R package). As gene expression studies develop and cross-referencing of RNA-Seq datasets becomes both more attractive and more powerful this issue of normalization and integration will become even more important, for example in initiatives such as the Human Cell Atlas (<https://humancellatlas.org>). Further, however well-integrated mathematically, such data ultimately demand experimental

validation—in this case a deeper analysis of the role of Tbx21 (and associated genes in the module) in inflationary memory would be valuable.

METHODS

PCA of expression profiles were performed in R (27) using *princomp/princomp* functions. Three-dimensional plots employing the first three principal components were generated using R package *rgl* (28), providing high-level functions for 3D interactive graphics.

The merging of data sets from different microarray platforms (Illumina, Affymetrix) is summarized in the following steps:

1. Independent arrays platform preprocessing and normalization of Affymetrix and Illumina data (*rma* function (*oligo*) and *lumiN(method="vsu")* function (*Lumi*), respectively).
2. Probes to gene collapse: in cases where multiple probes were designed to a gene, the unique best detected features for each gene symbol was selected (max signal probe) in Affymetrix and Illumina data (similarly to max probe collapse in GSEA algorithm, **Figure S2**).
3. Common gene intersection: a common set of collapsed symbols between Affymetrix and Illumina arrays were taken to build a merged data set. (14113 unique symbols x 62 array samples); in-house R code was written to perform probes collapse and unique gene symbol intersection between the two array platform (**Figure S2**).
4. Removal of platform batch affect: biases due by different arrays platform employed (Affymetrix and Illumina) using *ComBat* R function [*sva* package (29)] on the new merged data set using a common set of annotated gene symbols.
5. (Optional) Normalization of the new merged and batch effect corrected data set was performed to remove residual biases and make arrays more comparable [methods: “scale,” “quantile”(30), or “vsu”(31)].

Weighted Correlation Network Analysis [WGCNA, (32)] was executed using R package *WGCNA* (11) on a subset of highly variable genes (IQR > 0.5, 2231 features).

Hierarchical clustering analysis of samples using gene modules was performed using edited functions in *flashClust* R package (33).

Graphical visualization of gene module networks was generate on Cytoscape (34), with gene edges based on Topological Overlap Measures (TOM). Pavlidis matching template analysis (14) using Tbx21 gene as reference for expression pattern to match was performed on *MeV* java-based application (15).

Peak calling on Tbx21 binding profile data (16) was performed using *macs2* algorithm to identify genome-wide locations of transcription factor binding from ChIP-seq data (17). Tbx21 peaks were annotated using Homer bioinformatics tools [<http://homer.ucsd.edu/homer/ngs/annotation.html> (18)].

Effectiveness of batch value correction was performed using R package *pvca*, (10).

The versions of all the packages used is provided in the **Supplementary Materials and Methods**.

AUTHOR CONTRIBUTIONS

EM conceived the bioinformatic model, develop the idea, performed the bioinformatic analysis, and wrote the article. PK conceived the model, reviewed and wrote the article, and gave final approval. LNL contributed to the interpretation and design of the bioinformatics analysis and writing of the article. LNL and PK contributed equally.

ACKNOWLEDGMENTS

We are very grateful to the scientists who generated the original transcriptional datasets—particularly Stuart Sims and Bea Bolinger in Oxford and also John Wherry's team at the University of Pennsylvania, without whom this analysis would not have been possible. Thanks must also go to those who have collaborated on subsequent MCMV experiments including Julia Colston, Madeleine Zinser, Andy Highton, and Claire Gordon. This work was supported by the Wellcome Trust WT109965MA. LNL is supported by Cancer Research UK grant C30332/A23521.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00197/full#supplementary-material>

Figure S1 | (A) Clustering analysis of merged samples from Inflating and Exhausted models. Dendrogram showing samples clustering (equivalent to analysis in **Figure 3**) after removing outliers and filtering gene by expression variance (IQR > 0.5). **(B)** PCA analysis of merged samples from Inflating and Exhausted models. Samples PCA using first two principal components (equivalent to analysis in **Figure 3**) after removing outliers and filtering gene by expression variance (IQR > 0.5). **(C)** PCA analysis of merged samples from Inflating and Exhausted models. Samples PCA using first three principal components (equivalent to analysis in **Figure 3**) after removing outliers and filtering gene by expression variance (IQR > 0.5).

REFERENCES

1. Moskopidhis D, Lechner F, Pircher H, Zinkernagel RM. Virus persistence in acutely infected immunocompetent mice by exhaustion of antiviral cytotoxic effector T cells. *Nature*. (1993) 362:758–61. doi: 10.1038/362758a0
2. Mueller SN, Ahmed R. High antigen levels are the cause of T cell exhaustion during chronic viral infection. *Proc Natl Acad Sci USA*. (2009) 106:8623–8. doi: 10.1073/pnas.0809818106
3. Barber DL, Wherry EJ, Masopust D, Zhu B, Allison JP, Sharpe AH, et al. Restoring function in exhausted CD8 T cells during chronic viral infection. *Nature*. (2005) 439:682–7. doi: 10.1038/nature04444
4. Karrer U, Sierro S, Wagner M, Oxenius A, Hengel H, Koszinowski UH, et al. Memory inflation: continuous accumulation of antiviral CD8+ T cells over time. *J Immunol*. (2003) 170:2022–9. doi: 10.4049/jimmunol.170.4.2022
5. Sophie S, Robert R, Paul K. Evolution of diverse antiviral CD8+ T cell populations after murine cytomegalovirus infection. *Eur J Immunol*. (2005) 35:1113–23. doi: 10.1002/eji.200425534
6. Bolinger B, Sims S, O'Hara G, de Lara C, Tchilian E, Firner S, et al. A new model for CD8+ T cell memory inflation based upon a recombinant adenoviral vector. *J Immunol*. (2013) 190:4162–74. doi: 10.4049/jimmunol.120266
7. Bolinger B, Sims S, Swadling L, O'Hara G, de Lara C, Baban D, et al. Adenoviral vector vaccination induces a conserved program of CD8+ T cell memory differentiation in mouse and man. *Cell Rep*. (2015) 13:1578–88. doi: 10.1016/j.celrep.2015.10.034
8. Doering TA, Crawford A, Angelosanto JM, Paley MA, Ziegler CG, Wherry EJ. Network analysis reveals centrally connected genes and pathways involved in CD8+ T cell exhaustion versus memory. *Immunity*. (2012) 37:1130–44. doi: 10.1016/j.immuni.2012.08.021
9. Gentleman R, Carey V, Huber W, Irizarry R, Dudoit S. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor (Statistics for Biology and Health)*. Berlin: Springer-Verlag (2005).
10. Bushel P. *pvca: Principal Variance Component Analysis (PVCA)*. R package version 1.23.0 (2018).
11. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. (2008) 9:559. doi: 10.1186/1471-2105-9-559

Figure S2 | (A,B) Reactome pathways enrichment analysis. Most represented pathways in blue module **(A)** and in turquoise module **(B)**. Gene ratio is the proportion of pathways genes in the total number of module genes.

Figure S3 | Hierarchical clustering of Inflating/Exhausted samples based on turquoise module genes. Dendrogram plot showing sample clustering analysis (Euclidian distance) on Inflating-Exhausted merged sets, based on a gene set of 692 genes, detected as turquoise module in a repeated Gene co-expression network analysis of Inflating samples after removing outliers (Soft-thresholding power $\beta = 20$).

Figure S4 | Tbx21 and E2f2 Expression graphs. Normalized level of expressions of Tbx21 and E2f2 (in blue) across time points in Exhaustion and Inflation model experiments. Dash line marks the limit between early and late stage of infection.

Figure S5 | (A,B) GSEA enrichment plots of Reactome gene sets. First 9 top enriched gene sets pathways in Inflating samples **(A)** and in Exhausting samples **(B)** at late stages (IFNL: M38, 50 Days; EXHA: CI13, 30 Days).

Figure S6 | (A,B) Representative GSEA enrichment plots of Inflation vs. Exhaustion. Illustrative enrichments of a CD8 effector signature (exact source: GSE30962_1570_200_UP) in Inflation samples **(A)** and a CD8 memory signature (exact source: GSE1000002_1582_200_DN) in Exhaustion samples **(B)**.

Figure S7 | Assessment of batch effect contribution prior and after ComBat processing. An analysis by pvca R package was performed to estimate the variability of experimental effects.

Table S1 | (A,B) GO term enrichment analysis of WGCNA modules. GO categories enrichment analysis in modules detected including all Inflating samples **(A, $\beta = 9$)** and excluding possible outliers **(B, $\beta = 20$)**.

Table S2 | (A–C) TOM matrices of blue and turquoise modules. Table matrices with topological overlay measures (TOM) of blue modules detected including all samples **(A)** or excluding possible outliers **(B)**; TOM matrix of turquoise module including all inflating samples **(C)**.

Table S3 | GSEA report of Reactome gene sets enriched in Inflation. GSEA report of Reactome curated pathways found enriched (FDR < 0.25) in Inflating samples (M38, days 50) vs. Exhausting samples (CI13, days 30).

Table S4 | GSEA report of Reactome gene sets enriched in Exhaustion. GSEA report of Reactome curated pathways found enriched (FDR < 0.25) in Exhausting samples (CI13, days 30) vs. Inflating samples (M38, days 50).

Table S5 | (A,B) GSEA reports of MsigDB Immunological Signatures gene sets in Inflation and Exhaustion. GSEA reports of immunological signatures found enriched (FDR < 0.25) in Inflating samples **(A)** (M38, days 50) vs. Exhausting samples **(B)** (CI13, days 30).

Supplementary Materials and Methods | Relevant R and Bioconductor packages and versions used.

12. Yu G, He Q-Y. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol BioSyst.* (2016) 12:477–9. doi: 10.1039/C5MB00663E
13. Knox JJ, Cosma GL, Betts MR, McLane LM. Characterization of T-Bet and eomes in peripheral human immune cells. *Front Immunol.* (2014) 5:217. doi: 10.3389/fimmu.2014.00217
14. Pavlidis P, Noble WS. Analysis of strain and regional variation in gene expression in mouse brain. *Genome Biol.* (2001) 2:research0042.1. doi: 10.1186/gb-2001-2-10-research0042
15. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, et al. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques.* (2003) 34:374–8. doi: 10.2144/03342mt01
16. Dominguez CX, Amezcua RA, Guan T, Marshall HD, Joshi NS, Kleinstein SH, et al. The transcription factors ZEB2 and T-bet cooperate to program cytotoxic T cell terminal differentiation in response to LCMV viral infection. *J Exp Med.* (2015) 1:2041–56. doi: 10.1084/jem.20150186
17. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc.* (2012) 7:1728–40. doi: 10.1038/nprot.2012.101
18. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* (2010) 38:576–9. doi: 10.1016/j.molcel.2010.05.004
19. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. *Cell Syst.* (2015) 1:417–25. doi: 10.1016/j.cels.2015.12.004
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* (2005) 102:15545–50. doi: 10.1073/pnas.0506580102
21. West EE, Youngblood B, Tan WG, Jin H-T, Araki K, Alexe G, et al. Tight regulation of memory CD8+ T cells limits their effectiveness during sustained high viral load. *Immunity.* (2011) 35:285–98. doi: 10.1016/j.immuni.2011.05.017
22. Omilusik KD, Best JA, Yu B, Goossens S, Weidemann A, Nguyen JV, et al. Transcriptional repressor ZEB2 promotes terminal differentiation of CD8+ effector and memory T cell populations during infection. *J Exp Med.* (2015) 212:2027–39. doi: 10.1084/jem.20150194
23. Zhou X, Yu S, Zhao D-M, Harty JT, Badovinac VP, Xue H-H. Differentiation and persistence of memory CD8+ T cells depend on T cell factor 1. *Immunity.* (2010) 33:229–40. doi: 10.1016/j.immuni.2010.08.002
24. Gordon CL, Lee LN, Swadlow L, Hutchings C, Zinser M, Highton AJ, et al. Induction and maintenance of CX3CR1-intermediate peripheral memory CD8+ T cells by persistent viruses and vaccines. *Cell Rep.* (2018) 23:768–82. doi: 10.1016/j.celrep.2018.03.074
25. Gerlach C, Moseman EA, Loughhead SM, Alvarez D, Zwijsenburg AJ, Waanders L, et al. The chemokine receptor CX3CR1 defines three antigen-experienced CD8 T cell subsets with distinct roles in immune surveillance and homeostasis. *Immunity.* (2016) 45:1270–84. doi: 10.1016/j.immuni.2016.10.018
26. Kumar BV, Ma W, Miron M, Granot T, Guyer RS, Carpenter DJ, et al. Human tissue-resident memory T cells are defined by core transcriptional and functional signatures in lymphoid and mucosal sites. *Cell Rep.* (2017) 20:2921–34. doi: 10.1016/j.celrep.2017.08.078
27. R Core Team. *R: A Language and Environment for Statistical Computing* (2016). Available online at: <https://www.r-project.org/>
28. Adler D, Murdoch D. *rgl: 3D Visualization Using OpenGL* (2016). Available online at: <https://cran.r-project.org/package=rgl>
29. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics.* (2007) 8:118–27. doi: 10.1093/biostatistics/kxj037
30. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007
31. Huber W, von Heydebreck A. Many comments, suggestions by users are acknowledged, among them Dennis Kostka, Kreil D, Klein H-U, Gentleman R, Sarkar D, Smyth G. vsn: *Variance Stabilization and Calibration for Microarray Data* (2016).
32. Langfelder P, Chaochai Cai SH, Dong J, Miller J, Song L, Yip A, et al. *WGCNA: Weighted Correlation Network Analysis* (2016). Available online at: <https://cran.r-project.org/package=WGCNA>
33. Murtagh F, R development team, Langfelder P. *flashClust: Implementation of Optimal Hierarchical Clustering* (2012) Available online at: <https://cran.r-project.org/package=flashClust>
34. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past co-authorship with the authors EM and PK.

Copyright © 2019 Marchi, Lee and Klenerman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



RNA Structure—A Neglected Puppet Master for the Evolution of Virus and Host Immunity

Redmond P. Smyth^{1,2†}, Matteo Negroni^{3*†}, Andrew M. Lever^{4,5}, Johnson Mak⁶ and Julia C. Kenyon^{4,7,8*}

¹ Helmholtz Institute for RNA-based Infection Research, Würzburg, Germany, ² Faculty of Medicine, University of Würzburg, Würzburg, Germany, ³ Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR9002, F-67000, Strasbourg, France, ⁴ Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge, United Kingdom, ⁵ Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, ⁶ Institute for Glycomics, Griffith University, Gold Coast, QLD, Australia, ⁷ Department of Microbiology and Immunology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore, ⁸ Homerton College, Cambridge, United Kingdom

OPEN ACCESS

Edited by:

Gkikas Magiorkinis,
National and Kapodistrian University
of Athens, Greece

Reviewed by:

Subrata H. Mishra,
Johns Hopkins University,
United States
Jianzhong Zhu,
Yangzhou University, China

*Correspondence:

Matteo Negroni
m.negroni@ibmc-cnrs.unistra.fr
Julia C. Kenyon
jck33@cam.ac.uk

[†]These authors have contributed
equally to this work and are co-first
authors

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 10 June 2018

Accepted: 24 August 2018

Published: 19 September 2018

Citation:

Smyth RP, Negroni M, Lever AM,
Mak J and Kenyon JC (2018) RNA
Structure—A Neglected Puppet
Master for the Evolution of Virus and
Host Immunity.
Front. Immunol. 9:2097.
doi: 10.3389/fimmu.2018.02097

The central dogma of molecular biology describes the flow of genetic information from DNA to protein via an RNA intermediate. For many years, RNA has been considered simply as a messenger relaying information between DNA and proteins. Recent advances in next generation sequencing technology, bioinformatics, and non-coding RNA biology have highlighted the many important roles of RNA in virtually every biological process. Our understanding of RNA biology has been further enriched by a number of significant advances in probing RNA structures. It is now appreciated that many cellular and viral biological processes are highly dependent on specific RNA structures and/or sequences, and such reliance will undoubtedly impact on the evolution of both hosts and viruses. As a contribution to this special issue on host immunity and virus evolution, it is timely to consider how RNA sequences and structures could directly influence the co-evolution between hosts and viruses. In this manuscript, we begin by stating some of the basic principles of RNA structures, followed by describing some of the critical RNA structures in both viruses and hosts. More importantly, we highlight a number of available new tools to predict and to evaluate novel RNA structures, pointing out some of the limitations readers should be aware of in their own analyses.

Keywords: RNA structure, viral evolution, secondary structure, immune evasion, viral RNA

INTRODUCTION

Mutation rates of viral genomes are extremely high when compared with those of eukaryotic cells; RNA virus polymerases typically possess error rates of 10^{-4} to 10^{-6} per base (1). Such rapid mutation is a strategy by which they can evade host adaptive immune responses (2). Antiviral defenses of the innate immune system, which is less genetically flexible than the adaptive response, enable a very broad range of recognition from which it is difficult for viruses to escape, even given their high error rate. To counter the innate immune system, viruses have developed strategies to block its activation. This arms race prompts the immune system to develop counter measures to recognize and eliminate the virus, whilst viruses that survive and transmit successfully are those that have evolved to escape it. When considering the evolutionary pressures on both virus and host, research has often focused on the protein sequences needed by each. A recent explosion in

our understanding of the functions of RNA, however, leads us to consider instead the role of RNA itself in driving the evolution of viruses and of human immunity.

RNA is a truly multifunctional molecule. It directs ribosomes (themselves RNA based enzymes) to produce proteins, but also regulates cellular activity by interacting directly with proteins or nucleic acids and by catalyzing biochemical reactions (ribozymes). Indeed, RNA is now implicated in almost every cellular process, including immune defense. It is also recognized as being key to viral infection processes. RNA multifunctionality comes from its ability to fold into complex three-dimensional structures that can often switch conformation to effect different functions such as binding other RNA molecules or proteins. The initial fold of an RNA molecule depends primarily on its sequence and is established by Watson-Crick pairing of complementary bases into stem-loop structures that then orientate themselves relative to one another [for a general review of RNA structure see (3)]. This three-dimensional positioning can be stabilized by non-canonical interactions or structures such as pseudoknots, which occur where the nucleotides of a loop region base pair intramolecularly with complementary nucleotides. An example is shown in **Figure 1A**, within the complex structural element known as an IRES (internal ribosome entry site). As for proteins, single nucleotide mutations can alter the three-dimensional structure of the RNA, with corresponding deleterious or positive effects on its function; RNA structures are hence substrates for, and drivers of, viral evolution. For example, random mutation may confer a new beneficial function on a given structure that is then selectively favored by evolution.

Viruses are known to be extremely thrifty with their genomes to maximize replication speed, using strategies such as overlapping reading frames and polycistronic mRNAs. Similarly, they often contain functional RNA structures within both coding and non-coding regions. For example, the first 500 nucleotides of the HIV-1 genome is densely packed with structured domains that control key steps of the replication cycle including transcription, translation, export, packaging, and reverse transcription (7–9) and structural switches that aid their regulation (5, 10–12). Functionality encoded within an RNA structure is often a requisite for successful initiation or completion of viral replication. Many viruses, particularly those of the *picornaviridae*, initiate translation from IRES elements (13). RNA structures facilitating frameshifting enable viruses to encode multiple proteins from a single RNA (14). Additionally, viral RNA structures may directly or indirectly impact cellular immunity. In HIV-1, the transactivation response element (TAR, **Figure 1B**) regulates transcription of the genomic RNA and gene expression (15) thereby playing a central role in determining the level of virus detected by the immune system; at its most extreme leading to complete evasion of the immune response through latency (16). Frameshift structures or splicing regulators qualitatively and quantitatively manage the amounts of proteins produced by viruses, and hence those that are seen by the immune system (17). For example, some strains of Influenza A virus encode a second open reading frame (ORF) in segment three which is accessed by a low frequency

+1 ribosome frameshifting event (18). This ORF produces PA-X, a protein which modulates inflammatory, apoptotic, and T-lymphocyte signaling pathways (18). Viruses with larger genomes can even produce their own microRNAs (19) or long non-coding RNAs (lncRNAs) that control cellular functions (20, 21). Dengue virus expresses a subgenomic RNA that has been shown to inhibit interferon expression by binding to TRIM25 (22). This subgenomic RNA is produced when a cellular 5'-to-3' exoribonuclease stalls at a stable pseudoknot RNA structure in the 3' UTR; small substitutions within this structure modulate viral fitness and pathogenicity through their effects on the immune system. The dengue 3'UTR also folds differently in humans and insects (23) leading to production of different immunomodulatory non-coding RNAs in each host type (22, 23). This is one mechanism by which the same viral genome can both effectively replicate in human and insect cells and counteract these two divergent immune systems.

The immune system senses viral RNA using different mechanisms, including the recognition of viral single-stranded RNA by Toll-like receptors (24). The importance of RNA structures in viral replication is so fundamental that they are directly recognized by the innate immune response. Not only has the innate immune system evolved to recognize double-stranded RNA (dsRNA) within viral genomes or genomic replication intermediates, often via MDA5, it has also evolved to recognize the double-stranded parts of conserved RNA elements like IRESs (13) or 3' stem-loops, via RIG-I [reviewed in (25)]. The importance of this in maintaining broad antiviral defense is reflected in the fact that such RNA structures are also formed by DNA viruses in their protein-coding RNAs. The cellular double-stranded RNA recognition system not only leads to production of interferon, but was recently shown to upregulate NKG2D ligand, thus alerting NK cells to the presence of virus and enabling destruction of the infected cell (26). Viruses have evolved strategies to mask recognition of their RNAs, such as 5' cap-snatching (27), but their need to maintain certain critical RNA structures means that they struggle to avoid recognition entirely.

VIRAL EVOLUTION TO EVADE AN IMMUNE RESPONSE MAY BE CONSTRAINED OR FACILITATED BY RNA STRUCTURE

The presence of essential conserved viral RNA structures can constrain the ability of viruses to evolve and to evade the immune response. Some RNA viruses have optimized their genome structure and organization to facilitate viral evolution during co-infection of the same cell with different viral strains. One widespread strategy is genome segmentation leading to reassortment (seen in rotaviruses and influenza viruses) (28). Another common strategy is template switching during replication leading to recombination and the formation of genome chimeras (seen in retroviruses) (29). Reassortment and recombination are non-random processes that are known to depend on RNA sequence and structure, but the underlying mechanisms are often poorly understood. Both processes may be

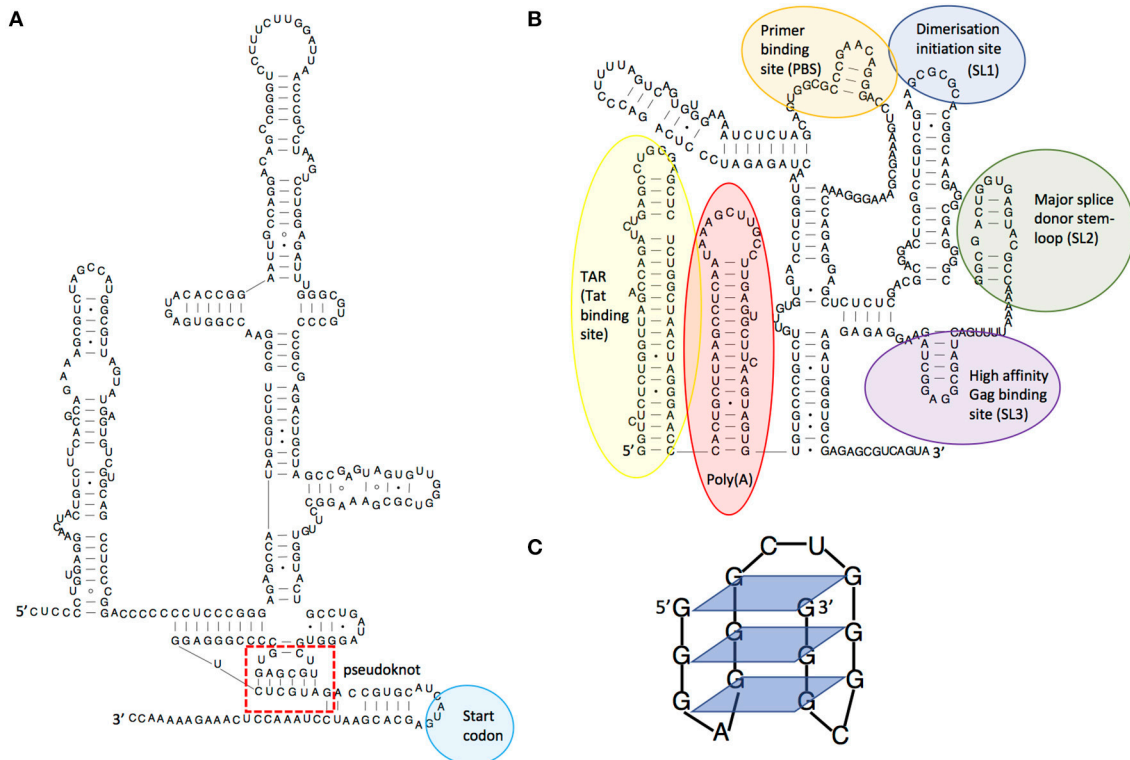


FIGURE 1 | Examples of RNA structures. **(A)** Structural diagram of the hepatitis C virus IRES structure, showing its pseudoknot. Adapted from Malygin et al. (4), **(B)** Structural model of the HIV-1 5' leader RNA, highlighting some of the important RNA structures it contains. Adapted from Kenyon et al. (5), **(C)** An RNA guanine quadruplex structure, showing the 4-way bonds between guanines in each plane. Adapted from Frees et al. (6).

facilitated by inter-molecular interactions (30, 31), and evolution can be both promoted or inhibited by intra-molecular RNA structure.

In HIV-1, each virion contains two copies of the RNA genome, which in the case of co-infection of the same cell can originate from two different proviruses. These are non-covalently joined as a dimer via stem-loop 1 (SL1, **Figure 1B**). During viral replication, reverse transcriptase (RT) switches template, thereby adding template switching generated errors to its inherently low fidelity, producing the genetic diversity that allows HIV-1 to rapidly escape the immune system and antiretroviral therapy (2). Sequence incompatibility preventing formation of heterodimers of genomic RNA at SL1 has been shown to be a major restriction to inter-subtype recombination, with only a 3-nt sequence difference being sufficient to disrupt this (32). When considering distantly-related strains of HIV with compatible SL1 sequences, the main factor governing recombination is the degree of local sequence similarity (33, 34). However, in more closely-related viral sequences, RNA structures strongly influence recombination locally (35–37). This has been shown to be the case for well-defined RNA structures within *env* such as the C2 hairpin and the Rev responsive element (RRE) (38). It has been suggested that this evolved to favor the stepwise folding of proteins during translation, but it has also been shown to favor the occurrence of recombination in these same regions (37).

As a consequence, recombination may shuffle whole domains of proteins thus generating structural variants that escape immune recognition, particularly for quaternary epitopes generated by the juxtaposition of different protein domains. RNA structures and sequences have also been shown to influence the fidelity of reverse transcription such that stable secondary structures enhance the number and type of mutations incorporated (39, 40). The regions of *env* encoding the external parts of the viral surface glycoprotein gp120 are under strong positive selection by the humoral immune system. Perhaps counterintuitively they present a lower degree of RNA structure; however, this can be accounted for by such rapid viral mutation that the RNA is unable to conserve base-pairing. As studies have shown, poorly structured RNA regions are reverse transcribed with higher fidelity, which paradoxically limits the rate of introduction of mutations in these highly variable regions of the genome (39).

RNA viruses with segmented genomes can undergo reassortment, leading to the exchange of entire gene segments, potentially giving rise to new viral strains to which humans have no previous immunity. In influenza A, reassortment has been historically associated with the emergence of pandemic strains, including the most recent H1N1 2009 pandemic which contained influenza gene segments from human, avian and swine lineages (41, 42). It is thought that packaging sequences within each gene segment direct their selective incorporation into virus particles

(43, 44) and recent work suggests a mechanism where packaging signals mediate RNA-RNA interactions that would guide their incorporation. It therefore follows that any preferential interactions or incompatibilities between vRNP segments would then regulate genetic reassortment and influenza evolution (45). It is possible that improved understanding of this process would help to better predict the emergence of pandemic influenza.

In addition to genome diversification through recombination and reassortment, RNA structures influence immune escape by modulating the viral proteome. The use of frameshifting or alternative start codons is often controlled by viral RNA structures, resulting in the translation of viral peptides or proteins in a different reading frame. The resulting peptides are often antigenic and may act to dilute the presentation of conventional viral peptides on the surface of the infected cell (46). In some viruses there is an evolutionary constraint to maintain translation-impeding RNA structures in genes that encode good T cell epitopes, thus maintaining their translation at levels too low to trigger T cell recognition and killing. The EBNA-1 RNA from Epstein-Barr virus for example regulates its own translation *in cis* (47) and the evolutionary pressure on this gene comes from the need to maintain G-quadruplex RNA structures (Figure 1C) that act as “steric blocks” to ribosomes. When these structures are destabilized, cells infected with the resulting mutant virus are more readily seen by T cells than those infected with wild-type virus (48).

Viruses also appear to be under pressure to maintain unstructured regions in their genome: there is a bias in HIV-1 toward the use of A's in the retroviral genome; this biases codon usage and ultimately even the amino acid composition of the viral proteins (49). Adenosines are vastly overrepresented in the single-stranded regions of RNA and underrepresented in double-stranded regions; their only binding partner, U, can also pair with G, which may explain the single-stranded nature of A-rich regions. Artificial introduction of extensive synonymous A to G mutations in *pol* led to increased stability of the dimeric genome inside the virion, and reduced reverse transcription as a result (50). The signature distribution of Adenosine frequency and its relation to local RNA structure was thought to be maintained by the influence of RNA secondary structures on reverse transcription. Changing the A ratio in local areas by only including codons found in natural isolates of HIV-1, did not affect replication efficiency *in vitro* (51) however it is possible that *in vivo* viruses use this strategy of maintaining parts of their genome as single-stranded in order to avoid innate immune recognition by RIG-I or MDA5. Many viruses also need to maintain a low number of CpG dinucleotides in their genome in order to avoid recognition by the zinc-finger antiviral protein (ZAP) (52). When HIV-1 codon usage was humanized, affecting the native RNA structure, a reduced IFN- α/β response was observed (53), suggesting that the maintenance of specific structures in the viral genome comes at the cost of greater recognition by the innate immune response. Despite this, the viral genome apparently undergoes positive selection for the maintenance of many specific RNA structures: when synonymous mutations were introduced extensively into the viral genome a decrease

in infectivity was observed that could be attributed to an expected alteration of splicing pattern and/or modification of RNA structures (50, 54).

THE ROLES OF VIRAL AND CELLULAR RNA STRUCTURE IN THE EVOLUTION OF HUMAN IMMUNE RESPONSES AND THE HUMAN IMMUNE SYSTEM

In terms of the evolution of the human genome, the degree of variation at the MHC class I locus is positively correlated with local pathogen richness, for which viruses are postulated to play an important role. This is particularly evident for HLA B (55). The importance of RNA structure in influencing the generation of T cell epitopes, either through translational enhancing, blocking, or frameshifting mechanisms, means that RNA structures within viruses must have influenced the evolution of the HLA locus. For example, macaques that have the correct MHC-I allele to present an antigenic cryptic peptide derived from the *env* ORF are better able to control simian immunodeficiency virus (SIV) infection (56) which would be expected to be a driver for the maintenance of this MHC allele within the population. As mentioned above, the mechanisms controlling the generation of cryptic translational products are often RNA structure-dependent.

Viral RNA structures have also influenced the evolution of the innate immune system; as previously discussed, hallmarks of viral RNAs are targeted by conserved RNA-binding proteins such as RIG-I. More specific antiviral proteins have also evolved, however. APOBEC3 proteins target retroviral genomes and are incorporated into viral particles (57). These host cell-derived cytidine deaminases bind to the viral RNA and mutate it during the reverse transcription process, leading to non-functional virus. It has been reported that regions of the genome under strong purifying selection present an underrepresentation of APOBEC3 target sequences, a signature of a strong pressure for limiting the occurrence of mutations in certain regions of the genome (39). Retroviruses have also developed direct strategies to counteract these proteins, often by encoding proteins that bind to them directly. Interestingly, APOBEC3s have recently been shown to bind to the same motifs in the viral RNA genome as the viral NC structural protein involved in genome packaging. This suggests a competitive relationship may have developed between host cell and viral factors, for binding to the same viral RNA structures (58).

NOVEL WAYS TO EXPLORE RNA STRUCTURE AND FUNCTION

RNA functionality is best understood through its structure, but RNA structure determination is extremely challenging. Although it is formed based on simple base pairing rules, for RNA molecules of biologically relevant sizes there are an astronomical number of possible structural permutations, meaning that RNA structure cannot be predicted easily from base pairing rules alone. Biophysical methods such as crystallography (59) and NMR (60)

are each able to determine RNA structure at atomic resolution but both have difficulty with large RNA substrates. This is evidenced by the paucity of atomic resolution RNA structures, compared to their protein equivalents. For example, the RCSB databank holds structural data for over 100,000 proteins, but contains only around 1000 RNA structures. This difficulty arises because RNA molecules tend to adopt long flexible shapes with weak tertiary interactions that are prone to misfolding (61). Furthermore, the negatively charged phosphate groups on the surface of an RNA molecule can impose technical challenges as they hinder crystal packing. Newer techniques are emerging to address this gap including small-angle X-ray scattering (62), single molecule FRET (63, 64), and atomic force microscopy (65).

RNA secondary structure is currently most commonly resolved using a combination of (i) phylogenetic approaches (ii), structure prediction algorithms, and (iii) experimental methods with chemical/enzymatic probes. Whilst these methodologies cannot determine RNA structure at atomic resolution, they are nevertheless able to generate models that provide useful biological insights (8, 66, 67). Indeed, RNA structure determination is currently undergoing a revolution thanks to advances in next generation sequencing technology that have transformed traditional biochemical assays into powerful tools that can characterize thousands of RNA structures in single experiments (68–71). The most widely used methods take advantage of chemical probes, such as dimethyl sulfate (DMS) and selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE) reagents, that differentially react with single stranded vs double stranded RNA. Knowledge of whether a nucleotide is likely to be base paired or not can significantly improve the accuracy of RNA structure predictions from thermodynamic folding algorithms when included as an energetic consideration in the modeling programme, known as a pseudo free energy parameter (72). For example, if chemical probes show a nucleotide is single-stranded, the modeling algorithm adds an energetic penalty to structures that include

it in a double-stranded region. The programme then displays the most energetically favorable structures, that fit all of the data best. Several chemical probes can penetrate cells and virions, which is important for understanding RNA function *in vivo*, such as the binding sites of regulatory proteins (9, 73). Further characterization of RNA structure-function relationships can be obtained using specialized approaches, such as mutational interference mapping experiment (MIME) (74) and cross-linking SHAPE (XL-SHAPE), where protein binding sites are mapped using UV cross-linking in parallel with SHAPE probing (75), and by CLIP (crosslinking-immunoprecipitation sequencing) related methodologies (76). More recently, RNA proximity ligation has emerged as a new class of RNA structural probing technique for the direct detection of long-range base pairing or inter-molecular interactions (77–84). These types of interactions are commonly found in viral genomes/regulatory RNAs and are difficult to identify with other methodologies. As the immune system is known to be regulated by non-coding RNAs (85), the ability to detect direct interactions between viral and cellular RNAs will be particularly important for future understanding of virus-host interactions.

AUTHOR CONTRIBUTIONS

JM conceived the review. RS, MN, AL, JM, and JK wrote the manuscript.

FUNDING

This work was supported by the Helmholtz Association (VH-NG-1347 to RS), Sidaction (AI25-1-02335 to MN), The Biomedical Research Centre UK and Clinical Academic Reserve UK (to AL), and the Medical Research Council UK (MR/N022939/1 to AL and JK). JM is recipient of funding from Australian National Health and Medical Research Council project grant A1121697.

REFERENCES

- Sanjuan R, Domingo-Calap P. Mechanisms of viral mutation. *Cell Mol Life Sci.* (2016) 73:4433–48. doi: 10.1007/s00018-016-2299-6
- Smyth RP, Davenport MP, Mak J. The origin of genetic diversity in HIV-1. *Virus Res.* (2012) 169:415–29. doi: 10.1016/j.virusres.2012.06.015
- Batey RT, Rambo RP, Doudna JA. Tertiary motifs in RNA structure and folding. *Angew Chem Int Ed Engl.* (1999) 38:2326–43.
- Malygin AA, Kossinova OA, Shatsky IN, Karpova GG. HCV IRES interacts with the 18S rRNA to activate the 40S ribosome for subsequent steps of translation initiation. *Nucleic Acids Res.* (2013) 41:8706–14. doi: 10.1093/nar/gkt632
- Kenyon JC, Prestwood LJ, Le Grice SF, Lever AM. In-gel probing of individual RNA conformers within a mixed population reveals a dimerization structural switch in the HIV-1 leader. *Nucleic Acids Res.* (2013) 41:e174. doi: 10.1093/nar/gkt690
- Frees S, Menendez C, Crum M, Bagga PS. QGRS-Conserved: a computational method for discovering evolutionarily conserved G-quadruplex motifs. *Hum Genomics* (2014) 8:8. doi: 10.1186/1479-7364-8-8
- Paillart JC, Dettenhofer M, Yu XF, Ehresmann C, Ehresmann B, Marquet R. First snapshots of the HIV-1 RNA structure in infected cells and in virions. *J Biol Chem.* (2004) 279:48397–403. doi: 10.1074/jbc.M408294200
- Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* (2009) 460:711–6. doi: 10.1038/nature08237
- Wilkinson KA, Gorelick RJ, Vasa SM, Guex N, Rein A, Mathews DH, et al. High-throughput SHAPE analysis reveals structures in HIV-1 genomic RNA strongly conserved across distinct biological states. *PLoS Biol.* (2008) 6:e96. doi: 10.1371/journal.pbio.0060096
- Keane SC, Van V, Frank HM, Sciandra CA, McCowen S, Santos J, et al. NMR detection of intermolecular interaction sites in the dimeric 5'-leader of the HIV-1 genome. *Proc Natl Acad Sci USA.* (2016) 113:13033–8. doi: 10.1073/pnas.1614785113
- Mailler E, Bernacchi S, Marquet R, Paillart JC, Vivet-Boudou V, Smyth RP. The Life-cycle of the HIV-1 gag-RNA complex. *Viruses* (2016) 8:E248. doi: 10.3390/v8090248
- Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, et al. NMR detection of structures in the HIV-1 5'-leader RNA that regulate genome packaging. *Science* (2011) 334:242–5. doi: 10.1126/science.1210460
- Fernandez N, Buddrus L, Pineiro D, Martinez-Salas E. Evolutionary conserved motifs constrain the RNA structure organization of picornavirus

- IRES. *FEBS Lett.* (2013) 587:1353–8. doi: 10.1016/j.febslet.2013.03.005
14. Houck-Loomis B, Durney MA, Salguero C, Shankar N, Nagle JM, Goff SP, et al. An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* (2011) 480:561–4. doi: 10.1038/nature10657
 15. Kao SY, Calman AF, Luciw PA, Peterlin BM. Anti-termination of transcription within the long terminal repeat of HIV-1 by tat gene product *Nature* (1987) 330:489–93.
 16. Ruelas DS, and Greene WC. An integrated overview of HIV-1 latency. *Cell* (2013) 155:519–29. doi: 10.1016/j.cell.2013.09.044
 17. Naphthine S, Ling R, Finch LK, Jones JD, Bell S, Brierley I, et al. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat Commun.* (2017) 8:15582. doi: 10.1038/ncomms15582
 18. Hayashi T, MacDonald LA, Takimoto T. Influenza A virus protein PA-X contributes to viral growth and suppression of the host antiviral and immune responses. *J Virol.* (2015) 89:6442–52. doi: 10.1128/JVI.00319-15
 19. Pfeffer S, Zavolan M, Grasser FA, Chien M, Russo JJ, Ju J, et al. Identification of virus-encoded microRNAs. *Science* (2004) 304:734–6. doi: 10.1126/science.1096781
 20. Reeves MB, Davies AA, McSharry BP, Wilkinson GW, Sinclair JH. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* (2007) 316:1345–8. doi: 10.1126/science.1142984
 21. Zhang X, Ma X, Jing S, Zhang H, Zhang Y. Non-coding RNAs and retroviruses. *Retrovirology* (2018) 15:20. doi: 10.1186/s12977-018-0403-8
 22. Manokaran G, Finol E, Wang C, Gunaratne J, Bahl J, Ong EZ, et al. Dengue subgenomic RNA binds TRIM25 to inhibit interferon expression for epidemiological fitness. *Science* (2015) 350:217–21. doi: 10.1126/science.aab3369
 23. Kieft JS, Rabe JL, Chapman EG. New hypotheses derived from the structure of a flaviviral Xrn1-resistant RNA: conservation, folding, and host adaptation. *RNA Biol.* (2015) 12:1169–77. doi: 10.1080/15476286.2015.1094599
 24. Jensen S, Thomsen AR. Sensing of RNA viruses: a review of innate immune receptors involved in recognizing RNA virus invasion. *J Virol.* (2012) 86:2900–10. doi: 10.1128/JVI.05738-11
 25. Kell AM, Gale M, Jr. RIG-I in RNA virus recognition. *Virology* (2015) 479–80:110–21. doi: 10.1016/j.virol.2015.02.017
 26. Estes G, Guerra S, Vales-Gomez M, Reyburn HT. Innate immune recognition of double-stranded RNA triggers increased expression of NKG2D ligands after virus infection. *J Biol Chem.* (2017) 292:20472–80. doi: 10.1074/jbc.M117.818393
 27. Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, et al. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* (2009) 458:914–8. doi: 10.1038/nature07745
 28. McDonald SM, Nelson MI, Turner PE, Patton JT. Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nat Rev Microbiol.* (2016) 14:448–60. doi: 10.1038/nrmicro.2016.46
 29. Levy DN, Aldrovandi GM, Kutsch O, Shaw GM. Dynamics of HIV-1 recombination in its natural target cells. *Proc Natl Acad Sci USA.* (2004) 101:4204–9. doi: 10.1073/pnas.0306764101
 30. Gavazzi C, Yver M, Isel C, Smyth RP, Rosa-Calatrava M, Lina B, et al. A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc Natl Acad Sci USA.* (2013) 110:16604–9. doi: 10.1073/pnas.1314419110
 31. Paillart JC, Shehu-Xhilaga M, Marquet R, Mak J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol.* (2004) 2:461–72. doi: 10.1038/nrmicro903
 32. Chin MP, Rhodes TD, Chen J, Fu W, Hu WS. Identification of a major restriction in HIV-1 intersubtype recombination. *Proc Natl Acad Sci USA.* (2005) 102:9002–7. doi: 10.1073/pnas.0502522102
 33. Baird HA, Gao Y, Galetto R, Lalonde M, Anthony RM, Giacomoni V, et al. Influence of sequence identity and unique breakpoints on the frequency of intersubtype HIV-1 recombination. *Retrovirology* (2006) 3:91. doi: 10.1186/1742-4690-3-91
 34. Magiorkinis G, Paraskevis D, Vandamme A.-M, Magiorkinis E, Sypsa V, Hatzakis A. *In vivo* characteristics of human immunodeficiency virus type 1 intersubtype recombination: determination of hot spots and correlation with sequence similarity. *J Gen Virol.* (2003) 84:2715–22. doi: 10.1099/vir.0.19180-0
 35. Baird HA, Galetto R, Gao Y, Simon-Loriere E, Abreha M, Archer J, et al. Sequence determinants of breakpoint location during HIV-1 intersubtype recombination. *Nucleic Acids Res.* (2006) 34:5203–16. doi: 10.1093/nar/gkl669
 36. Smyth RP, Schlub TE, Grimm AJ, Waugh C, Ellenberg P, Chopra A, et al. Identifying recombination hot spots in the HIV-1 genome. *J Virol.* (2014) 88:2891–902. doi: 10.1128/JVI.03014-13
 37. Simon-Loriere E, Martin DP, Weeks KM, Negroni M. RNA structures facilitate recombination-mediated gene swapping in HIV-1. *J Virol.* (2010) 84:12675–82. doi: 10.1128/JVI.01302-10
 38. Simon-Loriere E, Galetto R, Hamoudi M, Archer J, Lefeuvre P, Martin DP, et al. Molecular mechanisms of recombination restriction in the envelope gene of the human immunodeficiency virus. *PLoS Pathog.* (2009) 5:e1000418. doi: 10.1371/journal.ppat.1000418
 39. Geller R, Domingo-Calap P, Cuevas JM, Rossolillo P, Negroni M, Sanjuan R. The external domains of the HIV-1 envelope are a mutational cold spot. *Nat Commun.* (2015) 6:8571. doi: 10.1038/ncomms9571
 40. Harrison GP, Mayo MS, Hunter E, Lever AM. Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5' and 3' of the catalytic site. *Nucleic Acids Res.* (1998) 26:3433–42.
 41. Cohen J, Enserink M. Swine flu. After delays, WHO agrees: the 2009 pandemic has begun. *Science* (2009) 324:1496–7. doi: 10.1126/science.324_1496
 42. Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* (2009) 459:1122–5. doi: 10.1038/nature08182
 43. Gerber M, Isel C, Moules V, Marquet R. Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol.* (2014) 22:446–55. doi: 10.1016/j.tim.2014.04.001
 44. Hutchinson EC, von Kirchbach JC, Gog JR, Digard P. Genome packaging in influenza A virus. *J Gen Virol.* (2010) 91:313–28. doi: 10.1099/vir.0.017608-0
 45. Essere B, Yver M, Gavazzi C, Terrier O, Isel C, Fournier E, et al. Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc Natl Acad Sci USA.* (2013) 110:E3840–8. doi: 10.1073/pnas.1308649110
 46. Maness NJ, Wilson NA, Reed JS, Piaskowski SM, Sacha JB, Walsh AD, et al. Robust, vaccine-induced CD8(+) T lymphocyte response against an out-of-frame epitope. *J Immunol.* (2010) 184:67–72. doi: 10.4049/jimmunol.0903118
 47. Blake N. Immune evasion by gammaherpesvirus genome maintenance proteins. *J Gen Virol.* (2010) 91:829–46. doi: 10.1099/vir.0.018242-0
 48. Murat P, Zhong J, Lekieffre L, Cowieson NP, Clancy JL, Preiss T, et al. G-quadruplexes regulate Epstein-Barr virus-encoded nuclear antigen 1 mRNA translation. *Nat Chem Biol.* (2014) 10:358–64. doi: 10.1038/nchembio.1479
 49. van Hemert FJ, van der Kuyl AC, Berkhout B. The A-nucleotide preference of HIV-1 in the context of its structured RNA genome. *RNA Biol.* (2013) 10:211–5. doi: 10.4161/rna.22896
 50. Keating CP, Hill MK, Hawkes DJ, Smyth RP, Isel C, Le SY, et al. The A-rich RNA sequences of HIV-1 pol are important for the synthesis of viral cDNA. *Nucleic Acids Res.* (2009) 37:945–56. doi: 10.1093/nar/gkn1015
 51. Klaver B, van der Velden Y, van Hemert F, van der Kuyl AC, Berkhout B. HIV-1 tolerates changes in A-count in a small segment of the pol gene. *Retrovirology* (2017) 14:43. doi: 10.1186/s12977-017-0367-0
 52. Takata MA, Goncalves-Carneiro D, Zang TM, Soll SJ, York A, Blanco-Melo D, et al. CG dinucleotide suppression enables antiviral defence targeting non-self RNA. *Nature* (2017) 550:124–7. doi: 10.1038/nature24039
 53. Vabret N, Bailly-Bechet M, Najburg V, Muller-Trutwin M, Verrier B, Tangy F. The biased nucleotide composition of HIV-1 triggers type I interferon response and correlates with subtype D increased pathogenicity. *PLoS ONE* (2012) 7:e33502. doi: 10.1371/journal.pone.0033502
 54. Takata MA, Soll SJ, Emery A, Blanco-Melo D, Swanstrom R, Bieniasz PD. Global synonymous mutagenesis identifies cis-acting RNA elements that regulate HIV-1 splicing and replication. *PLoS Pathog.* (2018) 14:e1006824. doi: 10.1371/journal.ppat.1006824
 55. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol.* (2005) 15:1022–7. doi: 10.1016/j.cub.2005.04.050
 56. Yant LJ, Friedrich TC, Johnson RC, May GE, Maness NJ, Enz AM, et al. The high-frequency major histocompatibility complex class I allele Mamu-B*17 is associated with control of simian immunodeficiency virus SIVmac239 replication. *J Virol.* (2006) 80:5074–7. doi: 10.1128/JVI.80.10.5074-5077.2006

57. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* (2002) 418:646–50. doi: 10.1038/nature00939
58. York A, Kutluay SB, Errando M, Bieniasz PD. The RNA binding specificity of human APOBEC3 proteins resembles that of HIV-1 nucleocapsid. *PLoS Pathog.* (2016) 12:e1005833. doi: 10.1371/journal.ppat.1005833
59. Holbrook SR, Kim SH. RNA crystallography. *Biopolymers* (1997) 44:3–21.
60. Scott LG, Hennig M. RNA structure determination by NMR. *Methods Mol Biol.* (2008) 452:29–61. doi: 10.1007/978-1-60327-159-2_2
61. Woodson SA. Taming free energy landscapes with RNA chaperones. *RNA Biol.* (2010) 7:677–86. doi: 10.4161/rna.7.6.13615
62. Bhandari YR, Jiang W, Stahlberg EA, Stagno JR, Wang YX. Modeling RNA topological structures using small angle X-ray scattering. *Methods* (2016) 103:18–24. doi: 10.1016/j.jymeth.2016.04.015
63. Stephenson JD, Kenyon JC, Symmons ME, Lever AM. Characterizing 3D RNA structure by single molecule FRET. *Methods* (2016) 103:57–67. doi: 10.1016/j.jymeth.2016.02.004
64. Stephenson JD, Li H, Kenyon JC, Symmons M, Klennerman D, Lever AM. Three-dimensional RNA structure of the major HIV-1 packaging signal region. *Structure* (2013) 21:951–62. doi: 10.1016/j.str.2013.04.008
65. Schon P. Imaging and force probing RNA by atomic force microscopy. *Methods* (2016) 103:25–33. doi: 10.1016/j.jymeth.2016.05.016
66. Rausch JW, Sztuba-Solinska J, Le Grice SFJ. Probing the structures of viral RNA regulatory elements with SHAPE and related methodologies. *Front Microbiol.* (2017) 8:2634. doi: 10.3389/fmicb.2017.02634
67. Jayaraman D, Kenyon JC. New windows into retroviral RNA structures. *Retrovirology* (2018) 15:11. doi: 10.1186/s12977-018-0393-6
68. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures *in vivo*. *Nature* (2014) 505:701–5. doi: 10.1038/nature12894
69. Ding Y, Tang Y, Kwok CK, Zhang Y, Bevilacqua PC, Assmann SM. *In vivo* genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* (2014) 505:696–700. doi: 10.1038/nature12756
70. Underwood JG, Uzilov AV, Katzman S, Onodera CS, Mainzer JE, Mathews DH, et al. FragSeq: transcriptome-wide RNA structure probing using high-throughput sequencing. *Nat Methods* (2010) 7:995–1001. doi: 10.1038/nmeth.1529
71. Wan Y, Qu K, Ouyang Z, Kertesz M, Li J, Tibshirani R, et al. Genome-wide measurement of RNA folding energies. *Mol Cell.* (2012) 48:169–81. doi: 10.1016/j.molcel.2012.08.008
72. Deigan KE, Li TW, Mathews DH, Weeks KM. Accurate SHAPE-directed RNA structure determination. *Proc Natl Acad Sci USA.* (2009) 106:97–102. doi: 10.1073/pnas.0806929106
73. Watters KE, Abbott TR, Lucks JB. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Res.* (2016) 44:e12. doi: 10.1093/nar/gkv879
74. Smyth RP, Despons L, Huili G, Bernacchi S, Hijnen M, Mak J, et al. Mutational interference mapping experiment (MIME) for studying RNA structure and function. *Nat Methods* (2015) 12:866–72. doi: 10.1038/nmeth.3490
75. Kenyon JC, Prestwood LJ, Lever AM. A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Sci Rep.* (2015) 5:14369. doi: 10.1038/srep14369
76. Bieniasz PD, Kutluay SB. CLIP-related methodologies and their application to retrovirology. *Retrovirology* (2018) 15:35. doi: 10.1186/s12977-018-0417-2
77. Weidmann CA, Mustoe AM, Weeks KM. Direct duplex detection: an emerging tool in the RNA structure analysis toolbox. *Trends Biochem Sci.* (2016) 41:734–736. doi: 10.1016/j.tibs.2016.07.001
78. Helwak A, Kudla G, Dudnakova T, Tollervey D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* (2013) 153:654–65. doi: 10.1016/j.cell.2013.03.043
79. Kudla G, Granneman S, Hahn D, Beggs JD, Tollervey D. Cross-linking, ligation, and sequencing of hybrids reveals RNA-RNA interactions in yeast. *Proc Natl Acad Sci USA.* (2011) 108:10010–5. doi: 10.1073/pnas.1017386108
80. Ramani V, Qiu R, Shendure J. High-throughput determination of RNA structure by proximity ligation. *Nat Biotechnol.* (2015) 33:980–4. doi: 10.1038/nbt.3289
81. Lu Z, Zhang QC, Lee B, Flynn RA, Smith MA, Robinson JT, et al. RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* (2016) 165:1267–79. doi: 10.1016/j.cell.2016.04.028
82. Sharma E, Sterne-Weiler T, O'Hanlon D, Blencowe BJ. Global mapping of human RNA-RNA interactions. *Mol Cell.* (2016) 62:618–26. doi: 10.1016/j.molcel.2016.04.030
83. Aw JG, Shen Y, Wilm A, Sun M, Lim XN, Boon KL, et al. *In vivo* mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol Cell.* (2016) 62:603–17. doi: 10.1016/j.molcel.2016.04.028
84. Nguyen TC, Cao X, Yu P, Xiao S, Lu J, Biase FH, et al. Mapping RNA-RNA interactome and RNA structure *in vivo* by MARIO. *Nat Commun.* (2016) 7:12023. doi: 10.1038/ncomms12023
85. Chen YG, Satpathy AT, Chang HY. Gene regulation in the immune system by long noncoding RNAs. *Nat Immunol.* (2017) 18:962–72. doi: 10.1038/ni.3771

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Smyth, Negroni, Lever, Mak and Kenyon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Influenza Virus—Host Co-evolution. A Predator-Prey Relationship?

Konstantinos Voskarides, Eirini Christaki and Georgios K. Nikolopoulos*

Medical School, University of Cyprus, Nicosia, Cyprus

Influenza virus continues to cause yearly seasonal epidemics worldwide and periodically pandemics. Although influenza virus infection and its epidemiology have been extensively studied, a new pandemic is likely. One of the reasons influenza virus causes epidemics is its ability to constantly antigenically transform through genetic diversification. However, host immune defense mechanisms also have the potential to evolve during short or longer periods of evolutionary time. In this mini-review, we describe the evolutionary procedures related with influenza viruses and their hosts, under the prism of a predator-prey relationship.

Keywords: adaptation, antagonistic evolution, influenza virus, bottleneck, immune system, antigen, genetics, mutation

OPEN ACCESS

Edited by:

Tara Patricia Hurst,
Abcam, United Kingdom

Reviewed by:

Sarah Rowland-Jones,
University of Oxford, United Kingdom
Jianzhong Zhu,
Yangzhou University, China

*Correspondence:

Georgios K. Nikolopoulos
gknikolopoulos@gmail.com;
nikolopoulos.georgios@ucy.ac.cy

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 31 May 2018

Accepted: 15 August 2018

Published: 07 September 2018

Citation:

Voskarides K, Christaki E and
Nikolopoulos GK (2018) Influenza
Virus—Host Co-evolution. a
Predator-Prey Relationship?
Front. Immunol. 9:2017.
doi: 10.3389/fimmu.2018.02017

INTRODUCTION

Health disasters caused by influenza viruses are abundant in human history (1–3). The Spanish flu of 1918 is estimated to have claimed more than 50 million lives, a figure that was beyond the death toll of World War I (4, 5). Another two pandemics occurred later in the twentieth century when surveillance systems and laboratory capacity were in place to monitor them: the Asian flu (1957) and the Hong Kong flu (1968) (2, 6), which had lower fatality rates but still caused between half and 2 million deaths each. In 2009, a novel H1N1 influenza virus spread rapidly from Mexico to the rest of the globe (2) but had eventually moderate impact with less deaths than the previous pandemics (2, 7), partly due to our improved response. Seasonal epidemics may also be severe with high morbidity and mortality (8).

Viral infections are not a static phenomenon. Many viral genomes can adapt on their hosts' cellular environment while survivors of viral infections have probably special genetic characteristics that helped them remain alive. Therefore, both genomes change over time reminding us of the evolutionary race between predators and preys. Viral genomes however are smaller and more flexible, allowing viruses to rapidly adapt through random mutagenesis or due to transmission in other species. Pandemics of influenza are due to these random and sudden viral mutations or trans-species transmissions. Under this perspective, in this paper, we discuss about influenza viruses and the coevolution with the human genomes.

INFLUENZA VIRUS

Virology and Replication Cycle

Influenza viruses are members of the *Orthomyxoviridae* family. Their genome is comprised of a negative sense, segmented single-stranded RNA. Viral particles also contain essential viral proteins and a host cell-derived envelope (9). Influenza viruses are further classified into four types: A, B, C, and D. Influenza virus types A, B, and C infect and cause respiratory illness in humans. Influenza D viruses mainly affect cattle and are not known to infect humans (10). Influenza

A and B viruses cause seasonal epidemics whereas type C viruses usually cause a mild upper respiratory tract illness and associated epidemics have only been scarcely reported (11). Influenza A viruses can infect many animal species, including birds, pigs, horses, marine mammals, and other hosts, and can cause pandemics. Influenza A viruses are categorized into subtypes based on the molecular characteristics of their surface glycoproteins, hemagglutinin (HA), and neuraminidase (NA). Identification of at least 18 antigenically distinct HA subtypes and 11 distinct NA subtypes of influenza A virus strains infecting humans and animals have so far been determined (12, 13). Two genetically and antigenically distinct lineages (Victoria and Yamagata) of Influenza B viruses co-circulate in humans (14–16).

Hemagglutinin is comprised of a dimer HA1-HA2: HA1 is crucial for binding to the host cell receptor whereas HA2 for cell fusion. Viral endocytosis is followed by uncoating and release of viral RNA, which is imported into the nucleus where viral replication and protein synthesis take place using viral polymerase proteins and the host cell machinery (17). Virions are assembled in the cell surface and bud enclosed in an envelope originating from the host cell membrane. Neuraminidase allows the virus to leave the infected cell as it cleaves sialic acid (SA) from the cell surface receptors. Viral replication causes cell death with various mechanisms including disruption of protein synthesis and apoptosis. Since viral release continues for hours before cell death, many respiratory epithelial cells are affected and die within a few replication cycles (9, 18).

Influenza viruses target epithelial cells of the respiratory tract, which contain SA receptors. Epithelial cells across species express different SA receptors and Influenza A virus strains show a predilection for certain types of such receptors, making zoonotic transmission difficult. For example, human influenza strains have a predilection for SA α -2,6 galactose receptors, which are found in the respiratory epithelium of the upper airways in humans, while animal influenza A viruses bind to SA α -2,3 galactose, which is found on the epithelial cells of birds and pigs, but could also be expressed in the human lower respiratory tract epithelium (12, 19, 20). HA epitopes are the major determinants for the production of strain-specific neutralizing antibodies.

Clinical Manifestations

Influenza symptoms usually present abruptly, after an incubation period of 1–2 days. Systemic symptoms are characteristic and help differentiate influenza from other upper respiratory tract viral illnesses. These include high fever, chills, rigors, headache, myalgias, malaise, and anorexia. Fever and systemic symptoms commonly last for 3 days, however fever can last up to 8 days. Myalgias can be severe and usually involve the back and extremities. Respiratory symptoms include dry cough, sore throat, hoarseness, nasal congestion, and discharge (18).

Different subtypes of influenza have different ability to infect airway epithelial cells of the upper or lower respiratory tract, hence causing a milder infection or a more severe illness leading to severe pneumonia. For example, H5N1 infects alveolar epithelial cells as well as alveolar macrophages, triggering a significant pro-inflammatory response, which can result in severe lung injury (21–23).

HOST IMMUNE RESPONSE TO INFLUENZA VIRUS INFECTION

Cells of the innate immune response are the first and fast responders upon influenza virus infection, recruited by chemokines released by airway epithelial cells. Upon viral entry, intracellular viral ssRNA and other viral molecular patterns are recognized mainly by Toll-like receptors (TLR) 3,7,8,9 and retinoic acid-inducible gene-I protein (RIG-I) receptors. The downstream signaling triggered by the activation of these receptors results in the activation of transcription factors like nuclear factor kappa-B and interferon regulatory factor (IRF) 3 and 7, leading to the expression of pro-inflammatory cytokines and interferons (24–26). Moreover, NOD-like receptor family pyrin domain containing 3 (NALP3) inflammasome is also activated upon influenza virus infection promoting IL-1 β and IL-18 secretion, and pulmonary infiltration by neutrophils and macrophages (27).

Natural Killer (NK) cells, monocytes, neutrophils, and dendritic cells migrate to the site of infection and exhibit antiviral activity. NK cells have cytotoxic activity on cells infected with influenza virus, macrophages phagocytose infected cells and regulate adaptive immune responses, and dendritic cells present viral antigens bound to Major Histocompatibility Complex (MHC) molecules to naïve and memory T lymphocytes, initiating the specific adaptive immune response. In addition, immunoglobulins (mainly IgA) present in nasal secretions contribute to the anti-influenza immune response by preventing viral entry (26).

Both T and B cells are essential in the adaptive immune response to influenza virus infection. Naïve CD8⁺ T cells, upon activation by dendritic cells and facilitated by the action of cytokines, proliferate and differentiate to cytotoxic T lymphocytes (CTLs). CTLs are able to kill influenza virus-infected cells and also restrict viral replication via production of cytokines and effector molecules like perforin and granzymes (26, 28, 29). Memory CTLs cells can respond efficiently during a secondary infection and confer cross-protective heterosubtypic immunity (30, 31). Moreover, CD4⁺ T cells express co-stimulatory molecules that participate in antibody production by B cells. At the same time, antiviral cytokines, like INF- γ , TNE, and IL-2 are expressed from Th1 CD4⁺ T cells, which also help activate alveolar macrophages. Th2 CD4⁺ T cells contribute to B cell specific responses (32). Comprising an essential defense mechanism against influenza virus infection, B cells generate neutralizing (hemagglutinin-inhibiting antibodies) as well as non-neutralizing antibodies (anti-NA and antibodies against structural proteins). The latter, as well as specific antibody dependent cell-mediated cytotoxicity, play an important role in cross-protection against different influenza subtypes (31, 33). There is increased interest in the role of cross-reactive T cells and broadly neutralizing antibodies in the development of vaccines that could elicit “universal” immunity against different or novel strains of influenza viruses.

VIRAL EVOLUTION

Predator and Prey. Antagonistic Evolution

Predator-prey species relationship is of great interest in evolutionary biology since the balance of attack and defense mechanisms is very important for the long-term survival of both species. A simplified example that is often given is that of lions and their prey. The faster the lions are, the most successful they can be in acquiring food, surviving, and reproducing. The faster the zebras (lions' prey) are, the better they can escape the lions, so they survive and reproduce. It is all about Darwinian fitness, since survival and robustness are related with a higher probability of having more offspring (descendants). That is translated to reproducibility and to a better chance to evolve. Darwinian fitness cannot increase simultaneously for both predator and prey species. It increases for one species and reduces for the other. But that is not stable. Afterwards the opposite happens. This is the phenomenon of antagonistic evolution that is believed to push evolution forward and sometimes can increase phenotypic diversification (34).

Antagonistic evolution can be a main driver of multiple phenotypes, specialization, or even speciation. Multiple species evolution, those belonging to the same genus or family, is termed as evolutionary radiation. An example is that of marine mammals, dolphins, and whales. It is believed that multiple such species have been evolved due to prey (fish, squid) specialization (35, 36).

Antagonistic evolution also happens between host and various pathogens including viruses. Successful viruses are the ones that can cheat cells, go inside the cytoplasm, take advantage of the cell translation system (ribosomes), reproduce, and then exit the cell by exocytosis going to infect other cells. Viruses can be considered as the predators and cells as the preys, even though viruses do not necessarily kill the cells. This is probably a major difference with the classical predator-prey relationship. Evolutionary successful viruses are transmitted from one host to another (cell or a whole organism), making minimal damage. On the other site, cellular defense mechanisms, like the MHC proteins, have the potential for evolving in a quite short evolutionary time, this being a highly effective immunological shield.

Genetic Diversification of Influenza Viruses

Influenza viruses exhibit vast genetic variability sourced from three molecular mechanisms: mutation, genetic re-assortment, and genetic recombination. While the first two mechanisms have been extensively studied, the significance of the third one needs clarification (37). The results of these genetic procedures are antigenic shift or antigenic drift. Antigenic shift is a rapid change of the virus genome and antigenicity, coming from a combination of different viral strains. The eight RNA fragments of the influenza virus can re-assort between different strains, producing new subtypes. This happens usually inside livestock animals, in geographic regions where people live in proximity with animals (38). Antigenic shifts caused influenza pandemics in the past. On the other hand, mutations, usually nucleotide substitutions, can cause antigenic drifts. Antigenic drifts cause

more gradual changes to the viral HA and N proteins, this being the cause of the appearance of new influenza strains each year (39). Antigenic drifts are very important for vaccine production.

The major cause of the high evolutionary rate of influenza viruses is the lack of proofreading (correction) function of the viral RNA polymerase. Mutagenesis is close to 10^{-3} - 10^{-4} per nucleotide site, this being a very high mutagenesis rate and producing a high genetic variability in the viral genetic pool (40, 41). Another consequence of the continuously evolved viral genetic variability is the increased probability for widening the host-range of the virus due to HA protein changes (37).

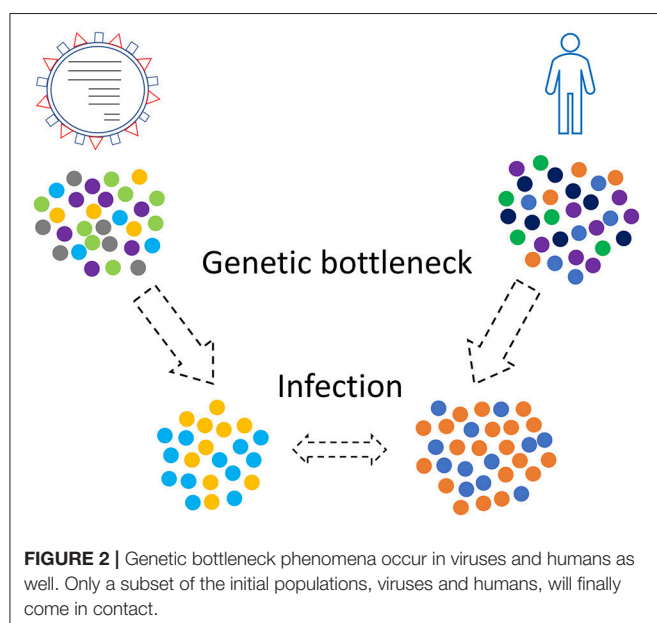
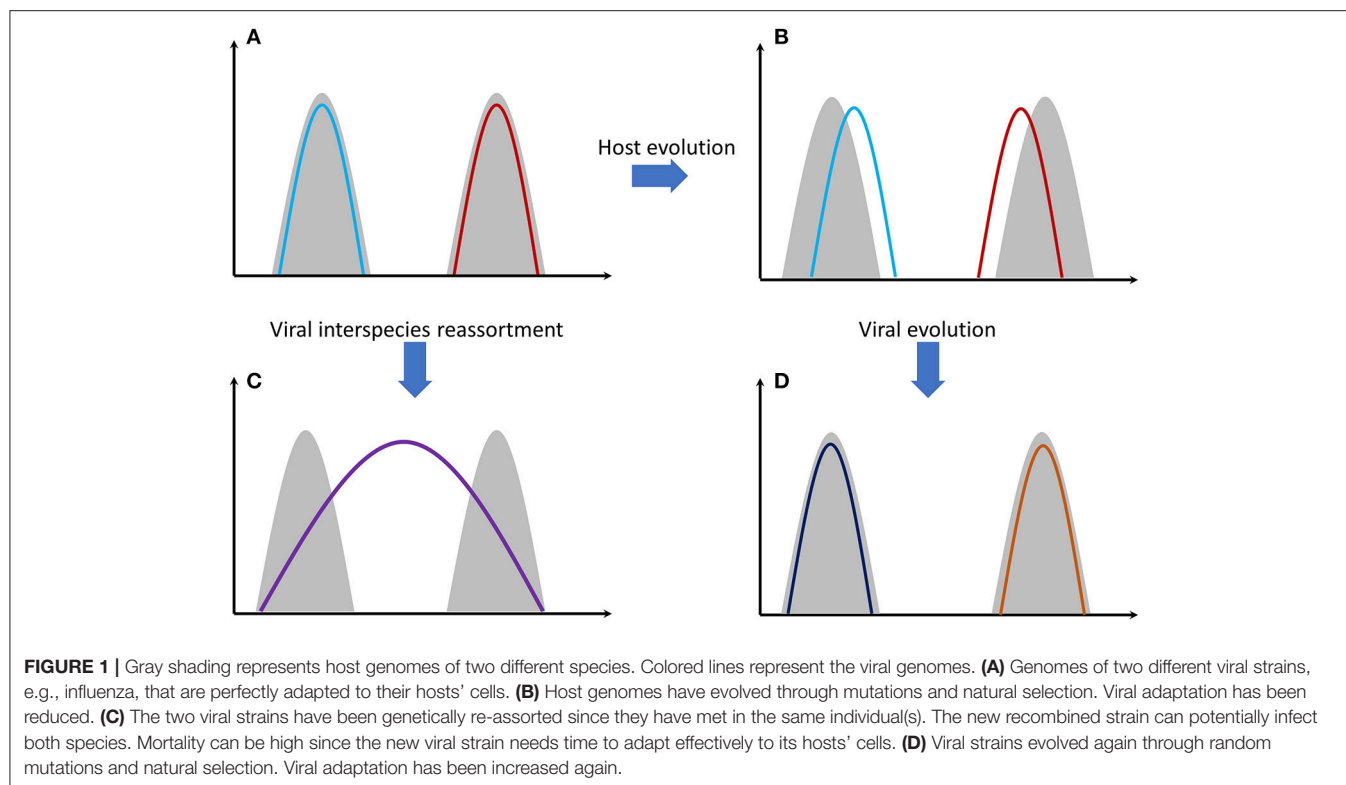
Influenza A probably evolves under a negative pressure for CpG oligonucleotides. This is considered to happen due to innate immune recognition, something that is not so vital for influenza B evolution (42, 43). Additionally, it has been shown that antigenic drifts are more frequent to influenza A viruses than influenza B viruses (44). Even though antigenic drifts are usually gradual changes regarding the viral properties, the possibility for a rapid evolutionary change due to a mutation (punctuated evolution) cannot be excluded.

One of the most interesting recent findings regarding influenza virus' genetic variability and evolution is that when people become infected, they are actually infected from a diversity of viruses and not from only one strain (45). This implies that vaccines target the dominant strain and not the whole population of viruses. Most of these viral strains are not well adapted but can stay for long inside a human (or other species) population, thus hiding the danger to gain a gradual adaptation through mutation accumulation processes (46–48). It is obvious that natural selection can determine which viral variants can survive and which cannot. Additionally, this multi-infection process may be related with cooperation or competition phenomena among viruses, that currently are not so well understood (39).

Further understanding of the mutational and evolutionary procedures related to influenza viruses can help us develop prediction models. Prediction models are very important for vaccine production. Most of the predictive modeling studies focus on the surface protein hemagglutinin (49). Hemagglutinin DNA sequences for thousands of influenza A strains, isolated over the last 40 years, are freely available in electronic databases (50). However, prediction studies and experimental evolution studies are very challenging. Currently, available tools are not adequate for a reliable approach. More details can be found in the recently published reviews (51) and (52).

A Predator—Prey Relationship Between Human and Influenza Virus

The relationship between viruses and eukaryotic cells has its origins millions of years ago. It is not so simple since many viruses complete their cycle inside a cell without harming it. Instead, sometimes their nucleic acid sequences have been embedded in the host cell DNA. In humans, about half of their genome is constituted of viral sequences, showing this long lasting ancient interaction (53, 54).



Antibody-mediated immunity to influenza viruses is well-understood. Vaccines success is based on this response. Despite this, vaccines are ineffective in many people. One reason is that cellular immunity through T cells is also a very important immunological response (32). Antibody variability is a genetically intrinsic mechanism of B cells. However, T-cell response is based on MHC (HLA) proteins efficiency that differs

among individuals and is an inherited trait. Recognition of viral proteins on the surface of infected cells is performed through MHC proteins. There are thousands of genetic polymorphisms in the MHC genes at the population level, producing millions of combinations. These combinations are inherited in the form of “genes packages” termed as haplotypes (each person carries two MHC haplotypes). It is unlikely for two individuals to carry a single common haplotype. Some MHC proteins (haplotypes) are more effective than others in recognizing viral proteins. Why is there this vast diversity of MHC proteins in human populations? The answer is that this increases the likelihood that some individuals will survive after a severe pandemic and this way a species extinction is avoided. On the other hand, we can assume that haplotypes that exist today are the ones that were successful in the past for combating specific viral or bacterial infections. Survivors passed these haplotypes onto next generations. Of course, as it was stated before, viruses are not stable entities. Mutations increase the number of viral types and some of them will finally “win” the previously successful MHC proteins causing again the (human) population to be susceptible. This is quite similar with the predator-prey relationship happening extensively in nature (Figure 1).

T4 cells (CD4 cells or T-helper cells) and T8 cells (CD8 cells or T-cytotoxic cells) are the most important lymphocytes that use MHC proteins for recognizing the virus-infected cells. Studies showed that the long-term evolution of T8 cells epitopes may have shaped the long-term evolution of influenza A virus (55, 56). On the other hand, influenza viruses seem to respond to this evolutionary pressure by mutations to their

nucleoprotein gene (nucleoprotein is an important viral capsid protein for T-cell recognition). It has also been proven that some combinations of MHC proteins and nucleoprotein variants result in immunological response and some not (57–59). This antagonistic evolutionary relationship is a factor of a vice-versa increasing genetic variability. In the same regard, TLRs may have remained evolutionary stable, possibly due to the fact that the pathogen-associated molecular patterns (PAMPs) they recognize have also remained unaltered, since they are fundamental for pathogen survival (60).

Another category of important proteins in human cells, defending against viral infections, are the IFITM (interferon-induced transmembrane) proteins. At least three human IFITM proteins exist (61). When upregulated (by both type I and type II interferons), they can restrict viruses outside cells, prohibiting viral penetration through the membrane lipid layers. IFITM3 seems to be more specific for influenza viruses (61). Gene polymorphisms in IFITM3 gene can increase or reduce susceptibility to influenza infections, underlining the evolutionary significance of these proteins (62, 63).

Genetic Bottlenecks

In population genetics, genetic bottleneck is the event where a sub-group of individuals originating from a larger population, migrate in another geographic region, carrying only a subset of the genetic information of the initial population. This is a standard procedure during evolution. Populations split, and a part of the initial genetic diversity is transferred in different geographic regions. Isolation and new mutations can dramatically change the genetic diversity of the new population.

Transmission of viruses in new hosts implies three different genetic bottleneck phenomena: (i) The viral transmission bottleneck, which determines how much of the viral diversity generated in one host passes to another during transmission (64, 65). Studies show that along with major variants, other minor viral variants can be transferred from one host to another (66). These minor variants can be transformed to major through natural selection pressures. Two different processes are related

with this bottleneck: virus entry and virus replication followed by production of new virions (ii) Individuals or sub-populations (e.g., humans) migrate in other geographic regions and transfer there only a group of viruses of the initial pool of viruses, and (iii) Genetic bottlenecks of the hosts, e.g., humans. Host genome is vital for the effectiveness of the immune response. Only a subset of the paternal population MHC genes' variability will be transferred in the new geographic region.

Combinations of those three kinds of genetic bottlenecks and evolutionary forces will shape the final virus-host relationship that will take place in a human population (**Figure 2**). This relationship is not static since new genetic bottlenecks take place, especially in present days that people travel a lot.

What determines the viral bottleneck size? Small viral founding populations are usually observed, even though this depends on the virus and the host (67). Experimental studies in ferret and guinea pig showed that influenza infections coming from limited dose of aerosol exposure had tighter bottlenecks than contact transmission (68). More studies are needed to elucidate the importance of this phenomenon in human populations. Viruses follow special routes in our planet and understanding this may contribute to prevention strategies.

CONCLUSION

Influenza viruses and humans have co-evolved in the past and will continue to co-evolve. Our understanding of influenza virus pathogenesis and evolution has increased, however, we are still unprepared to stop the next pandemic strain from emerging. Evolution prediction algorithms can help inform efforts toward the development of better preventive measures.

AUTHOR CONTRIBUTIONS

KV and EC: conception, drafting the manuscript, approval of the final version; GN: conception and design, critical revision of the manuscript, approval of the final version.

REFERENCES

- Ramiro D, Garcia S, Casado Y, Cilek L, Chowell G. Age-specific excess mortality patterns and transmissibility during the 1889-1890 influenza pandemic in Madrid, Spain. *Ann Epidemiol.* (2018) 28:267–72. doi: 10.1016/j.annepidem.2017.12.009
- Saunders-Hastings P, Krewski D. Reviewing the history of pandemic influenza: understanding patterns of emergence and transmission. *Pathogens* (2016) 5:66. doi: 10.3390/pathogens5040066
- Taubenberger JK, Morens DM. Pandemic influenza—including a risk assessment of H5N1. *Rev Sci Tech.* (2009) 28:187–202. doi: 10.1016/j.bbi.2008.05.010
- Chowell G, Sullivan P, Rothenberg R. Introduction to symposium: a century after the 1918 influenza pandemic. *Ann Epidemiol.* (2018) 28:265–6. doi: 10.1016/j.annepidem.2018.03.010
- Johnson NPAS, Mueller J. Updating the accounts: global mortality of the 1918-1920 “Spanish” influenza pandemic. *Bull Hist Med.* (2002) 76:105–15. doi: 10.1353/bhm.2002.0022
- Cockburn WC, Delon PJ, Ferreira W. Origin and progress of the 1968-69 Hong Kong influenza epidemic. *Bull World Health Organ.* (1969) 41:345–8. doi: 10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2
- Simonsen L, Spreeuwenberg P, Lustig R, Taylor RJ, Fleming DM, Kroneman M, et al. Global mortality estimates for the 2009 influenza pandemic from the GLaMOR project: a modeling study. *PLoS Med.* (2013) 10:e1001558. doi: 10.1371/journal.pmed.1001558
- Nikolopoulos G, Tsiodras S, Bonovas S. Influenza vaccination: Cassandra's prophecy. *Public Health* (2012) 126:82–3. doi: 10.1016/j.puhe.2011.09.028
- Fields B, Mahan N, Knipe D, Howley MP. *Fields Virology*. Philadelphia, PA: Lippincott Williams & Wilkins (2006)
- Zhai S-L, Zhang H, Chen S-N, Zhou X, Lin T, Liu R, et al. Influenza D virus in animal species in Guangdong province, Southern China. *Emerg Infect Dis.* (2017) 23:1392–6. doi: 10.3201/eid2308.170059
- Matsuzaki Y, Abiko C, Mizuta K, Sugawara K, Takashita E, Muraki Y, et al. A nationwide epidemic of influenza C virus infection in Japan in 2004. *J Clin Microbiol.* (2007) 45:783–8. doi: 10.1128/JCM.01555-06

12. Medina RA, García-Sastre A. Influenza A viruses: New research developments. *Nat Rev Microbiol.* (2011) 9:590–603. doi: 10.1038/nrmicro2613
13. Ma W, García-Sastre A, Schwemmler M. Expected and unexpected features of the newly discovered bat Influenza A-like viruses. *PLoS Pathog.* (2015) 11:e1004819. doi: 10.1371/journal.ppat.1004819
14. Lin YP, Gregory V, Bennett M, Hay A. Recent changes among human influenza viruses. *Virus Res.* (2004) 103:47–52. doi: 10.1016/j.virusres.2004.02.011
15. Shaw MW, Xu X, Li Y, Normand S, Ueki RT, Kunimoto GY, et al. Reappearance and global spread of variants of influenza B/Victoria/2/87 lineage viruses in the 2000–2001 and 2001–2002 seasons. *Virology* (2002) 303:1–8. doi: 10.1006/viro.2002.1719
16. Ikonen N, Pyhälä R, Axelin T, Kleemola M, Korpela H. Reappearance of influenza B/Victoria/2/87-lineage viruses: Epidemic activity, genetic diversity and vaccination efficacy in the Finnish Defence Forces. *Epidemiol Infect.* (2005) 133:263–71. doi: 10.1017/S0950268804003462
17. Sun X, Whittaker GR. Entry of influenza virus. *Adv Exp Med Biol.* (2013) 790:72–82. doi: 10.1007/978-1-4614-7651-1_4
18. Mandell G, Bennett J, Dolin R. *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, 8th ed.* Philadelphia, PA: Elsevier Health Sciences (2015).
19. de Graaf M, Fouchier RAM. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *EMBO J.* (2014) 33:823–41. doi: 10.1002/embj.201387442
20. Byrd-Leotis L, Cummings RD, Steinhauer DA. The interplay between the host receptor and influenza virus hemagglutinin and neuraminidase. *Int J Mol Sci.* (2017) 18:e18071541. doi: 10.3390/ijms18071541
21. Van Riel D, Munster VJ, De Wit E, Rimmelzwaan GF, Fouchier RAM, Osterhaus ADME, et al. Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *Am J Pathol.* (2007) 171:1215–23. doi: 10.2353/ajpath.2007.070248
22. Kuiken T, Riteau B, Fouchier RAM, Rimmelzwaan GF. Pathogenesis of influenza virus infections: The good, the bad and the ugly. *Curr Opin Virol.* (2012) 2:276–86. doi: 10.1016/j.coviro.2012.02.013
23. Nicholls JM, Chan MCW, Chan WY, Wong HK, Cheung CY, Kwong DLW, et al. Tropism of avian influenza A (H5N1) in the upper and lower respiratory tract. *Nat Med.* (2007) 13:147–9. doi: 10.1038/nm1529
24. Le Goffic R, Pothlichet J, Vitour D, Fujita T, Meurs E, Chignard M, et al. Cutting edge: influenza A virus activates TLR3-dependent inflammatory and RIG-I-dependent antiviral responses in human lung epithelial cells. *J Immunol.* (2007) 178:3368–72. doi: 10.4049/jimmunol.178.6.3368
25. Goubau D, Schlee M, Deddouche S, Puijssers AJ, Zillinger T, Goldeck M, et al. Antiviral immunity via RIG-I-mediated recognition of RNA bearing 59-diphosphates. *Nature* (2014) 514:372–5. doi: 10.1038/nature13590
26. Chen X, Liu S, Goraya MU, Maarouf M, Huang S, Chen J-L. Host immune response to influenza A virus infection. *Front Immunol.* (2018) 9:1–13. doi: 10.3389/fimmu.2018.00320
27. Ong JDH, Mansell A, Tate MD. Hero turned villain: NLRP3 inflammasome-induced inflammation during influenza A virus infection. *J Leukoc Biol.* (2017) 101:863–74. doi: 10.1189/jlb.4MR0616-288R
28. van Domselaar R, Bovenschen N. Cell death-independent functions of granzymes: Hit viruses where it hurts. *Rev Med Virol.* (2011) 21:301–14. doi: 10.1002/rmv.697
29. Grant EJ, Quiñones-Parra SM, Clemens EB, Kedzierska K. Human influenza viruses and CD8 + T cell responses. *Curr Opin Virol.* (2016) 16:132–42. doi: 10.1016/j.coviro.2016.01.016
30. Liang S, Mozdzanowska K, Palladino G, Gerhard W. Heterosubtypic immunity to influenza type A virus in mice. effector mechanisms and their longevity. *J Immunol.* (1994) 152:1653–61. doi: 10.4049/jimmunol.178.2.1030
31. Sridhar S. Heterosubtypic T-cell immunity to influenza in humans: Challenges for universal T-cell influenza vaccines. *Front Immunol.* (2016) 7:e00195. doi: 10.3389/fimmu.2016.00195
32. La Gruta NL, Turner SJ. T cell mediated immunity to influenza: mechanisms of viral control. *Trends Immunol.* (2014) 35:396–402. doi: 10.1016/j.it.2014.06.004
33. Neu KE, Henry Dunand CJ, Wilson PC. Heads, stalks and everything else: how can antibodies eradicate influenza as a human disease? *Curr Opin Immunol.* (2016) 42:48–55. doi: 10.1016/j.coi.2016.05.012
34. Brodersen J, Post DM, Seehausen O. Upward adaptive radiation cascades: predator diversification induced by prey diversification. *Trends Ecol Evol.* (2018) 33:59–70. doi: 10.1016/j.tree.2017.09.016
35. Morin PA, Archer FI, Foote AD, Vilstrup J, Allen EE, Wade P, et al. Complete mitochondrial genome phylogeographic analysis of killer whales (*Orcinus orca*) indicates multiple species. *Genome Res.* (2010) 20:908–16. doi: 10.1101/gr.102954.109
36. Hoelzel AR, Potter CW, Best PB. Genetic differentiation between parapatric “nearshore” and “offshore” populations of the bottlenose dolphin. *Proc R Soc B Biol Sci.* (1998) 265:1177–83. doi: 10.1098/rspb.1998.0416
37. Shao W, Li X, Goraya MU, Wang S, Chen JL. Evolution of influenza A virus by mutation and re-assortment. *Int J Mol Sci.* (2017) 18:e18081650. doi: 10.3390/ijms18081650
38. Ma W, Kahn RE, Richt JA. The pig as a mixing vessel for influenza viruses: human and veterinary implications. *J Mol Genet Med.* (2009) 03:158–66. doi: 10.4172/1747-0862.1000028
39. Johnson KEE, Song T, Greenbaum B, Ghedin E. Getting the flu: 5 key facts about influenza virus evolution. *PLoS Pathog.* (2017) 13:1–6. doi: 10.1371/journal.ppat.1006450
40. Ahlquist P. RNA-dependent RNA polymerases, viruses, and RNA silencing. *Science* (2002) 296:1270–3. doi: 10.1126/science.1069132
41. Chen R, Holmes EC. Avian influenza virus exhibits rapid evolutionary dynamics. *Mol Biol Evol.* (2006) 23:2336–41. doi: 10.1093/molbev/msl102
42. Jimenez-Baranda S, Greenbaum B, Manches O, Handler J, Rabadan R, Levine A, et al. Oligonucleotide motifs that disappear during the evolution of influenza virus in humans increase alpha interferon secretion by plasmacytoid dendritic cells. *J Virol.* (2011) 85:3893–904. doi: 10.1128/JVI.01908-10
43. Greenbaum BD, Cocco S, Levine AJ, Monasson R. Quantitative theory of entropic forces acting on constrained nucleotide sequences applied to viruses. *Proc Natl Acad Sci USA.* (2014) 111:5054–9. doi: 10.1073/pnas.1402285111
44. Bedford T, Suchard MA, Lemey P, Dudas G, Gregory V, Hay AJ, et al. Integrating influenza antigenic dynamics with molecular evolution. *Elife* (2014) 2014:e01914. doi: 10.7554/eLife.01914
45. Poon LLM, Song T, Rosenfeld R, Lin X, Rogers MB, Zhou B, et al. Quantifying influenza virus diversity and transmission in humans. *Nat Genet.* (2016) 48:195–200. doi: 10.1038/ng.3479
46. Dinis JM, Florek NW, Fatola OO, Moncla LH, Mutschler JP, Charlier OK, et al. Deep sequencing reveals potential antigenic variants at low frequencies in influenza A virus-infected humans. *J Virol.* (2016) 90:3355–65. doi: 10.1128/JVI.03248-15
47. Łuksza M, Lässig M. A predictive fitness model for influenza. *Nature* (2014) 507:57–61. doi: 10.1038/nature13087
48. Koelle K, Rasmussen DA. The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans. *Elife* (2015) 4:e07361. doi: 10.7554/eLife.07361
49. Morris DH, Gostic KM, Pompei S, Bedford T, Łuksza M, Neher RA, et al. Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol.* (2017) 26:102–18. doi: 10.1016/j.tim.2017.09.004
50. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* (2017) 22:30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494
51. Agor JK, Özalp OY. Models for predicting the evolution of influenza to inform vaccine strain selection. *Hum Vaccin Immunother.* (2018) 14:678–83. doi: 10.1080/21645515.2017.1423152
52. Klingens TR, Reimering S, Guzmán CA, McHardy AC. *In silico* vaccine strain prediction for human influenza viruses. *Trends Microbiol.* (2017) 26:119–31. doi: 10.1016/j.tim.2017.09.001
53. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* (2001) 409:860–921. doi: 10.1038/35057062
54. Katzourakis A, Gifford RJ. Endogenous viral elements in animal genomes. *PLoS Genet.* (2010) 6:1001191. doi: 10.1371/journal.pgen.1001191
55. Woolthuis RG, Van Dorp CH, Keşmir C, De Boer RJ, Van Boven M. Long-term adaptation of the influenza A virus by escaping cytotoxic T-cell recognition. *Sci Rep.* (2016) 6:33334. doi: 10.1038/srep33334
56. Fan K, Jia Y, Wang S, Li H, Wu D, Wang G, et al. Role of Itk signaling in the interaction between influenza A virus and T cells. *J Gen Virol.* (2012) 93(Pt 5):987–97. doi: 10.1099/vir.0.041228-0

57. Berkhoff EGM, Boon ACM, Nieuwkoop NJ, Fouchier RAM, Sintnicolaas K, Osterhaus ADME, et al. A mutation in the HLA-B*2705-restricted NP383-391 epitope affects the human influenza A virus-specific cytotoxic T-lymphocyte response *in vitro*. *J Virol*. (2004) 78:5216–22. doi: 10.1128/JVI.78.10.5216-5222.2004
58. Berkhoff EGM, Geelhoed-Mieras MM, Fouchier RAM, Osterhaus ADME, Rimmelzwaan GF. Assessment of the extent of variation in influenza A virus cytotoxic T-lymphocyte epitopes by using virus-specific CD8+ T-cell clones. *J Gen Virol*. (2007) 88:530–5. doi: 10.1099/vir.0.82120-0
59. Boon ACM, de Mutsert G, Graus YMF, Fouchier RAM, Sintnicolaas K, Osterhaus ADME, et al. Sequence variation in a newly identified HLA-B35-restricted epitope in the influenza A virus nucleoprotein associated with escape from cytotoxic T lymphocytes. *J Virol*. (2002) 76:2567–72. doi: 10.1128/Jvi.76.5.2567-2572.2002
60. Smith KD, Andersen-Nissen E, Hayashi F, Strope K, Bergman MA, Rassoulina Barrett SL, et al. Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat Immunol*. (2003) 4:1247–53. doi: 10.1038/ni1011
61. Bailey CC, Zhong G, Huang I-C, Farzan M. IFITM-family proteins: the cell's first line of antiviral defense. *Annu Rev Virol*. (2014) 1:261–83. doi: 10.1146/annurev-virology-031413-085537
62. Everitt AR, Clare S, Pertel T, John SP, Wash RS, Smith SE, et al. IFITM3 restricts the morbidity and mortality associated with influenza. *Nature* (2012) 484:519–23. doi: 10.1038/nature10921
63. Zhang YH, Zhao Y, Li N, Peng YC, Giannoulatou E, Jin RH, et al. Interferon-induced transmembrane protein-3 genetic variant rs12252-C is associated with severe influenza in Chinese individuals. *Nat Commun*. (2013) 4:1418. doi: 10.1038/ncomms2433
64. Zwart MP, Elena SF. Matters of size: genetic bottlenecks in virus infection and their potential impact on evolution. *Annu Rev Virol*. (2015) 2:161–79. doi: 10.1146/annurev-virology-100114-055135
65. Gutiérrez S, Michalakis Y, Blanc S. Virus population bottlenecks during within-host progression and host-to-host transmission. *Curr Opin Virol*. (2012) 2:546–55. doi: 10.1016/j.coviro.2012.08.001
66. Stack JC, Murcia PR, Grenfell BT, Wood JLN, Holmes EC. Inferring the inter-host transmission of influenza A virus using patterns of intra-host genetic variation. *Proc Biol Sci*. (2013) 280:20122173. doi: 10.1098/rspb.2012.2173
67. McCrone JT, Luring AS. Genetic bottlenecks in intraspecies virus transmission. *Curr Opin Virol*. (2018) 28:20–5. doi: 10.1016/j.coviro.2017.10.008
68. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, et al. Influenza a virus transmission bottlenecks are defined by infection route and recipient host. *Cell Host Microbe* (2014) 16:691–700. doi: 10.1016/j.chom.2014.09.020

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Voskarides, Christaki and Nikolopoulos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



APOBEC3G Regulation of the Evolutionary Race Between Adaptive Immunity and Viral Immune Escape Is Deeply Imprinted in the HIV Genome

Faezeh Borzooee, Krista D. Joris, Michael D. Grant and Mani Larijani*

Immunology and Infectious Diseases Program, Division of Biomedical Sciences, Faculty of Medicine, Memorial University of Newfoundland, St. John's, NL, Canada

OPEN ACCESS

Edited by:

Gkikas Magiorkinis,
National and Kapodistrian University
of Athens, Greece

Reviewed by:

Matteo Negroni,
Center for the National Scientific
Research (CNRS), France
Tara Patricia Hurst,
Abcam, United Kingdom
Jean-Christophe Paillart,
Université de Strasbourg, France

*Correspondence:

Mani Larijani
mlarijani@mun.ca

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 15 July 2018

Accepted: 07 December 2018

Published: 11 January 2019

Citation:

Borzooee F, Joris KD, Grant MD and
Larijani M (2019) APOBEC3G
Regulation of the Evolutionary Race
Between Adaptive Immunity and Viral
Immune Escape Is Deeply Imprinted in
the HIV Genome.
Front. Immunol. 9:3032.
doi: 10.3389/fimmu.2018.03032

APOBEC3G (A3G) is a host enzyme that mutates the genomes of retroviruses like HIV. Since A3G is expressed pre-infection, it has classically been considered an agent of innate immunity. We and others previously showed that the impact of A3G-induced mutations on the HIV genome extends to adaptive immunity also, by generating cytotoxic T cell (CTL) escape mutations. Accordingly, HIV genomic sequences encoding CTL epitopes often contain A3G-mutable “hotspot” sequence motifs, presumably to channel A3G action toward CTL escape. Here, we studied the depths and consequences of this apparent viral genome co-evolution with A3G. We identified all potential CTL epitopes in Gag, Pol, Env, and Nef restricted to several HLA class I alleles. We simulated A3G-induced mutations within CTL epitope-encoding sequences, and flanking regions. From the immune recognition perspective, we analyzed how A3G-driven mutations are predicted to impact CTL-epitope generation through modulating proteasomal processing and HLA class I binding. We found that A3G mutations were most often predicted to result in diminishing/abolishing HLA-binding affinity of peptide epitopes. From the viral genome evolution perspective, we evaluated enrichment of A3G hotspots at sequences encoding CTL epitopes and included control sequences in which the HIV genome was randomly shuffled. We found that sequences encoding immunogenic epitopes exhibited a selective enrichment of A3G hotspots, which were strongly biased to translate to non-synonymous amino acid substitutions. When superimposed on the known mutational gradient across the entire length of the HIV genome, we observed a gradient of A3G hotspot enrichment, and an HLA-specific pattern of the potential of A3G hotspots to lead to CTL escape mutations. These data illuminate the depths and extent of the co-evolution of the viral genome to subvert the host mutator A3G.

Keywords: CTL epitope, APOBEC3G (A3G), HIV, immune escape, viral evolution

INTRODUCTION

HIV, like other RNA viruses, evolves rapidly and continuously through the accumulation of mutations (1). The high rate of HIV genome mutation, between 10^{-4} and 10^{-5} mutations per nucleotide per replication cycle, is generated by HIV's error-prone reverse transcriptase (RT) (2–7). APOBEC3G (A3G) is a member of the apolipoprotein B mRNA-editing enzyme catalytic

polypeptide-like editing complex (APOBEC) family of cytidine deaminase enzymes. Malim and colleagues, in 2002, discovered that A3G is responsible for the prevalence of G to A mutations in HIV sequences from HIV infected individuals (8). The APOBEC family includes 11 members in humans: activation-induced cytidine deaminase (AID), APOBEC1, APOBEC2, APOBEC3A-H, and APOBEC4, which, through their cytidine deaminase activity, are involved in diverse physiological processes including lipid metabolism, antibody diversification, virus/retroelement restriction, and cancer genome hypermutation (9–13).

In general, the A3 branch family members are capable of impeding infectivity of HIV and several other viral infections such as hepatitis B, human T cell leukemia virus type 1, and human papillomavirus (14–18) though A3G is the most effective actor on the HIV genome (19). These enzymes exert their antiviral restriction activity by deamination of cytidine to uridine (C to U) in the minus-strand single-stranded DNA during reverse transcription of viral genomic RNA which mediates guanosine to adenosine (G to A) mutation in plus-strand DNA (20–26). A3G is constitutively expressed in resting CD4⁺ T cells, macrophages, and dendritic cells, but can be further induced by interferon (IFN) (27–29). It is packaged into the HIV virion in A3G-expressing producer cells and can act on the viral genome in the subsequently-infected cell (20). It has been shown by single-virion analysis that A3G can be co-packed with A3F, A3D, or A3H haplotype II and co-mutate the same viral genome in a single cycle of HIV replication (30).

A3G has a sequence preference for mutating C in CCC, TCC, and ACC motifs, but this preference is further modulated by the DNA secondary structure, together mediating accumulation of G to A mutations in viral cDNA (18, 21, 22, 24, 30–34). A3G-mediated mutations on the HIV genome also follow a twin gradient pattern, as determined by the central and end polypurine tracts (PPT)'s impact on the reverse transcription dynamics of the HIV genome altering the time that various regions are left single-stranded and available for A3G to act on (34, 35). Depending on the load and positions of A3G-induced mutations, this could lead to either degradation or G to A hypermutation in the viral genome (23, 25, 26). One of the early observations of the contribution of A3G to producing defective viral proteins was its ability to create premature stop codons (36, 37). For example, the codon encoding tryptophan (TGG) is converted by A3G to a stop codon (TGA) (37, 38). A3G is also able to physically interfere with HIV replication in a deamination-independent manner by blocking reverse transcription or binding to tRNA to prevent reverse transcription initiation (39–41). However, it appears that viral restriction has a higher requirement for deamination-dependent A3G activity (42–45).

Over the last decade and a half, several hundred published studies have focused on A3G's HIV restriction activity. A3G can induce as many as five mutations per kilobase (22, 26, 32, 38), at least an order of magnitude higher than RT's error rate. It has been suggested that HIV's RT is only responsible for 2% of HIV genome mutagenesis and the other 98% can be attributed to the action of A3G (46). While high A3G activity correlates with slower disease progression, lower A3G activity leading to

sub-lethal mutations might enhance HIV diversity and lead to more rapid disease progression (47). In contrast, other studies showed a much lower contribution of A3G to genetic variation of HIV, as compared with RT-driven mutations. One study reported negligible sub-lethal mutation frequencies as low as 4×10^{-21} and 1×10^{-11} for A3G and A3F mutations, respectively, which is significantly lower than the frequency of mutations arising from RT (39). Most reports suggest that HIV can experience both a beneficial and a harmful influence from A3G expression (47–50). Other studies reported that A3G is less likely to impose HIV diversification and facilitate viral diversification and adaptation *in vivo*, and that A3G, even at low expression levels, is lethal for HIV (36, 51, 52).

Most studies reporting the impressively high load of A3G mutation were carried out using Vif-deficient HIV because A3G is antagonized by the HIV encoded accessory protein Vif (53–56). Although Vif is necessary for HIV replication in A3G-expressing cells, it is not required in A3G-deficient cells (8, 53–58). Vif binding mediates proteasomal degradation of A3G, but it can also downregulate the translation of A3G (59–61). It has been reported that the accessory protein Vpr can also bind A3G and mediate its proteasomal degradation (62). Thus, in the presence of HIV's full complement of accessory anti-A3G factors, only low levels of mutations are induced by A3G (47, 49, 59, 63, 64).

HIV specific CD8⁺ cytotoxic T cell (CTL) responses and their human leukocyte antigen (HLA) restriction are crucial determinants of viral containment following the initial innate immune response (65, 66). Multiple parameters such as HLA genotype, virus sequence, and T cell receptor repertoire contribute to CTL response effectiveness (67–72). Despite the significant protective role of CTLs in limiting viral HIV replication, the immune system ultimately fails to clear HIV, at least in part because of mutations within or adjacent to CTL epitope during the early and chronic phases of disease progression (73). HIV is under intense evolutionary pressure for escape mutations that lead to evasion of CTL killing and CTL escape mutations are a major force in driving viral evolution in acute/early chronic infection (74, 75). These mutations can be located either inside or outside CTL epitopes, be fast or slow in appearance, but to be selected, they ought to maintain a balance with a cost for viral replication fitness and escape from CTL (76–80). Mutations that facilitate immune evasion are positively selected and become dominant in the viral population (73, 81, 82). On the other hand, sometimes amino acid alterations under immune pressure can even confer a *de novo* immune response (83).

CTL escape mutations can act through several mechanisms: by reducing or abrogating binding of viral epitopes to HLA Class I, disrupting intracellular epitope processing or altering recognition by T cell receptors (84–89). Viral proteins in infected cells are first proteolytically degraded in the cytosol by immunoproteasomes (86, 90, 91). Proteasomal degradation product peptides including epitope precursors can be up to 32 amino acids long; however, immunoproteasomes are inclined to generate longer peptides ending with C-terminal hydrophobic residues that are anchors for most HLA class I molecules. After post-proteasomal degradation, epitope precursors typically 8–16

amino acids in length, are transferred into the endoplasmic reticulum (ER) lumen where HLA class I molecules are folded and assembled, by the transporter-associated-with-antigen-processing (TAP1 and TAP2) (92). Further N-terminal trimming in the ER can occur by enzymes such as the ER aminopeptidases (ERAP1 and 2) to fit the groove of restricting HLA class I molecules (93, 94). The peptide-HLA complex is subsequently transported to the cell surface to be recognized by CTLs. Thus, proteasomal degradation and antigen processing are key determinants of epitope availability for the anti-HIV CTL response (95). In addition to mutations in epitopes, immune escape can also be altered through mutation of the flanking regions that impact proteasomal processing (96). It has been shown that the robustness of the anti-HIV CTL response correlates with the number of epitopes generated due to proteasomal cleavage, and thus, mutations that impact processing can drastically influence CTL escape (77–79, 89, 97–100).

The most restrictive step in antigen processing is the peptide's ability to bind to the specific expressed set of HLA class I molecules using N- and C-terminal anchor residues that bind into the groove of a specific HLA class I molecule (101). Mutations at anchor residues can disrupt HLA class I binding, whereas other mutations such as those at the central bulge of the peptide (normally residues, 4–6 in the canonical 9-mer peptide) interfere with TCR recognition of the HLA-peptide complex (73, 84, 96, 102). It remains to be fully understood how the host's condition shapes the availability of beneficial mutations; however, HLA profile is a major CTL escape driver (85, 103–105). HLA-B is the most protective among all three HLA class I loci (A, B, and C) (106). CTL escape and reversion pathways are more closely associated with epitopes restricted to protective HLA alleles such as HLA-B27 and -B57, which are associated with slower disease progression and more robust HIV-specific CTL responses (78, 107). Escape mutations in these HLA restricted epitopes incur a high fitness cost by reducing viral replicative capacity (108). In contrast, HLA-B35 and -B8 are associated with rapid disease progression based on their ability to present less effective epitopes to TCRs (109–113).

Mutational diversity of HIV's genome has a crucial role in evasion of immune recognition and multiple studies have implicated A3G as an important player in the interplay between the adaptive immune anti-viral CTL response viral adaptation and immune escape (114–119). A3G has been proposed to induce CTL escape in two ways, either by directly mutating CTL epitopes or by causing mutations outside epitopes which influence the peptide degradation and HLA presentation of wild-type CTL epitopes (117, 119–121). Several reports describe A3G-induced mutations located within or flanking CTL epitopes. One report found remarkable evidence for enrichment of non-synonymous amino acid substitutions by A3G in the anchor or proximal amino acid residues of HLA-restricted epitopes that are important in epitope processing leading to immune escape (115). Consistent with a role for A3G in CTL escape, an earlier bioinformatics study reported a reduction in CTL recognition as a result of A3G mutation in epitopes (119). On the other hand, it was demonstrated that increasing the turnover of truncated HLA-restricted peptides, generated due to the action of A3G, can

enhance the CTL response in a mouse model of CTL responses to HIV (120). We previously measured CTL recognition of wild-type or A3G-mutated epitopes *ex vivo* by CTL from HIV-infected individuals. We considered a limited subset of CTL epitopes known to elicit CTL recognition, and we focused on A3G-induced mutations in epitope residue positions 3, 5, and 7, which would mainly impact TCR recognition. We found that in the vast majority of instances, A3G-induced mutations in CTL epitopes abrogated CTL recognition of epitopes in an HLA-dependent manner (114, 117, 118). Moreover, we showed that A3G mutational hotspots are enriched in the viral genomic sequences encoding immunogenic CTL epitopes in Gag, Pol, and Nef (117). This is in agreement with the earlier study that also found enriched A3G hotspot motifs within the rapidly diversifying CTL escape sites in Env (119). Interestingly, and in contrast to our findings on A3G hotspot motif enrichment in CTL-encoding epitopes, another study reported that A3G hotspots are less frequently located at genomic locations encoding for the V1-V5 region, the most variable regions of the gp120 envelope protein, in order to hold in reserve its potential mutational capacity for long-term adaption of HIV to the antibody response (122).

Most of our knowledge about epitope-specific CTL responses in chronic HIV infection comes from studies using the standard IFN- γ ELISpot assay (123). Besides experimental approaches, a variety of computational tools for prediction of mutations and their impact have provided valuable information. Different sets of algorithms such as artificial neural networks (ANN), average relative binding (ARB), stabilizing matrix method (SMM) and others have enabled prediction of CTL epitopes within the viral proteome based on HLA-binding affinity (124, 125). The web algorithm HLA binding predictors have a broad allelic coverage with as much as 90–95% accuracy (124–128). HLA binding and subsequent recognition by TCR are the most selective steps in the peptide presentation pathway (129). However, other processes upstream of HLA binding such as proteasomal cleavage, TAP transporter and the stability of the peptide-HLA complex also shape viral epitope availability (130). Prediction tools, such as NetChop and PaProC have been developed based on protein degradation by purified proteasomes to predict potential cleavage sites (131–134). The reliability of these tools has been shown (135–138).

Thus, mutations that impact protein proteasomal processing and/or epitopes' HLA binding can lead to loss of CTL recognition and immune escape (88, 89, 96, 139, 140), but the extent to which A3G mutations could potentially impact each of the successive stages of CTL epitope generation and presentation is not known. Here we utilized the aforementioned computational tools to construct a comprehensive CTL epitope map of HIV based on the steps of antigen presentation: proteasomal cleavage, TAP transporter efficiency and HLA-binding affinity. We simulated all possible A3G-mediated mutations within and outside CTL epitope-encoding sequences of the HIV genome. We then examined predicted consequences for CTL epitope generation. We also probed whether the positions and predicted consequences of A3G-mediated mutations are random, or rather indicative of co-evolution of the HIV genome with the action of the host mutator. In cases where such co-evolution was observed,

we studied the depth and extent of the strategies used by the HIV genome to influence the outcomes of A3G activity. Since our experimental system is devoid of the immense *in vivo* selection pressure for CTL escape, and hence able to predict enhanced generation of CTL epitopes as well as the opposite scenario of immune evasion without bias, the analysis provides a unique lens for considering how viral genomes co-evolve with host restriction factors.

MATERIALS AND METHODS

A3G-Induced Mutation Simulation

Simulation of A3G-induced mutations was carried out as previously described (117). Briefly, the whole genome of the HIV-1 isolate HXB2 BRU was obtained from NCBI. This sequence was chosen since it was used in previous works and model systems that studied the role of A3G on HIV CTL escape (117, 119). A3G-induced mutations (G-to-A) on the 5'-most dG in A3G's trinucleotide hotspot motifs considering the sense of the +ve sense strand (GGG, GGA, and GGT) were manually simulated and translated to amino acid sequence. For this analysis, we considered first-round mutations. For multiple back-to-back A3G hotspots, all possible amino acid alteration consequences of A3G-induced mutations were considered.

Prediction of CTL Epitopes, HLA Binding, and Proteasomal Cleavage

To generate a comprehensive list of all potential CTL epitopes of HIV, we considered all HIV peptides that are predicted to be efficiently processed by proteasomes and also bind to HLA class I molecules. We identified the portions of the HIV genome encoding known CTL epitopes using the HIV Molecular Immunology Database (http://www.hiv.lanl.gov/content/immunology/tables/ctl_summary.html). We evaluated the predicted MHC binding affinity of wild-type and A3G-mutated variant CTL epitope sequences of Gag, Pol, Env, and Nef restricted to HLA-A02:01, -A03:01, -B57:01, and -B35:01. We utilized epitope prediction algorithms enabling us to investigate the impact of mutations at A3G hotspots within or in flanking regions of the predicted and known epitopes on HLA affinity binding, and epitope processing. Here we used NetMHCpan 4 (<http://www.cbs.dtu.dk/services/NetMHCpan/>) using artificial networks (ANN) to construct a fine CTL epitope map based on HLA-I binding. The NetMHCpan 4 server predicts binding of peptides to any HLA molecule of a known sequence using ANNs (127, 136, 141). Then the Immune Epitope Databases (IEDB) server (<http://tools.iedb.org/processing/>) was applied for further prediction based on proteasomal cleavage, TAP transporter efficiency and HLA binding affinity to improve the selection of potential epitopes. To evaluate the impact of A3G alterations on HLA-binding, we only considered predicted epitopes with a high-rank score between 0 and 0.5 percentile as strong HLA binders and 0.5–2.00 percentile as weak HLA binders (136). We then calculated a Delta from the wild-type sequence HLA binding score to evaluate the change in predicted HLA affinity caused by A3G-induced mutation. We set a Delta of 0.1 as a threshold of significant difference for enhanced or diminished

HLA binding affinity, based on distribution analysis of the difference values. We used NetChop 3.1 (<http://www.cbs.dtu.dk/services/NetChop/>) to display the impact of the mutation on proteasomal cleavage. The program C-term 3.0 network is trained with a database consisting of 1260 publicly available HLA class I ligands (using only the C-terminal cleavage site of the ligands). The highest probabilities of cleavage (threshold set at 0.5) were applied based on default program recommendation (142). To predict proteasomal cleavage sites, wild-type and A3G-induced mutated polypeptide of Gag, Pol, Env, and Nef were submitted to NetChop 4.

Analysis of A3G Hotspot Frequency in Sequences Encoding CTL Epitopes at the Nucleotide Level and Prediction of Amino Acid Alteration Consequences

To investigate the enrichment of A3G hotspots of CTL epitopes Gag, Pol, Env, and Nef restricted to HLA-A2:01, HLA-A3:01, HLA-B57:01, and HLA-B35:01, we counted the number of A3G hotspot motifs (GGA, GGG, and GGT) in CTL epitope-encoding regions vs. non-CTL epitope-encoding sequences. We normalized for gene size by dividing the frequency of hotspots by the total number of analyzed nucleotides in each gene. The normalized hotspot frequencies at the nucleotide levels were calculated for sequences encoding CTL epitopes restricted to each individual HLA, and non-CTL epitope encoding sequences. Then, the ratio of hotspot frequency was determined for inside to outside epitope-encoding sequences, for each A3G hotspot motif and each restricting HLA. As controls to evaluate potential A3G sequence enrichment in sequences encoding CTL epitopes, we conducted a parallel analysis with randomly shuffled HIV genomic sequence using the “Shuffle DNA” function of the Sequence Manipulation Suite (http://www.bioinformatics.org/sms2/shuffle_dna.html) resource. The HIV sequence was randomly shuffled six independent times, and A3G hotspot enrichment analysis was performed for each hotspot motif and restricting HLA, using the same border locations of sequences encoding CTL epitope sequences in the actual HIV genomic sequence. MATLAB was used to describe the distribution of amino acid alterations in epitope-surrounding regions, considering a 32-amino acid boundary around each 8–11 mer epitope, with a limit of either 2 or 4 A3G-induced amino acid changes on either the N- or C-terminal sides of each CTL epitope within this boundary considered to be a clustered pattern. Graphpad Prism 5 was used to generate the schematic graph to display the distribution of A3G-induced mutation within and in the flanking regions of CTL epitopes Gag, Pol, Env, and Nef.

For analysis of amino acid alteration consequences of A3G enrichment in CTL-encoding sequences, affected amino acids (as a result of non-synonymous A3G-mediated mutations), or non-affected amino acids (as a result of silent A3G-mediated mutations) were determined, and frequencies of amino acid consequences inside or outside epitope-encoding sequences were normalized to the total amino acid number of regions of each polypeptide (Gag, Pol, Env, and Nef) present in epitope or non-epitope regions, for each restricting HLA. Then, the ratio

of normalized non-synonymous to silent substitutions in CTL epitope encoding region to non-CTL epitope encoding region was calculated for A3G-induced mutations of Gag, Pol, Env, and Nef, restricted to each HLA. The ratio of non-synonymous and silent A3G-mediated amino acid substitutions in the CTL epitope encoding region to non-CTL epitope encoding region was calculated in two ways for each individual A3G hotspot motif. First, the ratio of A3G-induced mutations resulting in non-synonymous residue changes to silent mutations inside CTL epitopes was divided by the same ratio determined for regions of each polypeptide located outside CTL epitopes. Second, the frequency of A3G-mediated non-synonymous mutations inside CTL epitopes was divided by the frequency of A3G-mediated non-synonymous mutations outside CTL epitopes, ignoring silent consequences.

To analyze the distribution pattern of A3G hotspot positioning in the HIV genome, we divided the entire length of the genome into 60 bp stretches and counted the number of A3G hotspots whose mutation would result in non-synonymous amino acid changes or stop codon generation. This pattern was plotted and compared against the known mutational gradient of the HIV genome as previously described (35, 143). In the same manner, to analyze the distribution pattern of A3G hotspots whose mutation is predicted to result in CTL escape, we considered, for each HLA, A3G hotspots that fall in CTL epitope-encoding sequences and whose mutation caused a predicted decrease in HLA binding affinity, as described above. For each HLA, we also generated a map of the positions of CTL epitope-encoding sequences across the entire HIV genome. We derived a map of normalized escape potential which we defined as the number of CTL-escape inducing A3G hotspots in each 60 bp segment, normalized (divided) by the total number of A3G hotspots within the segment. Based on the normalized escape potential and the number of CTL epitopes encoded in each 60 bp segment, for each HLA, we derived a map of escape factor which we defined as the product of the number of CTL epitopes and normalized escape potential. Thus, the escape factor value represents the potential for A3G-induced mutations to generate CTL-escape in any given 60 bp increment of the HIV genome.

RESULTS

Potential Wild-Type and A3G-Mutated CTL Epitopes in Gag, Pol, Env, and Nef

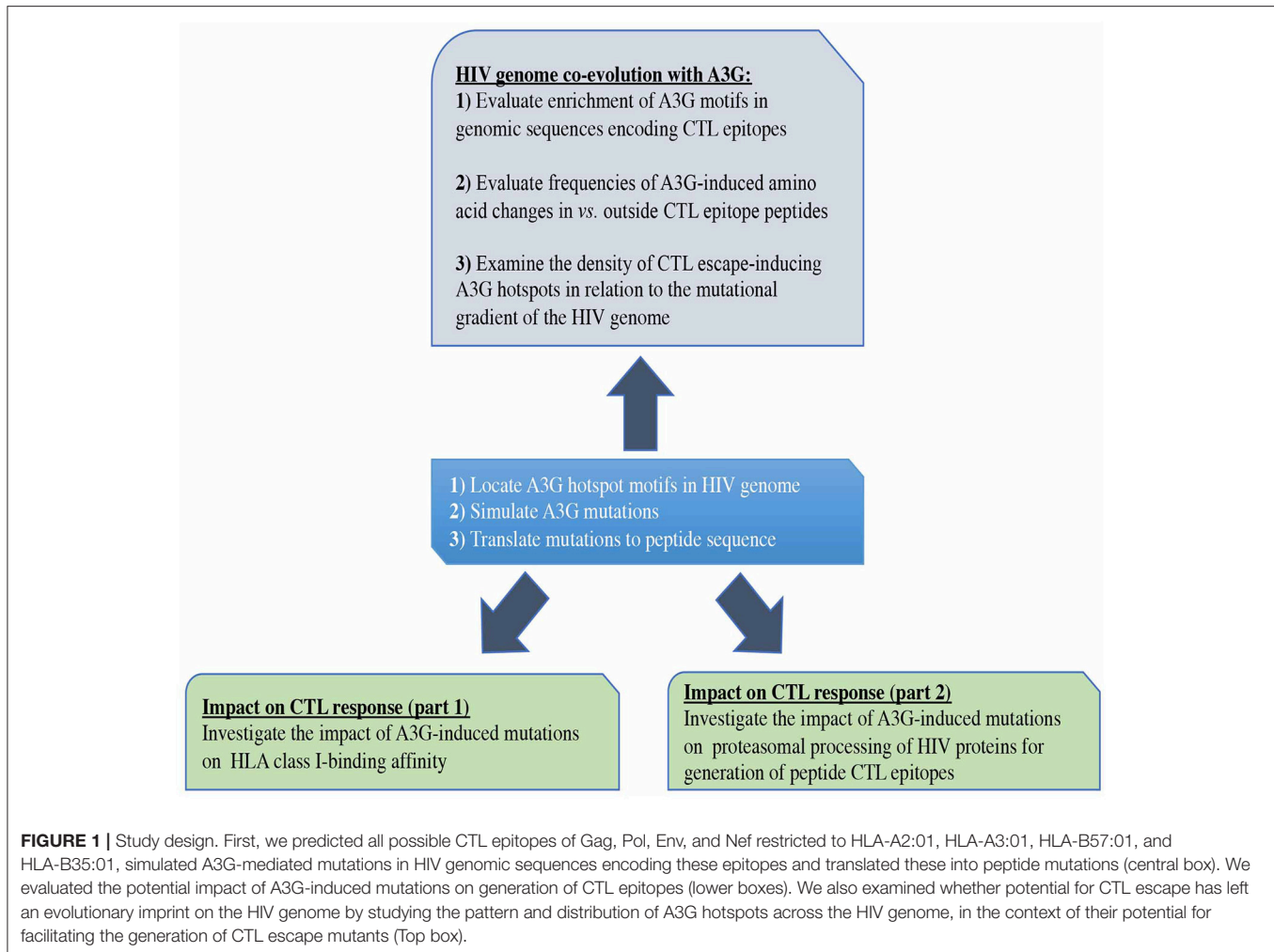
We utilized NetMHCpan 4 and IEDB using ANN to construct a CTL epitope map based on proteasomal cleavage, TAP transporter efficiency and HLA-binding affinity. Using entire peptide sequences Gag, Pol, Env, and Nef, we generated a list of all potential HLA-binding peptide epitopes. Although the binding affinity data covers 172 HLA molecules (136), we restricted our analysis to HLA-A02:01, -A03:01, -B57:01, and -B35:01 because HLA-A02:01 and A03:01 are frequent in the population, and HLA-B57:01 and -35:01 correlate with robust and weak HIV-specific CTL responses, respectively. Based on the potential definition of all possible epitopes through the presence of HLA-binding anchor residues in the HIV proteome, the number of

all putative potential epitopes is significantly higher than those that actually elicit CTL responses, due to limitation in either epitope processing, presentation to TCR, and the many complex physiological and immune response dynamics that underlie the CTL response that cannot be accounted for by epitope prediction algorithms (144). Nevertheless, we noted that the set of epitopes that we generated on the bases of HLA binding and proteasomal processing predictions included the majority (~70%) of experimentally-verified CTL epitopes listed in the HIV Molecular Immunology Database (Table S1). In addition to the predicted set of epitopes, we also included in our analyses experimentally-known CTL epitopes.

Thus, we generated an epitope list which includes all potential CTL epitopes restricted by HLA-A2:01, -A3:01, -B57:01, and -B35:01. In total, for Gag, Pol, Env, and Nef, we identified 14-12-14-10, 19-33-26-21, 22-14-20-8, and 8-6-9-8 epitopes restricted to HLA-A2:01, -A3:01, -B57:01, and -B35:01, respectively (Table 1 and Table S1). To dissect the role of A3G at each step of CTL epitope generation, we simulated A3G mutations in sequences encoding Gag, Pol, Env, and Nef. A3G mutations (G-to-A) on the 5'-most dG in A3G hotspot trinucleotide motifs (GGG, GGA, and GGT) were simulated and translated to the peptide sequence (Figure 1). We found 16-16-21-13 (Pol), 8-10-11-7 (Gag), 5-2-6-5 (Nef), and 13-8-15-3 (Env) restricted to HLA-A2:01, -A3:01, -B57:01, and -B35:01, respectively, whose encoding sequences contain A3G hotspots. After simulation of A3G-induced mutations at these motifs, we identified 33-33-44-20 possible mutated epitopes restricted to HLA-A2:01, -A3:01, -B57:01, and -B35:01 for Pol, of which 25 alterations include stop codons (Table 1 and Table S1). These numbers were 8-20-22-8 and 12 stop codons for Gag, 11-4-14-15 and 9 stop codons for Nef, 28-20-31-6, and 29 stop codons for Env. These results indicate that A3G-induced mutations can potentially alter CTL epitopes restricted to all four examined HLAs in Gag, Pol, Env, and Nef (Table 1 and Table S1). Considering all predicted wild-type epitopes, Gag, Pol, Env, and Nef contained 21, 40, 26, and 12% of the predicted CTL epitopes respectively. 25, 30, 26, and 19 of all CTL epitopes were restricted to HLA-A2:01, -A3:01, -B57:01, and -B35:01, respectively. After simulating A3G-induced mutations, 22, 42, 25, and 11% of all mutated epitopes came from Gag, Pol, Env, and Nef and 26, 23, 33, 18% of all epitopes were restricted to HLA-A2:01, -A3:01, -B57:01, and -B35:01, respectively. This initial comparison between the distribution of wild-type vs. A3G-mutated CTL epitopes suggested a bias for A3G-mediated mutations in HLA-B57-restricted epitopes, consistent with previous suggestions for a role of A3G in mediating decreased CTL recognition for peptides restricted to protective HLAs such as B57 (117, 118).

The Potential Impact of A3G-Induced Mutations on HLA Binding Affinity

To examine the specific impacts of A3G-induced mutations on HLA-binding, we considered predicted epitopes with a high-rank NetMHCpan 4 score between 0 and 0.5 percentile as strong HLA binders and 0.5–2.0 percentile as weak HLA binders, according to default parameters of the prediction algorithm



(136) (**Figure 1**, **Table S1**). However, we noted that 30% of experimentally-verified epitopes exhibit out of range and low HLA binding affinity scores; hence, their aforementioned absence in the total predicted pool of CTL epitopes (**Table S1**). To evaluate the change in predicted HLA affinity that occurred as a result of each A3G-induced mutation, we calculated a Delta value from the wild-type sequence HLA binding score. We set 0.1 as a threshold of difference for increased or reduced HLA binding affinity because below this value poor correlation was observed between the predicted HLA affinity rank and absolute nM affinities.

Although A3G-induced stop codons would not lead to infectious virus production, viral genomes containing stop codons can produce immunogenic truncated peptides which contain CTL epitopes (120). Thus, we considered all A3G-mutations, including stop codon generators (**Figure 2A**, top panel), or excluding stop codons (**Figure 2A**, bottom panel). Considering all A3G-induced mutations, 25, 46, 32, and 14% of HLA-A2:01-, -A3:01-, -B57:01-, and -B35:01-restricted epitopes exhibited increased HLA-binding affinity after A3G simulation mutation. Conversely, 75, 54, 68, and 86% of HLA-A2:01-, -A3:01-, -B57:01-, and -B35:01-restricted epitopes exhibited

decreased HLA-binding affinities as a result of A3G-induced mutations (**Figure 2A**, top panel). Excluding A3G-mediated stop codons, 32, 51, 62, and 24% of HLA-A2:01-, -A3:01-, -B57:01-, and -B35:01-restricted epitopes increased HLA-binding affinity after A3G simulation mutation, whilst 68, 49, 38, and 76% of A3G-induced mutations HLA-A2:01-, -A3:01-, -B57:01-, and -B35:01-restricted epitopes led to decreased HLA-binding affinities (**Figure 2A**, bottom panel).

Next, we sought to break down the effect of A3G-induced mutations at epitope-HLA anchor vs. non-anchor residues. A3G-induced mutations in non-anchor residues were predicted to lead to enhanced, diminished, or abolished HLA binding affinity for epitopes restricted to all 4 HLAs, with diminished/abolished HLA binding being the most common predicted outcome for HLA-A2:01-, -A3:01-, - and -B35:01-restricted epitopes (74, 53, 80% of all A3G mutations, respectively); for HLA-B57:01 however, that outcome was nearly equal to potential for enhanced HLA affinity (35%) (**Figure 2B**). Escape mutations that reduce class I HLA binding commonly occur at the N-terminal (amino acid position 2 in the peptide epitope) and/or the C terminal (e.g., amino acid position 9 in a 9mer peptide epitope) anchor residues in most epitopes (84, 96, 145). Thus, we then considered

TABLE 1 | Summary of the number of potential CTL epitopes restricted to HLA-A2:01, HLA-A3:01, HLA-B57:01, and HLA-B35:01 for wild-type and A3G-mediated mutated Gag, Pol, Env, and Nef proteins.

Hotspot	Wild type	Epitope containing hotspot	Altered	Altered containing stop codon
NUMBER OF HLA-A2:01, -A3:01, -B57:01, and -B35:01-RESTRICTED EPITOPES IN Gag				
Gag				
HLA2:01				
	14	8		
GGG			1	0
GGA			6	2
GGT			1	1
Total			8	3
Gag				
HLA-A3:01				
	12	10		
GGG			10	1
GGA			7	0
GGT			3	0
Total			20	1
Gag				
HLA-B57:01				
	14	11		
GGG			6	0
GGA			15	6
GGT			1	1
Total			22	7
Gag				
HLA-B35:01				
	10	7		
GGG			3	1
GGA			3	0
GGT			2	0
Total			8	1
NUMBER OF HLA-A2:01, -A3:01, -B57:01, and -B35:01-RESTRICTED EPITOPES IN Pol				
Pol				
HLA2:01				
	19	16		
GGG			10	1
GGA			15	4
GGT			4	0
Total			33	5
Pol				
HLA-A3:01				
	33	16		
GGG			16	3
GGA			15	0
GGT			2	0
Total			33	3
Pol				
HLA-B57:01				
	26	21		
GGG			16	4
GGA			22	10
GGT			6	1
Total			44	15
Total			20	2

(Continued)

TABLE 1 | Continued

Hotspot	Wild type	Epitope containing hotspot	Altered	Altered containing stop codon
Pol				
HLA-B35:01				
	21	13		
GGG			4	0
GGA			11	1
GGT			5	1
NUMBER OF HLA-A2:01, -A3:01, -B57:01, and -B35:01-RESTRICTED EPITOPES IN Env				
Env				
HLA2:01				
	22	13		
GGG			7	0
GGA			17	3
GGT			4	1
Total			28	4
Env				
HLA-A3:01				
	14	8		
GGG			5	0
GGA			13	3
GGT			2	0
Total			20	3
Env				
HLA-B57:01				
	20	15		
GGG			12	4
GGA			14	12
GGT			5	3
Total			31	19
Env				
HLA-B35:01				
	8	3		
GGG			0	0
GGA			6	3
GGT			0	0
Total			6	3
NUMBER OF HLA-A2:01, -A3:01, -B57:01, and -B35:01 RESTRICTED EPITOPES IN Nef				
Nef				
HLA2:01				
	8	5		
GGG			5	0
GGA			6	2
GGT			0	1
Total			11	3
Nef				
HLA-A3:01				
	6	2		
GGG			2	0
GGA			2	0
GGT			0	0
Total			4	0

(Continued)

TABLE 1 | Continued

Hotspot	Wild type	Epitope containing hotspot	Altered	Altered containing stop codon
Nef				
HLA-B57:01				
	9	6		
GGG			7	0
GGA			4	2
GGT			3	2
Total			14	4
Nef				
HLA-B35:01				
	8	5		
GGG			9	0
GGA			6	1
GGT			0	1
Total			15	2

A3G-induced mutations that target the anchor residues. A3G-induced mutations in anchor residues were predicted to lead to enhanced, diminished or abolished HLA binding affinity for epitopes restricted to all 4 HLAs, with diminished/abolished HLA binding being the most common predicted outcome for -B57:01-, - and -B35:01-restricted epitopes (33, 7% of all A3G mutations, respectively). In contrast, for HLA-A2:01- and -A3:01, A3G-induced mutations at anchor residues mediated diminished/abolished HLA binding affinity with the same or much lower levels than enhanced HLA binding affinities (2 and 1%, respectively) (**Figure 2C**).

Considering all epitopes that were within the aforementioned HLA-binding threshold range, 2 and 23% of A3G-induced amino acid changes targeted the N and C-terminal anchor positions, respectively. Twenty-three percent is approximately two-fold higher than expected by random chance (84, 146): if A3G-induced mutations were equally distributed amongst all residues in a pool of 8–11 mer peptides, then each residue ought to have a ~10% probability of being targeted. In total, A3G can potentially generate a stop codon in 30% of all epitopes that contain A3G hotspots. Of these, 43% led to a stop codon at the most C-terminal position, of which the overwhelming majority (92%) were restricted to HLA-B57:01, since tryptophan (the TGG codon) is the C-terminal anchor for HLA-B57:01 (**Figure 2A**, compare top and bottom panels). This codon, which has a high likelihood of containing an A3G hotspot motif dependent on the next downstream nucleotide, is the most susceptible codon for generating a stop due to G-to-A mutation (37, 38). Also, we observed 30% of all potential A3G-induced mutations in CTL epitopes were located at residue positions 3, 5, and 7 which are key for TCR recognition (147), while 40% of A3G-mediated substitutions targeted residue positions 1, 4, 6, 8, 9, and 10. These results indicate that the N- and C-terminal anchor residues are under- and over-targeted by A3G for mutation, whilst the middle positions are apparently equally targeted. Furthermore, the increased A3G targeting of the most C-terminal anchor residue reflects its overwhelming

propensity for stop codon generation in HLA-B57:01-restricted epitopes (**Figure 2A**). Based on these analyses, we conclude that A3G-induced mutations can increase or decrease HLA-binding affinities of potential CTL epitopes; however, the major outcome considering all mutations (non-synonymous amino acid changes and stop codons) at all residues (anchor and non-anchor) was decreased HLA-binding affinity. These results are consistent with previous observations that the CTL-epitope-encoding sequences of HIV have evolved to channel A3G-induced mutations to mediate CTL escape. On the other hand, we also observed the generation of 18 neo-epitopes based on enhanced HLA-binding affinity.

The Role of A3G-Induced Mutations on Proteasomal Processing of Epitopes

We utilized NetChop 3.1 to examine the impact of A3G-induced mutations on the proteasomal processing of HIV proteins, which is the step before peptide epitope generation for HLA binding. We submitted the entire sequences of Gag, Pol, Env, and Nef to NetChop, either in wild-type format, or including all possible A3G-induced mutations. On the entire peptide sequence, we overlaid the map of HLA-A2:01, -A3:01-, -B57:01, and -B35:01-restricted CTL epitopes. Each residue within wild-type or A3G-mutated Gag, Pol, Env, and Nef proteins was then assigned a cleavage prediction score (default threshold of 0.5 is considered a cleavage site) (142), and scores at each residue position were compared between wild-type and A3G-mutated proteins (**Supplementary File 1**: excel table). For this analysis, we considered two categories of A3G-induced mutations: those that fell within individual CTL epitopes, or those that fell outside but within six amino acids upstream or downstream of the N- or C- terminal residues of the epitope (148). A3G-induced mutations that generated new/enhanced cleavage sites within a CTL epitope, or abolished/decreased cleavage within the six amino acids adjacent to the epitope would likely lead to diminished proteasomal processing of the epitope. Conversely, A3G mutations that enhanced cleavage in the adjacent region of an epitope or abolished/decreased cleavage within the epitope itself would likely lead to enhanced proteasomal processing of the epitope.

In this manner, we quantified the impact of A3G-induced mutations on CTL epitope proteasomal cleavage (**Supplementary File 1, Table 2**). Considering all A3G-mediated mutations, for epitopes restricted to HLA-A2:01, -A3:01-, -B57:01, and -B35:01, respectively, 42, 53, 43, 45% of all A3G-induced mutation events resulted in decreased predicted proteasomal processing, whilst 58, 47, 57, and 55% of mutations resulted in generation of sites predicted to enhance proteasomal processing. If epitopes were categorized by protein of origin rather than restricting HLA, for Gag, Pol, Env, and Nef, respectively, 54, 43, 41, and 53% of all A3G-induced mutations resulted in decreased predicted proteasomal processing, whilst 46, 57, 59, and 47% resulted in enhanced predicted processing. Excluding A3G-mediated stop codon generation, for epitopes restricted to HLA-A2:01, -A3:01-, -B57:01, and -B35:01, respectively, 43, 56, 53, and 43% of all A3G-induced mutations

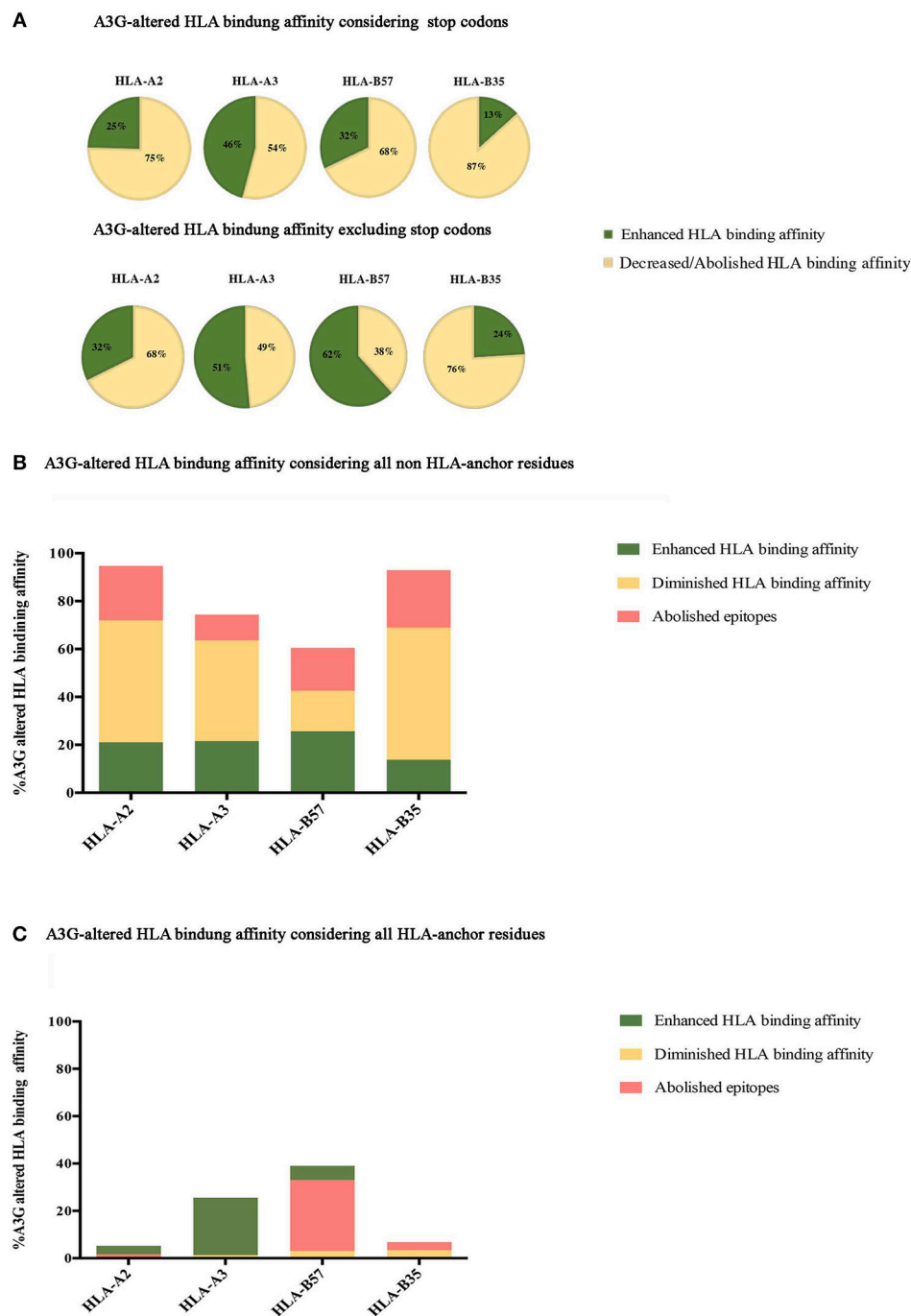


FIGURE 2 | Impact of A3G-induced mutations on HLA-binding. The impact of A3G-mediated mutations on HLA-binding affinity was predicted using NetMHCpan 4 to examine epitopes restricted to HLA-A2:01, HLA-A3:01, HLA-B57:01, and HLA-B35:01 for Gag, Pol, Env, and Nef proteins. Epitopes within the range of strong and weak HLA-binding affinity were considered as potential CTL epitopes. Delta value between A3G-mutated wild-type epitope scores were calculated with 0.1 rank difference as a threshold of meaningful difference for increased or reduced HLA binding affinity. **(A)** The impact of all A3G induced-mutations on HLA binding of epitopes in Gag, Pol, Env, and Nef restricted to HLA-A2:01, HLA-A3:01, HLA-B57:01, and HLA-B35:01 considering stop codon (top panel) and excluding stop codon (bottom panel). **(B)** The specific impact of A3G induced-mutations at non-anchor residues on HLA binding. **(C)** The specific impact of A3G-induced mutations on HLA-binding affinity considering only mutations at the N-terminal anchor position (amino acid residue 2 in a 8-11 mer) and the most C-terminal HLA anchor residue.

resulted in decreased predicted proteasomal processing, whilst, 57, 44, 43, and 57% resulted in enhanced proteasomal processing.

In total, including A3G-mediated stop codon generation 54% of all A3G-induced mutations that could potentially impact proteasomal cleavage were predicted to lead to enhanced

CTL epitope production, whilst 46% could potentially decrease CTL epitope production. These numbers are 51 and 49%, respectively, whilst A3G-mediated stop codon generation is excluded (Table 2). These results indicate that there has not been a strong evolutionary pressure maintained on the viral genome for utilizing A3G toward CTL escape at the level of modulating proteasomal processing for CTL epitope generation.

We then investigated whether mutations mediated by A3G are clustered around the CTL epitopes, with more A3G hotspots being present either near the N- or C-terminal boundaries of epitopes, than expected at random. Considering a limit of 4 mutational A3G hotspots (A3G hotspots whose mutation would lead to non-synonymous substitutions) in 32 residues around each epitope, there appeared to be a marked paucity of such clustering (Table 3, Figure S1); however, when this limit was lowered to 2 mutational A3G hotspots, instances of clustering expectedly rose to 50–70% of epitopes. When all A3G hotspots were considered, this clustering proportion rose to 70–80%; thus, if non-mutational A3G hotspots can alter aspects of epitope production pre-translation (e.g., splicing, expression, etc.) this could be considered a significant trend.

Patterns and Consequences of A3G Hotspot Distribution Within or Outside CTL Epitope-Encoding Regions

We then investigated the enrichment of A3G hotspots (GGA, GGG, and GGT) inside vs. outside genomic sequences encoding CTL epitopes restricted to HLA-A2:01, HLA-A3:01, HLA-B57:01, and HLA-B35:01 in Gag, Pol, Env, and Nef genes (Table S2, Figure 3). First, we normalized for total nucleotide length of each gene and calculated the ratio of normalized A3G hotspot frequencies inside to outside epitope-encoding sequences for each protein's CTL epitopes. Thus, an inside: outside ratio >1 would be indicative of A3G hotspot enrichment in CTL-encoding sequences. As a control, we subjected the entire HIV genomic sequence to a random shuffling process, six independent times, but retained the positions/borders of the CTL epitope-encoding sequences. We then conducted the same analysis and expectedly arrived at ratios of ~1 (Figure 3A). We did not observe a generalizable trend of hotspot enrichment (ratio >1) in CTL-encoding sequences; however, when compared to the hypothetical ratio of 1 and the randomly shuffled control analyses, we noted that sequences encoding epitopes restricted to HLA-A2:01 and HLA-B57:01 often exhibited the highest enrichment ratios of 2–2.5 for at least 1–2 out of the 3 A3G hotspot motifs (Table S2, Figure 3A). These data are consistent with our previous observations that viral genomic sequences encoding more immunogenic CTL epitopes (restricted to more common HLA alleles, or those that elicit a more effective CTL response) have evolved to maintain A3G hotspots. Since A3G-mediated mutations which lead to stop codon generation would most often lead to non-infectious genomes, it is difficult to envision how maintenance of such A3G motifs to alter CTL epitopes at the cost of producing a non-infectious virus could be advantageous for the virus. Thus, in our enrichment analyses which was conducted to measure the extent to which the viral genome has evolved to

TABLE 2 | Proteasomal cleavage site prediction using NetChop 3.1 for Gag, Pol, Env, and Nef for HLA-A2:01, HLA-A3:01, HLA-B57:01, HLA-B35:01-restricted epitopes.

Protein	HLA-A2:01				HLA-A3:01				HLA-B57:01				HLA-B35:01			
	New sites		Eliminated sites		New sites		Eliminated sites		New sites		Eliminated sites		New sites		Eliminated sites	
	Inside CTL epitope	Outside CTL epitope	Inside CTL epitope	Outside CTL epitope	Inside CTL epitope	Outside CTL epitope	Inside CTL epitope	Outside CTL epitope	Inside CTL epitope	Outside CTL epitope	Inside CTL epitope	Outside CTL epitope	Inside CTL epitope	Outside CTL epitope	Inside CTL epitope	Outside CTL epitope
Gag	7	9	3	2	2	1	11	3	3	0	4	1	9	6	3	1
Pol	6	10	6	4	2	5	9	14	5	3	10	5	10	10	6	10
Env	9	10	3	4	4	1	5	6	3	3	2	3	6	8	2	6
Nef	1	3	1	1	3	0	3	1	3	0	3	1	3	3	3	2

utilize A3G toward CTL escape, we excluded A3G hotspots that would lead to stop codons. Rather, we considered these separately by examining the frequency and positional distribution of stop codon-generating A3G motifs in Gag, Pol, Env, and Nef. We found that between 18 and 43% of stop codons are positioned in the first quarter of each peptide (**Figure S2**), and there was a general trend of more frequent A3G-mediated stop codon generation in Pol and Env, as compared to Gag and Nef.

Having examined gene sequence A3G hotspot enrichment, we sought to measure the potential consequences at CTL epitope protein level. To this end, all simulated A3G-induced mutations were translated to protein sequences as described above. Since it is known that A3G can mutate the entire viral genome at low levels, epitopes with multiple hotspots and multiple mutated versions were considered independently. We then quantified A3G-induced non-synonymous and silent substitutions that fell within or outside of CTL epitopes of Gag, Pol, Env, and Nef restricted to HLA-A2:01, HLA-A3:01, HLA-B57:01 and HLA-B35:01. Next, we determined the ratio of A3G-induced mutations that caused non-synonymous residue changes to A3G-induced mutations which resulted in silent mutations within CTL epitopes and divided this by the same ratio determined for regions of each polypeptide that fell outside CTL epitopes. This analysis was carried out for each individual A3G hotspot motif, and as a total for all amino acids affected by A3G mutations within each polypeptide (**Table S2** and **Figure 3B**). Thus, a ratio of >1 would indicate that the genomic sequence of HIV has evolved to channel A3G-induced mutations into amino acid changes, more often within CTL epitope-encoding regions as compared to sequences outside these portions. Indeed, we observed numerous instances of significant preferential channeling (ratios up to 4.5) toward non-synonymous residue changes in CTL epitopes of Gag, Pol, Env, and Nef as a result of A3G-induced mutations (**Figure 3B**). If we ignored A3G-driven silent consequences and evaluated the ratio of only A3G-mediated non-synonymous mutations inside CTL epitope-encoding regions to A3G-mediated non-synonymous mutations outside CTL epitope-encoding regions, we observed that in 28/48 graphed bars (58% of all measurements) the ratio was ≥ 1 , with some ratios in the 2–3 range (**Figure 3C**). In general, the bias for A3G mutations to translate to non-synonymous rather than silent amino acid mutations was more pronounced for Gag, Pol and Nef as compared to Env, consistent with the former three polypeptides housing the vast majority of HIV's CTL epitopes (**Figures 3B,C** compared to **Figure 3A**).

In principle, non-synonymous amino acid changes arising from A3G mutations can enhance or diminish antigen presentation as the proteasomal processing and HLA binding levels (**Figure 2**, **Table 2**). To examine the distribution patterns of A3G hotspots that could potentially lead to CTL escape, we generated a map of all A3G hotspots across the entire HIV genome and overlaid this map on the experimentally-determined and well-known gradient of G to A mutations across the HIV genome (**Figure 4A**). In the context of A3G action, this twin gradient has been suggested to be a consequence of HIV genome's replication dynamics. Certain portions of the HIV minus strand genome remain single-stranded for a

longer period compared to other segments, due to dynamics of RNA digestion by RT, the role of the Polypurine tracts (PPT), and subsequent positive sense strand polymerization. These segments are thus more available for A3G targeting resulting in a mutation gradient (34, 35, 143). We observed that regions near the central PPT and the C-terminal end (Nef) that are more highly mutated are rich in A3G hotspots, consistent with the notion that the viral genome has positioned hotspots in genomic locations that are more prone to being targeted by the A3G enzyme (**Figure 4A**). For each HLA, we plotted the number of CTL-epitope encoding sequences (in Gag, Pol, Env, and Nef) at incremental positions along the entire viral genome length (**Figure 4B**, top graph of each panel). We also plotted a normalized escape potential graph which represents the likelihood that an A3G hotspot located inside a CTL epitope-encoding sequence can generate a CTL-escape mutation. This was calculated by counting the A3G hotspots predicted to lower HLA binding affinities and normalizing these by the total abundance of A3G hotspots in the given CTL-epitope encoding region (**Figure 4B**, middle graph of each panel). Considering the number of CTL-epitopes encoded by a given genomic location (top panel), as well as the normalized escape potential of the sequences encoding this epitope (middle panel), we then generated an escape factor map which represents the compound potential for A3G to cause CTL escape across the HIV genome, for each HLA (**Figure 4B**, bottom graph of each panel). First, we noted that the potential for A3G-generated CTL escape was present throughout the length of the genome, for epitopes restricted to all 4 HLAs; however, it was generally more frequent in regions of the genome with a higher mutational potential and less frequent in regions known to be mutated at lower rates (comparing escape factor maps in **Figure 4B** to the mutational gradient in **Figure 4A**). Secondly, epitopes restricted to HLA-A2:01 and HLA-B57:01 exhibited overall higher abundance and frequent positioning of escape-inducing A3G hotspots, with HLA-B57:01-restricted epitopes also containing the highest escape factor values (**Figure 4B**). Thirdly, regions encoding for Gag, Pol and Nef contained generally a higher density of potentially CTL escape-inducing A3G hotspots, as compared to Env. The polypeptide and HLA-specific patterns observed are consistent with the A3G hotspot enrichment analysis (**Figure 3**) and taken together suggest the evolution of HIV genome to position A3G hotspot motifs in CTL-encoding regions, and highly mutable regions of the HIV genome, such that they preferentially yield CTL escape-inducing non-synonymous amino acid changes.

DISCUSSION

Here, we aimed to follow up on previous works suggesting that A3G is a source of CTL escape-inducing mutations. We first mapped all potential CTL epitopes within Gag, Pol, Env, and Nef, and considered the impact of A3G-induced mutations on these epitopes. To this end, we embarked on a two-pronged analysis: first, from the immune recognition perspective, we examined the effect of A3G-induced mutations

TABLE 3 | Enrichment of A3G hotspot motifs in flanking regions of sequence encoding CTL epitopes restricted to HLA-A2:01, HLA-A3:01, HLA-B57:01, and HLA-B35:01 for Gag, Pol, Env, and Nef.

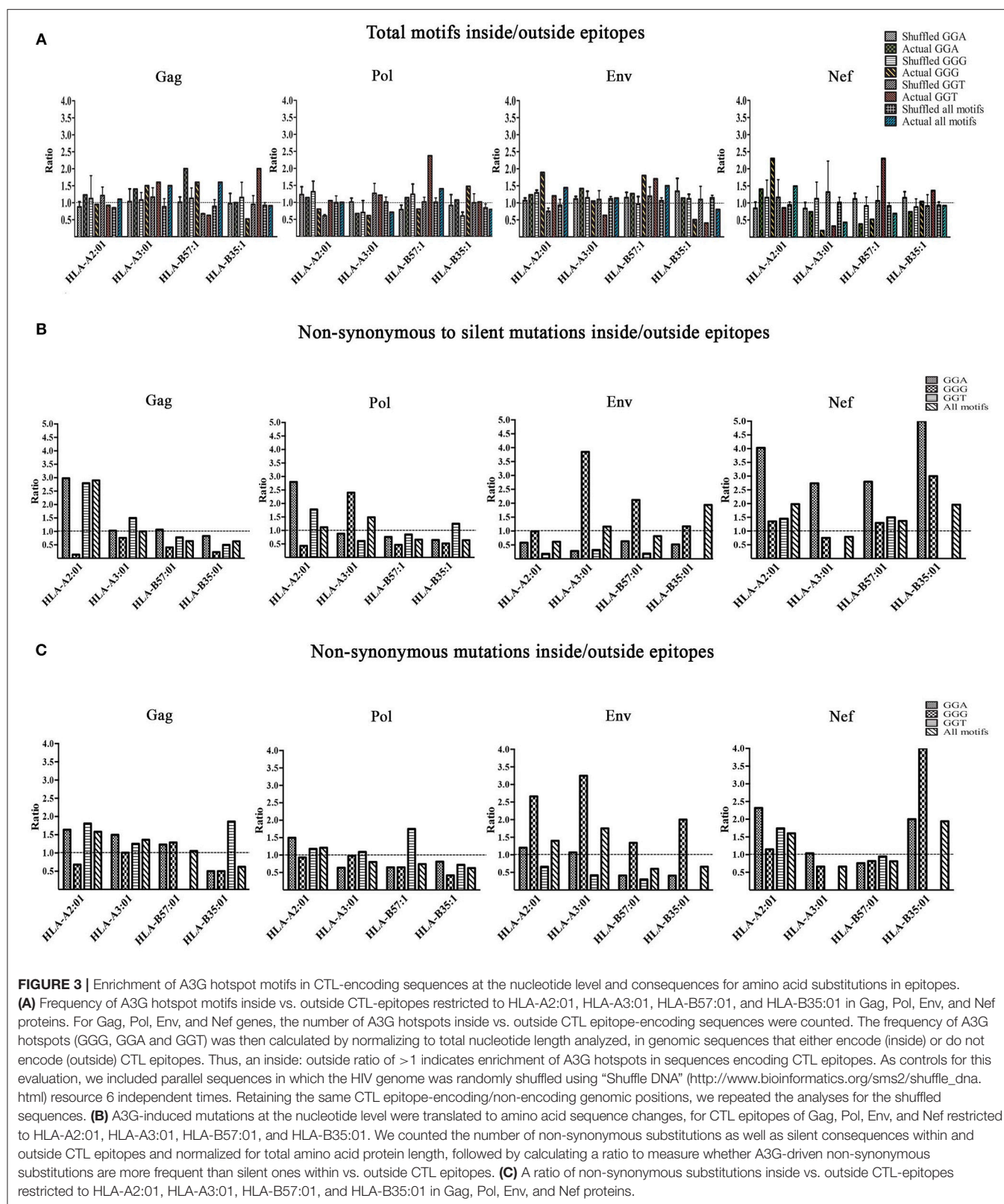
Protein	HLA	All instances epitope (surrounded by 2 mutations) (%)	All instances epitope (surrounded by 4 mutations) (%)	Non-synonymous mutation (surrounded by 2 mutations) (%)	Non-synonymous mutation (surrounded by 4 mutations) (%)
Gag	HLA-A2:01	69	7.6	61	0
	HLA-A3:01	53	0	30	0
	HLA-B57:01	78	7.1	64	0
	HLA-B35:01	80	10	50	0
Pol	HLA-A2:01	42	0	31	0
	HLA-A3:01	66	12	45	6
	HLA-B57:01	44	12	29	4
	HLA-B35:01	50	10	35	0
Env	HLA-A2:01	54	14	40	9
	HLA-A3:01	53	13	46	13
	HLA-B57:01	57	15	42	11
	HLA-B35:01	62	25	50	25
Nef	HLA-A2:01	75	12.5	75	0
	HLA-A3:01	83	33	60	0
	HLA-B57:01	75	25	50	0
	HLA-B35:01	75	37	50	12.5

on the various stages of CTL epitope production, including proteasomal processing and HLA-binding affinities. Second, from the viral genome evolution perspective, we examined whether, where and to what consequence, A3G hotspots have been maintained or enriched in genomic sequences that encode for CTL epitopes. At each stage of all analyses, we considered three individual A3G hotspots (GGA, GGG, and GGT), and potential impact on CTL epitopes restricted to four HLA alleles that have been previously shown to have differential abilities to present immunogenic CTL peptides of HIV. Furthermore, opposite to the notion of CTL escape mediated by A3G-induced mutations, we also considered A3G mutations that can potentially generate novel or more immunogenic CTL epitopes. Despite a wealth of information about the role of A3Gs in CTL escape, knowledge of novel CTL epitopes mediated by endogenous mutators remains poor. We found that although A3G-mediated mutations could potentially enhance or diminish the proteasomal cleavage of Gag, Pol, Env, and Nef into CTL epitopes, the overwhelming impact on HLA binding affinities of CTL epitopes as a result of A3G mutations was decreased affinity.

Here we also provide strong and novel lines of evidence for the co-evolution of the HIV genome with A3G, so as to utilize this host factor toward CTL escape. First, A3G hotspot motifs were positioned in CTL-encoding epitopes so as to preferentially cause non-synonymous mutations. Secondly, most A3G-induced mutations in CTL epitopes resulted in diminished/abrogated HLA binding capacity. Thirdly, the distribution pattern of CTL escape-inducing A3G hotspots across the HIV genome varies with restricting HLAs and generally correlates with the known mutational gradient across the entire HIV genome. These observations shed light on the multiple layers of depth to

which the HIV genome has resorted to position A3G hotspots for CTL escape. An earlier study which examined the overall pattern of A3G-mediated non-synonymous vs. silent mutations concluded that A3G has not left an evolutionary footprint on the HIV genome (149). This study broadly examined all A3G/F target motifs but not in the context of a specific biological force which may encourage genome evolution in response to A3G/F. In contrast, we argue that evidence for co-evolution of a pathogen's genome with a host factor may not be broadly apparent but must be sought for in the specific context of the pro/anti-viral biological impacts driven by the host factor and specific locations of the pathogen's genome impacted and under pressure by the host factor action. Thus, we considered HIV genome co-evolution with A3G/F in the context of the potential for CTL escape by focusing on sequences that encode for such epitopes. We find that when considered in this context, there is substantial evidence for the evolution of the HIV genome to subvert the activity of A3G/F toward its own gain. In our analyses of predicted immune response to HIV (HLA binding and proteasomal processing of CTL epitopes) we considered A3G-mediated stop codons; though these would yield non-infectious viruses, truncated proteins that are immunogenic could still be produced (120). On the other hand, we excluded A3G-mediated stop codons from our analyses of viral genome evolution (A3G hotspot enrichment and positioning) since they cannot be considered as an advantageous mode of utilizing A3G to the virus's benefit.

Rather than considering existing A3G hotspots as evidence for their role in selection as we have, the case can also be made for the opposite view; that if the usage of A3G hotspots toward advantageous outcomes for the virus was key, then current



sequences circulating at the population level ought to be rather devoid of A3G hotspots and instead rich in the mutated versions. Whilst this argument may hold true in several different contexts,

considering it in the context of CTL escape is more difficult. First, CTL escape mutations are usually not broadly selectable at the population level but are highly individual host-dependent

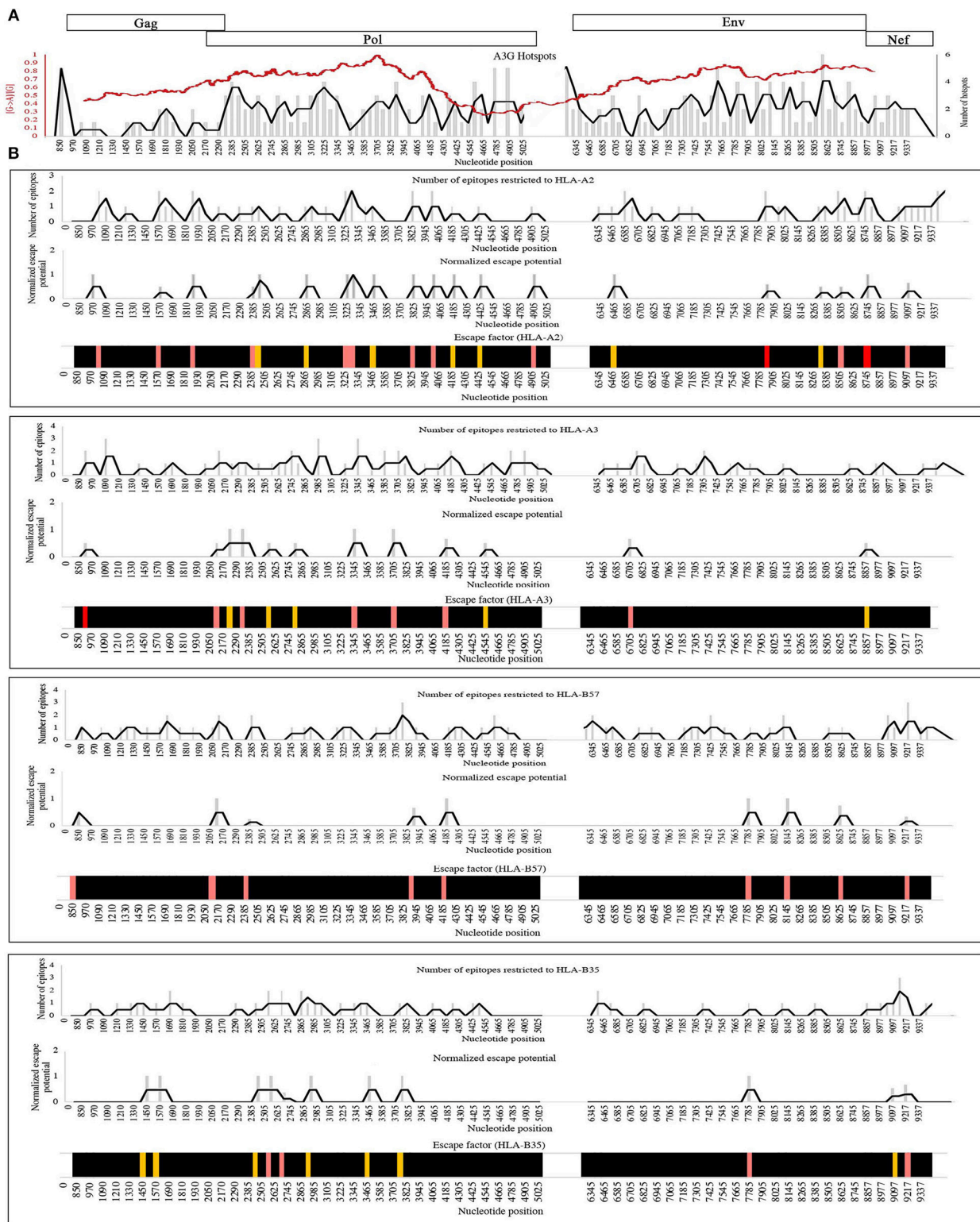


FIGURE 4 | Distribution pattern of CTL escape-inducing A3G hotspots in the context of the entire HIV genome and its mutational gradient. **(A)** To analyze the distribution pattern of A3G hotspots across the HIV genome, we counted the number of A3G hotspots that could result in non-synonymous amino acid changes within windows of 60 bp length across the entire length of Gag, Pol, Env, and Nef. The number of A3G hotspots in each 60 bp window is plotted as gray bars, and the overall pattern is shown as a black line. We then overlaid this pattern against the known mutational gradient of the HIV genome as previously described (red line) (143) (Continued)

FIGURE 4 | (Copy Right License Number: 4460241074666). **(B)** To analyze the distribution pattern of predicted CTL escape-inducing A3G for each HLA, a map of the positions of CTL epitope-encoding sequences for Gag, Pol, Env, and Nef, restricted to HLA-A2:01, HLA-A3:01, HLA-B57:01, and HLA-B35:01 was created (Top panel of each box). A map of normalized escape potential was also generated (middle panel of each box). We defined normalized escape potential as the number of CTL escape-inducing A3G hotspots in each 60 bp segment, normalized (divided) by the total number of A3G hotspots within the segment. Based on the normalized escape potential and the number of CTL epitopes encoded in each 60 bp segment, for each HLA we then constructed a map of escape factor which we defined as the product of the number of CTL epitopes and normalized escape potential (bottom panel). In this map, black indicates no potential for CTL escape (escape factor=0), orange indicates low potential (escape factor=1), pink indicates modest potential (escape factor range of 2–4), red indicates high potential (escape factor range of 5–7).

since they are intimately connected to HLA genotype. Second, from the virus's perspective, there may be two advantages to maintaining the A3G/F hotspots: first, conflicting demands of replication fitness on one hand and immune evasion on the other, which is best illustrated by the high rates at which certain CTL escape mutations revert to wildtype presumably fitter sequence, especially upon transmission to a new host with a different HLA genotype wherein CTL escape mutations from the previous host are no longer advantageous. Second, it may not be to the advantage of the virus to benefit from maximum CTL escape, as it would limit its replication capacity by quickly eliminating infected host cells (150, 151). Thus, it may be advantageous to conserve some CTL escape potential in the form of A3G hotspots to be available to use when it suits the virus. An example of this very conservation of A3G/F-mutational hotspots has been shown in terms of antibody epitopes in Env (122).

These findings bring to light novel aspects of the interplay between the host mutator A3G and the co-evolution of the viral genome. Overall, A3G-induced mutations were predicted to influence CTL epitope production and HLA binding, both toward the production of more immunogenic epitopes and conversely, toward CTL escape. It is important to consider these results in the context of two additional concepts: first, although the overall action of A3G on the HIV genome is predicted to result more often in CTL escape than in generation of new, more immunogenic epitopes, it is important to note that even if the latter and former occurred with equal probability, the escape mutations would be the dominant outcome under immune pressure *in vivo* (9, 87, 119). Second, in this analysis, we did not take into account the fitness consequences of A3G-induced mutations. Predicting the *in vitro* replicative fitness cost and peptide HLA binding affinity of clinically derived sequences has shown that escape mutations in CTL epitopes of Gag restricted to protective HLA class I alleles carried higher fitness costs and lower levels of reduction in HLA class I binding affinity compared to mutations in epitopes restricted to other HLA class I alleles. This suggests that one way by which protective HLA molecules act is by binding epitopes whose CTL escape mutations incur a high fitness cost with relatively low benefit in terms of HLA-binding affinity reduction (108).

The practical application of this work will lie in determining epitope choice for vaccine design. Epitope clusters and altered epitopes with the potential to be better processed or bound by HLA because of A3G mutations ought to be superior platforms for the development of prophylactic or post-infection CTL-based vaccines. Thus, accounting for and indeed exploiting

the action of endogenous genome mutators to design more effective vaccines would represent a strategic advance in HIV vaccine design. In addition, the analyses carried out here should be considered in the context of extensive A3 family enzyme mutations of tumor genomes, as understanding the mechanisms by which a tumor cell can escape, or boost CTL response is critical to developing vaccination and therapies based on CTL epitopes. We and others have postulated that the function of A3 family members in cancer genome mutagenesis may bear parallels to its role in viral genome mutagenesis, as tumor cells are also under pressure to avoid detection by CTL and could use A3-induced mutagenesis to this end (152–159). At present, whether and how frequently this may occur is unknown, and using similar analyses to gain insights will have important implications for the design of personalized anti-tumor CTL-based strategies.

AUTHOR CONTRIBUTIONS

FB generated all data, with help from KJ. FB and ML analyzed the data, generated the figures and wrote the manuscript. MG edited the manuscript. We thank our colleague Emma Quinlan for assistance with editing.

FUNDING

This work was supported by Canadian Institutes of Health Research (CIHR) operating grants (MOP111132 and OCH131580) to ML. FB is supported by a Dean's fellowship from the Faculty of Medicine Memorial University and a Ph.D. fellowship from the Beatrice Hunter Cancer Research Institute (BHCRI) with funds provided by the Cancer Research Training Program as part of the Terry Fox Foundation strategic Health Research Training Program in Cancer Research at CIHR.

ACKNOWLEDGMENTS

We are grateful to our colleague Emma M. Quinlan for reproducing the HIV mutational gradient (red trend line) in **Figure 4A**, from Kijak et al. (143) (Copy Right License Number: 4460241074666).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.03032/full#supplementary-material>

REFERENCES

- Coffin J, Swanstrom R. HIV pathogenesis: dynamics and genetics of viral populations and infected cells. *Cold Spring Harb Perspect Med.* (2013) 3:a012526. doi: 10.1101/cshperspect.a012526
- Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol.* (2010) 84:9864–78. doi: 10.1128/JVI.00915-10
- Kim T, Mudry RA Jr, Rexrode CAII, Pathak VK. Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. *J Virol.* (1996) 70:7594–602.
- Mansky LM. Forward mutation rate of human immunodeficiency virus type 1 in a T lymphoid cell line. *AIDS Res Hum Retroviruses* (1996) 12:307–14. doi: 10.1089/aid.1996.12.307
- Mansky LM, Temin HM. Lower *in vivo* mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol.* (1995) 69:5087–94.
- Menendez-Arias L. Molecular basis of human immunodeficiency virus type 1 drug resistance: overview and recent developments. *Antiviral Res.* (2013) 98:93–120. doi: 10.1016/j.antiviral.2013.01.007
- Pathak VK, Temin HM. Broad spectrum of *in vivo* forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. *Proc Natl Acad Sci USA.* (1990) 87:6019–23.
- Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* (2002) 418:646–50. doi: 10.1038/nature00939
- Conticello SG. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* (2008) 9:229. doi: 10.1186/gb-2008-9-6-229
- Imahashi M, Nakashima M, Iwatani Y. Antiviral mechanism and biochemical basis of the human APOBEC3 family. *Front Microbiol.* (2012) 3:250. doi: 10.3389/fmicb.2012.00250
- Malim MH. APOBEC proteins and intrinsic resistance to HIV-1 infection. *Philos Trans R Soc Lond B Biol Sci.* (2009) 364:675–87. doi: 10.1098/rstb.2008.0185
- Salter JD, Bennett RP, Smith HC. The APOBEC protein family: united by structure, divergent in function. *Trends Biochem Sci.* (2016) 41:578–94. doi: 10.1016/j.tibs.2016.05.001
- Feng Y, Goubran MH, Follack TB, Chelico L. Deamination-independent restriction of LINE-1 retrotransposition by APOBEC3H. *Sci Rep.* (2017) 7:10881. doi: 10.1038/s41598-017-11344-4
- Beggel B, Munk C, Daumer M, Hauck K, Haussinger D, Lengauer T, et al. Full genome ultra-deep pyrosequencing associates G-to-A hypermutation of the hepatitis B virus genome with the natural progression of hepatitis B. *J Viral Hepat.* (2013) 20:882–9. doi: 10.1111/jvh.12110
- Bonvin M, Greeve J. Hepatitis B: modern concepts in pathogenesis–APOBEC3 cytidine deaminases as effectors in innate immunity against the hepatitis B virus. *Curr Opin Infect Dis.* (2008) 21:298–303. doi: 10.1097/QCO.0b013e3282fe1bb2
- Derse D, Hill SA, Prinler G, Lloyd P, Heidecker G. Resistance of human T cell leukemia virus type 1 to APOBEC3G restriction is mediated by elements in nucleocapsid. *Proc Natl Acad Sci USA.* (2007) 104:2915–20. doi: 10.1073/pnas.0609444104
- Vartanian JP, Guetard D, Henry M, Wain-Hobson S. Evidence for editing of human papillomavirus DNA by APOBEC3 in benign and precancerous lesions. *Science* (2008) 320:230–3. doi: 10.1126/science.1153201
- Wiegand HL, Cullen BR. Inhibition of alpharetrovirus replication by a range of human APOBEC3 proteins. *J Virol.* (2007) 81:13694–9. doi: 10.1128/JVI.01646-07
- Chaipan C, Smith JL, Hu WS, Pathak VK. APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4+ T cells and macrophages. *J Virol.* (2013) 87:444–53. doi: 10.1128/JVI.00676-12
- Harris RS, Bishop KN, Sheehy AM, Craig HM, Petersen-Mahrt SK, Watt IN, et al. DNA deamination mediates innate immunity to retroviral infection. *Cell* (2003) 113:803–9. doi: 10.1016/S0092-8674(03)00423-9
- Klarmann GJ, Chen X, North TW, Preston BD. Incorporation of uracil into minus strand DNA affects the specificity of plus strand synthesis initiation during lentiviral reverse transcription. *J Biol Chem.* (2003) 278:7902–9. doi: 10.1074/jbc.M207223200
- Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* (2003) 424:99–103. doi: 10.1038/nature01709
- Mbisa JL, Barr R, Thomas JA, Vandegraaff N, Dorweiler JJ, Svarovskaia ES, et al. Human immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration. *J Virol.* (2007) 81:7099–110. doi: 10.1128/JVI.00272-07
- Shindo K, Takaori-Kondo A, Kobayashi M, Abudu A, Fukunaga K, Uchiyama T. The enzymatic activity of CEM15/Apobec-3G is essential for the regulation of the infectivity of HIV-1 virion but not a sole determinant of its antiviral activity. *J Biol Chem.* (2003) 278:44412–6. doi: 10.1074/jbc.C300376200
- Yang B, Chen K, Zhang C, Huang S, Zhang H. Virion-associated uracil DNA glycosylase-2 and apurinic/apyrimidinic endonuclease are involved in the degradation of APOBEC3G-edited nascent HIV-1 DNA. *J Biol Chem.* (2007) 282:11667–75. doi: 10.1074/jbc.M606864200
- Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* (2003) 424:94–8. doi: 10.1038/nature01707
- Chen K, Huang J, Zhang C, Huang S, Nunnari G, Wang FX, et al. Alpha interferon potentially enhances the anti-human immunodeficiency virus type 1 activity of APOBEC3G in resting primary CD4T cells. *J Virol.* (2006) 80:7645–57. doi: 10.1128/JVI.00206-06
- Peng G, Lei KJ, Jin W, Greenwell-Wild T, Wahl SM. Induction of APOBEC3 family proteins, a defensive maneuver underlying interferon-induced anti-HIV-1 activity. *J Exp Med.* (2006) 203:41–6. doi: 10.1084/jem.20051512
- Pillai SK, Abdel-Mohsen M, Guatelli J, Skasko M, Monto A, Fujimoto K, et al. Role of retroviral restriction factors in the interferon-alpha-mediated suppression of HIV-1 *in vivo*. *Proc Natl Acad Sci USA.* (2012) 109:3035–40. doi: 10.1073/pnas.1111573109
- Desimie BA, Burdick RC, Izumi T, Doi H, Shao W, Alvord WG, et al. APOBEC3 proteins can copackage and comutate HIV-1 genomes. *Nucleic Acids Res.* (2016) 44:7848–65. doi: 10.1093/nar/gkw653
- Holtz CM, Sadler HA, Mansky LM. APOBEC3G cytosine deamination hotspots are defined by both sequence context and single-stranded DNA secondary structure. *Nucleic Acids Res.* (2013) 41:6139–48. doi: 10.1093/nar/gkt246
- Lecossier D, Bouchonnet F, Clavel F, Hance AJ. Hypermutation of HIV-1 DNA in the absence of the Vif protein. *Science* (2003) 300:1112. doi: 10.1126/science.1083338
- Suspene R, Sommer P, Henry M, Ferris S, Guetard D, Pochet S, et al. APOBEC3G is a single-stranded DNA cytidine deaminase and functions independently of HIV reverse transcriptase. *Nucleic Acids Res.* (2004) 32:2421–9. doi: 10.1093/nar/gkh554
- Yu Q, Konig R, Pillai S, Chiles K, Kearney M, Palmer S, et al. Single-strand specificity of APOBEC3G accounts for minus-strand deamination of the HIV genome. *Nat Struct Mol Biol.* (2004) 11:435–42. doi: 10.1038/nsmb758
- Suspene R, Rusniok C, Vartanian JP, Wain-Hobson S. Twin gradients in APOBEC3 edited HIV-1 DNA reflect the dynamics of lentiviral replication. *Nucleic Acids Res.* (2006) 34:4677–84. doi: 10.1093/nar/gkl555
- Armitage AE, Deforche K, Chang CH, Wee E, Kramer B, Welch JJ, et al. APOBEC3G-induced hypermutation of human immunodeficiency virus type-1 is typically a discrete “all or nothing” phenomenon. *PLoS Genet.* (2012) 8:e1002550. doi: 10.1371/journal.pgen.1002550
- Ara A, Love RP, Follack TB, Ahmed KA, Adolph MB, Chelico L. Mechanism of enhanced hiv restriction by virion coencapsidated cytidine deaminases APOBEC3F and APOBEC3G. *J Virol.* (2017) 91:e02230-16. doi: 10.1128/JVI.02230-16
- Adolph MB, Love RP, Chelico L. Biochemical basis of APOBEC3 deoxycytidine deaminase activity on diverse DNA substrates. *ACS Infect Dis.* (2018) 4:224–38. doi: 10.1021/acsinfecdis.7b00221

39. Delviks-Frankenberry KA, Nikolaitchik OA, Burdick RC, Gorelick RJ, Keele BF, Hu WS, et al. Minimal contribution of APOBEC3-induced G-to-A hypermutation to HIV-1 recombination and genetic variation. *PLoS Pathog.* (2016) 12:e005646. doi: 10.1371/journal.ppat.1005646
40. Gillick K, Pollpeter D, Phalora P, Kim EY, Wolinsky SM, Malim MH. Suppression of HIV-1 infection by APOBEC3 proteins in primary human CD4(+) T cells is associated with inhibition of processive reverse transcription as well as excessive cytidine deamination. *J. Virol.* (2013) 87:1508–17. doi: 10.1128/JVI.02587-12
41. Wang X, Ao Z, Chen L, Kobinger G, Peng J, Yao X. The cellular antiviral protein APOBEC3G interacts with HIV-1 reverse transcriptase and inhibits its function during viral replication. *J. Virol.* (2012) 86:3777–86. doi: 10.1128/JVI.06594-11
42. Holmes RK, Malim MH, Bishop KN. APOBEC-mediated viral restriction: not simply editing? *Trends Biochem Sci.* (2007) 32:118–28. doi: 10.1016/j.tibs.2007.01.004
43. Iwatani Y, Chan DS, Wang F, Maynard KS, Sugiura W, Gronenborn AM, et al. Deaminase-independent inhibition of HIV-1 reverse transcription by APOBEC3G. *Nucleic Acids Res.* (2007) 35:7096–108. doi: 10.1093/nar/gkm750
44. Levin JG, Mitra M, Mascarenhas A, Musier-Forsyth K. Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.* (2010) 7:754–74. doi: 10.4161/rna.7.6.14115
45. Morse M, Huo R, Feng Y, Rouzina I, Chelico L, Williams MC. Dimerization regulates both deaminase-dependent and deaminase-independent HIV-1 restriction by APOBEC3G. *Nat Commun.* (2017) 8:597. doi: 10.1038/s41467-017-00501-y
46. Cuevas JM, Geller R, Garijo R, Lopez-Aldeguer J, Sanjuan R. Extremely high mutation rate of HIV-1 *in vivo*. *PLoS Biol.* (2015) 13:e1002251. doi: 10.1371/journal.pbio.1002251
47. Albin JS, Harris RS. Interactions of host APOBEC3 restriction factors with HIV-1 *in vivo*: implications for therapeutics. *Expert Rev Mol Med.* (2010) 12:e4. doi: 10.1017/S1462399409001343
48. Kim EY, Bhattacharya T, Kunstman K, Swantek P, Koning FA, Malim MH, et al. Human APOBEC3G-mediated editing can promote HIV-1 sequence diversification and accelerate adaptation to selective pressure. *J. Virol.* (2010) 84:10402–5. doi: 10.1128/JVI.01223-10
49. Sadler HA, Stenglein MD, Harris RS, Mansky LM. APOBEC3G contributes to HIV-1 variation through sublethal mutagenesis. *J. Virol.* (2010) 84:7396–404. doi: 10.1128/JVI.00056-10
50. Zanini F, Puller V, Brodin J, Albert J, Neher RA. *In vivo* mutation rates and the landscape of fitness costs of HIV-1. *Virus Evol.* (2017) 3:vex003. doi: 10.1093/ve/vex003
51. Armitage AE, Deforche K, Welch JJ, Van Laethem K, Camacho R, Rambaut A, et al. Possible footprints of APOBEC3F and/or other APOBEC3 deaminases, but not APOBEC3G, on HIV-1 from patients with acute/early and chronic infections. *J. Virol.* (2014) 88:12882–94. doi: 10.1128/JVI.01460-14
52. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci USA.* (2008) 105:7552–7. doi: 10.1073/pnas.0802203105
53. Conticello SG, Langlois MA, Yang Z, Neuberger MS. DNA deamination in immunity: AID in the context of its APOBEC relatives. *Adv Immunol.* (2007) 94:37–73. doi: 10.1016/S0065-2776(06)94002-4
54. Desimie BA, Delviks-Frankenberry KA, Burdick RC, Qi D, Izumi T, Pathak VK. Multiple APOBEC3 restriction factors for HIV-1 and one Vif to rule them all. *J. Mol. Biol.* (2014) 426:1220–45. doi: 10.1016/j.jmb.2013.10.033
55. Harris RS, Hultquist JF, Evans DT. The restriction factors of human immunodeficiency virus. *J. Biol. Chem.* (2012) 287:40875–83. doi: 10.1074/jbc.R112.416925
56. Kao S, Khan MA, Miyagi E, Plishka R, Buckler-White A, Strebel K. The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. *J. Virol.* (2003) 77:11398–407. doi: 10.1128/JVI.77.21.11398-11407.2003
57. Trono D. HIV accessory proteins: leading roles for the supporting cast. *Cell* (1995) 82:189–92.
58. von Schwedler U, Song J, Aiken C, Trono D. Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. *J. Virol.* (1993) 67:4945–55.
59. Mercenne G, Bernacchi S, Richer D, Bec G, Henriot S, Paillart JC, et al. HIV-1 Vif binds to APOBEC3G mRNA and inhibits its translation. *Nucleic Acids Res.* (2010) 38:633–46. doi: 10.1093/nar/gkp1009
60. Guerrero S, Libre C, Batisse J, Mercenne G, Richer D, Laumond G, et al. Translational regulation of APOBEC3G mRNA by Vif requires its 5'UTR and contributes to restoring HIV-1 infectivity. *Sci Rep.* (2016) 6:39507. doi: 10.1038/srep39507
61. Baig TT, Feng Y, Chelico L. Determinants of efficient degradation of APOBEC3 restriction factors by HIV-1 Vif. *J. Virol.* (2014) 88:14380–95. doi: 10.1128/JVI.02484-14
62. Zhou D, Wang Y, Tokunaga K, Huang F, Sun B, Yang R. The HIV-1 accessory protein Vpr induces the degradation of the anti-HIV-1 agent APOBEC3G through a VprBP-mediated proteasomal pathway. *Virus Res.* (2015) 195:25–34. doi: 10.1016/j.virusres.2014.08.021
63. Bruner KM, Murray AJ, Pollack RA, Soliman MG, Laskey SB, Capoferri AA, et al. Defective proviruses rapidly accumulate during acute HIV-1 infection. *Nat Med.* (2016) 22:1043–9. doi: 10.1038/nm.4156
64. Nowarski R, Britan-Rosich E, Shiloach T, Kotler M. Hypermutation by intersegmental transfer of APOBEC3G cytidine deaminase. *Nat Struct Mol Biol.* (2008) 15:1059–66. doi: 10.1038/nsmb.1495
65. Goulder PJ, Bunce M, Krausa P, McIntyre K, Crowley S, Morgan B, et al. Novel, cross-restricted, conserved, and immunodominant cytotoxic T lymphocyte epitopes in slow progressors in HIV type 1 infection. *AIDS Res Hum Retroviruses.* (1996) 12:1691–8. doi: 10.1089/aid.1996.12.1691
66. Wang YE, Li B, Carlson JM, Streeck H, Gladden AD, Goodman R, et al. Protective HLA class I alleles that restrict acute-phase CD8+ T-cell responses are associated with viral escape mutations located in highly conserved regions of human immunodeficiency virus type 1. *J. Virol.* (2009) 83:1845–55. doi: 10.1128/JVI.01061-08
67. Chen W, Anton LC, Bennink JR, Yewdell JW. Dissecting the multifactorial causes of immunodominance in class I-restricted T cell responses to viruses. *Immunity* (2000) 12:83–93. doi: 10.1016/S1074-7613(00)80161-2
68. Tschärke DC, Karupiah G, Zhou J, Palmore T, Irvine KR, Haeryfar SM, et al. Identification of poxvirus CD8+ T cell determinants to enable rational design and characterization of smallpox vaccines. *J. Exp. Med.* (2005) 201:95–104. doi: 10.1084/jem.20041912
69. Yewdell JW. Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. *Immunity* (2006) 25:533–43. doi: 10.1016/j.immuni.2006.09.005
70. Brumme ZL, John M, Carlson JM, Brumme CJ, Chan D, Brockman MA, et al. HLA-associated immune escape pathways in HIV-1 subtype B Gag, Pol and Nef proteins. *PLoS ONE* (2009) 4:e6687. doi: 10.1371/journal.pone.0006687
71. Brumme ZL, Kinloch NN, Sanchez S, Wong A, Martin E, Cobarrubias KD, et al. Extensive host immune adaptation in a concentrated North American HIV epidemic. *AIDS* (2018) 32:1927–38. doi: 10.1097/QAD.0000000000001912
72. Kinloch NN, Lee GQ, Carlson JM, Jin SW, Brumme CJ, Byakwaga H, et al. Genotypic and mechanistic characterization of subtype-specific HIV adaptation to host cellular immunity. *J. Virol.* (2018) 93:e01502-18. doi: 10.1128/JVI.01502-18
73. Allen TM, Altfield M, Geer SC, Kalife ET, Moore C, O'Sullivan K M, et al. Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *J. Virol.* (2005) 79:13239–49. doi: 10.1128/JVI.79.21.13239-13249.2005
74. Barouch DH, Kunstman J, Glowczwskie J, Kunstman KJ, Egan MA, Peyerl FW, et al. Viral escape from dominant simian immunodeficiency virus epitope-specific cytotoxic T lymphocytes in DNA-vaccinated rhesus monkeys. *J. Virol.* (2003) 77:7367–75. doi: 10.1128/JVI.77.13.7367-7375.2003
75. O'Connor DH, McDermott AB, Krebs KC, Dodds EJ, Miller JE, Gonzalez EJ, et al. A dominant role for CD8+ T-lymphocyte selection in simian immunodeficiency virus sequence variation. *J. Virol.* (2004) 78:14012–22. doi: 10.1128/JVI.78.24.14012-14022.2004

76. Domingo E, Holland JJ. RNA virus mutations and fitness for survival. *Annu Rev Microbiol.* (1997) 51:151–78. doi: 10.1146/annurev.micro.51.1.151
77. Mostowy R, Kouyos RD, Hoof I, Hinkley T, Haddad M, Whitcomb JM, et al. Estimating the fitness cost of escape from HLA presentation in HIV-1 protease and reverse transcriptase. *PLoS Comput Biol.* (2012) 8:e1002525. doi: 10.1371/journal.pcbi.1002525
78. Sunshine JE, Larsen BB, Maust B, Casey E, Deng W, Chen L, et al. Fitness-balanced escape determines resolution of dynamic founder virus escape processes in HIV-1 infection. *J Virol.* (2015) 89:10303–18. doi: 10.1128/JVI.01876-15
79. Tenzer S, Wee E, Burgevin A, Stewart-Jones G, Friis L, Lamberth K, et al. Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance. *Nat Immunol.* (2009) 10:636–46. doi: 10.1038/ni.1728
80. Yewdell JW, Haeryfar SM. Understanding presentation of viral antigens to CD8+ T cells *in vivo*: the key to rational vaccine design. *Annu Rev Immunol.* (2005) 23:651–82. doi: 10.1146/annurev.immunol.23.021704.115702
81. Barouch DH, Kunstman J, Kuroda MJ, Schmitz JE, Santra S, Peyerl FW, et al. Eventual AIDS vaccine failure in a rhesus monkey by viral escape from cytotoxic T lymphocytes. *Nature* (2002) 415:335–9. doi: 10.1038/415335a
82. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nat Rev Genet.* (2004) 5:52–61. doi: 10.1038/nrg1246
83. Han C, Kawana-Tachikawa A, Shimizu A, Zhu D, Nakamura H, Adachi E, et al. Switching and emergence of CTL epitopes in HIV-1 infection. *Retrovirology* (2014) 11:38. doi: 10.1186/1742-4690-11-38
84. Bronke C, Almeida CA, McKinnon E, Roberts SG, Keane NM, Chopra A, et al. HIV escape mutations occur preferentially at HLA-binding sites of CD8 T-cell epitopes. *AIDS* (2013) 27:899–905. doi: 10.1097/QAD.0b013e32835e1616
85. Goulder PJ, Phillips RE, Colbert RA, McAdam S, Ogg G, Nowak MA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med.* (1997) 3:212–7.
86. Klotzel PM. Antigen processing by the proteasome. *Nat Rev Mol Cell Biol.* (2001) 2:179–87. doi: 10.1038/35056572
87. Martin E, Carlson JM, Le AQ, Chopera DR, McGovern R, Rahman MA, et al. Early immune adaptation in HIV-1 revealed by population-level approaches. *Retrovirology* (2014) 11:64. doi: 10.1186/s12977-014-0064-1
88. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* (2002) 296:1439–43. doi: 10.1126/science.1069660
89. Zhang SC, Martin E, Shimada M, Godfrey SB, Fricke J, Locastro S, et al. Aminopeptidase substrate preference affects HIV epitope presentation and predicts immune escape patterns in HIV-infected individuals. *J Immunol.* (2012) 188:5924–34. doi: 10.4049/jimmunol.1200219
90. Tenzer S, Stoltze L, Schonfisch B, Dengjel J, Muller M, Stevanovic S, et al. Quantitative analysis of prion-protein degradation by constitutive and immuno-20S proteasomes indicates differences correlated with disease susceptibility. *J Immunol.* (2004) 172:1083–91. doi: 10.4049/jimmunol.172.2.1083
91. Toes RE, Nussbaum AK, Degermann S, Schirle M, Emmerich NP, Kraft M, et al. Discrete cleavage motifs of constitutive and immunoproteasomes revealed by quantitative analysis of cleavage products. *J Exp Med.* (2001) 194:1–12. doi: 10.1084/jem.194.1.1
92. Lauvau G, Kakimi K, Niedermann G, Ostankovitch M, Yotnda P, Firat H, et al. Human transporters associated with antigen processing (TAPs) select epitope precursor peptides for processing in the endoplasmic reticulum and presentation to T cells. *J Exp Med.* (1999) 190:1227–40.
93. Chang SC, Momburg F, Bhutani N, Goldberg AL. The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a “molecular ruler” mechanism. *Proc Natl Acad Sci USA.* (2005) 102:17107–12. doi: 10.1073/pnas.0500721102
94. Saveanu L, Carroll O, Lindo V, Del Val M, Lopez D, Lepelletier Y, et al. Concerted peptide trimming by human ERAP1 and ERAP2 aminopeptidase complexes in the endoplasmic reticulum. *Nat Immunol.* (2005) 6:689–97. doi: 10.1038/ni1208
95. Yusim K, Kesmir C, Gaschen B, Addo MM, Altfeld M, Brunak S, et al. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J Virol.* (2002) 76:8757–68. doi: 10.1128/JVI.76.17.8757-8768.2002
96. Draenert R, Le Gall S, Pfaffert KJ, Leslie AJ, Chetty P, Brander C, et al. Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection. *J Exp Med.* (2004) 199:905–15. doi: 10.1084/jem.20031982
97. Almeida JR, Price DA, Papagno L, Arkoub ZA, Sauce D, Bornstein E, et al. Superior control of HIV-1 replication by CD8+ T cells is reflected by their avidity, polyfunctionality, and clonal turnover. *J Exp Med.* (2007) 204:2473–85. doi: 10.1084/jem.20070784
98. Lazaro E, Godfrey SB, Stamegna P, Ogbechie T, Kerrigan C, Zhang M, et al. Differential HIV epitope processing in monocytes and CD4 T cells affects cytotoxic T lymphocyte recognition. *J Infect Dis.* (2009) 200:236–43. doi: 10.1086/599837
99. Ranasinghe SR, Kramer HB, Wright C, Kessler BM, di Gleria K, Zhang Y, et al. The antiviral efficacy of HIV-specific CD8(+) T-cells to a conserved epitope is heavily dependent on the infecting HIV-1 isolate. *PLoS Pathog.* (2011) 7:e1001341. doi: 10.1371/journal.ppat.1001341
100. Schmidt J, Iversen AK, Tenzer S, Gostick E, Price DA, Lohmann V, et al. Rapid antigen processing and presentation of a protective and immunodominant HLA-B*27-restricted hepatitis C virus-specific CD8+ T-cell epitope. *PLoS Pathog.* (2012) 8:e1003042. doi: 10.1371/journal.ppat.1003042
101. Meydan C, Otu HH, Sezerman OU. Prediction of peptides binding to MHC class I and II alleles by temporal motif mining. *BMC Bioinformatics* (2013) 14 (Suppl. 2):S13. doi: 10.1186/1471-2105-14-S2-S13
102. Klooverpris HN, Cole DK, Fuller A, Carlson J, Beck K, Schauenburg AJ, et al. A molecular switch in immunodominant HIV-1-specific CD8 T-cell epitopes shapes differential HLA-restricted escape. *Retrovirology* (2015) 12:20. doi: 10.1186/s12977-015-0149-5
103. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med.* (2003) 54:535–51. doi: 10.1146/annurev.med.54.101601.152346
104. Goulder PJ, Watkins DI. HIV and SIV CTL escape: implications for vaccine design. *Nat Rev Immunol.* (2004) 4:630–40. doi: 10.1038/nri1417
105. McMichael AJ, Borrow P, Tomaras GD, Goonetiilleke N, Haynes BF. The immune response during acute HIV-1 infection: clues for vaccine development. *Nat Rev Immunol.* (2010) 10:11–23. doi: 10.1038/nri2674
106. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* (2004) 432:769–75. doi: 10.1038/nature03113
107. Altfeld M, Kalife ET, Qi Y, Streeck H, Lichterfeld M, Johnston MN, et al. HLA alleles associated with delayed progression to AIDS contribute strongly to the initial CD8(+) T cell response against HIV-1. *PLoS Med.* (2006) 3:e403. doi: 10.1371/journal.pmed.0030403
108. Du Y, Zhang TH, Dai L, Zheng X, Gorin AM, Oishi J, et al. Effects of mutations on replicative fitness and major histocompatibility complex class I binding affinity are among the determinants underlying cytotoxic-T-lymphocyte escape of HIV-1 gag epitopes. *MBio* (2017) 8:e01050-17. doi: 10.1128/mBio.01050-17
109. Mlisana K, Werner L, Garrett NJ, McKinnon LR, van Loggenberg F, Passmore JA, et al. Rapid disease progression in HIV-1 subtype C-infected South African women. *Clin Infect Dis.* (2014) 59:1322–31. doi: 10.1093/cid/ciu573
110. Altfeld M, Addo MM, Rosenberg ES, Hecht FM, Lee PK, Vogel M, et al. Influence of HLA-B57 on clinical presentation and viral control during acute HIV-1 infection. *AIDS* (2003) 17:2581–91. doi: 10.1097/01.aids.0000096870.36052.b6
111. Amornkul PN, Karita E, Kamali A, Rida WN, Sanders EJ, Lakhi S, et al. Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. *AIDS* (2013) 27:2775–86. doi: 10.1097/QAD.0000000000000012
112. Dorak MT, Tang J, Tang S, Penman-Aguilar A, Coutinho RA, Goedert JJ, et al. Influence of human leukocyte antigen-B22 alleles on the course of human immunodeficiency virus type 1 infection in 3 cohorts of white men. *J Infect Dis.* (2003) 188:856–63. doi: 10.1086/378071
113. Kaslow RA, Duquesnoy R, VanRaden M, Kingsley L, Marrari M, Friedman H, et al. A1, Cw7, B8, DR3 HLA antigen combination associated with

- rapid decline of T-helper lymphocytes in HIV-1 infection. *A report from the multicenter AIDS cohort study*. *Lancet* (1990) 335:927–30.
114. Grant M, Larijani M. Evasion of adaptive immunity by HIV through the action of host APOBEC3G/F enzymes. *AIDS Res Ther.* (2017) 14:44. doi: 10.1186/s12981-017-0173-8
115. Kim EY, Lorenzo-Redondo R, Little SJ, Chung YS, Phalora PK, Maljkovic Berry I, et al. Human APOBEC3 induced mutation of human immunodeficiency virus type-1 contributes to adaptation and evolution in natural infection. *PLoS Pathog.* (2014) 10:e1004281. doi: 10.1371/journal.ppat.1004281
116. Monajemi M, Woodworth CE, Benkaroun J, Grant M, Larijani M. Emerging complexities of APOBEC3G action on immunity and viral fitness during HIV infection and treatment. *Retrovirology* (2012) 9:35. doi: 10.1186/1742-4690-9-35
117. Monajemi M, Woodworth CE, Zipperlen K, Gallant M, Grant MD, Larijani M. Positioning of APOBEC3G/F mutational hotspots in the human immunodeficiency virus genome favors reduced recognition by CD8+ T cells. *PLoS ONE* (2014) 9:e93428. doi: 10.1371/journal.pone.0093428
118. Squires KD, Monajemi M, Woodworth CE, Grant MD, Larijani M. Impact of APOBEC mutations on CD8+ T cell recognition of HIV epitopes varies depending on the restricting HLA. *J Acquir Immune Defic Syndr.* (2015) 70:172–8. doi: 10.1097/QAI.0000000000000689
119. Wood N, Bhattacharya T, Keele BF, Giorgi E, Liu M, Gaschen B, et al. HIV evolution in early infection: selection pressures, patterns of insertion and deletion, and the impact of APOBEC. *PLoS Pathog.* (2009) 5:e1000414. doi: 10.1371/journal.ppat.1000414
120. Casartelli N, Guivel-Benhassine F, Bouziat R, Brandler S, Schwartz O, Moris A. The antiviral factor APOBEC3G improves CTL recognition of cultured HIV-infected T cells. *J Exp Med.* (2010) 207:39–49. doi: 10.1084/jem.20091933
121. Champiat S, Garrison KE, Raposo RA, Burwitz BJ, Reed J, Tandon R, et al. T cells target APOBEC3 proteins in human immunodeficiency virus type 1-infected humans and simian immunodeficiency virus-infected Indian rhesus macaques. *J Virol.* (2013) 87:6073–80. doi: 10.1128/JVI.00579-12
122. Geller R, Domingo-Calap P, Cuevas JM, Rossolillo P, Negroni M, Sanjuan R. The external domains of the HIV-1 envelope are a mutational cold spot. *Nat Commun.* (2015) 6:8571. doi: 10.1038/ncomms9571
123. Gill DK, Huang Y, Levine GL, Sambor A, Carter DK, Sato A, et al. Equivalence of ELISpot assays demonstrated between major HIV network laboratories. *PLoS ONE* (2010) 5:e14330. doi: 10.1371/journal.pone.0014330
124. Nielsen M, Lundegaard C, Lund O. Prediction of MHC class II binding affinity using SMM-align, a novel stabilization matrix alignment method. *BMC Bioinformatics* (2007) 8:238. doi: 10.1186/1471-2105-8-238
125. Peters B, Sidney J, Bourne P, Bui HH, Buus S, Doh G, et al. The immune epitope database and analysis resource: from vision to blueprint. *PLoS Biol.* (2005) 3:e91. doi: 10.1371/journal.pbio.0030091
126. He Y, Rappuoli R, De Groot AS, Chen RT. Emerging vaccine informatics. *J Biomed Biotechnol.* (2010) 2010:218590. doi: 10.1155/2010/218590
127. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* (2009) 61:1–13. doi: 10.1007/s00251-008-0341-z
128. Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8:10 and 11 using prediction tools trained on 9mers. *Bioinformatics* (2008) 24:1397–8. doi: 10.1093/bioinformatics/btm128
129. Lundegaard C, Lund O, Buus S, Nielsen M. Major histocompatibility complex class I binding predictions as a tool in epitope discovery. *Immunology* (2010) 130:309–18. doi: 10.1111/j.1365-2567.2010.03300.x
130. Lundegaard C, Lund O, Kesmir C, Brunak S, Nielsen M. Modeling the adaptive immune system: predictions and simulations. *Bioinformatics* (2007) 23:3265–75. doi: 10.1093/bioinformatics/btm471
131. Hakenberg J, Nussbaum AK, Schild H, Rammensee HG, Kuttler C, Holzhueter HG, et al. MAPPP: MHC class I antigenic peptide processing prediction. *Appl Bioinformatics.* (2003) 2:155–8.
132. Kuttler C, Nussbaum AK, Dick TP, Rammensee HG, Schild H, Hader KP. An algorithm for the prediction of proteasomal cleavages. *J Mol Biol.* (2000) 298:417–29. doi: 10.1006/jmbi.2000.3683
133. Nussbaum AK, Kuttler C, Hader KP, Rammensee HG, Schild H. PAPROC: a prediction algorithm for proteasomal cleavages available on the WWW. *Immunogenetics* (2001) 53:87–94. doi: 10.1007/s002510100300
134. Saxova P, Buus S, Brunak S, Kesmir C. Predicting proteasomal cleavage sites: a comparison of available methods. *Int Immunol.* (2003) 15:781–7. doi: 10.1093/intimm/dxg084
135. Hertz T, Ahmed H, Friedrich DP, Casimiro DR, Self SG, Corey L, et al. HIV-1 vaccine-induced T-cell responses cluster in epitope hotspots that differ from those induced in natural infection with HIV-1. *PLoS Pathog.* (2013) 9:e1003404. doi: 10.1371/journal.ppat.1003404
136. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol.* (2017) 199:3360–8. doi: 10.4049/jimmunol.1700893
137. Lin HH, Ray S, Tongchusak S, Reinherz EL, Brusica V. Evaluation of MHC class I peptide binding prediction servers: applications for vaccine research. *BMC Immunol.* (2008) 9:8. doi: 10.1186/1471-2172-9-8
138. Peters B, Bui HH, Frankild S, Nielsen M, Lundegaard C, Kostem E, et al. A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Comput Biol.* (2006) 2:e65. doi: 10.1371/journal.pcbi.0020065
139. Iversen AK, Stewart-Jones G, Learn GH, Christie N, Sylvester-Hvid C, Armitage AE, et al. Conflicting selective forces affect T cell receptor contacts in an immunodominant human immunodeficiency virus epitope. *Nat Immunol.* (2006) 7:179–89. doi: 10.1038/ni1298
140. Kloverpris HN, Leslie A, Goulder P. Role of HLA adaptation in HIV evolution. *Front Immunol.* (2015) 6:665. doi: 10.3389/fimmu.2015.00665
141. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* (2016) 8:33. doi: 10.1186/s13073-016-0288-x
142. Nielsen M, Lundegaard C, Lund O, Kesmir C. The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* (2005) 57:33–41. doi: 10.1007/s00251-005-0781-7
143. Kijak GH, Janini LM, Tovanabutra S, Sanders-Buell E, Arroyo MA, Robb ML, et al. Variable contexts and levels of hypermutation in HIV-1 proviral genomes recovered from primary peripheral blood mononuclear cells. *Virology* (2008) 376:101–11. doi: 10.1016/j.virol.2008.03.017
144. Rucevic M, Boucau J, Dinter J, Kourjian G, Le Gall S. Mechanisms of HIV protein degradation into epitopes: implications for vaccine design. *Viruses* (2014) 6:3271–92. doi: 10.3390/v6083271
145. Allen TM, Altfield M, Yu XG, O'Sullivan KM, Lichterfeld M, Le Gall S, et al. Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. *J Virol.* (2004) 78:7069–78. doi: 10.1128/JVI.78.13.7069-78.2004
146. Carlson JM, Brumme CJ, Martin E, Listgarten J, Brockman MA, Le AQ, et al. Correlates of protective cellular immunity revealed by analysis of population-level immune escape pathways in HIV-1. *J Virol.* (2012) 86:13202–16. doi: 10.1128/JVI.01998-12
147. Guan P, Doytchinova IA, Flower DR. HLA-A3 supermotif defined by quantitative structure-activity relationship analysis. *Protein Eng.* (2003) 16:11–8. doi: 10.1093/proeng/gzg005
148. Steers NJ, Currier JR, Jobe O, Tovanabutra S, Ratto-Kim S, Marovich MA, et al. Designing the epitope flanking regions for optimal generation of CTL epitopes. *Vaccine* (2014) 32:3509–16. doi: 10.1016/j.vaccine.2014.04.039
149. Ebrahimi D, Anwar F, Davenport MP. APOBEC3 has not left an evolutionary footprint on the HIV-1 genome. *J Virol.* (2011) 85:9139–46. doi: 10.1128/JVI.00658-11
150. Fryer HR, Frater J, Duda A, Palmer D, Phillips RE, McLean AR. Cytotoxic T-lymphocyte escape mutations identified by HLA association favor those which escape and revert rapidly. *J Virol.* (2012) 86:8568–80. doi: 10.1128/JVI.07020-11
151. Fryer HR, Frater J, Duda A, Roberts MG, Investigators ST, Phillips RE, et al. Modelling the evolution and spread of HIV immune escape mutants. *PLoS Pathog.* (2010) 6:e1001196. doi: 10.1371/journal.ppat.1001196

152. Burns MB, Temiz NA, Harris RS. Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet.* (2013) 45:977–83. doi: 10.1038/ng.2701
153. Chan K, Roberts SA, Klimczak LJ, Sterling JF, Saini N, Malc EP, et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat Genet.* (2015) 47:1067–72. doi: 10.1038/ng.3378
154. Henderson S, Chakravarthy A, Su X, Boshoff C, Fenton TR. APOBEC-mediated cytosine deamination links PIK3CA helical domain mutations to human papillomavirus-driven tumor development. *Cell Rep.* (2014) 7:1833–41. doi: 10.1016/j.celrep.2014.05.012
155. Nik-Zainal S, Alexandrov LB, Wedge DC, Van Loo P, Greenman CD, Raine K, et al. Mutational processes molding the genomes of 21 breast cancers. *Cell* (2012) 149:979–93. doi: 10.1016/j.cell.2012.04.024
156. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* (2015) 160:48–61. doi: 10.1016/j.cell.2014.12.033
157. Swanton C, McGranahan N, Starrett GJ, Harris RS. APOBEC enzymes: mutagenic fuel for cancer evolution and heterogeneity. *Cancer Discov.* (2015) 5:704–12. doi: 10.1158/2159-8290.CD-15-0344
158. Venkatesan S, Rosenthal R, Kanu N, McGranahan N, Bartek J, Quezada SA, et al. Perspective: APOBEC mutagenesis in drug resistance and immune escape in HIV and cancer evolution. *Ann Oncol.* (2018) 29:563–72. doi: 10.1093/annonc/mdy003
159. Borzooee F, Asgharpour M, Quinlan E, Grant MD, Larijani M. Viral subversion of APOBEC3s: lessons for anti-tumor immunity and tumor immunotherapy. *Int Rev Immunol.* (2018) 37:151–64. doi: 10.1080/08830185.2017.1403596

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Borzooee, Joris, Grant and Larijani. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Human Endogenous Retroviruses Are Ancient Acquired Elements Still Shaping Innate Immune Responses

Nicole Grandi¹ and Enzo Tramontano^{1,2*}

¹ Laboratory of Molecular Virology, Department of Life and Environmental Sciences, University of Cagliari, Cagliari, Italy,

² Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Cagliari, Italy

OPEN ACCESS

Edited by:

Tara Patricia Hurst,
Abcam, United Kingdom

Reviewed by:

Alexander Emmer,
Martin Luther University of
Halle-Wittenberg, Germany
Renée Nicole Douville,
University of Winnipeg, Canada

*Correspondence:

Enzo Tramontano
tramont@unica.it

Specialty section:

This article was submitted to
Microbial Immunology,
a section of the journal
Frontiers in Immunology

Received: 25 May 2018

Accepted: 20 August 2018

Published: 10 September 2018

Citation:

Grandi N and Tramontano E (2018)
Human Endogenous Retroviruses Are
Ancient Acquired Elements Still
Shaping Innate Immune Responses.
Front. Immunol. 9:2039.
doi: 10.3389/fimmu.2018.02039

About 8% of our genome is composed of sequences with viral origin, namely human Endogenous Retroviruses (HERVs). HERVs are relics of ancient infections that affected the primates' germ line along the last 100 million of years, and became stable elements at the interface between self and foreign DNA. Intriguingly, HERV co-evolution with the host led to the domestication of activities previously devoted to the retrovirus life cycle, providing novel cellular functions. For example, selected HERV envelope proteins have been coopted for pregnancy-related purposes, and proviral Long Terminal Repeats participate in the transcriptional regulation of various cellular genes. Given the HERV persistence in the host genome and its basal expression in most healthy tissues, it is reasonable that human defenses should prevent HERV-mediated immune activation. Despite this, HERVs and their products (including RNA, cytosolic DNA, and proteins) are still able to modulate and be influenced by the host immune system, fascinatingly suggesting a central role in the evolution and functioning of the human innate immunity. Indeed, HERV sequences had been major contributors in shaping and expanding the interferon network, dispersing inducible genes that have been occasionally domesticated in various mammalian lineages. Also the HERV integration within or near to genes encoding for critical immune factors has been shown to influence their activity, or to be responsible for their polymorphic variation in the human population, such as in the case of an HERV-K(HML10) provirus in the major histocompatibility complex region. In addition, HERV expressed products have been shown to modulate innate immunity effectors, being therefore often related on the one side to inflammatory and autoimmune disorders, while on the other side to the control of excessive immune activation through their immunosuppressive properties. Finally, HERVs have been proposed to establish a protective effect against exogenous infections. The present review summarizes the involvement of HERVs and their products in innate immune responses, describing how their intricate interplay with the first line of human defenses can actively contribute either to the host protection or to his damage, implying a subtle balance between the persistence of HERV expression and the maintenance of a basal immune alert.

Keywords: HERV, endogenous retroviruses, innate immunity, interferon, evolution, autoimmunity, cancer

INTRODUCTION

Our genome includes an impressive proportion of repetitive elements, among which human endogenous retroviruses (HERVs) account for about the 8% (1). HERVs are DNA sequences of retroviral origin that have been acquired along the last 100 million of years through multiple integrations by now-extinct exogenous retroviruses (2, 3). Peculiarly, while known retroviruses target the somatic cells, these ancestral infections affected the primate germ line, leading to the vertical transmission of HERV relics through the offspring (**Figure 1**). It is however still not clear if the HERV-originating exogenous retroviruses had germ line cells as main/unique target or infected this population by chance (4). In general, the mechanism that formed HERV insertions is analogous to the one used by exogenous retrovirus replication. In both cases, once into the cytoplasm, the RNA genome is reverse transcribed into a double-stranded DNA (dsDNA) by the viral reverse transcriptase (5). The so-obtained proviral DNA is then integrated in the host chromosomes through the viral integrase that interacts with cellular cofactors (6). At this point, the provirus expression generates a set of mRNAs encoding for the different viral proteins. In the presence of a functional reverse transcriptase, the full-length mRNA can also be reverse transcribed, producing a proviral cDNA theoretically competent for new integration events. The action of the host editing systems and the genomic substitution rate, however, often made HERV proviruses defective, leaving only a residual protein coding capacity but more often producing non-coding RNAs. HERVs share with exogenous retroviruses the typical proviral structure, being normally composed of two long terminal repeats (LTRs) that flank the internal portion of the viral genes *gag*, *pro-pol* and *env* (**Figure 1**). The LTRs are formed during the reverse transcription and have a regulatory significance for viral genes' expression, including promoters, enhancers and polyadenylation signals. The retroviral genes encode for the structural components, i.e. matrix, capsid and nucleocapsid (*gag*) and the envelope surface and transmembrane subunits (*env*), as well as for the enzymes involved in the viral life cycle, namely protease (*pro*), reverse transcriptase and integrase (*pol*) (**Figure 1**). While only these simple retroviral gene products have been identified in the majority of ERVs, some groups are known to have a more complex genome that encodes for additional proteins. This is the case of HERV-K(HML2) sequences, whose *env* gene—depending on the presence or the absence of a characteristic 292-bp deletion—can originate two splicing variants, Np9 and Rec, respectively (7, 8). Both proteins have been investigated due to their possible oncogenic properties [for a recent review, see (9)]. Only very recently, the presence of a putative *rec* gene has been also reported in another HERV-K group, namely HML10 (10). Aside from protein-coding genes, HERV proviruses harbor a primer-binding site (PBS) and a polypurine tract (PPT) located between the 5'LTR and *gag* and between *env* and the 3'LTR, respectively (**Figure 1**). Both have an important role during reverse transcription: the PBS binds the cellular tRNA priming the synthesis of the (–)strand DNA, while the PPT acts as a primer for the (+)strand DNA production.

The PBS sequence had also traditionally been used for HERV classification, and the current nomenclature still often relies on the tRNA type recognizing the different HERV groups' PBS (e.g., HERV-K for lysine tRNA, HERV-W for tryptophan tRNA, etc.). However, HERV classification is still ongoing, and the use of different methods in the last decades led often to multiple designations for the same HERV sequences and groups (11). Beside the use of the above PBS-based taxonomy, a minority of HERV groups has been named following unconventional criteria, such as the presence of a proximal cellular gene (e.g., HERV-ADP) or a peculiar amino acid motif (e.g., HERV-FRD). Clearly, such a lack of established rules generated confusion in the field, underlining the need of univocal guidelines for the naming of HERV groups, their single members and the latter expressed products (11, 12). Hence, the above mentioned nomenclature criteria are now considered inadequate, being not based on phylogenetic aspects and, regarding the PBS-based one, not taking into account the frequent occurrence of alternative PBS types (3, 13, 14). Currently, HERVs are broadly divided into three classes according to their similarity to exogenous members: class I (*gammaretrovirus*- and *epsilon**retrovirus*-like), class II (*betaretrovirus*-like) and class III (*spumaretrovirus*-like). The classification of the various HERV groups is instead based on the phylogenetic relationships among the different groups - considering above all the highly-conserved *pol* gene - and can be corroborated by structural features found in all the members of the same genus or class (3, 11, 13). For what concerns the HERV nomenclature, a revised naming system has been introduced by the Human Genome Organization Gene Nomenclature Committee (12). Hence, for clarity, we provide a list of aliases in the presence of multiple HERV designation (**Table S1**).

An updated analysis of HERV sequences within the human genome has been recently performed (3) with the software RetroTector (15). A multi-step approach allowed to classify about 3200 HERV insertions in 39 “canonical” groups plus 31 “non-canonical” clades showing a mosaic structure arisen from recombination and secondary integrations (3). This classification provided a comprehensive overview of HERV diversity, being also a useful background for the still-ongoing characterization of the different HERV groups at the genomic level. Such an exact knowledge of the individual HERV sequences' localization and coding capacity could in fact represent a great advance in the understanding of their potential effects in both physiological and pathological contexts (16). In particular, while in the former some HERV locus-specific activities have already been demonstrated in placenta (9, 17–19), the tentative link between HERVs and human diseases has not lead to any definitive association, yet. In this regard, while many reports investigated the various HERV groups' overall expression in diseased tissues, very few studies tried to assign it to specific proviral loci of origin and to analyze their differential expression between patients and healthy individuals. This, together with the above-mentioned lack of knowledge about individual HERV sequences and the absence of standardized methodologies for HERV association studies,

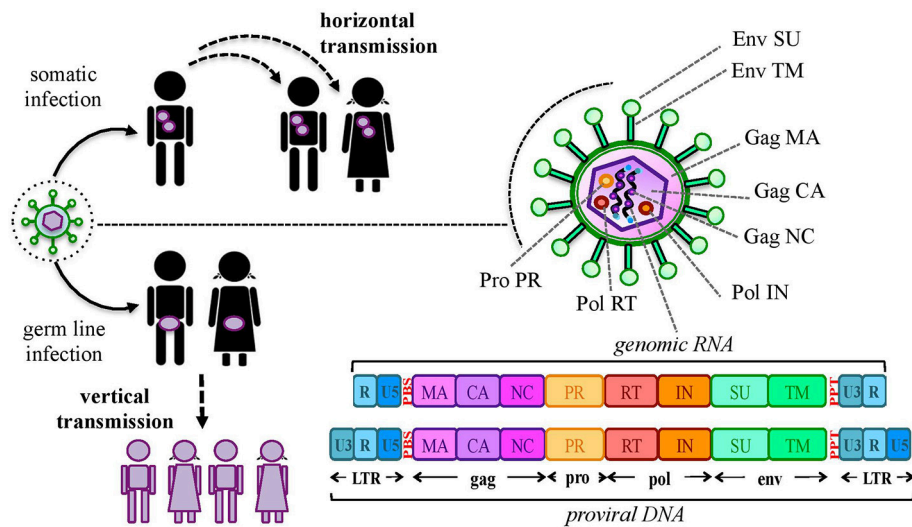


FIGURE 1 | Origin and general structure of HERV sequences. Exogenous retroviruses normally infect a specific type of somatic cells, being diffused from a host to new individuals by a horizontal transmission. In the case of HERVs, the ancestral retroviral infection affected the germ line cells: in this way, the proviral sequences have been endogenized and vertically transmitted to all the cells of descendant individuals. HERVs have been so inherited in a Mendelian fashion to the offspring, being fixed in the human population. The general structure of a full-length HERV provirus is represented: the two Long Terminal Repeats (LTRs) are formed during the reverse transcription of the viral RNA genome and flank the *gag*, *pro*, *pol*, and *env* genes. The primer binding site (PBS) and the polypurine tract (PPT) are located between 5′LTR and *gag* and between *env* and 3′LTR, respectively. The viral genes encode for the structural and non-structural proteins found in the viral particle: *gag* matrix (MA), capsid (CA) and nucleocapsid (NC); *pro-pol* protease (PR)—reverse transcriptase (RT) and integrase (IN); *env* surface (SU) and transmembrane (TM) subunits. While in exogenous retroviral infections the integrated provirus is transcribed by the cellular machinery to release new virions, the HERV persistence within the host genome and the action of cellular editing systems led to the accumulation of mutations that often made the proviruses coding-defective and thus unable to produce infectious particles.

prevented until now the identification of precise molecular mechanisms of pathogenesis and, thus, the unambiguous link to any disorder (16, 20). Nevertheless, even if their significance in disease etiology is still uncertain, it is by now clear that HERVs can have an impact on the host in several ways. While some effects depend on the mere presence of HERV sequences within the human genome (16), in some other instances, HERV expression can provide RNAs and proteins potentially able to *trans*-regulate human genes and to influence the host immunity (21–24). It has been shown that HERV expression products can act as pathogen-associated molecular patterns (PAMPs), triggering the cellular receptors responsible for the first line of defenses (25–27). They can moreover provide antigenic epitopes recognized by lymphocytes (especially through molecular mimicry with exogenous viral molecules) and stimulating the onset of specific T- and B- cells (28–30). These mechanisms might be particularly relevant for the tentative involvement of HERVs in autoimmunity and inflammatory diseases. On the contrary, HERVs have also been involved in the downregulation of the host immunity, starting from their role in maternal immune tolerance to the fetus to their suggested protective action against excessive immune activation (31–34). The present review will discuss the interplay between HERV insertions and their expression products and our defenses, with a special attention to their contribution in shaping and influencing the human innate immunity.

HERV ROLE IN THE EVOLUTION AND SHAPING OF THE HUMAN GENOME

About 75 years ago, the pioneer studies of Barbara McClintock suggested that transposable elements (TEs)—now known to constitute >40% of our DNA (1)—were not useless “junk DNA” but normal components of eukaryotic genomes that can have important regulatory roles (35). Nowadays, growing evidences confirm that TEs had a crucial role in the shaping and evolution of vertebrates’ genomes, contributing to the establishment of lineage-specific patterns of gene expression (36, 37). TE insertion appear to be highly conserved among mammals, showing an increased density in the proximity of cellular genes and being a main source of transcription factor binding sites and regulatory signals (37–42). This points out their relevance for the human development and transcriptional modulation throughout evolution (36, 38). Of course, the majority of TEs have been acquired several million of years ago, being by now silenced due to the accumulation of mutations in their coding sequence and to various cellular mechanism of transcriptional repression, such as histone hypermethylation. Despite this, many TEs are likely involved in a long-term co-evolution with the host, that exerted firstly a selective pressure against their detrimental effects and led then in some instances to their cooptation for biological processes (43, 44). Intriguingly, even the general epigenetic silencing of TEs could represent an ancestral adaptation of the harboring organism that, initially, evolved defense systems to downregulate

them and, then, developed strategies to control gene expression through their exaptation (44). In this view, the sporadic loss of epigenetic regulation could be not only linked to certain developmental stages and disorders, but might also constitute an opportunity to boost evolvability (44). Accordingly, despite controlling mechanisms, many HERV sequences still retain a residual expression capacity, leading to the production of either coding or non-coding RNA transcripts that can both influence the host biology. One of the most remarkable examples of the HERV impact on vertebrate physiology is represented by “syncytins,” an ensemble of Env proteins encoded by different HERV sequences in all eutherian mammals through a process of convergent evolution (45). Syncytins have in fact been domesticated independently by the various species, providing common and important functions for placenta development and physiology. In the case of the human genome, two *env* loci, namely *ERVWE1* (HERV-W, 7q21.2) and *ERVWE2* (HERV-FRD, 6p24.1), encode for the coopted Env proteins syncytin-1 and -2, respectively (17, 18). While syncytin-1 has a pivotal role in placental syncytiotrophoblast development and homeostasis (17, 18, 46, 47), syncytin-2 is thought to be involved in the maternal immune tolerance to the fetal allograft (32). The role of these and other HERV-derived Env proteins have been recently reviewed elsewhere (9).

Aside from protein production, the thousands of HERV sequences dispersed in our DNA contributed to the evolution of the primates' genome by providing an abundant source of regulatory elements. It is well known that our genetic information is organized in regulatory networks, involving both *cis*-regulatory sequences and *trans*-acting genes, and that their interaction is at the base of cellular plasticity and evolution (23, 44). Also HERVs participate to this complex interplay, being able to regulate the host genes' activity in several ways and at different expression levels (**Figure 2**). In fact, even the sole presence of HERV proviruses and solitary LTRs can influence cellular genes' activity (**Figure 2a**). Repetitive elements' insertion is a source of genomic modifications, possibly leading to the disruption or insertional mutagenesis of co-localized genes or promoting chromosomal rearrangements through non-allelic homologous recombination (**Figure 2a**). This was reported in male infertility, in which the intra-chromosomal recombination between two homologous HERV-I sequences located on chromosome Y is responsible for the microdeletion of the azoospermia factor a (48). Moreover, HERV LTRs can provide *cis*-regulatory activity to nearby cellular genes by enhancing their transcription or even providing alternative promoters and splicing signals, also with a remarkable tissue-specificity (**Figure 2a**). A representative example is the HERV-E provirus integrated upstream of an ancestral amylase gene and acting as a specific enhancer in parotid glands (49, 50). Furthermore, despite the frequent loss of protein-coding capacity, also the abundant production of HERV-derived non-coding RNAs (ncRNAs) such as microRNA and long ncRNA, may likely provide *cis*-regulatory elements able to modulate the expression of the host genes, either alone (e.g., providing a recognition motif for an RNA-binding protein) or in concert with cellular transcription factors (**Figure 2b**). This occurs for example in human embryonic stem cells, whose

pluripotency depends on nuclear long ncRNAs expressed by a HERV-H element and recruiting specific cellular transcriptional activators (51). In addition, HERV ncRNAs can act as “RNA sponges” binding and dampening microRNA families involved in the post-transcriptional regulation of gene expression, given that several human microRNAs have high sequence homology with HERVs (52–54). Such a miRNA sponge activity has been reported in the positive regulation of pluripotency in embryonic stem cells, which depends on the interaction of a HERV-H long ncRNA (HPAT5, locus 6q27) to complementary sequences in the let-7 microRNA family (54, 55) (**Figure 2b**). Finally, if a HERV protein is produced, it can be able to modulate the host genic expression through the biological activities previously involved in the virus life cycle and now providing new cellular functions (21) (**Figure 2c**). As an example, HERV Gag and Rec proteins are known to influence the stability, localization and translation of cellular transcripts (21). Accordingly, HML2 Rec was reported to interact with ~1,600 cellular mRNAs in embryonic cells and, especially, to influence their ribosome occupancy, possibly suggesting regulatory function coopted for early development (56). Similarly, the Arc Gag-like protein derived from an ancient Ty3/gypsy retrotransposon has been repurposed during brain evolution to mediate communications between neural cells, having an important role in the development and plasticity of the nervous system (57–59). In particular, Arc has been shown to assemble into capsids that include mRNA sequences to be transferred from a neuron to new recipient cells through extracellular vesicles, then undergoing activity-dependent translation (60).

Beside the role of individual HERV sequences, it has been proposed that the ensemble of TEs widespread in the human DNA can have a more extensive role. By constituting a sort of parallel regulatory network, TEs can possibly influence multiple host genes and, thus, shape whole pathways involved in complex cellular processes (22, 36, 61). Accordingly, more than one third of p53 binding sites in the human genome have been dispersed by class I HERV sequences, which have become major components of the p53 regulatory network (62). Hence, the acquisition of HERV insertions could have acted as a driving force for human genome plasticity and cellular networking, and, as described below, such TE-dependent modeling and functional renewal has been particularly critical in the evolution of innate immunity. Vertebrate genomes have been subjected to substantial rearrangements through the acquisition of HERVs and, intriguingly, such colonization has been concomitant to the development of important immune pathways (24). Major shifts have occurred in both innate and adaptive immune systems, increasing the complexity and specificity of vertebrate antiviral defenses (24). The loci of primate MHC (major histocompatibility complex) are in fact characterized by a high density of HERV integrations, which contributed to their remarkable plasticity (63) and led to genic variations among species. As an example, a HERV-K(HML10) provirus inserted within the ninth intron of human complement *C4A* gene (MHC class III), also called HERVKC4, is responsible for its dichotomous size variation (10, 64–66). Besides generating physical changes among species, such TE

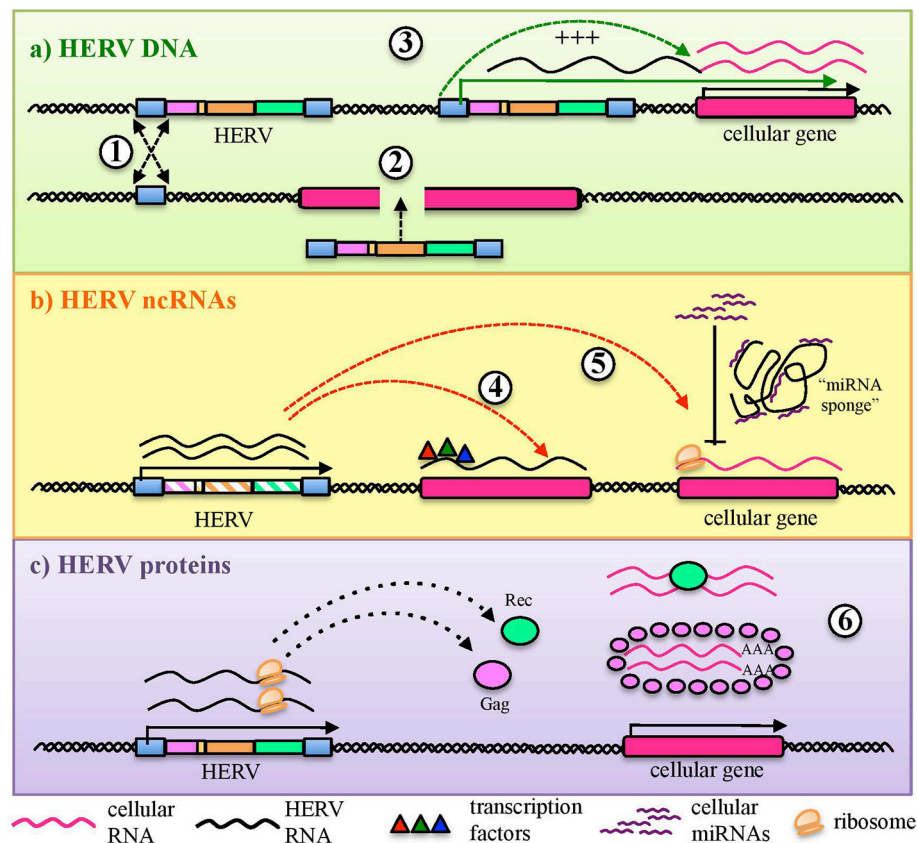


FIGURE 2 | HERV role in the regulation and shaping of human gene expression. HERVs can influence the host gene expression at several levels. **(a)** Integrated DNA sequences can trigger chromosomal rearrangements by non-allelic homologous recombination (1) or disrupt co-localized genes through insertional mutagenesis (2). Moreover, HERV sequences integrated in proximity to a cellular gene can provide alternative promoters or enhance its expression through LTR *cis*-regulatory elements (3). **(b)** HERV non-coding RNAs (ncRNAs) can also be able to *cis*-regulate cellular genes, even through the recruitment of cellular regulators (e.g.: transcription and splicing factors) (4). In addition, HERV ncRNA have been reported to act as “microRNA sponges,” binding and dampening microRNA families responsible for post-transcriptional modifications (5). **(c)** Finally, some HERV proteins can also regulate genic expression through their interaction with cellular mRNAs and the modulation of their transfer and ribosome occupancy (6).

insertions in immune gene introns could also account for regulatory effects, being present mostly in antisense orientation and subjected to bidirectional transcription (10). Noteworthy, mRNA sequences originated by classes of genes with a relatively recent expansion, such as the ones involved in immunity, are enriched in TEs, which are not present in the transcripts arising from highly conserved genes with basic functions (38). Overall, one of the soundest evidences about the role of HERVs in the shaping of pivotal immune systems regards the interferon (IFN) network, a crucial antiviral pathway for innate immunity and a fundamental effector to initiate and maintain adaptive responses (67). Intriguingly, Chuong and coauthors showed that HERV insertions greatly contributed to the evolution and amplification of IFN transcriptional network, dispersing independently a wide number of IFN-inducible enhancers in many mammalian genomes (22). In particular, the experimental deletion of a subset of these endogenous elements in human cell lines affected the activity of neighboring genes induced by IFN, impairing thus important immune pathway such as the AIM2 inflammasome

(22). Furthermore, due to their residual regulatory activities, HERV LTRs can act as promoter and/or enhancers after IFN-mediated stimulation (68). This has been reported for HERV-K(HML2) LTRs, which harbor two IFN-stimulated response elements (ISREs) activated by the IFN signaling, leading to the increased HML2 expression in response to inflammation (69, 70). In line with this, HERV-K *env* genes transcription can be stimulated by IFN α , encoding Env superantigens responsible for polyclonal T-cell activation (71).

HERVS AS ACTIVATORS OF ANTIVIRAL INNATE IMMUNITY

Innate immunity is the first and most ancient line of defenses against microbial infections, acting by a complex network that is conserved throughout the animal kingdom (72). When an infectious agent overcomes the organism physical barriers, the presence of conserved PAMPs (i.e., lipids, proteins, glycans,

and nucleic acids) allows their prompt recognition by innate immunity sensors, namely pattern recognition receptors (PRRs). PRRs are germ line-encoded receptors with a pivotal role in antiviral defenses, as well as in the response to self-injuries (73), recognizing PAMPs and danger-associated molecular patterns (DAMPs), i.e., molecules present in damaged or stressed tissues (74).

In vertebrates, PRRs include five major classes of receptors that are localized either transmembrane or in the cytosol (tm- and cytPRRs, respectively) of different cell types (73). tmPRRs are represented by the Toll Like Receptors (TLRs), detecting PAMPs either on the cell surface or in the endosomal compartments. cytPRRs include RIG-I-like receptors (RLRs), NOD-like receptors (NLRs), C-type lectin receptors (CLRs) and DNA sensors, all recognizing intracellular PAMPs (26, 75–77). cytPRRs are usually present in cells that can be actively infected by a given class of microbial agents, while tmPRR cell-extrinsic recognition is independent from the cell infection and usually occurs in cells devoted to pathogen detection (78). In both cases, PRR stimulation by microbial PAMPs activates a complex cascade of signaling that triggers the production of various pro-inflammatory molecules, including cytokines, chemokines and type I IFN (78, 79). These effectors have a double role. On the one side, they quickly establish an antimicrobial environment to counteract the infection. On the other side, the activation of PRRs expressed on antigen-presenting cells (APCs), especially dendritic cells (DCs), evokes the development of a long-lasting adaptive immunity, leading to the onset of specific cellular and humoral defenses (78). The latter are mainly represented by the clonal expansion of naïve cytotoxic T cells and the production of antibodies by B lymphocytes. Other important APCs are macrophages and B lymphocytes, which are albeit mostly involved in the stimulation of already-activated T cells. The overall innate immune signaling has been reviewed elsewhere (72, 79–81): in the following sections we will focus our attention on the specific pathways relevant to exogenous retrovirus sensing and, thus, possibly involved also in HERV immune recognition.

HERVs are integral parts of the human genome since millions of years, are highly transcribed during embryonic development and expressed at variable levels in several adult tissues. Hence, from the immunological point of view, HERV expressed products somehow stay at the interface between self-molecules and microbial antigens. In fact, also depending on the cellular and immunological surrounding, HERV-derived molecules can be either tolerated by human defenses or able to stimulate the human immunity, leading to their intensive investigation in various autoimmune diseases. In addition, the presence of other immune stimuli has been shown to be capable to influence HERV expression, suggesting a complex and multifaceted interplay that is still not completely clarified. In principle, the innate immune pathways activated by HERV-derived products are the same involved in the first line antiviral defenses counteracting exogenous retroviruses, including both tm- and cytPRRs (26) (**Figure 3**). In humans, TLRs from 1 to 10 are inserted in plasmatic or endosomal membranes and expressed on both innate immune cells and non-immune cell types, while cyt-PRRs

are soluble elements normally found in the cytosol or migrating to it when stimulated by the presence of retroviral molecules following cellular infection (80, 82). Both tm- and cytPRRs can potentially recognize HERV nucleic acids and proteins either due to their molecular identity with exogenous viral PAMPs or as DAMPs (26, 74). The upregulation of HERV transcription, even in the absence of a causal role in the initiation of immune-related pathogenesis, can thus provide numerous HERV-derived PAMPs or DAMPs able to further prompt inflammation, contributing to the symptomatology. This could likely occur in diseases such as autoimmunity and cancer, which have in common a general epigenetic de-regulation known to strongly liberate the non-specific retroelements' expression (9, 16). Hence, even if no definitive associations have been described yet, a growing body of studies is confirming that HERVs may have an impact on the human immune homeostasis, starting from their interaction with PRRs known to have a central role in the early response to HIV-1 infection, i.e., TLRs, RLRs, and cytoplasmic DNA-sensors (83).

Sensing of HERVs by Transmembrane PRRs

As mentioned above, different tmPRRs can potentially be activated by either HERV nucleic acids or proteins (84) (**Figure 3**). TLRs are the first PRRs to be identified and have been studied in greater detail: they can sense viral molecules either on the cellular surface, through plasmatic membrane TLRs (1, 2, 4, 5, 6, and 10), or in the endosome, through TLRs localized in the endosomal membranes (3, 7, 8, 9) (80). All TLRs have a common structure, characterized by a first domain that binds specific ligands, protruding outside the cell surface or inside the endosome, and a second conserved cytosolic domain that is required for the intracellular signaling. Their activation relies albeit on different types of molecules. Plasmatic membrane TLRs can recognize retroviral proteins, either as individual molecules or as components of viral-like structures (**Figure 3**). Accordingly, TLR2 and TLR4 are both able to detect the Env proteins of retroviruses, including HIV-1 gp120 that induce NF- κ B activation and proinflammatory cytokine secretion (85). In line with this, both receptors are upregulated in DCs, macrophages and peripheral blood mononuclear cells during HIV infection (86–88). Endosomal TLRs can instead be involved in the sensing of different retroviral nucleic acids, which should however be auto-phagocytized from the cytoplasm to the endosomal lumen to be detected (26, 74) (**Figure 3**). In particular, one of the most immunogenic nucleic acids PAMPs is dsRNA, which is normally not found in uninfected cells. Viral dsRNA can stimulate an innate response through TLR3 signaling (72) and, interestingly, the immune activity elicited by TLR3 agonists has been shown to provide protection from different viral infections, among which certain HIV strains (89). TLR3 has been also implicated in the HIV-1 transactivation (90), replication (91), and inflammation (92). Following proviral transcription, HERV ssRNA could be recognized by both TLR7 and -8. The same TLRs, in fact, are important for the recognition of the HIV RNA genome during acute infection, probably through the recognition of guanosine-

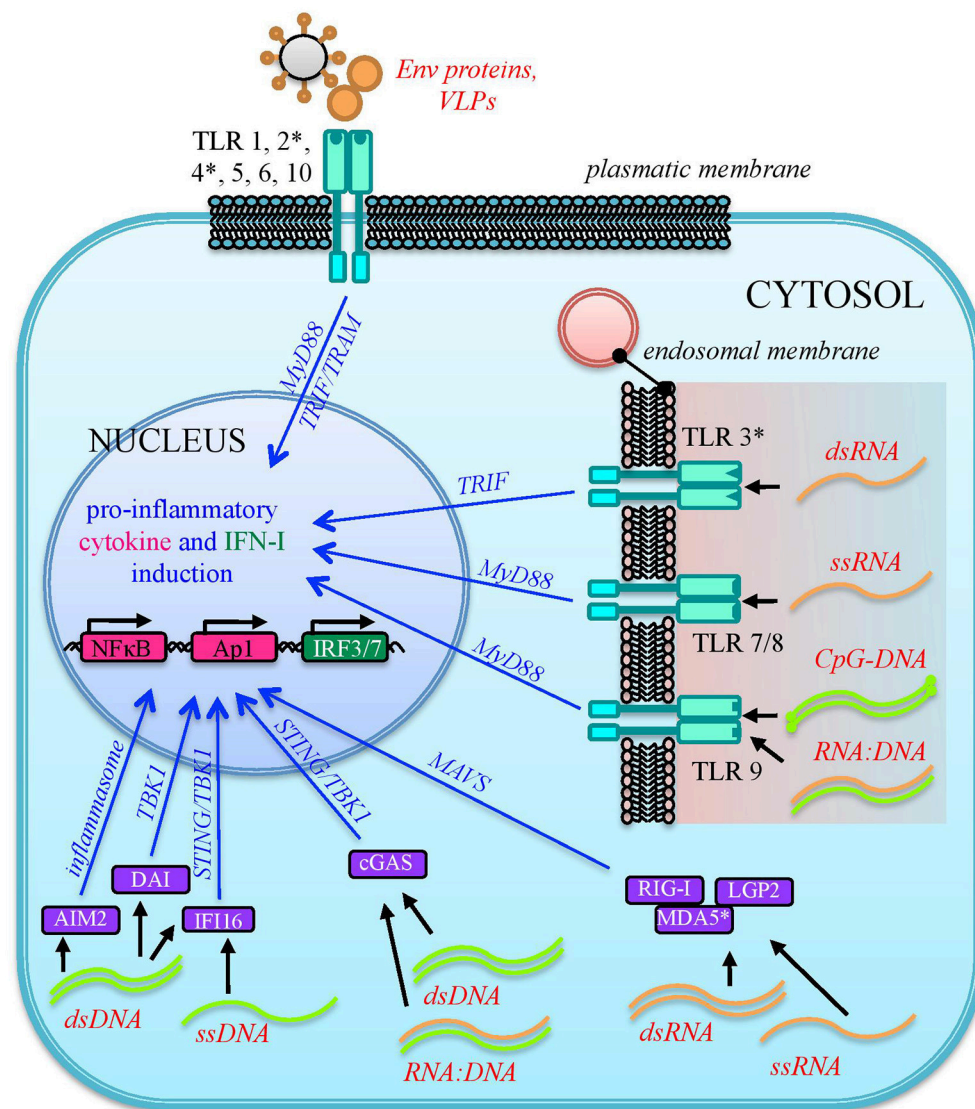


FIGURE 3 | Sensing of HERV molecules by innate immunity PRRs. Different HERV proteins and nucleic acids (overall indicated in red) can theoretically be detected as PAMPs or DAMPs by cellular sensors localized at the plasmatic or endosomal membranes (transmembrane PRRs or Toll Like Receptors, cyano) or present as soluble factors in the cytosol (cytosolic PRRs, violet). The sensing of these viral molecules by both kind of PRRs triggers a signaling cascade (blue arrows) that leads to the nuclear activation of immune genes encoding for pro-inflammatory effectors, represented by cytokines and type I IFN. The individual PRRs for which a direct interaction with HERV molecules has been reported are marked with an asterisk.

and uridine- rich ssRNA that stimulates the production of IFN- α and proinflammatory cytokines by DCs and macrophages (93). In the presence of reverse transcriptase activity, the HERV ssRNA can serve as a template to the synthesis of RNA:DNA hybrids, which can be sensed by TLR9 (94). The latter was in fact shown to sense RNA:DNA hybrids presenting viral-derived sequences, leading to the secretion of IFN-I and pro-inflammatory cytokines in DCs (94). TLR9 is also the sole DNA-sensing TLR and specifically binds unmethylated CpG-rich DNA (72, 82) that, however, is present mostly in bacteria and DNA viruses (72, 80). Recently, TLR9 agonists have been reported to induce HIV production in latently infected cells, being investigated as

possible therapeutic strategies to attack the proviral reservoir (95, 96).

In any of the above pathways, once activated, tmPRR dimeric receptors prime a complex intracellular cascade of signaling that, through several kinases and ubiquitinases, could lead to the activation and nuclear translocation of transcription factors that stimulate the expression of cytokines, chemokines and type I IFN (IFN-I) establishing an initial antiviral status and stimulating the adaptive immune defenses. Currently, even if all the mentioned tmPRRs could be potentially able to sense HERV-derived molecules, there are yet limited findings reporting the direct interaction between these ligands and few plasmatic

membrane TLRs. Contrarily, endosomal TLRs were widely shown to be involved in the control of murine ERVs (97) albeit still remaining poorly characterized for their interplay with human TEs.

In one of the most investigated groups, the HERV-W, the Env surface subunit has been characterized to be a potent stimulator of TLR4, harboring remarkable pro-inflammatory properties that might contribute to multiple sclerosis immunopathogenesis (98–100) (**Table 1**). The HERV-W Env was shown to interact with both TLR4 and CD14, inducing proinflammatory molecules that are prevalent in multiple sclerosis, such as IL-1, IL-6 and tumor necrosis factor α [(99, 109, 110), **Table 1**]. These immune effectors can likely contribute to the damage of brain populations, including DCs, oligodendrocytes and astrocytes (111). HERV-W Env was in fact shown to activate DCs and promote a Th1-like immune response (99), adding mechanistic insights to the tentative link between HERV-W sequences and multiple sclerosis [recently reviewed in Grandi and Tramontano (16)]. The interaction between HERV-W Env and TLR4 was also investigated in both murine and human oligodendrocyte precursors, leading to an increased production of cytokines and inducible nitric oxide synthase (101) (**Table 1**). In line with this, HERV-W Env overexpression in mice led to the development of experimental allergic encephalomyelitis (112) and inflammatory hallmarks typical of myelin injuries (113, 114). Such a HERV-W Env mediated pro-inflammatory effect was responsible for reduced oligodendrocytes differentiation and affected myelin expression and renewal (101), as also demonstrated by the fact that treatment with a specific antibody for HERV-W Env (GNbAC1) was able to rescue myelin expression (115). In addition, HERV-W Env hold a significant superantigen activity (116), that has also been proposed to play a role in demyelination, by evoking a polyclonal non-specific T-cell activation and the massive release of multiple cytokines (98, 109). A similar interaction between HERV-W Env and the TLR4 of pancreatic β cells has been proposed to promote autoimmune reactions and to affect insulin secretion in type I diabetes patients (117).

A second evidence of HERV pro-inflammatory potential comes from the investigation of HERV-K(HML2) group expression in psoriasis, another poorly understood autoimmune disorder. In this context, wild type and mutated HML2 dUTPases (metallo-enzymes that hydrolyzes dUTP preventing its incorporation into the viral DNA) were shown to interact *ex vivo* with TLR2 (108). Such interaction stimulated the expression of NF- κ B and induced Th1 and Th17 cytokine production in DCs and Langerhans-like cells as well as, even if at lower levels, in keratinocytes (108) (**Table 1**). This could suggest a possible role of HML2 dUTPase in the immunopathogenesis of psoriatic lesions, even if these observations have been performed in primary cells from healthy individuals and not confirmed *in vivo*. A subsequent study explored the HML2 dUTPase locus in a large number of psoriatic patients, reporting the association of some single nucleotides variants with a lower susceptibility to the disease (118). A small subset of these patients was also tested for dUTPase-induced cellular immunity, showing frequently increased B- and T-cell responses that were albeit not further characterized at the molecular level (118).

Finally, the stimulation of endosomal TLR3 by HERV dsRNA molecules has been observed after the treatment with DNA methyltransferase inhibitors, anticancer agents that were shown to remove methylation from the HERV promoter regions, causing their reactivation and subsequently triggering IFN α and β responses (27). Interestingly, such a HERV-mediated immune-stimulation is probably at the base of demethylating agent anticancer effect (see below).

Sensing of HERVs by Cytosolic PRRs

Besides transmembrane TLRs, other PRRs devoted to the antiviral response are present in the cytosol as soluble factors or reach this compartment after the sensing of microbial nucleic acids. Among them, the most relevant for antiviral immunity include the RLRs retinoic acid inducible gene I (RIG-I), sensing short dsRNA and ssRNA with a 5'-triphosphate moiety, and MDA5 (melanoma differentiated associated gene 5), that detects long dsRNA molecules (72). RIG-I was shown to be a crucial intracellular sensor for HIV nucleic acids, recognizing complex secondary structures in HIV RNA and stimulating IFN-I release by mononuclear cells (83). Accordingly, as a viral immune escape mechanism, HIV protease was shown to target RIG-I for lysosome degradation (119). Contrarily, MDA5 was not affected by HIV protease-mediated degradation (119), even if its upregulation has also been observed in HIV infection (120). Other cytPRRs, namely DNA sensors, recognize instead DNA of viral origin that, differently from nuclear DNA, can be found in the cytoplasm during retroviral replication. Particularly, in HIV-1 infected cells, the replication intermediates cDNA, ssDNA, and RNA:DNA hybrids can all be detected by cytoplasmic DNA sensing proteins (83, 121). Viral dsDNA can be sensed by the DNA-dependent activator of IFN (DAI) and the IFN γ inducible protein 16 (IFI16), both triggering the STING-TBK1-IRF3 signaling, as well as by the absent in melanoma protein 2 (AIM2), activating the inflammasome pathway (26, 82, 122). Another important sensor for cytoplasmic dsDNA is cGAS (cyclic GMP-AMP synthase), which can also recognize DNA-RNA hybrids and mtDNA released after mitochondrial damage (82). The main dsDNA sensors relevant to HIV infection are IFI16 and cGAS, which are both upregulated in the absence of antiretroviral therapy and associated with chronic immune activation (123). For what concerns foreign ssDNA, stem-rich secondary structures in HIV-1 ssDNA were shown to induce IFI16 in macrophages (124).

Given that retrovirus replication takes place mostly in the cytoplasm, cytPRRs are known to be crucial to the detection of RNA and DNA originated by exogenous retroviruses, while very few evidences about the molecular sensing of HERV nucleic acids are available yet (26). Particularly, similarly to TLR3, also MDA5 has been shown to sense HERV dsRNA triggered by the treatment with DNA methyltransferase inhibitors, evoking IFN production (27). A role in the prevention of HERV DNA/RNA accumulation has been proposed for certain enzymes involved in the cytoplasmic homeostasis of nucleic acids, which could possibly provide some protection against HERV-mediated immune activation (9). Such hypothesis derives from observations made in mice deficient for the

TABLE 1 | HERVs for which a direct role in the modulation of innate immunity has been reported.

Group	Locus	Name	Type	Effect	Mechanism	Reference
HERV-W	?	–	Env SU	immune-activation	stimulation of CD14 and TLR4, monocyte and dendritic cell activation, inflammatory cytokines production	(99)
	?	–	Env	immune-activation	stimulation of cytokine and inducible nitric oxide synthase production	(101)
	HERWE1	syncytin-1	Env TM	immune-suppression	Th1 to Th2 shift in maternal immune-tolerance	(102)
HERV-FRD	HERWE2	syncytin-2	Env TM	immune-suppression	Th1 to Th2 shift in maternal immune-tolerance	(32)
HERV-H	HERV-H3	Env-59	Env TM	immune-suppression	inhibition of inflammation effectors	(103, 104)
	HERV-H17	Env-60	Env TM ISD	immune-suppression	tumor immune escape by stimulation of CCL19 chemokine and expansion of pluripotent CD271 ⁺ cells	(105)
HERVP71	?	HERVP71A	LTR	immune-suppression	enhancer for HLA-G in human extravillous trophoblasts, inhibiting natural killer cytotoxicity	(106)
HERV-K(HML2)	HERV-K108	–	Env TM	immune-suppression	inhibition of T cells activation, cytokine release and immune gene expression	(107)
	6p21.33 (PSORS1 gene)	–	Pro dUTPase	immune-activation	stimulation of TLR2, induction of NF- κ B and Th1+Th17 cytokines	(108)

3' \rightarrow 5' exonuclease 1 (Trex1), in which the accumulation of endogenous retroelements' cDNA led to the activation of innate DNA sensors and the production of IFN (125–127). Even if a similar function is still to be confirmed for human Trex1, this enzyme was shown to restrict the reverse-transcribed DNA from endogenous retroelements which instead accumulated in Trex-1 deficient cells (125). Moreover, patients with Aicardi-Goutières autoimmune syndrome lack this enzyme and show inflammatory and IFN-I responses that could possibly be sustained by HERV nucleic acids accumulation. Hence, Trex1 could have a role in the control of HERV cDNA-mediated immune activation (125, 128), also considering that this cellular protein was recently reported to prevent cGAS or IFI16-mediated recognition of HIV-1 replication intermediates, being possibly involved in viral immune escape mechanisms (129). Similarly, the 3' \rightarrow 5' RNA exoribonuclease SKIV2L has been recently involved in the exosomal degradation of endogenous RNA molecules in the cytosol, thus possibly preventing also HERV RNA accumulation that could trigger viral RNA-sensing receptors (130).

Consequences of Innate Immunity Stimulation

Independently from the cytosolic or membrane-associated localization, the sensing of viral molecules by innate immune receptors evokes the production of pro-inflammatory effectors (such as IFN, cytokines and chemokines), which promptly establish an antiviral status. This immune activation contributes to contain the infection and prepares the ground for subsequent adaptive immune responses, mediated by T- and B-lymphocytes that elicit specific cellular and humoral defenses, respectively. Overall, such innate-adaptive signaling is fundamental to counteract exogenous viruses, and the resulting immune activation is generally able to eliminate the infectious triggers and

then to shut down. In the case of HERV molecules, however, their stable presence and expression in the organism could provide continuous triggers to the host immune sensors. In fact, the main findings linking HERVs to autoimmune and inflammatory disorders rely on the chronic stimulus exerted by endogenous retroviral molecules, which could sustain molecular mimicry events based on the sequence identity between HERV products and either exogenous viruses or body components (16, 28, 30, 131, 132). Worth to note, after the initial immune activation, the production of IFN establishes a positive-feedback that further upregulates IFN-stimulated genes by both a paracrine and an autocrine loop, fostering chronic inflammation and autoimmunity. The same antiviral status prompted by HERV expression can hence create a vicious circle in which inflammatory molecules and epigenetic dysregulation further upregulate HERV expression (26, 69, 84). For all these reasons, even if no definitive etiological evidences have been reached yet, the role of HERVs in triggering innate defenses could contribute to autoimmune pathological developments and, hence, be a valuable therapeutic target. Accordingly, monoclonal antibodies against HERV-W Env proteins are currently under clinical trials as innovative approach against multiple sclerosis (133, 134) and type I diabetes (117). In addition, even in the absence of a direct immunogenic role of HERV-derived proteins, the presence and nucleotide variability of HERV loci in proximity to key immune genes should be taken into account for their possible influence on the latter, especially in the context of genome wide association studies. These analyses are by now an established method to identify risk loci linked to autoimmune conditions, but rarely take into account HERV sequences and their networking with cellular genes. As an example, MS-associated SNPs showed a significant enrichment of HERV insertion and potential HERV ORFs in their genetic neighborhood as compared to control SNPs

(135, 136). Accordingly, SNPs in the antiviral gene TRIM5 were shown to be negatively correlated to MS, while SNPs around a HERV-Fc1 locus on chromosome X had a significant association with the disease, with a risk that was strongly influenced by the two genes in an additive fashion (137). Hence, besides protein production, the genetic analysis of HERV loci in the neighborhood of immune genes, especially if presenting key immune roles, can provide insights on inter-individual variants concurring to autoimmunity risks.

HERVS AS INHIBITORS OF INNATE IMMUNITY

In parallel to pro-inflammatory effects, HERV-derived peptides have also been implicated in immune-suppressive mechanisms. The latter mainly involve Env transmembrane subunits, which hold a characteristic immunosuppressive domain (ISD) conserved among retroviral Env proteins. In animal exogenous retroviruses, the ISD counteracts the host antiviral responses (138, 139) while, in the case of HERVs, such element has been occasionally coopted for the maternal immune tolerance during pregnancy (32, 45). In this physiological status, a subtle immune balance must allow the invasion of fetal trophoblasts, which present also paternal antigens, albeit maintaining the ability to protect the organism from microbial infections. Pregnancy is hence accompanied by the repression of cellular immunity through the shift from Th1 inflammatory cytokine production (TNF- α , IFN- γ , and IL-2) toward an anti-inflammatory Th2 cytokines response (IL-4, IL-5, IL-10), to prevent cytotoxic processes potentially harmful to the fetus (102, 140, 141). Such shift is thought to be mainly due to syncytin-2 ISD, which is highly preserved and shows strong immunosuppressive potential (32). However, also syncytin-1 was shown to be able to inhibit cytokine production in blood, suggesting a possible contribution in the maternal immune shift (102).

Besides these domesticated proteins, other HERV Envs have shown immunosuppressive activity. The transmembrane subunit of a HERV-K(HML2) sequence was reported to inhibit T cells activation in a similar way to the HIV ISD, influencing cytokine release and immune gene expression (107). Similarly, a HERV-H Env protein (Env-59) ISD showed anti-inflammatory potential in a mouse model of arthritis (103), being inversely related to the level of pathogenic effectors (such as IL-6 and TLR7) in human autoimmune rheumatic diseases (104). Furthermore, a recently described HERV LTR (HERVP71A, locus 6p22.1) was shown to serve as a tissue-specific enhancer for HLA-G gene expression in human extravillous trophoblasts, at the fetal-maternal interface, inhibiting the natural killer cell cytotoxicity and conferring immune tolerance to the developing placenta (106).

HERVS AND EXOGENOUS INFECTIONS

HERVs have been proposed to have a role during exogenous viral infections, and such role could be either beneficial or harmful (16). It is worth to note that the interplay between exogenous and endogenous viruses is still not fully characterized, even if

some evidences suggest that exogenous infection sustained by different viral species—among which HIV, herpesviruses and influenza—are able to modulate HERV expression (142–144). In the case of an upregulation of HERV expression, such cooperative action could increase the immune triggering exerted by HERV products and, especially in the presence of retroviral infections, possibly account for complementation of defective viruses and recombination events (16). Here, however, we will focus on the possible protective effects against exogenous infections exerted by HERV expression products, which can theoretically be able to restrict any step of the viral cycle (21).

As any organism or biological entity, viruses are subjected to an ensemble of selective pressures, being exerted by the host defenses as well as by the surrounding environment, including other microbes threatening the same host. Due to this, the different viral species evolved strategies to avoid the host antiviral systems and to compete with other viral populations, to assure their replication. In line with this, it is well known that a cell infected by a certain virus often become resistant to superinfection, developing a virus-induced viral resistance against members of the same species as well as different viruses (24). Considering HERVs, such cross-protection has been investigated mostly for exogenous retroviruses that share identity in their protein and nucleic acids, being more prone to interact with HERV products (16) (**Figure 4**). A partial resistance to exogenous infection could depend on the interference and blocking of the same cellular receptor by HERV-derived proteins or pseudo-viral particles (145, 146) (**Figure 4**). Inside the cell, the expression of HERV antisense transcripts has been proposed to confer protection against infections through the complementary interaction with homologous RNA sequences originated by exogenous retroviruses' expression, forming dsRNA molecules that can be recognized as PAMPs by human PRRs (26, 72) (**Figure 4**). Finally, in the case of HERV proteins' production, their identity with exogenous viral proteins could led to complementation events possibly affecting the formation of viral particles (**Figure 4**). Accordingly, a HERV-K(HML2) Gag protein was reported to co-assemble with HIV-1 Gag and to subsequently impair HIV-1 capsid formation as well as HIV-1 particles release and infectivity (147, 148) (**Figure 4**).

On the one side, all these direct interactions between HERV and exogenous viral products has been argued to had a role in the restriction and extinction of HERV-originating ancestral exogenous retroviruses (21). In line with this, Blanco-Melo and coauthors “resuscitated” two ancestral Env proteins: one was reconstructed from various HERV-T insertions in primate species, being more ancient, and the other was encoded by a human HERV-T provirus acquired about 20 million years later (149). The “older” Env was shown to bind the human receptor MTC1, and to be able to generate infectious pseudotyped MLV particles and syncytia (149). The “modern” HERV-T Env, albeit being still able to bind MTC1, lost these abilities and even downregulated MTC1 either at the cell surface (causing its internalization) or during secretion (blocking its transport), leading to its degradation (149). On the other side, the same interactions could have also contributed to the rapid evolution and adaptation of HERV genes, driving the selection of elements

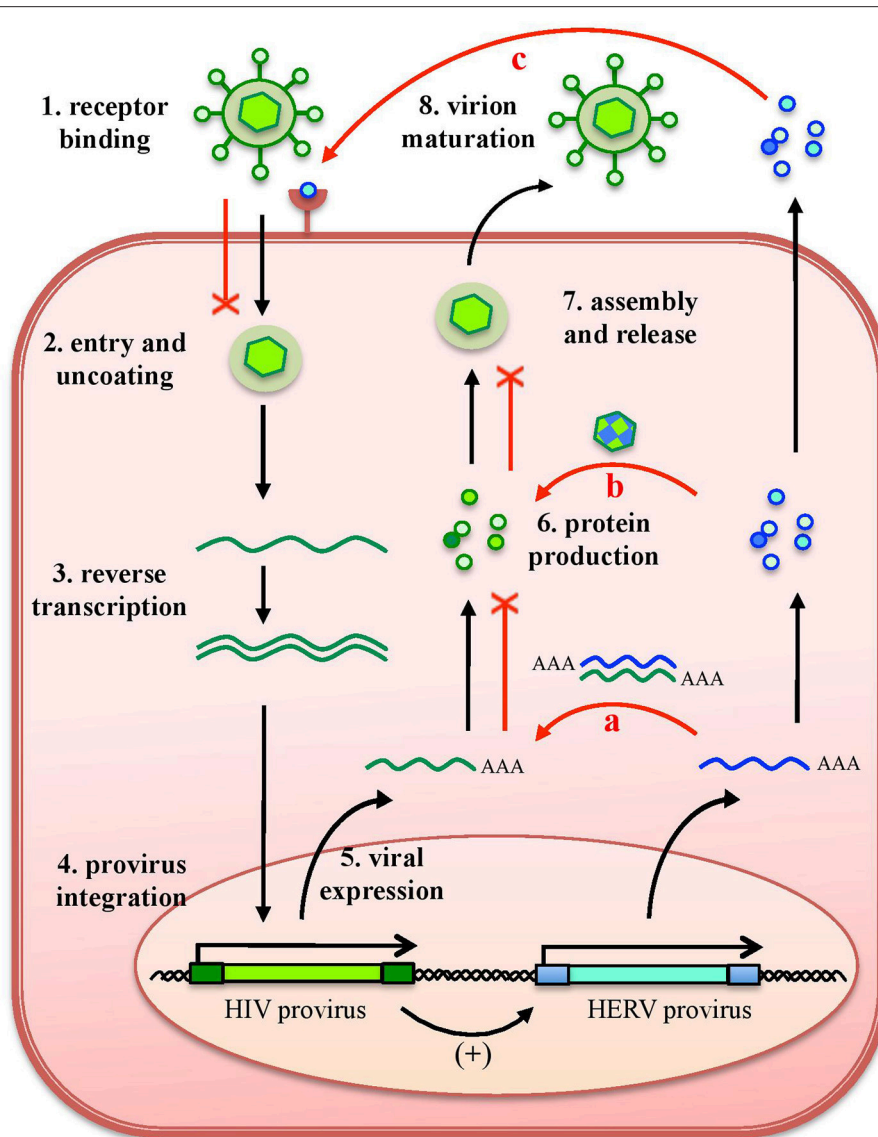


FIGURE 4 | Protective effects exerted by HERVs against exogenous infections. The effects of HERV expression in the impairment of exogenous viruses (red lines) are represented in relation to HIV infection. The various steps of HIV replication are also schematized. Even if HERV expression could theoretically affect any of them, the major implications regard (a) the expression of HERV mRNAs, that can interact by complementarity with HIV RNA, could form dsRNA that is detected as a PAMP by cellular innate immunity sensors; (b) the complementation of HIV structural components with HERV proteins, possibly affecting HIV particles' assembly and release; and (c) the binding of HERV proteins or pseudoparticles to the same cellular receptor, preventing HIV binding and entry. Of note, these effects can be eventually sustained also by the possible upregulation of HERV expression by HIV infection.

that could increase the range of restricted microbes and confer a sort of broad-viral protection (21). In this regard, HERV-K(HML2) Rec proteins upregulation has recently been reported during embryogenesis, leading to the specific stimulation of the IFN-induced viral restriction factor IFITM126 in epiblast and embryonic stem cells (56, 150). This finding led to the fascinating idea that Rec and the HML2 mRNAs, associated to it for nuclear export, might be detected in the cytosol, eliciting an innate anti-viral response thought to broadly inhibit embryonic viral infections (56).

HERVS, CANCER AND ANTICANCER STRATEGIES

In the context of cancer development, the ability of HERVs to modulating the immune system may have opposite effects, being potentially relevant to both the oncogenic processes and the anticancer defenses (23).

On the one side, HERV immunosuppressive functions might contribute to cancer progression by reducing the immune recognition and attack of tumor cells. As an example, the ISD

peptide of an HERV-H Env (env60), namely H17, was shown to induce tumor cells' epithelial-to-mesenchymal transition, stimulating CCL19 chemokine expression and the subsequent recruitment and expansion of pluripotent immunoregulatory CD271⁺ cells (105). This mechanism was proposed to be critical for cancerous cell immune escape as well as for metastatic invasion and adhesion (105).

On the other side, the immunogenic properties of HERV expression combined with their general upregulation in cancer tissues, especially due to a broad epigenetic dysregulation, could hence represent an innovative therapeutic target (16). In fact, HERV products have been investigated in the field of anticancer immunotherapy, especially if expressed to higher extents (tumor-associated) or exclusively (tumor-specific) in transformed cells (16). As mentioned above, demethylating agents—commonly used in anticancer therapy—are known to induce a hypomethylated status liberating retrotransposon expression, which is the base of their therapeutic activity (25). In fact, such a stimulation led to the production of HERV bidirectional transcripts that can form dsRNA, sensed by cellular PRRs and directing thus a IFN-I and -III response against colorectal tumor cells (29). Accordingly, the individual knock-down of key immune players like MDA5, MAVS, and IRF7 in cancer cells significantly reduced the anticancer activity of DNA methyltransferase inhibitors (29). The activation of dsRNA sensors by HERV nucleic acids following the treatment with demethylating agents has been reported at the same time also by Chiappinelli et al. in ovarian cancer cells (27). Authors underlined that a major anticancer mechanism of demethylating agents is an induced IFN-I immune response mediated by the cytosolic dsRNA sensing of the multiple upregulated HERV transcripts (27). In particular, the cellular PRRs TLR3 and MDA5 have a pivotal role in such immune triggering, given that tumors showing high HERV expression present a concomitant significant activation of these viral sensors (27). Therefore, demethylating agents have been proposed either in association with active immunization therapies, to produce synergistic anticancer effects (151), or as a strategy to overcome primary resistance to immune checkpoint blocking therapies (25).

CONCLUSIONS

A growing body of evidences suggests that the relationship between our genome and HERVs constitutes an intricate

and multifaceted co-evolution spanning million of years throughout mammalian development. During such a long liaison, the detrimental effects exerted by HERVs have been balanced by beneficial activities that brought innovation and diversity to the human genome and physiology. Among these, the establishment of diverse HERV-mediated regulatory networks and the co-option of HERV proteins for pregnancy functions provided pivotal features characterizing our biological nature. Intriguingly, while HERVs are products of ancestral exogenous viral infections pervading primates, they became major contributors in shaping and improving the human antiviral immunity. Nowadays, they are still able to modulate it in an ambivalent way, suggesting that some new adaptive interplay between HERVs and our genome might still evolve *de novo* through sequence variation and fortuitous interactions (21). Additional studies are needed to characterize in detail the interaction between individual HERV products and specific immune receptors as well as the pathways involved in HERV-mediated modulation of innate responses. The characterization of such a complex interplay would allow to better understand the persistence of these long-time genomic residents and to finally clarify their role in human immune physiology and pathogenesis.

AUTHOR CONTRIBUTIONS

NG and ET participated to the conception, drafting and revision of the manuscript and approved the final version.

FUNDING

This work was supported by a POR-FESR 2014-2020 grant.

ACKNOWLEDGMENTS

We would like to thank the colleagues involved in the studies reported in the present review, and apologize to the ones whose work has not been referenced here.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.02039/full#supplementary-material>

REFERENCES

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature* (2001) 412:860–921. doi: 10.1038/35057062
2. Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet.* (2006) 7:149–73. doi: 10.1146/annurev.genom.7.080505.115700
3. Vargiu L, Rodriguez-Tomé P, Sperber GO, Cadeddu M, Grandi N, Blikstad V, et al. Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* (2016) 13:7. doi: 10.1186/s12977-015-0232-y
4. Mayer J, Meese E. Human endogenous retroviruses in the primate lineage and their influence on host genomes. *Cytogenet Genome Res.* (2005) 110:448–56. doi: 10.1159/000084977
5. Esposito F, Corona A, Tramontano E. HIV-1 reverse transcriptase still remains a new drug target: structure, function, classical inhibitors, and new inhibitors with innovative mechanisms of actions. *Mol Biol Int.* (2012) 2012:586401. doi: 10.1155/2012/586401
6. Esposito F, Tramontano E. Past and future. Current drugs targeting HIV-1 integrase and reverse transcriptase-associated ribonuclease H activity: single and dual active site inhibitors. *Antivir Chem Chemother.* (2013) 23:129–44. doi: 10.3851/IMP2690

7. Magin C, Löwer R, Löwer J. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. *J Virol.* (1999) 73:9496–507.
8. Armbruster V, Sauter M, Krautkraemer E, Meese E, Kleiman A, Best B, et al. A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res.* (2002) 8:1800–7.
9. Grandi N, Tramontano E. HERV envelope proteins: physiological role and pathogenic potential in cancer and autoimmunity. *Front Microbiol.* (2018) 9:462. doi: 10.3389/fmicb.2018.00462
10. Grandi N, Cadeddu M, Pisano MP, Esposito F, Blomberg J, Tramontano E. Identification of a novel HERV-K(HML10): Comprehensive characterization and comparative analysis in non-human primates provide insights about HML10 proviruses structure and diffusion. *Mob DNA* (2017) 8. 15. doi: 10.1186/s13100-017-0099-7
11. Blomberg J, Benachenhou F, Blikstad V, Sperber G, Mayer J. Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. *Gene* (2009) 448:115–23. doi: 10.1016/j.gene.2009.06.007
12. Mayer J, Blomberg J, Seal RL. A revised nomenclature for transcribed human endogenous retroviral loci. *Mob DNA* (2011) 2:7. doi: 10.1186/1759-8753-2-7
13. Jern P, Sperber GO, Blomberg J. Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy. *Retrovirology* (2005) 2:50. doi: 10.1186/1742-4690-2-50
14. Grandi N, Cadeddu M, Blomberg J, Tramontano E. Contribution of type W human endogenous retrovirus to the human genome: characterization of HERV-W proviral insertions and processed pseudogenes. *Retrovirology* (2016) 13:1–25. doi: 10.1186/s12977-016-0301-x
15. Sperber G, Airola T, Jern P, Blomberg J. Automated recognition of retroviral sequences in genomic data - RetroTector®. *Nucleic Acids Res.* (2007) 35:4964–76. doi: 10.1093/nar/gkm515
16. Grandi N, Tramontano E. Type W human endogenous retrovirus (HERV-W) integrations and their mobilization by L1 machinery: contribution to the human transcriptome and impact on the host physiopathology. *Viruses* (2017) 9:162. doi: 10.3390/v9070162
17. Blond JL, Lavillette D, Cheynet V, Bouton O, Oriol G, Chapel-Fernandes S, et al. An envelope glycoprotein of the human endogenous retrovirus HERV-W is expressed in the human placenta and fuses cells expressing the type D mammalian retrovirus receptor. *J Virol.* (2000) 74:3321–9. doi: 10.1128/JVI.74.7.3321-3329.2000
18. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, et al. Syncytin is a captive retroviral envelope protein involved. *Nature* (2000) 403:785–9. doi: 10.1038/35001608
19. Blaise S, de Parseval N, Bénit L, Heidmann T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc Natl Acad Sci USA.* (2003) 100:13013–8. doi: 10.1073/pnas.2132646100
20. Voisset C, Weiss R a, Griffiths DJ. Human RNA “rumor” viruses: the search for novel human retroviruses in chronic disease. *Microbiol Mol Biol Rev.* (2008) 72:157–96. doi: 10.1128/MMBR.00033-07
21. Frank JA, Feschotte C. Co-option of endogenous viral sequences for host cell function. *Curr Opin Virol.* (2017) 25:81–89. doi: 10.1016/j.coviro.2017.07.021
22. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* (2016) 351:1083–7. doi: 10.1126/science.aad5497
23. Göke J, Ng HH. CTRL+INSERT: retrotransposons and their contribution to regulation and innovation of the transcriptome. *EMBO Rep.* (2016) 17:1131–44. doi: 10.15252/embr.201642743
24. Villarreal LP. Viral ancestors of antiviral systems. *Viruses* (2011) 3:1933–1958. doi: 10.3390/v3101933
25. Wolff F, Leisch M, Greil R, Risch A, Pleyer L. The double-edged sword of (re)expression of genes by hypomethylating agents: from viral mimicry to exploitation as priming agents for targeted immune checkpoint modulation. *Cell Commun Signal* (2017) 15:13. doi: 10.1186/s12964-017-0168-z
26. Hurst TP, Magiorkinis G. Activation of the innate immune response by endogenous retroviruses. *J Gen Virol.* (2015) 96:1207–1218. doi: 10.1099/vir.0.000017
27. Chiappinelli KB, Strissel PL, Desrichard A, Chan TA, Baylin SB, Correspondence S. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* (2015) 162:974–986. doi: 10.1016/j.cell.2015.07.011
28. Trela M, Nelson PN, Rylance PB. The role of molecular mimicry and other factors in the association of human endogenous retroviruses and autoimmunity. *APMIS* (2016) 124:88–104. doi: 10.1111/apm.12487
29. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* (2015) 162:961–73. doi: 10.1016/j.cell.2015.07.056
30. Ramasamy R, Joseph B, Whittall T. Potential molecular mimicry between the human endogenous retrovirus W family envelope proteins and myelin proteins in multiple sclerosis. *Immunol Lett.* (2017) 183:79–85. doi: 10.1016/j.imlet.2017.02.003
31. Dupressoir A, Lavialle C, Heidmann T. From ancestral infectious retroviruses to bona fide cellular genes: role of the captured syncytins in placentation. *Placenta* (2012) 33:663–671. doi: 10.1016/j.placenta.2012.05.005
32. Mangeney M, Renard M, Schlecht-Louf G, Bouallaga I, Heidmann O, Letzelter C, et al. Placental syncytins: Genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc Natl Acad Sci USA.* (2007) 104:20534–20539. doi: 10.1073/pnas.0707873105
33. Hummel J, Kämmerer U, Müller N, Avota E, Schneider-Schaulies S. Human endogenous retrovirus envelope proteins target dendritic cells to suppress T-cell activation. *Eur J Immunol.* (2015) 45:1748–59. doi: 10.1002/eji.201445366
34. Meylan F, De Smedt M, Leclercq G, Plum J, Leupin O, Marguerat S, et al. Negative thymocyte selection to HERV-K18 superantigens in humans. *Blood* (2005) 105:4377–82. doi: 10.1182/blood-2004-07-2596
35. McClintock B. Intracellular systems controlling gene action and mutation. *Brookhaven Symp Biol.* (1956) 58–74.
36. Feschotte C. The contribution of transposable elements of the evolution of regulatory networks. *Nat Rev Genet.* (2008) 9:397–405. doi: 10.1038/nrg2337
37. Trizzino M, Park Y, Holsbach-Beltrame M, Aracena K, Mika K, Caliskan M, et al. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Res.* (2017) 27:1623–33. doi: 10.1101/gr.218149.116
38. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. *Trends Genet.* (2003) 19:530–6. doi: 10.1016/j.tig.2003.08.004
39. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* (2008) 18:1752–62. doi: 10.1101/gr.080663.108
40. Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, et al. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. *PLoS Genet.* (2017) 13:e1006883. doi: 10.1371/journal.pgen.1006883
41. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, et al. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* (2014) 24:1963–76. doi: 10.1101/gr.168872.113
42. Duperray A, Barbe D, Raguenez G, Weksler BB, Romero IA, Couraud PO, et al. Inflammatory response of endothelial cells to a human endogenous retrovirus associated with multiple sclerosis is mediated by TLR4. *Int Immunol.* (2015) 27:545–53. doi: 10.1093/intimm/dxv025
43. Feschotte C, Gilbert C. Endogenous viruses: insights into viral evolution and impact on host biology. *Nat Rev Genet.* (2012) 13:283–96. doi: 10.1038/nrg3199
44. Hua-Van A, Le Rouzic A, Boutin TS, Filée J, Capy P. The struggle for life of the genome's selfish architects. *Biol Direct* (2011) 6:19. doi: 10.1186/1745-6150-6-19
45. Lavialle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, et al. Paleovirology of “syncytins”, retroviral env genes exapted for a role in placentation. *Philos Trans R Soc Lond B Biol Sci.* (2013) 368:20120507. doi: 10.1098/rstb.2012.0507
46. Frendo JL, Olivier D, Cheynet V, Blond JL, Bouton O, Vidaud M, et al. Direct involvement of HERV-W Env glycoprotein in human

- trophoblast cell fusion and differentiation. *Mol Cell Biol.* (2003) 23:3566–74. doi: 10.1128/MCB.23.10.3566-3574.2003
47. Malassiné A, Handschuh K, Tsatsaris V, Gerbaud P, Cheynet V, Oriol G, et al. Expression of HERV-W Env glycoprotein (syncytin) in the extravillous trophoblast of first trimester human placenta. *Placenta* (2005) 26:556–62. doi: 10.1016/j.placenta.2004.09.002
 48. Kamp C, Hirschmann P, Voss H, Huellen K, Vogt PH. Two long homologous retroviral sequence blocks in proximal Yq11 cause AZFa microdeletions as a result of intrachromosomal recombination events. *Hum Mol Genet.* (2000) 9:2563–72. doi: 10.1093/hmg/9.17.2563
 49. Samuelson LC, Wiebauer K, Snow CM, Meisler MH. Retroviral and pseudogene insertion sites reveal the lineage of human salivary and pancreatic amylase genes from a single gene during primate evolution. *Mol Cell Biol.* (1990) 10:2513–20. doi: 10.1128/MCB.10.6.2513
 50. Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* (1992) 6:1457–65. doi: 10.1101/gad.6.8.1457
 51. Lu X, Sachs F, Ramsay L, Jacques PÉ, Göke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol.* (2014) 21:423–5. doi: 10.1038/nsmb.2799
 52. Hakim ST, Alsayari M, McLean DC, Saleem S, Addanki KC, Aggarwal M, et al. A large number of the human microRNAs target lentiviruses, retroviruses, and endogenous retroviruses. *Biochem Biophys Res Commun.* (2008) 369:357–62. doi: 10.1016/j.bbrc.2008.02.025
 53. Hadjiargyrou M, Delihias N. The intertwining of transposable elements and non-coding RNAs. *Int J Mol Sci.* (2013) 14:13307–28. doi: 10.3390/ijms140713307
 54. Wang Y, Xu Z, Jiang J, Xu C, Kang J, Xiao L, et al. Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human embryonic stem cell self-renewal. *Dev Cell.* (2013) 25:69–80. doi: 10.1016/j.devcel.2013.03.002
 55. Durruthy-Durruthy J, Sebastiano V, Wossidlo M, Cepeda D, Cui J, Grow EJ, et al. The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. *Nat Genet.* (2016) 48:44–52. doi: 10.1038/ng.3449
 56. Grow EJ, Flynn RA, Chavez SL, Bayless NL, Wesche D, Martin L, et al. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* (2015) 522:221–5. doi: 10.1038/nature14308
 57. Plath N, Ohana O, Dammermann B, Errington ML, Schmitz D, Gross C, et al. Arc/Arg3.1 Is essential for the consolidation of synaptic plasticity and memories. *Neuron* (2006) 52:437–44. doi: 10.1016/j.neuron.2006.08.024
 58. Mikuni T, Uesaka N, Okuno H, Hirai H, Deisseroth K, Bito H, et al. Arc/Arg3.1 is a postsynaptic mediator of activity-dependent synapse elimination in the developing cerebellum. *Neuron* (2013) 78:1024–35. doi: 10.1016/j.neuron.2013.04.036
 59. Zhang W, Wu J, Ward MD, Yang S, Chuang YA, Xiao M, et al. Structural basis of arc binding to synaptic proteins: implications for cognitive disease. *Neuron* (2015) 86:490–500. doi: 10.1016/j.neuron.2015.03.030
 60. Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, et al. The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. *Cell* (2018) 172:275–88.e18. doi: 10.1016/j.cell.2017.12.024
 61. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet.* (2017) 18:71–86. doi: 10.1038/nrg.2016.139
 62. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, et al. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci USA.* (2007) 104:18613–8. doi: 10.1073/pnas.0703637104
 63. Andersson G, Svensson AC, Setterblad N, Rask L. Retroelements in the human MHC class II region. *Trends Genet.* (1998) 14:109–114. doi: 10.1016/S0168-9525(97)01359-0
 64. Dangel AW, Mendoza AR, Baker BJ, Daniel CM, Carroll MC, Wu LC, et al. The dichotomous size variation of human complement C4 genes is mediated by a novel family of endogenous retroviruses, which also establishes species-specific genomic patterns among Old World primates. *Immunogenetics* (1994) 40:425–436. doi: 10.1007/BF00177825
 65. Chu X, Rittner C, Schneider PM. Length polymorphism of the human complement component C4 gene is due to an ancient retroviral integration. *Exp Clin Immunogenet.* (1995) 12:74–81.
 66. Mack M, Bender C, Schneider PM. Detection of retroviral antisense transcripts and promoter activity of the HERV-K(C4) insertion in the MHC III region. *Immunogenetics* (2004) 56:321–32. doi: 10.1007/s00251-004-0705-y
 67. Nehyba J, Hrdličková R, Bose HR. Dynamic evolution of immune system regulators: the history of the interferon regulatory factor family. *Mol Biol Evol.* (2009) 26:2539–50. doi: 10.1093/molbev/msp167
 68. Katoh I, Kurata S. Association of endogenous retroviruses and long terminal repeats with human disorders. *Front Oncol.* (2013) 3:234. doi: 10.3389/fonc.2013.00234
 69. Manghera M, Douville RN. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology* (2013) 10:16. doi: 10.1186/1742-4690-10-16
 70. Manghera M, Ferguson-Parry J, Lin R, Douville RN. NF- κ B and IRF1 induce endogenous retrovirus K expression via interferon-stimulated response elements in its 5' long terminal repeat. *J Virol.* (2016) 90:9338–49. doi: 10.1128/JVI.01503-16
 71. Stauffer Y, Marguerat S, Meylan F, Ucla C, Sutkowski N, Huber B, et al. Interferon-alpha-induced endogenous superantigen. a model linking environment and autoimmunity. *Immunity* (2001) 15:591–601. doi: 10.1016/S1074-7613(01)00212-6
 72. Gürtler C, Bowie AG. Innate immune detection of microbial nucleic acids. *Trends Microbiol.* (2013) 21:413–20. doi: 10.1016/j.tim.2013.04.004
 73. Shekarian T, Valsesia-Wittmann S, Brody J, Michallet MC, Depil S, et al. Pattern recognition receptors: immune targets to enhance cancer immunotherapy. *Ann Oncol.* (2017) 28:1756–66. doi: 10.1093/annonc/mdx179
 74. Tang D, Kang R, Coyne CB, Zeh HJ, Lotze MT. PAMPs and DAMPs: signal 0s that spur autophagy and immunity. *Immunol Rev.* (2012) 249:158–75. doi: 10.1111/j.1600-065X.2012.01146.x
 75. Kawai T, Akira S. Toll-like receptor and RIG-I-like receptor signaling. *Ann N Y Acad Sci.* (2008) 1143:1–20. doi: 10.1196/annals.1443.020
 76. Paludan S, Bowie A. Immune sensing of DNA. *Immunity* (2013) 38:870–880. doi: 10.1016/j.immuni.2013.05.004
 77. Pålsson-McDermott EM, O'Neill LAJ. Building an immune system from nine domains. *Biochem Soc Trans.* (2007) 35(Pt 6):1437–44. doi: 10.1042/BST0351437
 78. Iwasaki A, Medzhitov R. Control of adaptive immunity by the innate immune system. *Nat Immunol.* (2015) 16:343–53. doi: 10.1038/ni.3123
 79. Gasteiger G, D'osualdo A, Schubert DA, Weber A, Bruscia EM, Hartl D, et al. Cellular innate immunity: an old game with new players journal of innate immunity. *J Innate Immun.* (2017) 9:111–25. doi: 10.1159/000453397
 80. Lester SN, Li K. Toll-like receptors in antiviral innate immunity. *J Mol Biol.* (2014) 426:1246–64. doi: 10.1016/j.jmb.2013.11.024
 81. Rusek P, Wala M, Druszczyńska M, Fol M. Infectious agents as stimuli of trained innate immunity. *Int J Mol Sci.* (2018) 19:E456. doi: 10.3390/ijms19020456
 82. Xia P, Wang S, Gao P, Gao G, Fan Z. DNA sensor cGAS-mediated immune recognition. *Protein Cell* (2016) 7:777–91. doi: 10.1007/s13238-016-0320-3
 83. Sandstrom TS, Ranganath N, Angel JB. Impairment of the type I interferon response by HIV-1: potential targets for HIV eradication. *Cytokine Growth Factor Rev.* (2017) 37:1–16. doi: 10.1016/j.cytogfr.2017.04.004
 84. Hurst TP, Magiorkinis G. Epigenetic control of human endogenous retrovirus expression: focus on regulation of long-terminal repeats (LTRs). *Viruses* (2017) 9:1–13. doi: 10.3390/v9060130
 85. Nazli A, Kafka JK, Ferreira VH, Anipindi V, Mueller K, Osborne BJ, et al. HIV-1 gp120 induces TLR2- and TLR4-mediated innate immune activation in human female genital epithelium. *J Immunol.* (2013) 191:4246–58. doi: 10.4049/jimmunol.1301482
 86. Lester RT, Yao XD, Ball TB, McKinnon LR, Kaul R, Wachihi C, et al. Toll-like receptor expression and responsiveness are increased in viraemic HIV-1 infection. *AIDS* (2008) 22:685–94. doi: 10.1097/QAD.0b013e3282f4de35
 87. Hernandez JC, Arteaga J, Paul S, Kumar A, Latz E, Urcuqui-Inchima S. Up-regulation of TLR2 and TLR4 in dendritic cells in response to HIV type 1

- and coinfection with opportunistic pathogens. *AIDS Res Hum Retroviruses* (2011) 27:1099–109. doi: 10.1089/AID.2010.0302
88. Hernández JC, Stevenson M, Latz E, Urcuqui-Inchima S. HIV type 1 infection up-regulates TLR2 and TLR4 expression and function *in vivo* and *in vitro*. *AIDS Res Hum Retroviruses* (2012) 28:1313–28. doi: 10.1089/aid.2011.0297
 89. Zhou Y, Wang X, Liu M, Hu Q, Song L, Ye L, et al. A critical function of toll-like receptor-3 in the induction of anti-human immunodeficiency virus activities in macrophages. *Immunology* (2010) 131:40–9. doi: 10.1111/j.1365-2567.2010.03270.x
 90. Bhargavan B, Woollard SM, Kanmogne GD. Toll-like receptor-3 mediates HIV-1 transactivation via NF κ B and JNK pathways and histone acetylation, but prolonged activation suppresses Tat and HIV-1 replication. *Cell Signal* (2016) 28:7–22. doi: 10.1016/j.cellsig.2015.11.005
 91. Buitendijk M, Eszterhas SK, Howell AL. Toll-like receptor agonists are potent inhibitors of human immunodeficiency virus-type 1 replication in peripheral blood mononuclear cells. *AIDS Res Hum Retroviruses* (2014) 30:457–67. doi: 10.1089/AID.2013.0199
 92. Kim H, Yang E, Lee J, Kim SH, Shin JS, Park JY, et al. Double-stranded RNA mediates interferon regulatory factor 3 activation and interleukin-6 production by engaging Toll-like receptor 3 in human brain astrocytes. *Immunology* (2008) 124:480–8. doi: 10.1111/j.1365-2567.2007.02799.x
 93. Heil F. Species-specific recognition of single-stranded rna via toll-like receptor 7 and 8. *Science* (2004) 303:1526–9. doi: 10.1126/science.1093620
 94. Rigby RE, Webb LM, Mackenzie KJ, Li Y, Leitch A, Reijns MAM, et al. RNA:DNA hybrids are a novel molecular pattern sensed by TLR9. *EMBO J* (2014) 33:542–58. doi: 10.1002/embj.201386117
 95. Winckelmann AA, Munk-Petersen LV, Rasmussen TA, Melchjorsen J, Hjelholt TJ, Montefiori D, et al. Administration of a Toll-like receptor 9 agonist decreases the proviral reservoir in virologically suppressed HIV-infected patients. *PLoS ONE* (2013) 8:e62074. doi: 10.1371/journal.pone.0062074
 96. Offersen R, Nissen SK, Rasmussen TA, Østergaard L, Denton PW, Søgaard OS, et al. A novel toll-like receptor 9 agonist, MGN1703, enhances HIV-1 transcription and NK cell-mediated inhibition of HIV-1-infected autologous CD4⁺ T Cells. *J Virol* (2016) 90:4441–53. doi: 10.1128/JVI.00222-16
 97. Yu P, Lübken W, Slomka H, Gebler J, Konert M, Cai C, et al. Nucleic acid-sensing Toll-like receptors are essential for the control of endogenous retrovirus viremia and ERV-induced tumors. *Immunity* (2012) 37:867–79. doi: 10.1016/j.immuni.2012.07.018
 98. Perron H, Jouvin-Marche E, Michel M, Ounanian-Paraz A, Camelo S, Dumon A, et al. Multiple sclerosis retrovirus particles and recombinant envelope trigger an abnormal immune response *in vitro*, by inducing polyclonal V β 16 T-lymphocyte activation. *Virology* (2001) 287:321–32. doi: 10.1006/viro.2001.1045
 99. Rolland A, Jouvin-Marche E, Viret C, Faure M, Perron H, Marche PN. The envelope protein of a human endogenous retrovirus-W family activates innate immunity through CD14/TLR4 and promotes Th1-like responses. *J Immunol* (2006) 176:7636–44. doi: 10.4049/jimmunol.176.12.7636
 100. Saresella M, Rolland A, Marventano I, Cavarretta R, Caputo D, Marche P, et al. Multiple sclerosis-associated retroviral agent (MSRV)-stimulated cytokine production in patients with relapsing-remitting multiple sclerosis. *Mult Scler* (2009) 15:443–7. doi: 10.1177/1352458508100840
 101. Kremer D, Schichel T, Förster M, Tzekova N, Bernard C, Van Der Valk P, et al. Human endogenous retrovirus type W envelope protein inhibits oligodendroglial precursor cell differentiation. *Ann Neurol* (2013) 74:721–32. doi: 10.1002/ana.23970
 102. Tolosa JM, Schjenken JE, Clifton VL, Vargas A, Barbeau B, Lowry P, et al. The endogenous retroviral envelope protein syncytin-1 inhibits LPS/PHA-stimulated cytokine responses in human blood and is sorted into placental exosomes. *Placenta* (2012) 33:933–41. doi: 10.1016/j.placenta.2012.08.004
 103. Laska MJ, Trolldborg A, Hauge EM, Bahrami S, Stengaard-Pedersen K. AB0082 anti-inflammatory activity of a human endogenous retroviral genetic element in experimental arthritis. *Ann Rheum Dis* (2016) 75:924.2–5. doi: 10.1136/annrheumdis-2016-eular.4267
 104. Laska MJ, Trolldborg A, Hauge EM, Bahrami S, Stengaard-Pedersen K. Human endogenous retroviral genetic element with immunosuppressive activity in both human autoimmune diseases and experimental arthritis. *Arthritis Rheumatol* (2017) 69:398–409. doi: 10.1002/art.39867
 105. Kudo-Saito C, Yura M, Yamamoto R, Kawakami Y. Induction of immunoregulatory CD271+ cells by metastatic tumor cells that express human endogenous retrovirus H. *Cancer Res* (2014) 74:1361–70. doi: 10.1158/0008-5472.CAN-13-1349
 106. Ferreira LMR, Meissner TB, Mikkelsen TS, Mallard W, O'Donnell CW, et al. A distant trophoblast-specific enhancer controls HLA-G expression at the maternal-fetal interface. *Proc Natl Acad Sci USA* (2016) 113:5364–9. doi: 10.1073/pnas.1602886113
 107. Morozov VA, Dao Thi VL, Denner J. The transmembrane protein of the human endogenous retrovirus - K (HERV-K) modulates cytokine release and gene expression. *PLoS ONE* (2013) 8:e70399. doi: 10.1371/journal.pone.0070399
 108. Ariza ME, Williams MV. A human endogenous retrovirus K dUTPase triggers a TH1, TH17 cytokine response: does it have a role in psoriasis? *J Invest Dermatol* (2011) 131:2419–27. doi: 10.1038/jid.2011.217
 109. Rolland A, Jouvin-Marche E, Saresella M, Ferrante P, Cavaretta R, Créange A, et al. Correlation between disease severity and *in vitro* cytokine production mediated by MSRV (Multiple Sclerosis associated RetroViral element) envelope protein in patients with multiple sclerosis. *J Neuroimmunol* (2005) 160:195–203. doi: 10.1016/j.jneuroim.2004.10.019
 110. Mameli G, Astone V, Khalili K, Serra C, Sawaya BE, Dolei A. Regulation of the syncytin-1 promoter in human astrocytes by multiple sclerosis-related cytokines. *Virology* (2007) 362:120–30. doi: 10.1016/j.virol.2006.12.019
 111. Antony JM, Deslauriers AM, Bhat RK, Ellestad KK, Power C. Human endogenous retroviruses and multiple sclerosis: innocent bystanders or disease determinants? *Biochim Biophys Acta* (2011) 1812:162–76. doi: 10.1016/j.bbdis.2010.07.016
 112. Perron H, Dougier-Reynaud HL, Lomparski C, Popa I, Firouzi R, Bertrand JB, et al. Human endogenous retrovirus protein activates innate immunity and promotes experimental allergic encephalomyelitis in mice. *PLoS ONE* (2013) 8:e80128. doi: 10.1371/journal.pone.0080128
 113. Antony JM, van Marle G, Opii W, Butterfield DA, Mallet F, Yong VW, et al. Human endogenous retrovirus glycoprotein-mediated induction of redox reactants causes oligodendrocyte death and demyelination. *Nat Neurosci* (2004) 7:1088–95. doi: 10.1038/nn1319
 114. Antony JM, Ellestad KK, Hammond R, Imaizumi K, Mallet F, Warren KG, et al. The human endogenous retrovirus envelope glycoprotein, syncytin-1, regulates neuroinflammation and its receptor expression in multiple sclerosis: a role for endoplasmic reticulum chaperones in astrocytes. *J Immunol* (2007) 179:1210–24. doi: 10.4049/jimmunol.179.2.1210
 115. Kremer D, Förster M, Schichel T, Göttle P, Hartung HP, Perron H, et al. The neutralizing antibody GNBAC1 abrogates HERV-W envelope protein-mediated oligodendroglial maturation blockade. *Mult Scler* (2014) 21:1200–3. doi: 10.1177/1352458514560926
 116. Emmer A, Staeger MS, Kornhuber ME. The retrovirus/superantigen hypothesis of multiple sclerosis. *Cell Mol Neurobiol* (2014) 34:1087–96. doi: 10.1007/s10571-014-0100-7
 117. Levett S, Medina J, Joanou J, Demolder A, Queruel N, Réant K, et al. An ancestral retroviral protein identified as a therapeutic target in type-1 diabetes. *JCI Insight* (2017) 2:1–18. doi: 10.1172/JCI.INSIGHT.94387
 118. Hultman GK, Ariza M, Williams M V, Mariana D. Protective effect of human endogenous retrovirus K dUTPase variants on psoriasis susceptibility. *J Invest Dermatol* (2013) 132:1833–40. doi: 10.1038/jid.2012.69
 119. Solis M, Nakhaei P, Jalalirad M, Lacoste J, Douville R, Arguello M, et al. RIG-I-mediated antiviral signaling is inhibited in HIV-1 infection by a protease-mediated sequestration of RIG-I. *J Virol* (2011) 85:1224–36. doi: 10.1128/JVI.01635-10
 120. Ghosh M, Shen Z, Fahey JV, Crist SG, Patel M, Smith JM, et al. Pathogen recognition in the human female reproductive tract: expression of intracellular cytosolic sensors NOD1, NOD2, RIG-1, and MDA5 and response to HIV-1 and *Neisseria gonorrhea*. *Am J Reprod Immunol* (2013) 69:41–51. doi: 10.1111/aji.12019
 121. Iwasaki A. Innate immune recognition of HIV-1. *Immunity* (2013) 37:389–98. doi: 10.1016/j.immuni.2012.08.011

122. Unterholzner L, Keating SE, Baran M, Horan KA, Jensen SB, Sharma S, et al. IFI16 is an innate immune sensor for intracellular DNA. *Nat Immunol.* (2010) 11:997–1004. doi: 10.1038/ni.1932
123. Nissen SK, Højten JF, Andersen KLD, Kofod-Olsen E, Berg RK, Paludan SR, et al. Innate DNA sensing is impaired in HIV patients and IFI16 expression correlates with chronic immune activation. *Clin Exp Immunol.* (2014) 177:295–309. doi: 10.1111/cei.12317
124. Jakobsen MR, Bak RO, Andersen A, Berg RK, Jensen SB, Tengchuan J, et al. IFI16 senses DNA forms of the lentiviral replication cycle and controls HIV-1 replication. *Proc Natl Acad Sci USA.* (2013) 110:E4571–80. doi: 10.1073/pnas.1311669110
125. Stetson DB, Ko JS, Heidmann T, Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* (2008) 134:587–98. doi: 10.1016/j.cell.2008.06.032
126. Gall A, Treuting P, Elkon KB, Loo YM, Gale M, Barber GN, et al. Autoimmunity initiates in nonhematopoietic cells and progresses via lymphocytes in an interferon-dependent autoimmune disease. *Immunity* (2012) 36:120–31. doi: 10.1016/j.immuni.2011.1.018
127. Xu J, Zoltick PW, Gamero AM, Gallucci S. TLR ligands up-regulate Trex1 expression in murine conventional dendritic cells through type I Interferon and NF- κ B-dependent signaling pathways. *J Leukoc Biol.* (2014) 96:93–103. doi: 10.1189/jlb.2A0713-393RR
128. Yang YG, Lindahl T, Barnes DE. Trex1 exonuclease degrades ssDNA to prevent chronic checkpoint activation and autoimmune disease. *Cell* (2007) 131:873–86. doi: 10.1016/j.cell.2007.10.017
129. Wheeler LA, Trifonova RT, Vrbanac V, Barteneva NS, Liu X, Bollman B, et al. TREX1 knockdown induces an interferon response to HIV that delays viral infection in humanized mice. *Cell Rep.* (2016) 15:1715–27. doi: 10.1016/j.celrep.2016.04.048
130. Eckard SC, Rice GI, Fabre A, Badens C, Gray EE, Hartley JL, et al. The SKIV2L RNA exosome limits activation of the RIG-I-like receptors. *Nat Immunol.* (2014) 15:839–45. doi: 10.1038/ni.2948
131. Nelson P, Rylance P, Roden D, Trela M, Tugnet N. Viruses as potential pathogenic agents in systemic lupus erythematosus. *Lupus* (2014) 23:596–605. doi: 10.1177/0961203314531637
132. Mameli G, Erre GL, Caggiu E, Mura S, Cossu D, Bo M, et al. Identification of a HERV-K env surface peptide highly recognized in Rheumatoid Arthritis (RA) patients: a cross-sectional case-control study. *Clin Exp Immunol.* (2017) 189:127–31. doi: 10.1111/cei.12964
133. Curtin F, Perron H, Kromminga A, Porchet H, Lang AB. Preclinical and early clinical development of GNBAC1, a humanized IgG4 monoclonal antibody targeting endogenous retroviral MSRV-Env protein. *MAbs* (2015) 7:265–75. doi: 10.4161/19420862.2014.985021
134. Curtin F, Perron H, Faucard R, Porchet H, Lang AB. Treatment against human endogenous retrovirus: a possible personalized medicine approach for multiple sclerosis. *Mol Diagn Ther.* (2015) 19:255–65. doi: 10.1007/s40291-015-0166-z
135. Brütting C, Emmer A, Kornhuber M, Staeger MS. A survey of endogenous retrovirus (ERV) sequences in the vicinity of multiple sclerosis (MS)-associated single nucleotide polymorphisms (SNPs). *Mol Biol Rep.* (2016) 43:827–36. doi: 10.1007/s11033-016-4004-0
136. Emmer A, Brütting C, Kornhuber M, Staeger MS. Genetic determinants of antibody levels in cerebrospinal fluid in multiple sclerosis: possible links to endogenous retroviruses. *Int J Mol Sci.* (2018) 19:E786. doi: 10.3390/ijms19030786
137. Nexø BA, Christensen T, Frederiksen J, Møller-Larsen A, Oturai AB, Villesen P, et al. The etiology of multiple sclerosis: genetic evidence for the involvement of the human endogenous retrovirus HERV-Fc1. *PLoS ONE* (2011) 6:e16652. doi: 10.1371/journal.pone.0016652
138. Mangeney M, Heidmann T. Tumor cells expressing a retroviral envelope escape immune rejection *in vivo*. *Proc Natl Acad Sci USA.* (1998) 95:14920–5. doi: 10.1073/pnas.95.25.14920
139. Blaise S, Mangeney M, Heidmann T. The envelope of Mason-Pfizer monkey virus has immunosuppressive properties. *J Gen Virol.* (2001) 82:1597–600. doi: 10.1099/0022-1317-82-7-1597
140. Haraguchi S, Good RA, Cianciolo GJ, Day NK. A synthetic peptide homologous to retroviral envelope protein down-regulates TNF- α and IFN- γ mRNA expression. *J Leukoc Biol.* (1992) 52:469–72. doi: 10.1002/jlb.52.4.469
141. Haraguchi S, Good RA, James-Yarish M, Cianciolo GJ, Day NK. Differential modulation of Th1- and Th2-related cytokine mRNA expression by a synthetic peptide homologous to a conserved domain within retroviral envelope protein. *Proc Natl Acad Sci USA.* (1995) 92:3611–5. doi: 10.1073/pnas.92.8.3611
142. Li F, Nellåker C, Sabuncian S, Yolken RH, Jones-Brando L, Johansson AS, et al. Transcriptional derepression of the ERVWE1 locus following influenza A virus infection. *J Virol.* (2014) 88:4328–37. doi: 10.1128/JVI.03628-13
143. Young GR, Mavrommatis B, Kassiotis G. Microarray analysis reveals global modulation of endogenous retroelement transcription by microbes. *Retrovirology* (2014) 11:59. doi: 10.1186/1742-4690-11-59
144. Nellaker C, Yao Y, Jones-Brando L, Mallet F, Yolken RH, Karlsson H, et al. Transactivation of elements in the human endogenous retrovirus W family by viral infection. *Retrovirology* (2006) 3:44. doi: 10.1186/1742-4690-3-44
145. Melder DC, Pankratz VS, Federspiel MJ. Evolutionary pressure of a receptor competitor selects different subgroup A avian leukosis virus escape variants with altered receptor interactions. *J Virol.* (2003) 77:10504–14. doi: 10.1128/JVI.77.19.10504-10514.2003
146. Spencer TE, Mura M, Gray CA, Griebel PJ, Palmarini M. Receptor usage and fetal expression of ovine endogenous betaretroviruses: implications for coevolution of endogenous and exogenous retroviruses. *J Virol.* (2003) 77:749–753. doi: 10.1128/JVI.77.1.749-753.2003
147. Monde K, Contreras-Galindo R, Kaplan MH, Markovitz DM, Ono A. Human endogenous retrovirus K Gag coassembles with HIV-1 Gag and reduces the release efficiency and infectivity of HIV-1. *J Virol.* (2012) 86:11194–208. doi: 10.1128/JVI.00301-12
148. Monde K, Terasawa H, Nakano Y, Soheilian F, Nagashima K, Maeda Y, et al. Molecular mechanisms by which HERV-K Gag interferes with HIV-1 Gag assembly and particle infectivity. *Retrovirology* (2017) 14:27. doi: 10.1186/s12977-017-0351-8
149. Blanco-Melo D, Gifford RJ, Bieniasz PD. Co-option of an endogenous retrovirus envelope for host defense in hominid ancestors. *Elife* (2017) 6:1–19. doi: 10.7554/eLife.22519
150. Yan L, Yang M, Guo H, Yang L, Wu J, Li R, et al. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat Struct Mol Biol.* (2013) 20:1131–9. doi: 10.1038/nsmb.2660
151. Chiappinelli KB, Zahnow CA, Ahuja N, Baylin SB. Combining epigenetic and immune therapy to combat cancer. *Cancer Res.* (2016) 76:1683–9. doi: 10.1158/0008-5472.CAN-15-2125

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Grandi and Tramontano. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Exaptation of HERV-H: Evolutionary Analyses Reveal the Genomic Features of Highly Transcribed Elements

Patrick Gemmell¹, Jotun Hein² and Aris Katzourakis^{1*}

¹ Department of Zoology, University of Oxford, Oxford, United Kingdom, ² Department of Statistics, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:

Tara Patricia Hurst,
Abcam, United Kingdom

Reviewed by:

Martin Sebastian Staeger,
Martin Luther University of
Halle-Wittenberg, Germany
Nicole Grandi,
University of Cagliari, Italy

*Correspondence:

Aris Katzourakis
aris.katzourakis@zoo.ox.ac.uk

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 05 July 2018

Accepted: 28 May 2019

Published: 09 July 2019

Citation:

Gemmell P, Hein J and Katzourakis A
(2019) The Exaptation of HERV-H:
Evolutionary Analyses Reveal the
Genomic Features of Highly
Transcribed Elements.
Front. Immunol. 10:1339.
doi: 10.3389/fimmu.2019.01339

HERV-H endogenous retroviruses are thought to be essential to stem cell identity in humans. We embrace several decades of HERV-H research in order to relate the transcription of HERV-H loci to their genomic structure. We find that highly transcribed HERV-H loci are younger, more fragmented, and less likely to be present in other primate genomes. We also show that repeats in HERV-H LTRs are correlated to where loci are transcribed: type-I LTRs associate with stem cells while type-II repeats associate with embryonic cells. Our findings are generally in line with what is known about endogenous retrovirus biology but we find that the presence of the zinc finger motif containing region of *gag* is positively correlated with transcription. This leads us to suggest a possible explanation for why an unusually large proportion of HERV-H loci have been preserved in non-solo-LTR form.

Keywords: HERV-H, endogenous retrovirus, transcription, stem cell, exaptation

BACKGROUND

Endogenous retroviruses (ERVs) are the result of germ line retroviral integrations that are passed from one generation to another in a Mendelian fashion. At integration, a typical ERV locus contains the viral genes *gag*, *pol*, and *env*, as well as two flanking long terminal repeats (LTRs) that are identical. Once present in a population, ERVs may increase in number via reinfection or retrotransposition (1) until all active members of the family are silenced by host defenses (1, 2), degraded by mutations, or truncated by solo-LTR formation (1). Most human endogenous retroviruses (HERVs) are highly degraded, with a small number of described cases where they have been co-opted by their host, preserving their sequence, and now play roles in human biology. HERV derived interferon-inducible enhancers regulate human innate immune pathways (3), with the widespread recruitment of endogenous viruses in host immunity across animals pointing to a systematic evolutionary process (4). Furthermore, the ERV derived *syncytins* are important for placentation (5), and recent studies suggest that members of the HERV-H family are important to the maintenance of stem cell identity (6, 7).

The HERV-H family—formerly known as RTVL-H, for retrovirus like sequence primed with a primer binding site homologous to histidine tRNA—has been studied down to the level of individual insertions on many occasions in the past 30 years. The LTRs of HERV-H promote its transcription, and variation in these regions have been used to divide the family into three subtypes: type-I, type-Ia, and type-II (8, 9). The recombinant type-Ia subtype was originally thought to have

expanded recently and to have stronger transcription than loci with pure type-I or type-II LTRs (9); however, the relative youth of type-Ia repeats was later brought into question by Anderssen et al. (10), who discovered type-Ia repeats in the marmoset, a New World monkey species. Thus, considerable effort has been made to identify the structural variations of HERV-H that are important to its biological activity.

The sequencing of the human genome was important to a more comprehensive analysis of HERV-H. A study of the majority of full-length HERV-H integrations was performed by Jern et al. (11), who clustered HERV-H into two groups, the larger HERV-H group of 926 full-length elements and the smaller HERV-H like group of 92 full-length elements. Within the larger HERV-H group, Jern et al. (11) used previously studied variants of HERV-H (12, 13) to define two further subgroups: an older group of 77 RGH2-like elements (13) having a fairly intact *pol* and more frequently containing *env*; and a younger group of 705 RTVLH2-like elements (12) having more *pol* deletions and less frequently containing *env*. The 926 bona-fide HERV-H elements in the human genome were subsequently studied in some detail resulting in an annotated consensus sequence (14). In an unrelated study on ERV replication mechanisms, a small number of relatively intact HERV-H (presumably the RGH2-like elements) were found to have low *env* dN/dS. It was suggested that these elements had replicated via re-infection and were also responsible for copying other less intact members of the HERV-H family (15). These studies provided further important information on the origin and diversity of HERV-H within the human genome.

In the last few years interest in HERV-H has redoubled as the relationship between retrotransposons and stem cells has become a focus of research. In the case of HERV-H, there is evidence that HERV-H transcripts are upregulated in, and necessary for, the maintenance of human stem cell identity. This evidence derives from studies including those of Wang et al. (7) who show that half of the full-length HERV-H in the human genome are bound by pluripotency associated transcription factors NANOG, OCT5, and LBP9. These loci produce chimeric hESC (human embryonic stem cell) and hiPSC (human induced pluripotent stem cell) specific transcripts and long non-coding RNAs. By disrupting HERV-H or LBP9, Wang et al. (7) could demonstrate that some HERV-H elements play an essential biological role: differential markers were upregulated while pluripotency-associated transcription factors were downregulated, so that an ability for self-renewal was shown to be impaired. The same year, Lu et al. (6) reached a similar conclusion on the importance of HERV-H after discovering that interfering with HERV-H transcripts in hESCs led to modified cells becoming more fibroblast like, with concomitant changes in the appropriate transcriptional markers. More recently, Gemmell et al. (16) showed that the most highly transcribed HERV-H have diverged quickly, and are presumably under directional selection, while Göke et al. (17) studied HERV-H transcription in different cell types and concluded that HERV-H loci are stage-specific regulatory elements in humans. Combined, these studies suggest that HERV-H sequences have been exapted by their hosts, whereby retrovirally derived sequences have been functionally

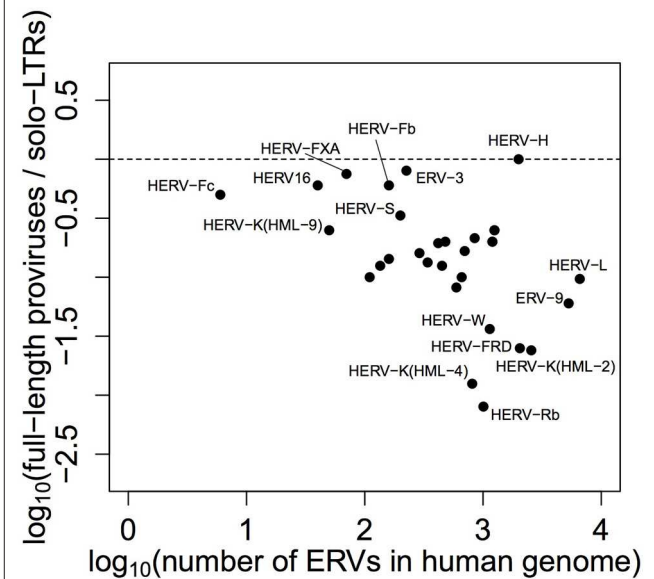


FIGURE 1 | Ratio of ERVs found in full-length form to solo-LTR form, plotted against total number of ERVs, by family—in humans, HERV-H has the highest full-length provirus to solo-LTR ratio (1:1) and is particularly unusual among larger retroviral families. The data plotted in this figure were originally presented by Bannert and Kurth (18) in tabular form.

co-opted for their roles in host biology. These exciting recent results have yet to be integrated with the tradition of studying HERV-H from an evolutionary perspective.

Given the current interest in HERV-H, we think it important to capitalize on several decades of detailed work on the family, including the aforementioned consensus and the LTR repeat types that have previously been shown to affect transcription. Additionally, we think it notable that an especially large number of HERV-H loci have been maintained in non-solo-LTR form (see **Figure 1**), an oddity that has been raised but not resolved in the recent literature (19–21). For these reasons, the crux of this study is as follows: what is the relationship between the genomic characteristics of a HERV-H locus and the level at which it is transcribed? Below we identify the features of HERV-H that are significantly correlated with transcription and show how age, integrity, and the LTR repeat type of a locus affect its transcription levels. We also demonstrate that there is a relationship between *gag* and transcription, and conjecture that the presence of zinc finger motifs in the terminal region of this gene might be related to why so many HERV-H loci are maintained in a full-length state.

RESULTS

We produced multiple sequence alignments of the HERV-H genes *gag* (1,080 sequences), *pol* (1,126 sequences), and *env* (1,081 sequences), using the consensus HERV-H (14) as a guide. (Complete procedures are provided in the Methods section). We used the same consensus to form alignments of the 5'

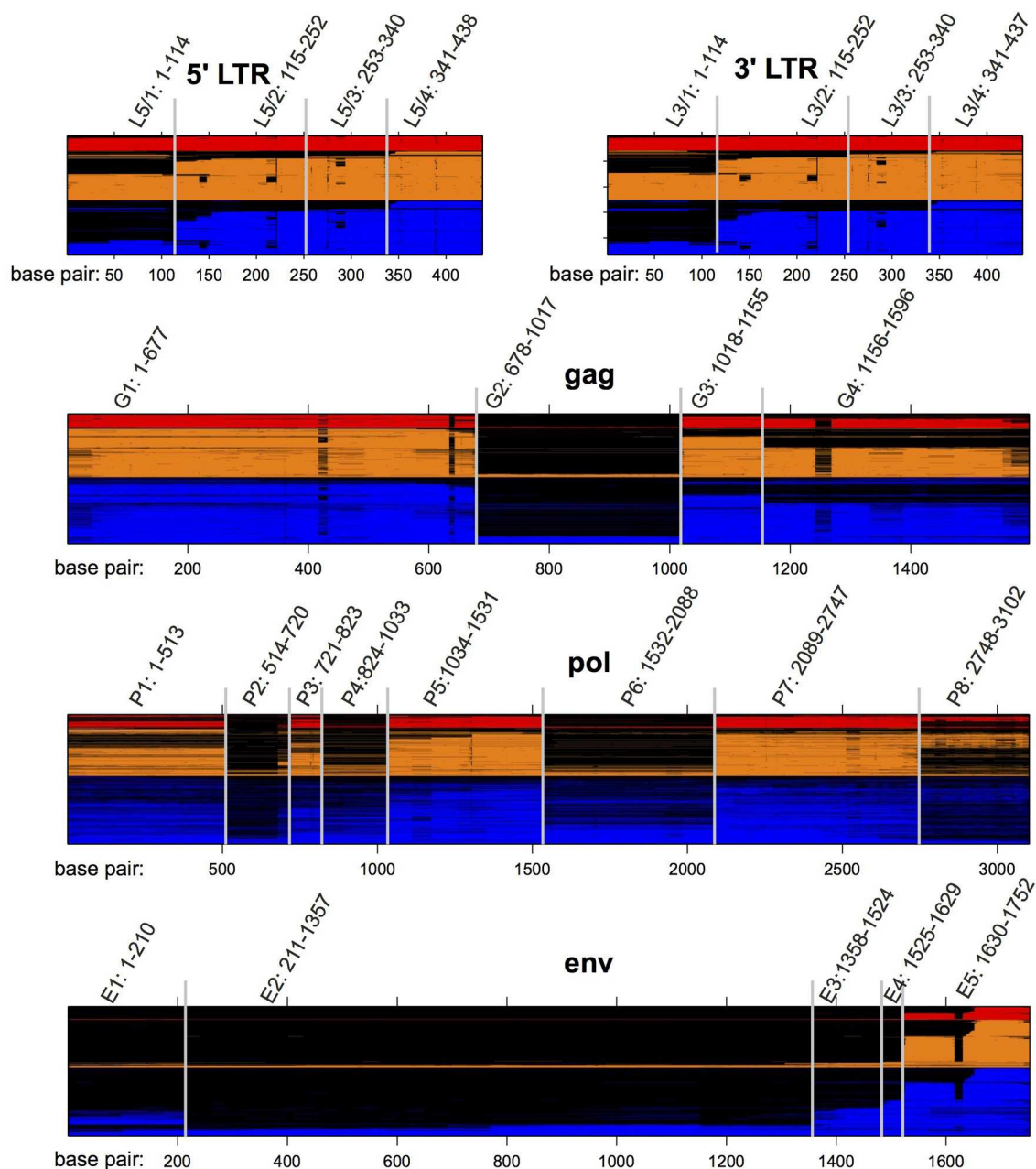


FIGURE 2 | HERV-H sequences from the human genome aligned to the published consensus of Jern et al. (14). Loci are colored according to the categorical transcription levels of Wang et al. (7): red (highly active); orange (moderately active); blue (inactive). Gaps in the alignment are colored black. The locations of regions L5/1–L5/4, G1–G4, P1–P8, E1–E5, and L3/1–3/4 are also noted.

(908 sequences) and 3' (932 sequences) LTRs of HERV-H. Examining the alignments we identified regions that were shared deletions across subsets of viral sequences. We refer to these regions via enumeration from 5' to 3' and they are depicted in **Figure 2**.

BLASTn was used to search for the previously characterized type-I and type-II repeats within the HERV-H LTRs. To place the HERV-H loci in context we obtained their distance from

the nearest gene, and also located 847 of the loci in a six-way alignment of primate genomes, thereby gaining some perspective on their structural state in other primates. Finally, we were able to determine the genetic distance (K80 substitution model) between the paired LTRs of 627 of the HERV-H loci—as the divergence between paired LTRs accumulates at a neutral rate, both genetic distance and orthology contribute information about the age of the loci under investigation.

TABLE 1 | PGLS applied to transcription data from embryonic cells for 409 HERV-H loci: $\lambda = 0.48$ (0.17, 0.78); $R^2 = 0.24$.

Coefficient	Estimate	2.5%	97.5%	Std. err.	t value	p-value
(Intercept)	-5.26	-6.92	-3.60	0.85	-6.22	$p < 0.01$
L5/1	0.76	0.23	1.29	0.27	2.82	$p < 0.01$
G4	0.46	0.10	0.82	0.18	2.50	$p = 0.01$
P6	-0.98	-1.72	-0.23	0.38	-2.57	$p = 0.01$
Type-II	2.24	1.32	3.16	0.47	4.76	$p < 0.01$
Type-Ia	0.71	-0.55	1.97	0.64	1.11	$p = 0.27$

TABLE 2 | PGLS applied to transcription data from stem cells for 409 HERV-H loci: $\lambda = 0.50$ (0.27, 0.73); $R^2 = 0.13$.

Coefficient	Estimate	2.5%	97.5%	Std. err.	t value	p-value
(Intercept)	-6.23	-8.17	-4.29	0.99	-6.30	$p < 0.01$
L5/1	1.28	0.72	1.85	0.29	4.46	$p < 0.01$
G4	0.75	0.21	1.29	0.27	2.72	$p = 0.01$
P6	-0.91	-1.75	-0.06	0.43	-2.11	$p = 0.04$
LTR divergence	-21.21	-32.54	-9.89	5.78	-3.67	$p < 0.01$
Type-I	0.88	0.44	1.33	0.23	3.87	$p < 0.01$

Because HERV-H loci are related, they do not represent independent data points for statistical analysis. To quantify the relationship between HERV-H transcription data (7) and the aforementioned per locus genomic features of HERV-H, we controlled for the non-independence between measurements due to shared evolutionary history by incorporating phylogenetic information. A maximum likelihood (ML) phylogeny of the genes of the HERV-H loci was constructed, and we used a phylogenetic generalized least squares (PGLS) method to control for the non-independence of measurements across the HERV-H family. Of the 834 loci in the phylogeny, 409 had a complete collection of genomic features (i.e., no missing values) and could therefore be used in our PGLS analysis (Tables 1, 2). Akaike information criterion (AIC) based model selection and plots were used to explore model space and we formulated minimum adequate models to explain the features of our dataset i.e., we carefully explored regression space and include here models containing statistically informative explanatory variables. We provide our data in **Additional File 1**. The response variable in our analyses was the logarithm of per locus mean transcription in reads per kilobase per million reads (RPKM) and PGLS regressions were performed using measurements from three distinct cell types defined by Wang et al. (7): somatic (32 sample types); embryonic (114 sample types); and stem (55 hESC and 25 hiPSC sample types). An overview of the methods of Wang et al. is included in our own methods section, which also describes how to link our data to theirs.

Qualitative Analysis

In this section we discuss what we notice about our results by looking at them. This motivates the quantitative analysis that follows in the next section.

The relationship between HERV-H phylogeny, transcription, and genomic structure is visually summarized in **Figure 3**. To aid the reader, a simplified and annotated partial version of this figure is also provided in **Figure 4**. The tips of the tree in **Figure 3** are colored according to the categorical transcription level assigned to them by Wang et al. (7) (see Methods). Upon visual inspection, it can be seen that loci annotated as highly transcribed by Wang et al. (7) are located toward the bottom of the ladderized tree, and we confirmed that this phenomenon was also true of phylogenies built using LTRs and from amino-acid or nucleotide trees of individual HERV-H genes (some examples are contained in **Additional File 2**). Thus, HERV-H transcription appears to have a phylogenetic component.

In which cells is HERV-H active? **Figures 3, 4** represent the average transcription of HERV-H loci across cells categorized as somatic, embryonic, or stem. The figure shows a hierarchy of HERV-H transcription such that transcription is typically lowest in somatic cells, and higher in embryonic cells and stem cells in turn. Note that this comparison between different cell types involves contrasting data gathered in different ways (e.g., cell culture vs. primary sources) by different labs and that any systematic effect this leads to is not something that we correct for.

LTRs identified as type-I outnumber those identified as type-Ia/II. However, considering transcription in conjunction with LTR subtypes, one can also see that type-Ia/II loci are distributed throughout the tree and that type-II repeats are associated with higher levels of embryonic transcription, even if a locus is categorically characterized as inactive or moderately active by Wang et al. (7). Therefore, in agreement with Göke et al. (17), we seem to identify stage-specific transcription, though in this case we relate it to particular well-studied repeat types [see e.g., (10)].

If transcription is positively correlated with function, our results suggest that many parts of consensus HERV-H might have deleterious effects that prevent their exaptation. Columns L5, gag, pol, env, and L3 of **Figure 3** indicate the presence or absence of the viral regions L5/1–L5/4, G1–G4, P1–P8, E1–E5, and L3/1–3/4 at particular loci. Highly active loci toward the bottom of the tree can be seen to have highly intact LTRs, while some less transcribed loci toward the top of the tree have less intact LTRs. Less active older loci toward the top of the tree tend to have relatively complete internal regions when compared to the active loci at the bottom of the figure.

The status of active HERV-H in other primates is also of interest. The sequences used to build the phylogeny in **Figure 3** come from the human genome. A rough indication of the age and status of HERV-H loci in other primates is given by the column *cgom*. Loci present in chimpanzee, gorilla, orangutan, or macaque at a minimum of 25% of the level that they appear in human, and that therefore presumably consist of more than a solo-LTR, are marked with a dot as appropriate. Visual inspection of **Figures 3, 4** shows that the less derived and less active HERV-H loci that are located toward the top of the tree appear more often present in a substantive way in other primates; the active and derived loci toward the bottom of the tree appear to be absent or degraded in distant primates. This suggests that if HERV-H is important to stem cell biology in non-human primates, then the loci that are involved might be different to those that are involved in humans.

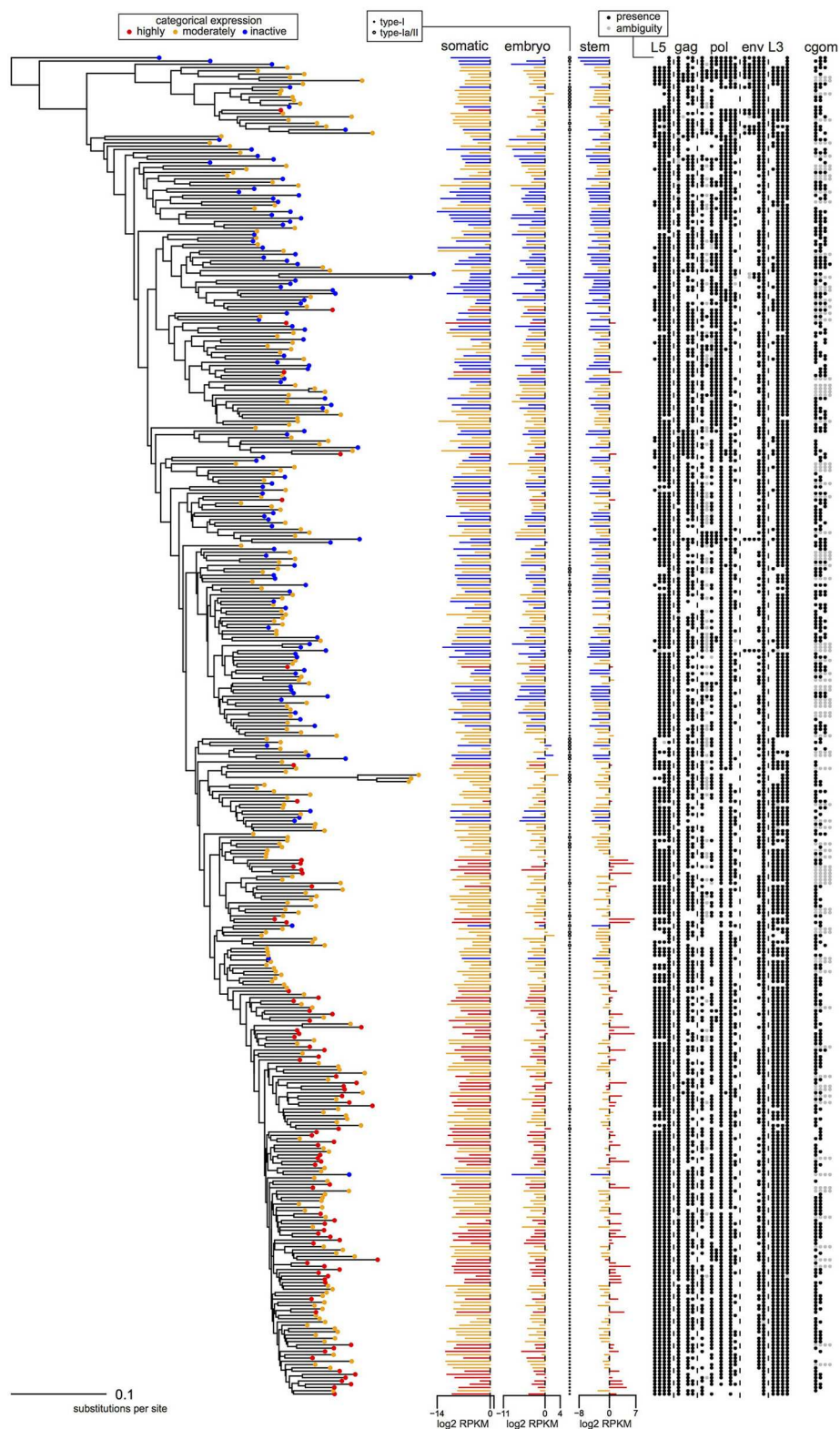


FIGURE 3 | Phylogeny, transcription levels, and per locus genomic features of the 409 HERV-H loci from the human genome that were analyzed using PGLS. Loci are colored according to the categorical transcription levels of Wang et al. (7): highly active (red); moderately active (orange); inactive (blue). Columns *L5*, *gag*, *pol*, *env*, and *L3* indicate the presence/absence of regions L5/1–L5/4, G1–G4, P1–P8, E1–E5, and L3/1–L3/4, respectively; column *cgom* indicates the substantive presence/absence of each locus in chimpanzee, gorilla, orangutan, and macaque.

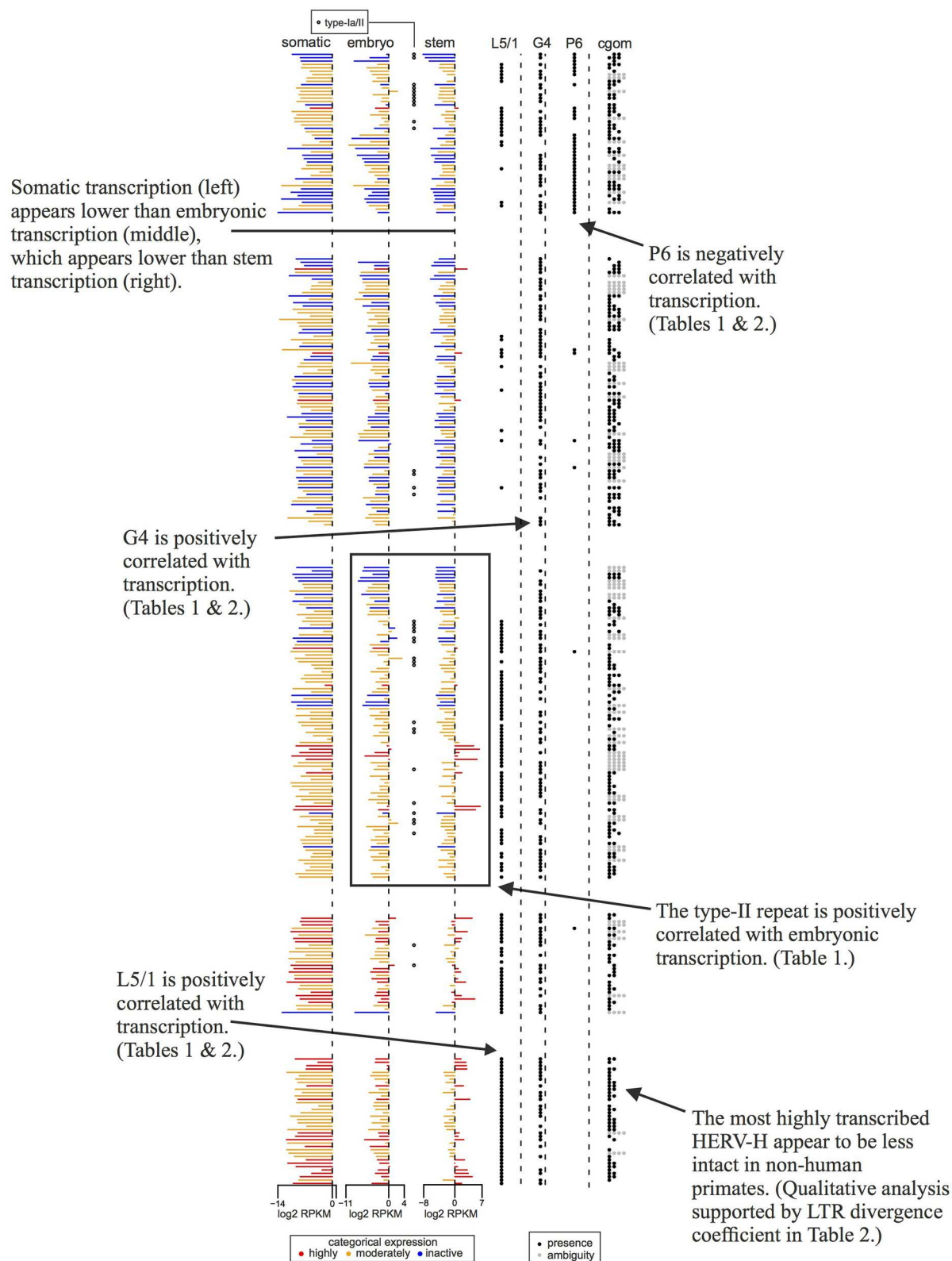


FIGURE 4 | Concise guide to **Figure 3** showing transcription levels and per locus genomic features of a subset of the 409 HERV-H loci. Rows are ordered, colored, and sourced as per **Figure 3**, though some rows are omitted to aid clarity. Annotations highlight aspects of our results that are discussed in detail in the main article text.

Statistical Results

The above observations—younger loci are more transcribed; loci with less intact genes are more transcribed; loci with more intact LTRs are more transcribed; the repeat type of an LTR is relevant to where the corresponding HERV-H is transcribed—are supported by a statistical analysis and robust to phylogenetic non-independence. Two PGLS regressions are reported in detail: **Table 1** displays a regression for transcription measurements from embryonic cells and **Table 2** shows a similar analysis of stem cell data. Both regressions were constructed with respect to the same 409 loci though each model contains only the explanatory variables that were found to be important from a statistical perspective.

In embryonic cells, the presence of region P6 of *pol* is negatively correlated with transcription. The P6 region of *pol* has previously been identified with an RNaseH motif and the 5' end of the integrase domain (14). Positive correlations between embryonic transcription and the presence of the LTR region L5/1, as well as between embryonic transcription and type-II repeats, are also seen. For stem cells, the situation is similar, in that P6 is negatively correlated with transcription, whereas L5/1 is positively correlated with transcription. However, **Table 2** also shows that in stem cells it is type-I LTRs (indicator TI + UI) that are positively correlated with transcription, and that the older an ERV is the less active it is likely to be. Surprisingly, given that **Figures 3, 4** suggest that intact loci tend to be less transcribed, the presence of the G4 region of *gag* is positively correlated with transcription in both embryonic cells and stem cells.

A coefficient of determination corrected to take account of correlation structure [see (22), p. 224] can be applied to the regressions presented in **Tables 1, 2**. The regression describing embryonic data has $R^2 = 0.24$, and therefore explains nearly a quarter of the variance in the transcription of HERV-H using only a few features of the loci themselves. The model describing stem cell data has $R^2 = 0.13$, and seems to underfit the most highly transcribed HERV-H loci. Both PGLS models possess intermediate levels of phylogenetic signal which suggests that the evolutionary relationship between HERV-H loci must be taken into account in order to make correct statistical inferences relating structure and transcription. As expected, given that there is no reported role for HERV-H in somatic cells, we were unable to produce a PGLS regression with $R^2 > 0.01$ for somatic cell measurements—therefore we do not think there is a meaningful relationship between the structure of the majority of HERV-H loci and their somatic transcription.

DISCUSSION AND CONCLUSIONS

We found the repeat class of LTRs to be related to where the corresponding HERV-H is transcribed. HERV-H loci classified as type-I were roughly twice as highly transcribed in stem cells when compared to other loci. On the other hand, we found the presence of a type-II repeat increased transcription of HERV-H loci by a factor of four in embryonic cells. The finding that type-II repeats are preferentially transcribed in embryonic cells is potentially important in light of the widespread interest in cultivating naive human cells. Mouse ESC cultures have been shown to

contain sub-populations of naive-like cells that have totipotent properties (23), and these totipotent cells have transcriptional characteristics that are usually associated with 2-cell embryos. Our analyses show that the type-II repeat is correlated with HERV-H transcription in early human embryonic cells. At the same time, these loci with type-II repeats do not appear to be a focus of Wang et al. (7) as they do not usually cluster into the highly active category (**Figures 3, 4**). As the work of Macfarlan et al. (23) shows that the natural timing of the transcription of repeat loci is relevant for identifying naive-like cells in mouse, it is worth considering whether a similar situation is true for humans also.

We also demonstrated that the integrity of HERV-H LTRs is important to the magnitude of HERV-H transcription. In particular, transcription of HERV-H is consistently correlated with the presence of the first 114 bp (L5/1) of the consensus LTR. This short sequence is part of the larger U3 region and is known to contain a MYB binding site (76–80) as well as to finish with an Sp1 binding site (105–114). The correlation makes sense given Sp1 binding sites have been shown to be important in previous assays (10, 24).

A related feature of our results is the fragmented nature of the *pol* and *env* genes of many HERV-H loci (**Figures 2, 3**). It has previously been shown that the decay of *env* is a common feature of the most prolific families of endogenous retroviruses (25) and HERV-H seems to be typical in this respect. In addition it is sensible to assume that a more complete (i.e., longer) *pol* or *env* will contain more sequence that triggers host defenses (2) or is more likely to be involved in harmful ectopic recombination events [e.g., (26)]. In particular, we find here that the P6 region of *pol* is negatively correlated with transcription to the extent that the presence of P6 is associated with a roughly 2-fold decrease in HERV-H transcription. It will be interesting to determine whether the P6 region attracts repressive chromatin/DNA marks that could explain this observation. This result shows that, even today, more complete HERV-H loci are subject to reduced transcription in their host.

In light of the preceding discussion it is striking that the *gag* gene is relatively complete in the majority of HERV-H, and that the presence of the G4 region of *gag* is associated with a 38 and 68% increase in HERV-H transcription in embryonic and stem cells, respectively. The consensus HERV-H *gag* has previously been annotated and was described as containing two zinc finger motifs (14). These motifs are located in region G4 (positions 1390–1431 and 1459–1497) and would have originally enabled proteins encoded by the exogenous progenitor of HERV-H to bind single-stranded viral RNA as part of a packaging process. The association between these motifs and HERV-H transcription levels may be a statistical coincidence, but it is unexpected enough to encourage speculation on the alternative that there might be a causal explanation for the relationship.

We draw upon recent research that shows that endogenous viral sequences have been frequently repurposed by their hosts to provide functional roles in host immunity. This process results in virally derived genes that are now functioning as host immune genes, which have been termed endogenous viral element-derived immunity (EDIs) (27). We propose that just

as exogenous zinc finger motifs enable retroviral proteins to package RNA to the benefit of an exogenous virus, so too might endogenous zinc finger motifs have at one time enabled HERV-H proteins to bind viral RNA to the benefit of the host. The general phenomenon of viral material being captured by a host for immune benefits is well-established and occurs via many mechanisms and in many organisms (27). In this case our suggestion is that insofar as zinc finger motif-complete products of HERV-H were in other ways defective, or bound related viruses that they did not package, they would act in competition with genuinely effective packaging proteins and therefore as antagonists against viral reinfection. This scenario invokes a new role for zinc finger domains in host defense as compared to their (unrelated) known role as part of host KRAB zinc finger based defenses (28–31).

Reflecting on the above discussion we conclude with the following narrative. In the distant past exogenous HERV-H was a complete retrovirus adapted to horizontal transmission and at some point roughly 35 million years ago an endogenization occurred (32). It is possible that HERV-H then drifted to fixation. The intact nature of HERV-H *gag* is in stark contrast to the fragmented *pol* and *env* genes that are the hallmarks of prolific endogenous retroviruses. If we suppose the relationship between the zinc finger motif containing region of *gag* and transcription is important, this would go some way toward explaining why HERV-H is present at a roughly 1:1 full-length to solo-LTR ratio in the human genome. This ratio is extraordinary, but it could be that while HERV-H LTRs drive transcription, at some point in time the zinc finger motifs in the last third of *gag* made some HERV-H loci useful to the host, and consequently more likely to be at first tolerated and then later co-opted for a role in stem cell identity. This would suggest we view co-opted HERV-H as a biologically convenient aggregation of LTRs and the G4 region of *gag*, both embedded in an ancestral ERV packaging, the remainder of which was of little functional consequence over evolutionary time.

METHODS

Multiple Sequence Alignments of HERV-H Loci

We obtained the genomic sequence underlying the 1,225 full-length HERV-H loci described by (7) from the UCSC Genome Browser database available at <http://genome.ucsc.edu> (33), and used Genome Reference Consortium Human Build 37 (GRCh37, h19) for the human genome. We also obtained the corresponding RepeatMasker (<http://www.repeatmasker.org>) gene annotations from the same database. For the non-human primates, we used the following assembly versions: Chimp: CHIMP2.1.4, Gorilla: gorGor3.1, Orangutan: PPYG2, Macaque: MMUL1.0, Marmoset: C_jacchus3.2.1.

Before aligning the HERV-H loci we first used RepeatMasker annotations to identify any non-HERV-H repeats within the underlying sequences. Such non-HERV-H repeats, for example SINEs or LINEs, would, if ignored, introduce confounding indels into our analysis. For this reason all but the outermost 20

bp of these repeats were removed from the 1,225 HERV-H sequences before constructing alignments. The remaining 40 bp or less of non-HERV-H repeats were flagged so that we could remove them by hand, thereby ensuring that we did not remove mistakenly RepeatMasked genuine HERV-H sequence. Non-HERV-H repeats were found to be rare, so that 80% of the sequence underlying the 1,225 loci remained completely unmodified. The remaining 20% of loci had a median of 12% of their underlying sequence removed.

To identify the *gag*, *pol*, and *env* genes within the 1,225 HERV-H sequences we used tBLASTn (34). We searched each of the 1,225 loci using a previously published (14) HERV-H consensus sequence as a query. Hits of at least 25 bp in length and with an expect value no more than 10^{-6} were merged in a way that maintained fragment order and sense. The result of this reconstruction was 1,080 *gag*, 1,126 *pol*, and 1,081 *env* genes.

Only 20 of the 1,225 HERV-H loci were not matched by one of the three tBLASTn searches. An inspection using Dfam (35) revealed that these loci were sequences with short internal regions, that contained other retroviral insertions, or that were short overall (one full-length region was only 44 bp).

A similar search and merge process was performed for the 5' and 3' LTRs of each of the 1,225 HERV-H loci. In this case the search was performed using BLASTn (34) and resulted in the reconstruction of 908 5' LTRs and 932 3' LTRs.

To construct multiple sequence alignments of the genes and LTRs of the HERV-H loci we first pairwise aligned the reconstructed sequences to the appropriate part of the consensus. We then progressively combined these alignments to create five multiple sequence alignments, one for each of the three genes and one for each LTR. Pairwise alignment was conducted with Stretcher (36) and progressive multiple alignment was conducted using MUSCLE (37).

Identifying Regions Containing Common Deletions

Examining the sequence in each alignment we identified regions that demarcated deletions common to subsets of viral sequences. The presence of sequence in these regions was bimodally distributed and is represented in **Figure 2**, where the regions are named via enumeration from 5' to 3': we identified eight regions in the alignment of HERV-H LTRs (L5/1–L5/4 for the four regions in the 5' LTR and L3/1–L3/4 for the four regions in the 3' LTR); four regions in *gag* (G1–G4); eight regions in *pol* (P1–P8); and four regions in *env* (E1–E4). As per Belshaw et al. (15), if <5% of a region was aligned for a particular HERV-H the region was marked as absent at that locus, while if more than 55% of a region was aligned it was marked as present. Ambiguous regions were coded as missing values. The 25 regions we defined were included as indicator variables (i.e., 1 or 0) when constructing statistical models (below).

Distance to Nearest Gene

The RefSeq gene annotation track was downloaded from the appropriate UCSC Genome Browser database (as detailed above). The distance between the centroid of each HERV-H locus and its nearest neighboring gene was calculated and recorded.

Characterization of LTR Subtypes

Several examples of HERV-H LTR subtypes (9, 32) are provided by Anderssen et al. (10). We constructed consensus type-I (TI) and type-II (TII) repeats as well as consensus unique-I (UI) and unique-II (UII) sequences based on **Figure 2** of the study by Anderssen et al. (10). These consensus sequences were used as queries in a BLASTn search against the 1,225 HERV-H loci. Hits with expect values of no more than 10^{-6} were treated as indicating the presence of the appropriate sequence. Data on the presence of LTR subtype sequences at HERV-H loci are recorded in **Additional File 1**.

Pairing Loci to EPO Multiple Alignments

To examine the status of the 1,225 HERV-H full-length loci in other primates we obtained the Enredo-Pecan-Ortheus (EPO) genome scale multiple sequence alignment from Ensembl Release 71 (38). We then used BLASTn to locate the HERV-H loci in the human row of the EPO alignments. Of the 1,225 HERV-H loci, 847 were unambiguously located (unique and exact matches) within a six-way EPO alignment. The remaining loci were not present in the alignments in an unambiguous form. This could be because the region of the human genome they are located in is not included in the EPO alignments, or because the region they are located in was identified as having been duplicated in a primate other than human.

Some uncertainty remains over loci that were unambiguously located. This is because the EPO methodology could fail to identify genuine orthologs and paralogues in the non-human primates. We note that orthology information from the EPO alignments contributes qualitatively to our results but was not used in our statistical analyses.

Tree Building and Phylogenetic Regression

A supermatrix concatenation of *gag*, *pol*, and *env* alignments was produced. The alignment was edited by hand to remove short or badly aligned regions and sequences that could not reasonably be assumed to be homologous. The tree building software RAxML 8.2.3 (39) was used to produce a maximum likelihood (ML) tree relating the 834 sequences in the supermatrix alignment. Tree inference was performed under the GTR + γ substitution model.

The supermatrix based tree was rooted by constructing an auxiliary phylogeny of the reverse-transcriptase (RT) region of *pol* (nucleotides 82–576 of the *pol* consensus). The 569 HERV-H sequences with relatively complete RT, having over 140 of 165 possible codons, were first translated and then combined with a panel of 15 HERV-W RTs. An ML tree was constructed from the resulting alignment using the PROTGAMMAUTO option.

All regression analysis was conducted with the R system (40). Regressions taking into account phylogeny were performed using a phylogenetic generalized least squares (PGLS) approach as introduced by (41). The λ method (42) was used to assess

phylogenetic signal. Analyses were conducted using the R package APE (43).

Transcription Data

Wang et al. (7) previously curated HERV-H transcription data from a large panel of RNA sequencing experiments performed on a variety of cell types. We have subsequently related this data to genomic features in our own study.

In overview, Wang et al. (7) compared uniquely mapped HERV-H reads to other uniquely mapped reads and RPKM (reads per kilobase million) normalized. Wang et al. (7) also used hierarchical clustering to apply one of three labels (highly active, moderately active, or inactive) to each HERV-H locus as a way of summarizing the transcription of a locus across many cell types and experiments. These data are fully available in Supplementary Table 7 of Wang et al. (7) while the provenance of the reads themselves is available in Supplementary Table 4 of the same article. Our data can be related in full to the data of Wang et al. (7) via the first column of our **Additional File 1**: our serial number 1 to 1,225 corresponds to what Wang et al. refer to as Repeat_ID, which takes values HERVH_1 to HERVH_1225.

The data used by Wang et al. (7) come from a variety of cell types which they grouped into five broad categories including somatic tissue (both tissue specific and mixture), embryonic cells, hESC (human embryonic stem cells), and hiPSC (induced pluripotent stem cells). This processing and grouping of RNA transcription experiment data by Wang et al. (7) was convenient for our study and we relied on their data without subsequent relabeling or adjustment.

AUTHOR CONTRIBUTIONS

PG and AK conceived the study, interpreted the results, and participated in the editing of the manuscript. PG performed the analysis. PG was supervised by AK and JH. All authors read and approved the final manuscript.

FUNDING

PG was supported by the Engineering and Physical Sciences Research Council and AK is funded by the Royal Society.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.01339/full#supplementary-material>

Additional File 1 | This file contains information on the structure of primate HERV-H loci in a machine readable tabular form.

Additional File 2 | This file contains trees built using HERV-H LTRs and individual HERV-H genes.

REFERENCES

1. Katourakis A, Rambaut A, Pybus OG. The evolutionary dynamics of endogenous retroviruses. *Trends Microbiol.* (2006) 13:463–8. doi: 10.1016/j.tim.2005.08.004
2. Wolf G, Greenberg D, Macfarlan TS. Spotting the enemy within: targeted silencing of foreign DNA in mammalian genomes by the Krüppel-associated box zinc finger protein family. *Mob DNA.* (2015) 6:1–20. doi: 10.1186/s13100-015-0050-8

3. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science*. (2016) 351:1083–7. doi: 10.1126/science.aad5497
4. Katourakis A, Aswad A. Endogenous viruses provide shortcuts in antiviral immunity. *Curr Biol*. (2016) 26:R427–9. doi: 10.1016/j.cub.2016.03.072
5. Lavielle C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, et al. Paleovirology of ‘syncytins’, retroviral *env* genes exapted for a role in placentation. *Philos Trans R Soc B Biol Sci*. (2013) 368:20120507. doi: 10.1098/rstb.2012.0507
6. Lu X, Sachs F, Ramsay L, Jacques PÉ, Göke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol*. (2014) 21:423–5. doi: 10.1038/nsmb.2799
7. Wang J, Xie G, Singh M, Ghanbarian AT, Raskó T, Szvetnik A, et al. Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature*. (2014) 516:405–9. doi: 10.1038/nature13804
8. Mager DL. Polyadenylation function and sequence variability of the long terminal repeats of the human endogenous retrovirus-like family RTVL-H. *Virology*. (1989) 173:591–9. doi: 10.1016/0042-6822(89)90570-9
9. Goodchild NL, Wilkinson DA, Mager DL. Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology*. (1993) 196:778–88. doi: 10.1006/viro.1993.1535
10. Anderssen S, Sjøttem E, Svineng G, Johansen T. Comparative analyses of LTRs of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset, African green monkey, and man. *Virology*. (1997) 234:14–30. doi: 10.1006/viro.1997.8590
11. Jern P, Sperber GO, Blomberg J. Definition and variation of human endogenous retrovirus H. *Virology*. (2004) 327:93–110. doi: 10.1016/j.virol.2004.06.023
12. Mager DL, Freeman JD. Human endogenous retroviruslike genome with type C *pol* sequences and *gag* sequences related to human T-cell lymphotropic viruses. *J Virol*. (1987) 61:4060–6.
13. Hirose Y, Takamatsu M, Harada F. Presence of *env* genes in members of the RTVL-H family of human endogenous retrovirus-like elements. *Virology*. (1993) 192:52–61. doi: 10.1006/viro.1993.1007
14. Jern P, Sperber GO, Ahlsén G, Blomberg J. Sequence variability, gene structure, and expression of full-length human endogenous retrovirus H. *J Virol*. (2005) 79:6325–37. doi: 10.1128/JVI.79.10.6325-6337.2005
15. Belshaw R, Katourakis A, Paçes J, Burt A, Tristem M. High copy number in human endogenous retrovirus families is associated with copying mechanisms in addition to reinfection. *Mol Biol Evol*. (2005) 22:814–7. doi: 10.1093/molbev/msi088
16. Gemmell P, Hein J, Katourakis A. Orthologous endogenous retroviruses exhibit directional selection since the chimp-human split. *Retrovirology*. (2015) 12:52. doi: 10.1186/s12977-015-0172-6
17. Göke J, Lu X, Chan YS, Ng HH, Ly LH, Sachs F, et al. Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell*. (2015) 16:135–41. doi: 10.1016/j.stem.2015.01.005
18. Bannert N, Kurth R. The evolutionary dynamics of human endogenous retroviral families. *Annu Rev Genomics Hum Genet*. (2006) 7:149–73. doi: 10.1146/annurev.genom.7.080505.115700
19. Izsvák Z, Wang J, Singh M, Mager DL, Hurst LD. Pluripotency and the endogenous retrovirus HERVH: conflict or serendipity? *Bioessays*. (2016) 38:109–17. doi: 10.1002/bies.201500096
20. Gemmell P, Hein J, Katourakis A. Phylogenetic analysis reveals that ERVs “die young” but HERV-H is unusually conserved. *PLoS Comput Biol*. (2016) 12:e1004964. doi: 10.1371/journal.pcbi.1004964
21. Römer C, Singh M, Hurst LD, Izsvák Z. How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency. *Curr Opin Virol*. (2017) 25:49–58. doi: 10.1016/j.coviro.2017.07.001
22. Paradis E. *Analysis of Phylogenetics and Evolution With R*. New York, NY: Springer Science & Business Media (2012).
23. Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, et al. Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature*. (2012) 487:57–63. doi: 10.1038/nature11244
24. Sjøttem E, Anderssen S, Johansen T. The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3’ to the TATA box. *J Virol*. (1996) 70:188–98.
25. Magiorkinis G, Gifford RJ, Katourakis A, De Ranter J, Belshaw R. *Env*-less endogenous retroviruses are genomic superspreaders. *Proc Natl Acad Sci USA*. (2012) 109:7385–90. doi: 10.1073/pnas.1200913109
26. Gonzalez J, Petrov DA. Evolution of genome content: population dynamics of transposable elements in flies and humans. In: Anisimova M, editor. *Evolutionary Genomics: Statistical and Computational Methods*, Vol. 1. Totowa, NJ: Humana Press (2012). p. 361–83.
27. Aswad A, Katourakis A. Paleovirology and virally derived immunity. *Trends Ecol Evol*. (2012) 27:627–36. doi: 10.1016/j.tree.2012.07.007
28. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, et al. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res*. (2014) 24:1260–70. doi: 10.1101/gr.172833.114
29. Thomas JH, Schneider S. Coevolution of retroelements and tandem zinc finger genes. *Genome Res*. (2011) 21:1800–12. doi: 10.1101/gr.121749.111
30. Lukic S, Nicolas JC, Levine AJ. The diversity of zinc-finger genes on human chromosome 19 provides an evolutionary mechanism for defense against inherited endogenous retroviruses. *Cell Death Differ*. (2014) 21:381–7. doi: 10.1038/cdd.2013.150
31. Jacobs FM, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katman S, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature*. (2014) 516:242–5. doi: 10.1038/nature13760
32. Mager DL, Freeman JD. HERV-H endogenous retroviruses: presence in the new world branch but amplification in the old world primate lineage. *Virology*. (1995) 213:395–404. doi: 10.1006/viro.1995.0012
33. Rosenbloom KR, Armstrong J, Barber GP, Casper J, Clawson H, Diekhans M, et al. The UCSC genome browser database: 2015 update. *Nucleic Acids Res*. (2015) 43:D670–81. doi: 10.1093/nar/gku1177
34. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. (1990) 215:403–10. doi: 10.1016/S0022-2836(05)80360-2
35. Wheeler TJ, Clements J, Eddy SR, Hubley R, Jones TA, Jurka J, et al. Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res*. (2013) 41:D70–82. doi: 10.1093/nar/gks1265
36. Rice P, Longden I, Bleasby A. EMBOSS: the European molecular biology open software suite. *Trends Genet*. (2000) 16:276–7. doi: 10.1016/S0168-9525(00)00204-2
37. Edgar RC. Muscle: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. (2004) 32:1792–7. doi: 10.1093/nar/gkh340
38. Flicek P, Amodé MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, et al. Ensembl 2012. *Nucleic Acids Res*. (2012) 40:D84–90. doi: 10.1093/nar/gkr991
39. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. (2014) 30:1312–3. doi: 10.1093/bioinformatics/btu033
40. R Core Team. *A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing (2014).
41. Grafen A. The phylogenetic regression. *Philos Trans R Soc Lond Series B Biol Sci*. (1989) 326:119–57. doi: 10.1098/rstb.1989.0106
42. Pagel M. Inferring evolutionary processes from phylogenies. *Zool Scripta*. (1997) 26:331–48. doi: 10.1111/j.1463-6409.1997.tb00423.x
43. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics*. (2004) 20:289–90. doi: 10.1093/bioinformatics/btg412

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Gemmell, Hein and Katourakis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Related Endogenous Retrovirus-K Elements Harbor Distinct Protease Active Site Motifs

Matthew G. Turnbull^{1†} and Renée N. Douville^{1,2*}

¹ Department of Biology, University of Winnipeg, Winnipeg, MB, Canada, ² Department of Immunology, University of Manitoba, Winnipeg, MB, Canada

Background: Endogenous retrovirus-K is a group of related genomic elements descending from retroviral infections in human ancestors. HML2 is the clade of these viruses which contains the most intact provirus copies. These elements can be transcribed and translated in healthy and diseased tissues, and some of them produce active retroviral enzymes, such as protease. Retroviral gene products, including protease, contribute to illness in exogenous retroviral infections. There are ongoing efforts to test anti-retroviral regimens against endogenous retroviruses. Herein, we examine the potential activity and diversity of human endogenous retrovirus-K proteases, and their potential for impact on immunity and human disease.

Results: Sequences similar to the endogenous retrovirus-K HML2 protease and reverse transcriptase were identified in the human genome, classified by phylogenetic inference and compared to Repbase reference sequences. The topologies of trees inferred from protease and reverse transcriptase sequences were similar and agreed with the classification using reference sequences. Surprisingly, only 62/480 protease sequences identified by BLAST were classified as HML2; the remainder were classified as other HML groups, with the majority (216) classified as HML3. Variation in functionally significant protease motifs was explored, and two major active site variants were identified – the DTGAD variant is common in all groups, but the DTGVD motif appears limited to HML3, HML5, and HML6. Furthermore, distinct RNA expression patterns of protease variants are seen in disease states, such as amyotrophic lateral sclerosis, breast cancer, and prostate cancer.

Conclusion: Transcribed ERVK proteases exhibit a diversity which could impact immunity and inhibitor-based treatments, and these facets should be considered when designing therapeutic regimens.

Keywords: endogenous retrovirus-K (ERVK), protease, protease inhibitor, active site motifs, RNAseq, amyotrophic lateral sclerosis, breast cancer, prostate cancer

INTRODUCTION

Retroviridae is a diverse family composed of both exogenous infectious viral species whose life cycle includes stages with a ssRNA genome inside virions which is converted into a dsDNA provirus, and endogenous proviral species transmitted in a Mendelian fashion. The evolutionary pressures on ERVs and exogenous retroviruses (XRVs) are very different; ERVs experience much stronger

OPEN ACCESS

Edited by:

Tara Patricia Hurst,
Abcam, United Kingdom

Reviewed by:

Antoinette Van Der Kuyl,
University of Amsterdam, Netherlands
Larance Ronsard,
Ragon Institute of MGH, MIT and
Harvard, United States

*Correspondence:

Renée N. Douville
r.douville@uwinnipeg.ca
orcid.org/0000-0002-7854-007X
[†]orcid.org/0000-0001-8603-4882

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 01 May 2018

Accepted: 25 June 2018

Published: 18 July 2018

Citation:

Turnbull MG and Douville RN (2018)
Related Endogenous Retrovirus-K
Elements Harbor Distinct Protease
Active Site Motifs.
Front. Microbiol. 9:1577.
doi: 10.3389/fmicb.2018.01577

negative selection against pathogenicity since their survival depends directly on reproduction of their host (Holmes, 2011; Stewart et al., 2011; Barbeau and Mesnard, 2015). In fact, some ERVs are directly involved in placentation (Mi et al., 2000), and some host genes descend from retroviral ancestors (Kaneko-Ishino and Ishino, 2012). Indeed, the role of ERVs in the evolution and function of their host genome is becoming more clear (Bannert and Kurth, 2004; Cowley and Oakey, 2013). The apparent inactivity of many ERVs is probably a direct consequence of this pressure, as many ERV integrations contain inactivating mutations that disrupt the function or expression of viral genes. The most severe inactivating mutation is the case of solo LTRs, where the entire coding region of the ERV is deleted. Despite this, a surprising number of ERVs appear to lack obvious inactivating mutations, and some are even infectious, blurring the line between ERV and XRV (Boller et al., 2008; Jha et al., 2011; Denner and Young, 2013; Kozak, 2014). Although no infectious ERV has been proven to exist in the human genome, some loci are polymorphic and population genetics analysis and functional studies have not ruled out low-level ongoing infectious replication of some human ERVs (Turner et al., 2001; Jha et al., 2011; Naveira et al., 2014; Wildschutte et al., 2014, 2016).

Here we focus on the *Betaretrovirus*-like human ERVK, and particularly the HML2 clade (HK2), which includes loci encoding functional enzymes, and whose youngest members may be less than 200,000 years old (Franklin et al., 1988; Boller et al., 2008). Actively transcribed ERVs, such as *ERVK-10*, are assigned names by the Human Gene Nomenclature Committee as recommended by Mayer et al. (2011). The biological role of transcribed ERVK elements is becoming more clear in recent years (Ruda et al., 2004; Macfarlan et al., 2012; Manghera et al., 2014, 2015; Michaud et al., 2014; Schlesinger et al., 2014; Grow et al., 2015; Li et al., 2015; Bray et al., 2016; Christensen, 2016; Becker et al., 2017; Prudencio et al., 2017), but the effects of individual viral proteins is understudied as compared to their XRV counterparts. Protease-mediated maturation of retroviral structural proteins and enzymes is critical to the XRV life-cycle; viruses lacking PR activity produce virions with an immature morphology and are not infectious (Kohl et al., 1988). Even though ERVs are not known to replicate by infection of new cells, they can increase in copy number by reinfection of their

host cell; this process called reintegration requires the activity of all the viral enzymes (Dewannieux et al., 2006). Because PR function is conserved across *Retroviridae*, a virus with an inactive protease may be complemented by another retroviral protease. The factors which permit or exclude this complementation must be context dependent; for example, the HK2 protease encoded by *ERVK-10* cleaves HIV-1 Gag *in vitro*, although not *in vivo* (Towler et al., 1998). The contextual nature of complementation is not perfectly understood, so care must be taken in applying findings across model systems, and particularly in the context of human disease (Towler et al., 1998; Contreras-Galindo et al., 2012; Morandi et al., 2015).

Structural analysis proceeds most easily using crystallographic structural determination, which is not yet available for the HK2 PR; however, biochemical and genetic analyses reveal a typical A2 aspartic protease, with each 106-residue monomer of the mature dimer processed autocatalytically from a larger precursor (Mueller-Lantzsch et al., 1993; Towler et al., 1998; Kuhelj et al., 2001; Dunn et al., 2002). The active HK2 PR is moderately sensitive to pepstatin, with a pH optimum around 4.5 (Towler et al., 1998; Kuhelj et al., 2001). Some HK2 elements (such as *ERVK-10*) encode a functional protease, and HK2 virions with condensed cores have been observed budding from human cells (Mueller-Lantzsch et al., 1993; Sauter et al., 1995; Towler et al., 1998; Boller et al., 2008). Furthermore, the two reconstructed HK2 viruses *Phoenix* and *ERVK_{CON}* have a PR-dependent infectivity (Dewannieux et al., 2006; Lee and Bieniasz, 2007). Although the substrate affinity of the HK2 PR has not been extensively studied, the PR cleavage sites in the HK2 Gag polyprotein are known (Kraus et al., 2011), as is its susceptibility to select PIs (Towler et al., 1998; Kuhelj et al., 2001).

In addition to catalyzing maturation of viral proteins, PR may play a role in immunity and pathology of human diseases. Protease pathogenicity has been best studied in HIV-1, whose PR is known to cleave a variety of host proteins (Shoeman et al., 1991; Impens et al., 2012), and to trigger apoptosis under specific conditions (Shoeman et al., 1991; Rumlova et al., 2014). HIV-1 PR also directly combats innate immunity by trafficking the dsRNA sensor RIG-I to the lysosome (Solis et al., 2011). Another example of HIV PR inhibition of innate immunity is its ability to cleave RIPK1 and RIPK2 proteins, thus successfully abrogating NF- κ B signaling (Wagner et al., 2015). The effect of proteases is not limited to innate immune signaling proteins, but also effector proteins such as intrinsic restriction factors. Both FIV and MLV cleave APOBEC3 in their respective hosts as a means to evade this antiviral defense (Abudu et al., 2006; Yoshikawa et al., 2017). Given the known pathogenic potential of retroviral proteases, it seems prudent to explore the potential effects of PRs encoded by ERVK which, in contrast to HIV-1, are present in all human beings – and which have been associated with specific human diseases, such as ALS, schizophrenia, rheumatic disease and cancer (Frank et al., 2005; Seifarth et al., 2005; Reynier et al., 2009; Douville et al., 2011; Li et al., 2015; Christensen, 2016).

Here, we undertake genomic, sequence-function, and transcriptomic analysis of the diverse ERV PRs in the human genome. We draw on the existing ERV and XRV literature to evaluate the phylogenetic distribution of sequence motifs

Abbreviations: AA, peptide/amino acid; ALS, amyotrophic lateral sclerosis; BLAST, basic local alignment search tool; CA, Capsid; DNA, deoxyribonucleic acid; dsRNA, double stranded RNA; DU, dUTPase; ERV, endogenous retrovirus; ERVK, endogenous retrovirus-K; FTP, file transfer protocol; GIMP, GNU image manipulation program; GRCh38, genome research consortium human genome release 38; GTR, general time-reversible; HIV, human immunodeficiency virus; HK, human endogenous retrovirus-K HML; HML, homology to mouse mammary tumor virus-like; HMM, hidden Markov model; IN, integrase; JTT, Jones, Taylor, and Thornton; LTR, long terminal repeat; MA, matrix; MACSE, multiple alignment of coding sequences; NCB, National Computational Biology Institute; NC, nucleocapsid; NT, nucleotide/nucleic acid; PI, protease inhibitor; PNG, portable network graphics; PR, protease; RAXML, randomized accelerated maximum likelihood; RH, RNase H; RIG, retinoic acid inducible gene; RNA, ribonucleic acid; RNA-Seq, RNA sequencing; RT, reverse transcriptase; sALS, sporadic ALS; SRA, sequence read archive; ssRNA, single stranded RNA; SU, surface subunit of envelope; WAG, Whelan and Goldman; XIST, X inactive specific transcript; XRV, exogenous retrovirus.

with potential functional relevance, focusing on HK2 PRs. We accomplish this by identifying human ERV protease sequences, inferring their phylogenetic affiliation to each other, and comparing them to representative sequences. The classification process was carried out in parallel using RT as a standard comparator. Here we show that distinct ERVK PR variants with predicted differences in their functional activity are differentially transcribed in disease-relevant human tissues.

RESULTS

Representative Retroviral Motifs

With the goal of using the most recent human genome build GRCh38 to identify loci encoding retroviral PR and RT enzymes, HMMs were used as a search tool. Pfam-A was searched for all relevant protein families related to retroviruses. **Table 1** describes the output from this search, revealing 92 HMMs in total, with 40 directly related to retroviral core and accessory proteins. Most important amongst these are the retroviral protease domain RVP, and the RT core domain RVT_1. Retrovirus-associated HMMs (**Table 1**), the PR sequence of *ERVK-10* (Towler et al., 1998), and the RVT_1 domain of *Phoenix* (Dewannieux et al., 2006), were subsequently used to identify *pro* and *pol* sequences in representative retroviral sequences and in the human genome.

Representative Endogenous Retroviral Sequences

As expected, BLAST identified more, but less diverse sequences, than HMMER in both the human genome and Repbase consensa. tBLASTn for HK2 PR matched every ERVK clade except HK10, but no other ERV. HMMER identified an RVP domain in every ERVK clade, and most other ERVs. The regions identified by each method overlap. tBLASTn for the *Phoenix* RVT_1 domain matched most ERVs. HMMER also identified an RVT_1 domain in most ERVs. The results of these methods mostly overlapped, with some notable differences. BLAST identified only part of the RVT_1 domain on ERVE, and the RVT_1 domain on ERV9 and ERVfc covers only part of the BLAST result. Each of the BLAST and HMMER results from ERVfb only partially overlapped. HMMER identified an RVT_1 domain on ERVT, but BLAST found no match. Both methods disproportionately detected HML2 loci; BLAST did so because the query sequences were HML2-derived, whereas HMMER did so because HML2 loci are more intact and therefore more easily discovered by LTRharvest.

The nucleotide sequences from each search were aligned using their translation and their phylogeny was inferred using RAXML from both nucleotide and amino acid (AA) alignments (see **Figure 1** for PR tBLASTn results). The resulting alignments and trees were used to classify the genomic sequences identified by the corresponding search. These reference trees have very low bootstrap support values; however, *pol* and *pro* trees have similar topology and the classifications based on evolutionary placement agree with the tree topology inferred directly from genomic *pro* sequences.

TABLE 1 | Retroviral HMMs from Pfam sorted according to the retroviral protein product from which they were derived.

Protein	Pfam HMMs			
• MA	• Gag_p10	• Gag_p12	• zf-CCHC_5	
• CA	• zf-CCHC	• Gag_p30	• Gag_p17	
• NC	• Gag_p24	• zf-H2C2	• Gag_p19	
	• Gag_MA	• gag_pre-integrs	• Retro_M	
	• Retrotrans_gag			
• PR	• Asp_protease_2	• RVP	• Spuma_A9PTase	
	• gag-asp_protease	• RVP_2		
• RT	• RVT_1	• RVT_3	• RVT_N	
	• RVT_2	• RVT_connect	• RVT_thumb	
• RH	• RNase_H			
• IN	• IN_DBD_C	• Rve	• Integrase_Zn	
• SU	• GP41	• HERV-K_env_2	• Env_gp36	
	• Avian_gp85	• MMTV_SAg		
• DU	• dUTPase			
• Accessory proteins	• Deltaretro_Tax	• Myc-N	• Tax	
	• F-protein	• Myc-LZ	• VPR	
• Other	• SPX	• Ribosomal_L22	• UBN2_2	
	• Borrelia_orfX	• Chromo	• UBN2_3	
	• Gypsy	• PHD	• Exo_endo_phos_2	
	• Peptidase_A17	• PHO4	• Ets	
	• MRI	• G-patch	• Flu_PA	
	• TYA	• PhoH	• RdRP_1	
	• Orbi_VP4	• EXS	• Mononeg_RNA_pol	
	• EVI2A	• Exo_endo_phos	• Flavi_NS5	
	• Peptidase_A2B	• Tctex-1	• Bromo_coat	
	• ALIX_LYPXL_bnd	• MRV1	• RdRP4	
	• Reo_sigma1	• Pkinase_Tyr	• Peptidase_A3	
	• TLV_coat	• DUF1725	• Flavi_glycop_C	
	• DUF1011	• DUF2155	• BAF	
	• HLH	• AA_permease_2	• RNA_replicase_B	
	• Pkinase	• Cupin_8	• Birna_RdRp	
	• Ras	• AA_permease_C	• Peptidase_C34	
	• bZIP_1	• DUF4219	• BLVR	
	• UQ_con			

HMMs in the “other” category correspond to common functional motifs, non-retroviral proteins, or domains with unknown function.

Classification of ERV Sequences in the Human Genome

Having established a system to identify PR and RT elements, ERV nucleotide sequences were identified in GRCh38 using BLAST and HMMER as directed by GenomeTools. These were then curated to eliminate highly divergent sequences with rare insertions (less frequent than 1/20 sequences) and to minimize the size of their multiple alignment before phylogenetic inference. LTRdigest hits that did not co-occur with another retroviral domain were eliminated. The genomic distribution of the resulting ERV annotations is presented in **Table 2**, along with curation data.

When examining PR hits, tBLASTn for *ERVK-10* PR identified 480 sequences in GRCh38 (**Additional Files 1, 2**). Forty-two are identical to another PR sequence, of which 15 are from an alternative assembly. LTRdigest identified 150 RVP domains in GRCh38 (**Additional Files 3, 4**). Only 8 of these were shorter than half of the expected length of an RVP domain. Occasionally,

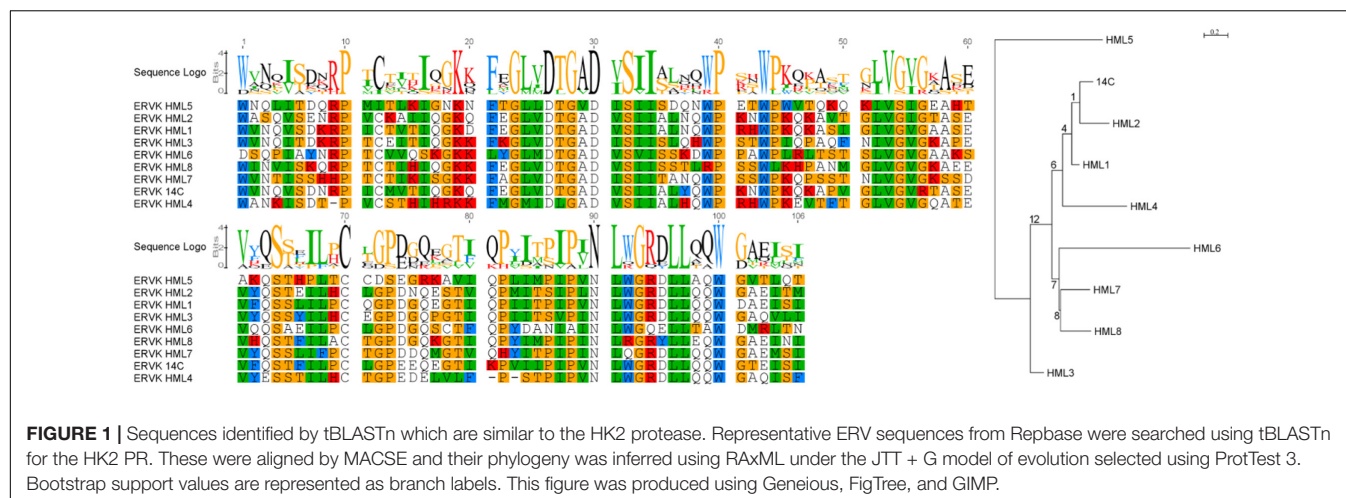


TABLE 2 | Distribution of PR and RT search results in the human genome GRCh38.

Chromosome	tBLASTn <i>ERVK-10</i> PR	LTRdigest RVP	tBLASTn <i>Phoenix</i> RVT_1		LTRdigest RVT_1	
			Raw	Curated	Raw	Curated
1	32	15	97	34	115	15
2	17	6	79	36	95	7
3	28	14	109	57	136	13
4	35	14	113	53	144	15
5	23	11	74	39	140	9
6	30	6	95	37	104	9
7	18	3	58	24	92	3
8	24	13	91	32	85	13
9	8	1	39	13	49	3
10	19	5	52	23	80	8
11	18	12	71	40	104	12
12	15	10	75	37	82	10
13	5	2	34	19	30	1
14	10	4	35	13	57	4
15	7	3	17	7	26	3
16	10	4	23	12	14	2
17	11	3	21	17	10	2
18	2	3	19	5	28	2
19	42	6	76	38	18	6
20	5	0	15	0	20	0
21	3	0	12	6	10	0
22	7	1	14	6	12	1
X	22	7	108	14	267	7
Y	37	7	85	10	48	4
Alternative	52	N/A	64	N/A	N/A	N/A
Total	480	150	1412	572	1766	149

"Raw" refers to all sequences identified, and "Curated" refers only to sequences that were aligned for further analysis. Credible RT sequences are those which were not eliminated for causing gappy sites. Curated RVT_1 sequences are those which co-occurred with another retroviral HMM match. LTRharvest and LTRdigest were not used to search alternative assemblies, so there are no results to report. Since this curation workflow did not exclude any protease search results, the number is shown only once.

sequences from different loci were indistinguishable; twelve were identical to another RVP locus. RVP positions 38–42 are absent in 83 PRs; 48 of these were confidently classified into one of ERVW, ERV9, ERVE, ERVT, or ERV3. The remaining 35 PR sequences

were not confidently classified. In contrast, the PR sequences identified by LTRdigest are less numerous and more diverse than those identified by BLAST. There is a substantial overlap in the sequences recognized by each method, but LTRdigest identified

HK2 PR with a disproportionately higher frequency than other ERVs, given their relative occurrence in the human genome (Gifford and Tristem, 2003; Bannert and Kurth, 2006).

When examining RT hits, tBLASTn for the RVT_1 domain of *Phoenix* identified 1412 sequences in GRCh38, of which 840 were removed because they induced gaps (**Additional Files 5, 6**). LTRdigest identified 1766 RVT_1 domains in GRCh38, of which 1617 were removed either because they did not co-occur with another core retroviral domain or they contained uncommon insertions which induced gaps in the alignment (**Additional Files 7, 8**). Standalone RVT_1 domains were eliminated. This is because non-ERV retroelements, such as LINEs (Boissinot et al., 2000), also encode a RT with an RVT_1 domain, but lack other retroviral proteins. These non-ERV retroelements would skew the analysis, since they are much more numerous; retroelements as a whole make up 42.2% of the human genome, but ERVs make up less than 8.3% [47]. Although LINEs are not flanked by LTRs, they are nonetheless annotated when LTRharvest mistakes other appropriately spaced repetitive sequences for LTRs.

In order to assign appropriate nomenclature to the identified sequences, we classified 489 loci (one or more PR or RT encoding regions separated by no more than 10,000 bp) (**Table 3**). We observe 53 loci contain only one BLAST or LTRdigest result, 324 loci contain 2, 57 loci contain 3, 53 loci contain 4, and one each contain 5 and 6 results. Four loci had all results placed on internal nodes by RAxML and were not classified at all. Eighteen loci contained sequences whose classification varied; in most cases these sequences were truncated or insertion of one retroelement inside another resulted in spurious association of sequences belonging to separate elements.

Totals are shown for each ERV supergroup. The total number of unplaced sequences is also shown along with the number of sequences placed on internal branches, which in both cases were between the H/F and W/9 clades. Sequences were considered unplaced if the probability of their placement into the tree was less than 90%. One sequence was placed on an internal node outside these clades. AA alignments are translated from nucleotide (NT) alignments.

Phylogenetic trees constructed from genomic sequences (**Figure 2**) shared similar trends as those observed in the reference trees (**Figure 1** and **Additional Files 9–16**). The classifications assigned by placement of reference sequences into the phylogeny (**Additional Files 17–32**) agrees with the branching of the phylogeny inferred from genomic sequences directly. This can be clearly seen in **Figure 2**, in which the leaves of these phylogenies are colored according to their classification by evolutionary placement. Each phylogenetic tree (**Additional Files 33–40**) constructed from aligned nucleic acid sequences of either PR or RT from reference ERV sequences contained a bipartition segregating ERVK clades, although it should be noted that BLAST PR results include exclusively ERVK sequences. HK1, 2, 4, 9, and 10 commonly formed a clade within ERVK. In all four trees in **Figure 2**, ERVK leaves form a distinct clade, and likewise ERVW and ERV9 form a clade which is associated with ERVF. ERVS forms a clade with ERVL. ERVI forms a clade with ERVADP. ERVE forms a clade with ERV3. **Figure 2** is intended to allow the reader to quickly see the broad trends; if the reader is

TABLE 3 | Number of sequences classified into each category from each search method.

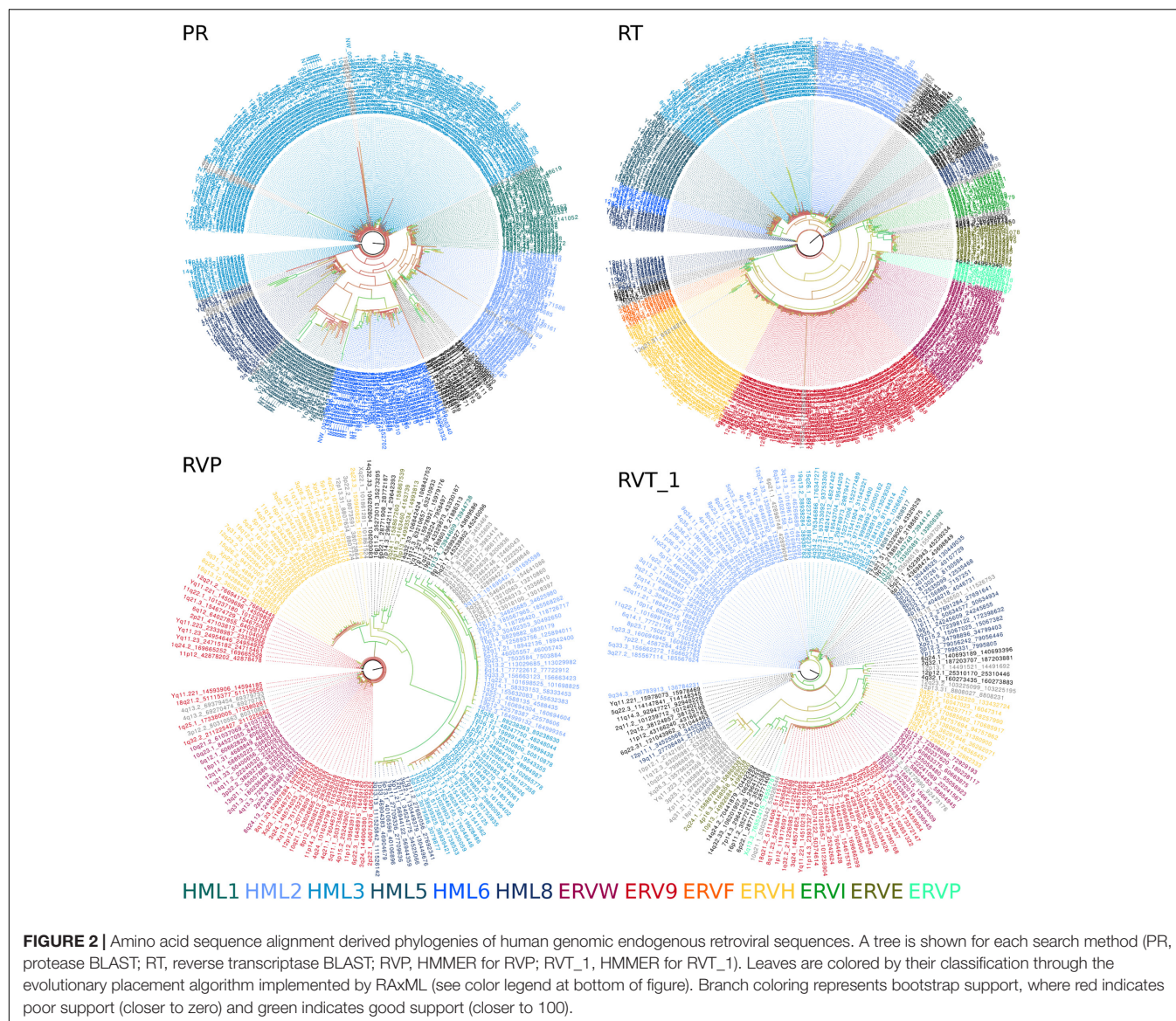
	PR		RVP		RT		RVT_1	
	NT	AA	NT	AA	NT	AA	NT	AA
ERV9					12	12	1	1
ERV9			34	31	114	114	22	22
ERVW			13	13	53	53	10	9
ERVH			24	22	62	61	11	9
ERVF					2	2		
ERVFXA					7	7		
ERVFB					1	1		
ERVVB			1		2	2		
ERVI					20	18		
ERVE			3	3	20	20	3	3
ERV3			2	2	2	2	1	1
ERV7			4	4			4	4
Gammaretrovirus total			81	75	283	280	51	48
HK1	39	39	1	1	16	16	2	2
HK2	62	60	32	20	53	53	36	35
HK4	7	7	2	2	6	6	2	2
HK9 (K14C)	15	12	2	2	5	5	2	2
HK10 (KC4)					5	1	1	
HK3	216	215	22	22	81	80	14	14
HK5	36	36			38	38	1	1
HK6	52	54			11	11		
HK7	3	3	2	2	10	10	7	7
HK8	34	35	8	8	35	34	17	16
Betaretrovirus total	464	461	69	57	260	254	82	79
ERVL					6	6	5	4
UNPLACED	16	19	0	18	10	19	10	17
INTERNAL					1	1		
Total	480		150		572		149	

PR, results from BLAST for HK2 protease; RT, results from BLAST for Phoenix RVT_1 domain of reverse transcriptase; RVP, results from LTRdigest for RVP domain; RVT_1, results from LTRdigest for RVT_1 domain; NT, results classified by their nucleic acid sequence; AA, results classified by their translation.

interested in the placement of specific elements, the phylogenetic trees output by RAxML are included in **Additional Files 33–40**.

Alternate Genome Assemblies Reveal Additional PR Sequences

Although it fails to capture the real diversity of humanity, the reference human genome contains variable regions with multiple valid assemblies. New loci can be distinguished from loci which were duplicated by inclusion in both assemblies by examining the flanking genomic sequences, which should also be identical for technical duplicates. The 15 kilobases flanking the start site of each PR identified in GRCh38 by BLAST and GenomeTools was extracted and compared to alternative assemblies using blastn. Eight regions had a perfect match, of which five were originally identified by BLAST. Many BLAST hits surrounding identical PR sequences had partial matches indicative of closely similar insertions in divergent genomic loci. Moreover, 31 of the 52



regions identified on alternative assemblies had no high scoring matches. The fact that many sequences could not be paired up is consistent with previous observation of unique insertions in a recent analysis of the ERV content within alternative assemblies (Wildschutte et al., 2014).

Diversity of Protease Sequences

The consensus sequences for each ERVK clade from Repbase contain many conserved sites (Figure 1). It is clear that PR functional motifs are less diverse than the surrounding sequences (Figure 3), and that variants occur with different frequencies (Table 4). The α -helix (C2) has only one common variant (GRDLL). The active site loop (B1) has two; the DTGAD motif is most common and occurs in all clades, whereas the second most frequent (DTGVD) occurs only in the HK3, HK5, and HK6 clades. The tree in Figure 4 is displayed to clearly show the abundance of HK3 sequences (top of figure) versus all

other clades. Additionally, Figure 4 highlights the distribution of common active site variants (DTGAD in blue, DTGVD in orange).

Active site motifs were extracted from MACSE aligned protease sequences identified by BLAST in the human genome (GRCh38). Sequences containing stop codons and/or frameshift mutations were removed and their frequency was determined. This was done using standard UNIX utilities. Sequences were considered inactive if they contained mutations not observed in active proteases of other XRVs; otherwise, they were considered potentially active.

ERVK HML2 Proteases Exhibit Great Diversity

Sequences classified as HK2 by RAxML were examined in more detail to determine if their variability could have functional

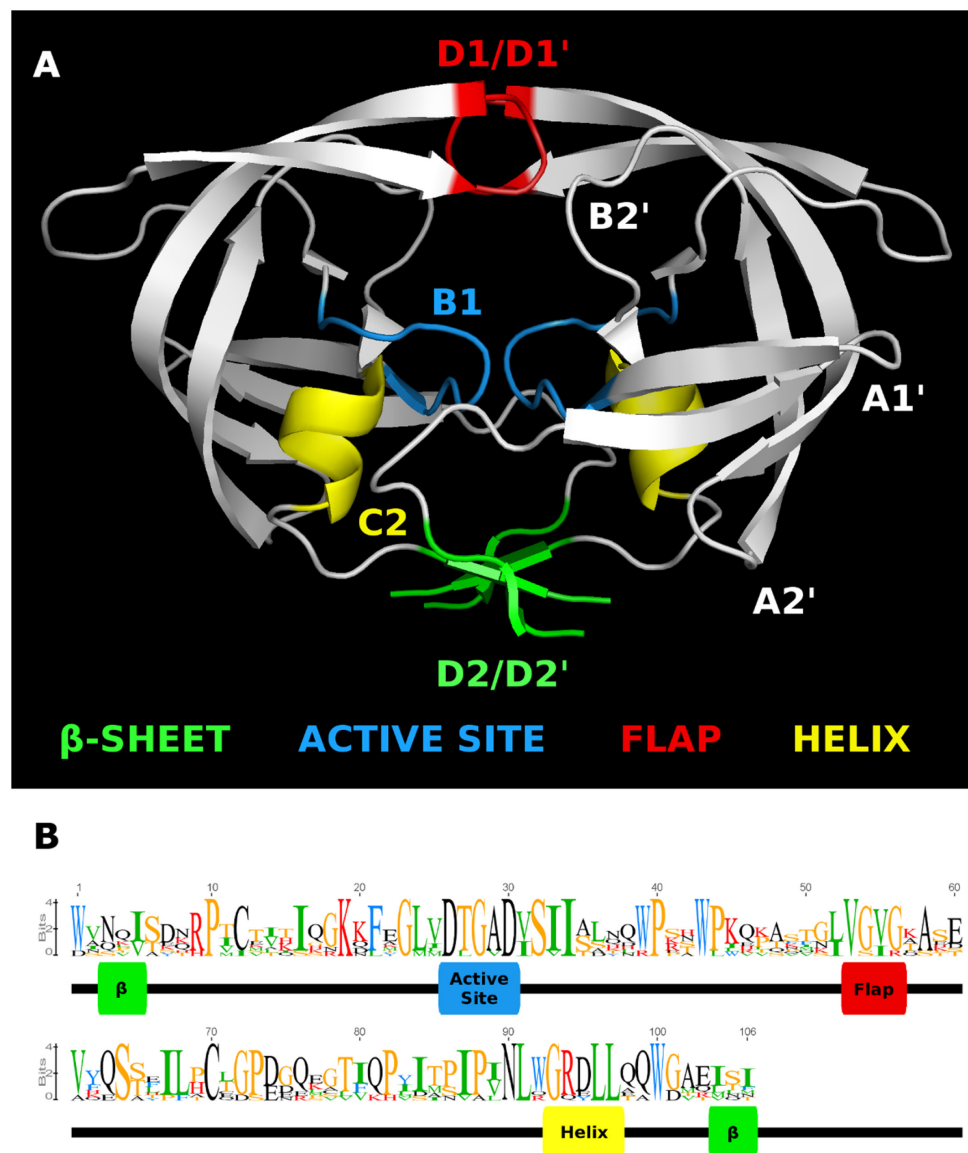


FIGURE 3 | HMM logo of representative mature endogenous retrovirus K proteases. **(A)** The structure of the HIV-1 protease 3A2O (Hidaka et al., 2009) is shown with the major regions referenced in the text highlighted. **(B)** HMM logo of the aligned mature protease sequences of each ERVK entry in Repbase (HK1 through 10) with referenced regions highlighted in the same colors as the homologous regions in 3A2O. This figure was made with Pymol, Geneious, and GIMP.

consequences if the protease is translated, especially if it could lead to differential drug susceptibility because the variable residue intrudes into the substrate binding site. Several columns of the alignment in **Figure 5** represent variable sites which clearly subdivide the clade in two.

The most notable variation within HK2 sequences (**Figure 5**) is the variable AA (L, V or I) at position 89 which is predicted to help form the S3 and S1' binding subsites by homology (Menendez-Arias et al., 1994). Columns 55, 65, and 66 are also notable. The AA (V or I) at position 55 follows the flap interface (columns 51–54) predicted by homology (Kuhelj et al., 2001). Residues 65 and 66 are near the surface of PR in the N-terminal strand of A2; from this position, their

variability could impact protein-protein interactions within the viral polyproteins or with host partners, and could potentially influence flap mobility during substrate binding via interactions with α -helix C1. Thus, the cellular complement of HK2 PR variants represent a continuum of enzymatic affinities and protein-protein interactions.

Variant Proteases Are Expressed in Human Health and Disease

We looked in publicly available RNA-Seq datasets (**Table 5**) from diseases with some known association with ERV expression to establish how genomic diversity is transcribed in healthy and

TABLE 4 | Frequency of co-occurrence for observed ERVK active site and associated helix motifs.

Number	B1	C2	Number	B1	C2	Number	B1	C2
Predicted active if translated								
52	DTGAD	GRDLL	17	DTGVD	GRDLL			
Potentially active								
4	DTRAD	GRDLL	1	DTWVD	GRDLL	1	DTGDD	GRDLL
4	DTEAD	GRDLL	1	DTVAD	GRDLL	1	DTGAN	GRDLL
3	DTGSD	GRDLL	1	DTGVD	GRHLL	1	DTGAN	GKDLL
3	DTGED	GRDLL	1	DTGVD	GKDLL	1	DTGAD	GRELL
3	DTGAD	GQDLL	1	DTGTD	GRDLL	1	DTGAD	GRDIL
3	DTGAD	GKDLL	1	DTGMD	GRDLL	1	DTGAD	GQELL
3	DTDAD	GRDLL	1	DTGID	GRDLL	1	DTAAD	GRDLL
2	DTEVD	GRDLL	1	DTGGD	GRDLL			
Predicted inactive if translated								
4	DTGAD	RRDLL	1	DTRSD	GRDLL	1	DTGAD	GHLL
3	DTGAD	ERDLL	1	DTRMD	GKEYY	1	DTEAD	GQDLL
2	HTGAD	GRDLL	1	DTRAD	GWDPL	1	DRGMD	GRDLL
2	DTWVD	GGTLL	1	DTGVD	LSPHTFI	1	DMGAD	DQDLL
2	WAWAV	GWDLL	1	DTGVA	GRDLL	1	DMGAD	
1	VTGVD	GRDLL	1	DTGPD		1	DIGVD	GRDLL
1	ITWGR	GVDN	1	DTGVD	RRDLL	1	DIGAD	GGDLL
1		GWDLL	1	DTGAD	VWDL	1	DIGAD	ERDLL
1	ETGVD	GWDLL	1	DTGAD	GVDLL	1	DAGAD	GRDLL
1	DTVAD	GGDLL	1	DTGAD	GTRPI	1	ATGAD	GRDLL
1	DTRTD	GRDVL	1	DTGAD	GRYLL	1	ALGAD	RRDLL

disease states (Seifarth et al., 2005; Douville et al., 2011; Ren et al., 2012; Agoni et al., 2013; Wildschutte et al., 2014; Abba et al., 2015; Li et al., 2015; Brohawn et al., 2016). RNA-Seq results were narrowed to examine the expression of the two most highly expressed and well-studied groups HML2 and HML3. Further, ERVK loci (detectable by both BLAST and LTRharvest as previously described) were limited to proteases without gross inactivating mutations. Reanalysis of transcriptomics data from three independent studies focused on ALS, breast cancer and prostate cancer are shown in **Figure 6**; expression of other HML groups are in **Additional File 42**. Overall, the expression of loci encoding DTGAD and DTGVD motifs is significantly different, due to low expression of DTGVD loci, and each of these is divergent from the pooled expression of other less common motifs.

Expression of protease encoding transcripts alone does not prove that these will be properly translated or proteolytically processed (Bauerova-Zabranska et al., 2005; Tien et al., 2018). Additional supporting evidence comes from examination of the upstream *gag* gene by searching reads aligning to the same locus using tBLASTn for the Gag sequence of ERV-K113. Several, but not all, loci expressed *gag* as well as *pro*. Individual experiments are required to confirm Gag-pro-pol polyprotein processing for each transcribed locus.

There is a trend toward increased ERVK transcription in cervical spinal tissues of ALS patients versus neuro-normal controls (**Figure 6**); indeed, this enhancement is evident when patients are stratified by sex, with female cases having an increased burden of ERVK (**Additional File 43**, $p < 0.01$).

When tissues from cancer patients were examined, breast cancer biopsies revealed increased PR transcripts with DTGAD ($p < 0.05$) and alternate PR active site motifs ($p < 0.05$), as compared to cosmetic breast resection controls. Overall, HML-2 expression was enhanced in breast cancer ($p < 0.01$). This ERVK load difference in breast cancer appears to be driven by the increased expression of three loci in 8p23.1 flanked by LTR5A, and encoding a DTGSD active site motif. Interestingly, an HIV-1 PR variant which encodes a DTGSD motif is known to be active, although less so than the wild-type enzyme (Hong et al., 1998). Similarly, HML-2 transcripts are increased in prostate cancer specimens as compared to autologous non-cancerous adjacent tissue ($p < 0.01$). These findings point toward a mixed profile of PR variants in disease states with elevated ERVK expression.

DISCUSSION

The human genome is home to a diversity of ERVs; this necessitates much effort in discriminating those of biological and/or clinical importance. Our identification of transcribed ERVK loci with a potential to encode functional PR enzymes will inform future studies examining their biological impact and drug susceptibility.

There remain technical challenges in assessing ERVs, as the results from LTRharvest depict that this paradigm is not ideal for analysis of more complicated loci containing recombination, serial insertions, or substantial deletions. Theoretically, each ERV should encode exactly 1 *pro* and 1 *pol* at the moment of insertion,

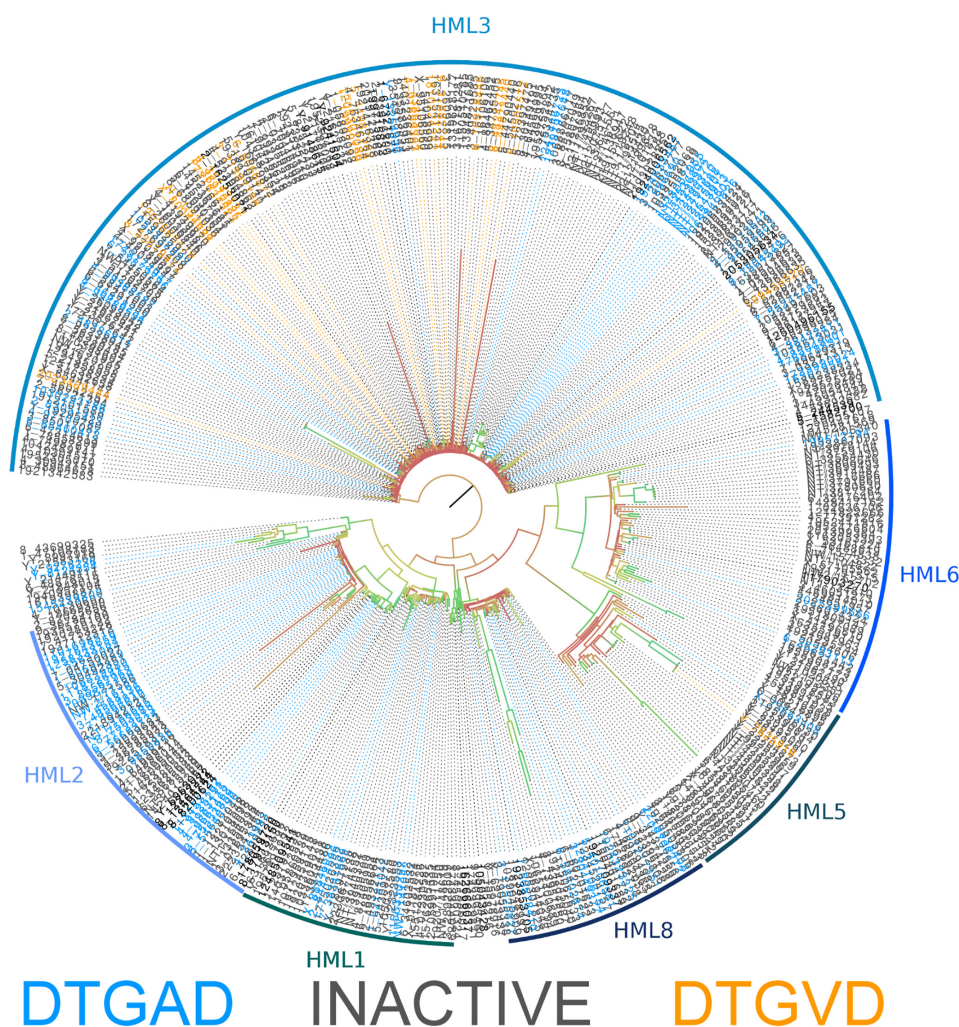


FIGURE 4 | Phylogeny of human genomic nucleotide sequences derived using BLAST and the ERVK HK2 PR. Potentially active sequences with a DTGAD motif in their active site are colored blue, and those with DTGVD are colored orange. Sequences containing inactivating mutations are colored gray. The classification of each area of the tree is indicated by the name associated with the arc around the perimeter. Branch coloring represents bootstrap support, where red indicates poor support (closer to zero) and green indicates good support (closer to 100). Placement of the root at the branch connecting HK3 is only a matter of visual convenience – the real root branch is more likely that leading to HK8. This alignment is available in **Additional File 41**.

but over time deletion and insertion events can remove or obscure these signatures. Despite this issue, the use of HMMs to identify ERVs within the human genome was effective. Indeed, RVP and RVT_1 annotations are more commonly identified within elements annotated by LTRharvest, since BLAST did not detect loci distantly related to the query HK2 sequences. The search results were improved by the merging of results from the same paralog and alignment-based extension. These sequences were grouped both by direct phylogenetic inference and by comparison to a reference tree. The phylogenetic patterns apparent in **Figures 1, 2** broadly agree with previously recognized relationships between ERVs (Vargiu et al., 2016). Although trees from *pro* and *pol* differ in some respects, they generally agree at deeper nodes which are important for classification, despite the lower phylogenetic signal of protease due to its comparatively shorter length.

There is value in examining both reference genomes and alternate assemblies. The reference genome was constructed from few people (Osoegawa et al., 1998; Lander et al., 2001), suggesting that the ERV content of individual human genomes could vary considerably and deserves closer scrutiny; especially since unfixed loci are likely to be younger (Belshaw et al., 2005), and could encode functional enzymes. Future use of data from platforms such as the 1000 Genomes Project (2018) will provide an improved appreciation of the diversity of ERV content in human DNA.

ERVK Protease Sequence Diversity

The diversity of human genomic retroviral proteases reflects their evolutionary history and conserved structure-function relationships. Residues flanking the active site and those maintaining the dimer interface are less variable, a pattern clearly

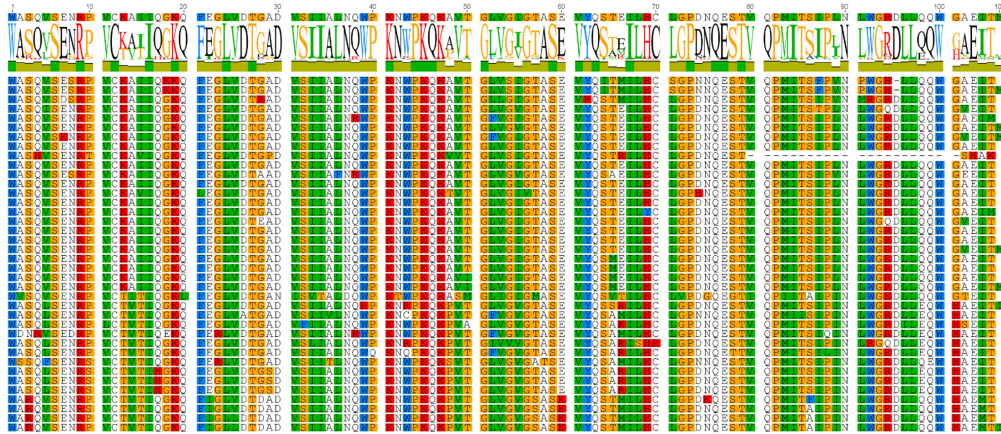


FIGURE 5 | Aligned protein sequences of human genomic HML2 proteases. The subset of the full protein alignment of BLAST results for the ERVK HK2 PR in the human genome which were classified as HK2 by the evolutionary placement algorithm are shown here. Two clear patterns emerge, suggesting distinct protease sequence groups. Figure produced using Geneious.

TABLE 5 | RNA-Seq Libraries from the Sequence Read Archive analyzed in this study.

Accession	Condition	Tissue	Library	Sequencing platform	Samples	Reference
SRP064478	sALS	Cervical spinal cord	Truseq Stranded Total RNA HT Kit (Illumina)	Nextseq 500 from ~260 bp fragments	7 sALS, 8 non-ALS, without technical replication	Brohawn et al., 2016
ERP000550	Prostate cancer	Prostate	Oligo dT purification, fragmentation, and random hexamer PCR	HiSeq 2000 2 × 90 bp from ~200 bp fragments	14 tumor/non-tumor from same patient, without technical replication	Ren et al., 2012
SRP058722	High grade ductal carcinoma <i>in situ</i>	Breast	ScriptSeq v2 RNA-Seq Kit (Epicenter)	HiSeq 2000 2 × 76 bp, fragment length unreported	25 ductal carcinoma <i>in situ</i> , 10 cultured primary normal breast tissue, without technical replication	Abba et al., 2015

discernible in the Pfam family RVP (Finn et al., 2014) and evident within **Figure 1**. The β -sheet at the “bottom” of the enzyme is an exception to this rule, as its residues are not as conserved, perhaps because this region of tertiary structure is driven by backbone interactions (Louis et al., 2007). Overall, the most notable variation in genomic protease sequences is the absence of RVP positions 38–42 in 83 PRs, none of which were classified as ERVK.

The sequence of the retroviral protease active site (B1) and the active-site associate helix (C2) are conserved, as seen in **Figures 1, 3, 5** for ERVK. The most common active site motif of ERVK, DTGAD, is identical to that predominantly observed in exogenous retroviruses. The second most common motif, DTGVD (A29V), is observed in functional retroviral proteases; in fact, there are many examples of HIV proteases bearing this variation (e.g., Uniprot: P15833). Given the wide phylogenetic distribution of A29V and the similar chemical properties of alanine and valine residues, this is probably a functional variant of the ERVK protease. This motif could therefore represent pre-integration variation of the ancestral XRV which gave rise to ERVK, which is further supported by the concentration of DTGVD motif containing sequences in HK3, seen in **Figure 4**. Alternatively, the motif could represent a post-integration mutation predating the amplification of HK3 (Mayer and Meese, 2002). We also predict several α -helix (C2)

variants could be functional in ERVK. In the HIV-1 PR (e.g., PDB: 1HVC), the C2 arginine residue (R94) fills the space between B1 and the loop connecting A1' to D2' while forming hydrogen bonds to residues in both. Another large polar residue might fill this role; indeed, lysine is known to do so in HIV-1 protease (Louis et al., 2007). Several genomic motifs fit these requirements, such as the HK2 R94Q and R94K variants, and the HK8 D95E variant. Systematic experimental verification of the activity of individual PR variants is thus warranted.

Nonetheless, we predict that most of the less common active site loop variants are inactive (**Figure 4**). This prediction is pulled from our understanding of the composition of this motif. Many variations are of a residue which catalyzes the hydrolytic reaction, or which is predicted to be essential to the dimer interface. The catalytic residues must be aspartic acid. The threonine residue which helps form the dimer interface is uncommonly replaced by serine (e.g., PDB: 2JYS, 2RSP) (Jaskolski et al., 1990; Hartl et al., 2008). Others could disrupt the active site by substituting a side-chain of very different size. The glycine residue must be small, as distortions to the structure of this region could disrupt the catalytic structure. This is also true for the alanine residue, although it is known to be replaced by serine in HIV-1 without a total loss of activity (Ido et al., 1991). The final aspartic acid residue participates in hydrogen bonds with the nearby arginine residues (Louis et al., 2007), and might be replaced by another

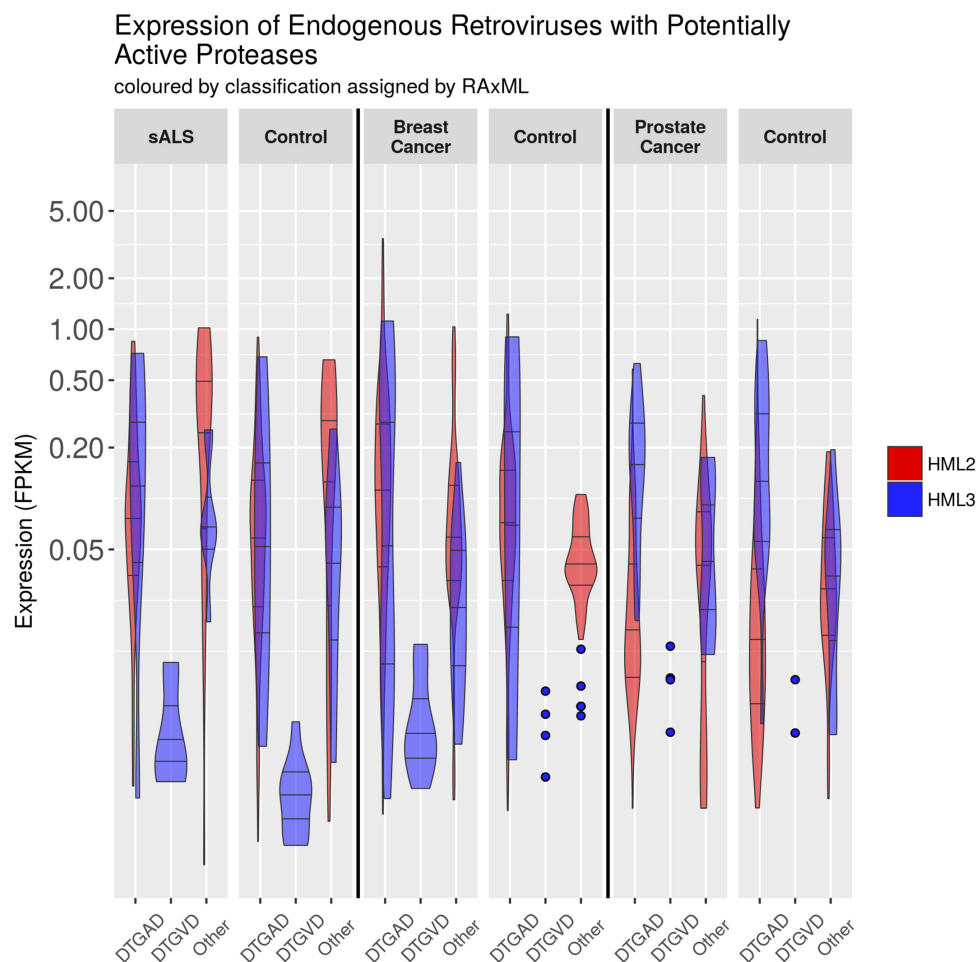


FIGURE 6 | Expression of HML2 and HML3 by RNA-Seq. The expression of ERV loci in GRCh38 containing an uninterrupted protease was measured with bowtie2 using RNA-Seq data from different conditions (SRA accession numbers for each condition are: ALS – SRP064478, Breast Cancer – SRP058722, Prostate Cancer – ERP000550). Normalization to fragments per kilobase of exon per million mapped reads (FPKM) is relative to the entire locus. This graphic was made using R with the tidyverse library, and GIMP. The N's of each column are, from left to right, HML2 and then HML3: 104, 8, 0, 7, 29, 7, 118, 9, 0, 10, 35, 8; 369, 27, 0, 17, 108, 25, 146, 11, 0, 4, 40, 5; 195, 14, 0, 4, 64, 13, 177, 14, 0, 2, 52, 13. Counts are also provided in CSV and Excel format in **Additional Files 44, 45**.

small charged or polar residue; however, the character of this residue may not be critical, since it can vary with glycine in HIV-1 protease (Louis et al., 2007).

The young, polymorphic HK2 element ERVK113 is well studied and was reported to produce immature virions, perhaps due to a non-functional protease (Boller et al., 2008). This is in contrast to the known active HK2 protease of *ERVK-10* (Ono, 1986; Towler et al., 1998). This is further supported by reported differences in the sequences of these two mature proteases; the G56S (reported as G234S) substitution in ERVK113 may disrupt the flap interface, compromising PR activity and conferring an immature particle morphology (Boller et al., 2008).

Implications of ERVK PR Diversity on Immunity

Retroviral proteases have numerous and varied protein targets which are difficult to predict, with correspondingly broad cellular

effects stemming from their expression. Many proteases include targets with immune functions, but these vary based on host-virus pairings. It remains to be seen how ERVK PRs may modulate innate immunity signaling in humans, as seen with XRVs (Abudu et al., 2006; Solis et al., 2011; Wagner et al., 2015; Yoshikawa et al., 2017). Based on our knowledge of HIV (Konvalinka et al., 1995; Snasel et al., 2000; Lin et al., 2003), it is even conceivable that each ERVK PR variant may target distinct panels of cellular proteins, due to their different substrate specificities.

The potential impact of ERVK PR not limited to the innate immune system. Observations from the recovery of immune function in patients on HAART containing or not containing PIs suggests that HIV PR can directly impact the functioning of adaptive immune cells independently of other viral proteins (Ananworanich et al., 2003; Chiodi, 2006). Furthermore, differential ERVK PR substrate specificities could lead to changes in the peptide fragment population from which MHC class

I epitopes are drawn, in addition to the impact that PI use alters this system (Kourjian et al., 2014; Kourjian et al., 2016). Our observation of distinct ERVK PR variants expressed in disease highlights the need for future studies to consider their overlapping and distinct impacts on cell signaling and proteome profiles related to immunity.

Implications for Drug-Based Treatment of ERVK-Associated Disease

Many protease coding sequences could be translated into an active enzyme, and their sequence variability may impact their activity, and likely their susceptibility to antiretroviral drugs. Variation between DTGAD and DTGVD in the active site of sequences classified as HK3 is notable. This variation could have biological relevance when both HK2 and HK3 elements are expressed, such as in ALS (Douville et al., 2011) and schizophrenia (Frank et al., 2005). The HK2 PR is susceptible to some clinically relevant HIV-1 PIs, although not to the same degree as the HIV-1 PR (Towler et al., 1998; Kuhelj et al., 2001; Tyagi et al., 2017). Since ERVK expression is associated with many disease states, and the activity of retroviral enzymes can be pathogenic, it is possible that PIs may be useful in the future treatment of ERVK-associated conditions. In fact, at least one ongoing clinical trial (NCT02437110) plans to administer the PI Darunavir, along with other drugs, to patients with ALS. Another clinical trial (NCT01528865) targeting ERVK in lymphoma was withdrawn – but the intent to explore the clinical utility of ERVK PIs is becoming clear.

The rapid development of drug resistance is a major hurdle in the treatment of HIV-1 infections, but since ERVK is fixed in the host genome, it cannot rapidly evolve such drug resistance. However, ERVK elements are diverse and likely respond differently to PIs. This diversity is not represented in the ERVK literature, which is focused on the consensus HK2 PR (Towler et al., 1998; Kuhelj et al., 2001; Dewannieux et al., 2006; Lee and Bieniasz, 2007; Tyagi et al., 2017). One recent paper states that the active site loop of the HIV-1 and ERVK PR are identical (Tyagi et al., 2017), a finding which this report directly contradicts. We predict this diversity may necessitate the simultaneous use of multiple PIs in the treatment of ERVK-associated diseases. Consequently, if the drug regimen in the above-mentioned clinical trial is not effective in treating ALS, this cannot be taken as evidence that such therapy would not be effective using multiple PIs in a combination therapy.

We have established that a multitude of ERVK protease variants exist in the human genome, and that some of these variants can reasonably be predicted to have differing drug binding profiles. However, many ERVs are transcriptionally repressed (Bogerd et al., 2006; Schlesinger and Goff, 2013; Schlesinger et al., 2014); our analysis of publicly available RNA-Seq data shows not only that ERVs are expressed, but that the expression of loci with differing biochemical and evolutionary characteristics varies between and within different disease conditions, as well as in healthy controls. Due to the relatively recent insertion of ERVK sequences within the human genome, and the pathogenicity of XRV PRs, it is warranted

that future studies consider their role in modulating immunity. Since distinct biochemical and ancestral features may impact the susceptibility of these enzymes to PIs, we propose that patient-specific drug regimens may be required in treatment of ERVK associated disease. Moreover, given the potential importance of protease sequence variability, the sequences of other ERVK proteins (particularly the drug targets reverse transcriptase and integrase) should also be explored.

CONCLUSION

Within the human genome, ERVK proteases exhibit a high degree of diversity. Specifically, two predominant PR active sites emerge, the DTGAD and DTGVD variants, which are differentially expressed in disease states. This study will also be an asset for inferring how ERVK PRs impact the human proteome, specifically as it pertains to immune function. It is possible to interpret this paper as casting a negative light on the prospect of inhibitor-based treatment of ERVK-associated disease, but this is not our intention. We do, however, caution against treatments and techniques that treat the numerous and diverse elements called ERVK as though they were a single invariable element. This variability is not endless; biomedical and clinical techniques to account for this enzymatic variation exist. We recommend that the results of inhibition assays against the specific ERVK proteases expressed in each disease state should be considered if anti-PR drug regimens are to be implemented for inflammatory disease.

MATERIALS AND METHODS

Sources for databases and software employed in the course of this study are listed in **Table 6**.

Sequences homologous to retroviral protease (*pro*) and reverse transcriptase (*pol*) were automatically extracted from the last major build of the reference human genome and classified with published sequences from Repbase. Structure-function analysis was applied to ERV PR sequences using the limited published data on them, and by transferring knowledge of better studied proteases.

Reference Sequences

The 38th official assembly of the human genome (GRCh38/hg38 published December 2013) was obtained from NCBI. Representatives for ERVADP, ERV9, ERV3, ERVE, ERVFRD, ERVE, ERVFXA, ERVFC, ERVFB, ERVH, ERVI, ERVL, ERVRB, ERVP, ERVS, ERVT, ERV16, ERVW, and HML groups 1 through 10 (HK1 through HK10) were derived from Repbase v19.07 (Jurka et al., 2005). HK9 and HK10 are also called K14C and KC4, respectively. The sequences of ERVK *Phoenix* (Dewannieux et al., 2006), and *ERVK-10* (Kuhelj et al., 2001) were taken from their respective publications. These sequences can be found in **Additional Files 46, 47**, respectively. Genes were located using BLAST v2.2.28 (Altschul et al., 1990) and HMMER v3.1b1 (Eddy, 1998). Cytological bands were assigned by a perl script using co-ordinates from NCBI.

TABLE 6 | Databases and software used in this study.

Name	URL	Version or accession date
Bowtie2	https://www.github.com/BenLangmead/bowtie2	2.3.3
BLAST	ftp.ncbi.nlm.nih.gov/blast/executables/blast+/	2.2.28
Cytological bands	ftp.ncbi.nlm.nih.gov/genomes/MapView/Homo_sapiens/objects/	2015-11-20
FASTQC	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/	0.11.5
Figtree	http://tree.bio.ed.ac.uk/software/figtree/	1.4.1
Geneious	https://www.geneious.com/	
GenomeTools	http://genometools.org/pub/	1.5.1
GIMP	http://www.gimp.org/downloads	2
GRCh38	ftp://ftp.ncbi.nlm.nih.gov/genomes/archive/old_genbank/Eukaryotes/vertebrates_mammals/Homo_sapiens/GRCh38/	2014-08-27
HMMER	http://hmmerr.janelia.org	3.1b1
Htslib	https://github.com/samtools	1.5-22-g7a6854b
JAVA		1.8.0-144
jModelTest	(D Darriba) https://github.com/ddarriba/jmodeltest2	2
MACSE	(V Ranwez) http://www.mybiosoftware.com/macse-1-01b-multiple-alignment-of-coding-sequences.html	1.01b
Pfam-A	ftp.ebi.ac.uk/pub/databases/Pfam/current_release/Pfam-A.hmm.gz	2014-09-05
ProtTest	(D Darriba) https://www.github.com/ddarriba/prottest3	3
RAxML	https://github.com/stamatak/standard-RAxML	7.2.8
Repbase	(J Jurka) http://www.girinst.org/repbase/	19.07
RSYNC	https://rsync.samba.org/	v3.0.6 protocol v30
Samtools	https://github.com/samtools	1.5-9-g473d6a4
Sratoolkit	https://github.com/ncbi/sra-tools	2.8.2-1
Trimmomatic	https://github.com/timflutre/trimmomatic	0.36

BLAST

Each GRCh38 chromosome and representative retroviral sequence was searched by tBLASTn for the amino acid sequences of the mature *ERVK-10* PR and the RVT_1 domain of *Phoenix pol* identified by HMMER. BLAST hits from same gene in different frames were merged, aligned by MACSE v1.01b (Ranwez et al., 2011), and then sequences 5' and/or 3' of each hit which would make the alignment flush were retrieved. Proteases identified in the reference genome assembly were matched to alternative assemblies using a perl script.

HMMER and GenomeTools

HMMs for which to search were identified in Pfam-A using the search term “retrovirus” (Finn et al., 2014). ERV reference sequences were searched by HMMER with an *e*-value of 1.0. Putative LTR retroelements were identified in GRCh38 by

LTRharvest, and these were subsequently analyzed by LTRdigest, which ran HMMER. LTRharvest and LTRdigest are part of Genometools v1.5.1 (Gremme et al., 2013). Hits from the same locus in different reading frames were merged and aligned with MACSE, then the regions which would make each sequence flush to the alignment were retrieved.

Phylogenetic Inference and Classification

Results from each of the four searches were curated by eliminating HMM matches not associated with another core retroviral domain and by removing BLAST results which induced gaps in $\geq 95\%$ of alignment rows with MACSE. Sequences were then re-aligned with MACSE and models of evolution were selected using jModelTest (Darriba et al., 2012) and ProtTest (Guindon et al., 2010; Darriba et al., 2011). Maximum likelihood phylogenies were inferred using RAxML v7.2.8 (Stamatakis, 2014) for nucleotide and protein alignments of each search (8 trees in total). GRCh38 results were classified using RAxML's evolutionary placement algorithm with reference to the inferred phylogeny of representative ERV sequences.

Misclassified or recombinant elements were sought by comparing the assignment of search results separated by less than 10 kbp. This locus-based definition is computationally simple, but potentially error prone; a more accurate element-based analysis would require algorithmic reconstruction of insertion architecture, as was undertaken in the construction of Dfam (Wheeler et al., 2013).

Measuring Expression in Publicly Available RNA-Seq Data

Publicly available RNA-Seq data from tissues associated with ERVK expression in human disease were obtained from the SRA using sratools. The resulting FASTQ files were aligned to the human genome (GRCh38) using bowtie2 following examination with FASTQC. The resulting SAM file was then indexed using samtools and finally the resulting indexed BAM file was queried for expression of ERV loci identified using the methods above. The raw expression values thus obtained were normalized by library size and locus length to produce FPKM values.

The resulting FPKM values were examined on density and QQ plots to confirm that they did not conform to a normal distribution. Non-parametric tests of statistical significance were used in this paper. For unpaired expression data (breast cancer, ALS) the Mann–Whitney *U* test was used, and for paired data (prostate cancer) the Wilcoxon Signed-Rank Test was used. In all cases tests were two-sided and *p*-values were corrected using the Bonferroni procedure, considering each set of contrasts involving the same variables to be a family.

AVAILABILITY OF DATA AND MATERIAL

The datasets used to generate the findings of this paper are available from NCBI by their accession numbers. SRA accession numbers for each RNA-Seq data are: ALS – SRP064478, Breast Cancer – SRP058722, Prostate Cancer – ERP000550. The human

genome version 38 was obtained from the NCBI FTP site [79]. Datasets generated during the current study which are not included as additional files are available from the corresponding author upon request. Important scripts used in the course of this study are included in **Additional File 48**.

AUTHOR CONTRIBUTIONS

MT and RD conceived the experiments. MT performed the bioinformatics analysis. MT and RD wrote the manuscript.

FUNDING

Funding support was provided by the University of Winnipeg and a Discovery Grant from Natural Sciences and Engineering Research Council of Canada (NSERC #RGPIN-2016-05761).

ACKNOWLEDGMENTS

We would like to thank the Douville lab team for their input on this project. This work was supported by WestGrid (www.westgrid.ca) and Compute Canada (www.computeCanada.ca) by use of the computer cluster Orcinus. We acknowledge that the University of Winnipeg is in Treaty 1 territory and in the heart of the Métis nation. Orcinus is located on the traditional, ancestral, and unceded territory of the Musqueam people. MT currently resides in the traditional territory of the Three Fire Peoples: the Ojibwa, Odawa, and Potawatomi.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.01577/full#supplementary-material>

FILE 1 in FASTA (fa) format | Aligned translated genomic protease BLAST results. The sequences found by BLAST for the HK2 protease in the human genome aligned and translated by MACSE.

FILE 2 in FASTA (fa) format | Aligned genomic protease BLAST results. The sequences found by BLAST for the HK2 protease in the human genome aligned by MACSE.

FILE 3 in FASTA (fa) format | Aligned translated genomic RVP HMMER results. The sequences found by HMMER directed by LTRdigest for the protein family HMM RVP in the human genome aligned and translated by MACSE.

FILE 4 in FASTA (fa) format | Aligned genomic RVP HMMER results. The sequences found by HMMER directed by LTRdigest for the protein family HMM RVP in the human genome aligned by MACSE.

FILE 5 in FASTA (fa) format | Aligned translated genomic reverse transcriptase BLAST results. The sequences found by BLAST for the HK2 reverse transcriptase in the human genome aligned and translated by MACSE.

FILE 6 in FASTA (fa) format | Aligned genomic reverse transcriptase BLAST results. The sequences found by BLAST for the HK2 reverse transcriptase in the human genome aligned by MACSE.

FILE 7 in FASTA (fa) format | Aligned translated genomic RVT_1 HMMER results. The sequences found by HMMER directed by LTRdigest for the protein family HMM RVT_1 in the human genome aligned and translated by MACSE.

FILE 8 in FASTA (fa) format | Aligned genomic RVT_1 HMMER results. The sequences found by HMMER directed by LTRdigest for the protein family HMM RVT_1 in the human genome aligned by MACSE.

FILE 9 in Newick (nwk) format | Phylogeny inferred from genomic BLAST results for the HK2 protease amino acid sequences. The phylogeny of amino acid sequences identified in the human genome GRCh38 by BLAST using the mature HK2 protease as a query as inferred by RAxML under the JTT + G model of evolution and aligned by MACSE.

FILE 10 in Newick (nwk) format | Phylogeny inferred from genomic BLAST results for the HK2 protease nucleic acid sequences. The phylogeny of nucleic acid sequences identified in the human genome GRCh38 by BLAST using the mature HK2 protease as a query as inferred by RAxML under the GTR model of evolution and aligned by MACSE.

FILE 11 in Newick (nwk) format | Phylogeny inferred from genomic BLAST results for the HK2 reverse transcriptase amino acid sequences. The phylogeny of amino acid sequences identified in the human genome GRCh38 by BLAST using the HK2 reverse transcriptase as a query as inferred by RAxML under the JTT + G model of evolution and aligned by MACSE.

FILE 12 in Newick (nwk) format | Phylogeny inferred from genomic BLAST results for the HK2 reverse transcriptase nucleic acid sequences. The phylogeny of nucleic acid sequences identified in the human genome GRCh38 by BLAST using the HK2 reverse transcriptase as a query as inferred by RAxML under the GTR model of evolution and aligned by MACSE.

FILE 13 in Newick (nwk) format | Phylogeny inferred from genomic HMMER results for RVP amino acid sequences. The phylogeny of amino acid sequences identified in the human genome GRCh38 by HMMER directed by LTRdigest using the protein family HMM RVP as a query as inferred by RAxML under the JTT + G model of evolution and aligned by MACSE.

FILE 14 in Newick (nwk) format | Phylogeny inferred from genomic HMMER results for RVP nucleic acid sequences. The phylogeny of nucleic acid sequences identified in the human genome GRCh38 by HMMER directed by LTRdigest using the protein family HMM RVP as a query as inferred by RAxML under the GTR + G model of evolution and aligned by MACSE.

FILE 15 in Newick (nwk) format | Phylogeny inferred from genomic HMMER results for RVT_1 amino acid sequences. The phylogeny of amino acid sequences identified in the human genome GRCh38 by HMMER directed by LTRdigest using the protein family HMM RVT_1 as a query as inferred by RAxML under the JTT + G model of evolution and aligned by MACSE.

FILE 16 in Newick (nwk) format | Phylogeny inferred from genomic HMMER results for RVT_1 nucleic acid sequences. The phylogeny of nucleic acid sequences identified in the human genome GRCh38 by HMMER directed by LTRdigest using the protein family HMM RVT_1 as a query as inferred by RAxML under the GTR + G model of evolution and aligned by MACSE.

FILE 17 formatted as described in the RAxML user manual (txt) | Classification likelihood weights for evolutionary placement of genomic protease BLAST amino acid sequences onto the corresponding reference tree. The placement of genomic sequences onto the tree of Repbase consensus was done using RAxML's evolutionary placement algorithm under the JTT + G model of evolution, and this file contains the likelihood of each potential placement for each sequence.

FILE 18 formatted as described in the RAxML user manual (txt) | Classification likelihood weights for evolutionary placement of genomic protease BLAST nucleic acid sequences onto the corresponding reference tree. The placement of genomic sequences onto the tree of Repbase consensus was done using RAxML's evolutionary placement algorithm under the GTR + G model of evolution, and this file contains the likelihood of each potential placement for each sequence.

FILE 19 formatted as described in the RAxML user manual (txt) |

Classification likelihood weights for evolutionary placement of genomic reverse transcriptase BLAST amino acid sequences onto the corresponding reference tree. The placement of genomic sequences onto the tree of Repbase consensa was done using RAxML's evolutionary placement algorithm under the RtREV + G model of evolution, and this file contains the likelihood of each potential placement for each sequence.

FILE 20 formatted as described in the RAxML user manual (txt) |

Classification likelihood weights for evolutionary placement of genomic reverse transcriptase BLAST nucleic acid sequences onto the corresponding reference tree. The placement of genomic sequences onto the tree of Repbase consensa was done using RAxML's evolutionary placement algorithm under the GTR + G model.

FILE 21 formatted as described in the RAxML user manual (txt) |

Classification likelihood weights for evolutionary placement of genomic RVP HMMER amino acid sequences onto the corresponding reference tree. The placement of genomic sequences onto the tree of Repbase consensa was done using RAxML's evolutionary placement algorithm under the WAG + G + F model of evolution, and this file contains the likelihood of each potential placement for each sequence.

FILE 22 formatted as described in the RAxML user manual (txt) |

Classification likelihood weights for evolutionary placement of genomic RVP HMMER nucleic acid sequences onto the corresponding reference tree. The placement of genomic sequences onto the tree of Repbase consensa was done using RAxML's evolutionary placement algorithm under the GTR + G model of evolution, and this file contains the likelihood of each potential placement for each sequence.

FILE 23 formatted as described in the RAxML user manual (txt) |

Classification likelihood weights for evolutionary placement of genomic RVT_1 HMMER amino acid sequences onto the corresponding reference tree. The placement of genomic sequences onto the tree of Repbase consensa was done using RAxML's evolutionary placement algorithm under the RtREV + G model of evolution, and this file contains the likelihood of each potential placement for each sequence.

FILE 24 formatted as described in the RAxML user manual (txt) |

Classification likelihood weights for evolutionary placement of genomic RVT_1 HMMER nucleic acid sequences onto the corresponding reference tree. The placement of genomic sequences onto the tree of Repbase consensa was done using RAxML's evolutionary placement algorithm under the GTR + G model of evolution, and this file contains the likelihood of each potential placement for each sequence.

FILE 25 in Newick (nwk) format | Original labeled reference tree for evolutionary placement of genomic protease BLAST amino acid sequences. The reference tree for placement of genomic protease BLAST results labeled for use with the corresponding classification likelihood file.

FILE 26 in Newick (nwk) format | Original labeled reference tree for evolutionary placement of genomic protease BLAST nucleic acid sequences. The reference tree for placement of genomic protease BLAST results labeled for use with the corresponding classification likelihood file.

FILE 27 in Newick (nwk) format | Original labeled reference tree for evolutionary placement of genomic reverse transcriptase BLAST amino acid sequences. The reference tree for placement of genomic reverse transcriptase BLAST results labeled for use with the corresponding classification likelihood file.

FILE 28 in Newick (nwk) format | Original labeled reference tree for evolutionary placement of genomic reverse transcriptase BLAST nucleic acid sequences. The reference tree for placement of genomic reverse transcriptase BLAST results labeled for use with the corresponding classification likelihood file.

FILE 29 in Newick (nwk) format | Original labeled reference tree for evolutionary placement of genomic RVP HMMER amino acid sequences. The reference tree for placement of genomic RVP HMMER directed by LTRdigest results labeled for use with the corresponding classification likelihood file.

FILE 30 in Newick (nwk) format | Original labeled reference tree for evolutionary placement of genomic RVP HMMER nucleic acid sequences. The reference tree for placement of genomic RVP HMMER directed by LTRdigest results labeled for use with the corresponding classification likelihood file.

FILE 31 in Newick (nwk) format | Original labeled reference tree for evolutionary placement of genomic RVT_1 HMMER amino acid sequences. The reference tree for placement of genomic RVT_1 HMMER directed by LTRdigest results labeled for use with the corresponding classification likelihood file.

FILE 32 in Newick (nwk) format | Original labeled reference tree for evolutionary placement of genomic RVT_1 HMMER nucleic acid sequences. The reference tree for placement of genomic RVT_1 HMMER directed by LTRdigest results labeled for use with the corresponding classification likelihood file.

FILE 33 in Newick (nwk) format | Phylogeny inferred from amino acid sequences from representative Repbase consensa searched using BLAST with the HK2 protease as a query. The phylogeny of amino acid sequences identified in representative Repbase consensa by BLAST using the mature HK2 protease as a query as inferred by RAxML under the JTT + G model of evolution and aligned by MACSE.

FILE 34 in Newick (nwk) format | Phylogeny inferred from nucleic acid sequences from representative Repbase consensa searched using BLAST with the HK2 protease as a query. The phylogeny of nucleic acid sequences identified in representative Repbase consensa by BLAST using the mature HK2 protease as a query as inferred by RAxML under the GTR + G model of evolution and aligned by MACSE.

FILE 35 in Newick (nwk) format | Phylogeny inferred from amino acid sequences from representative Repbase consensa searched using BLAST with the HK2 reverse transcriptase as a query. The phylogeny of amino acid sequences identified in representative Repbase consensa by BLAST using the HK2 reverse transcriptase as a query as inferred by RAxML under the RtREV + G model of evolution and aligned by MACSE.

FILE 36 in Newick (nwk) format | Phylogeny inferred from nucleic acid sequences from representative Repbase consensa searched using BLAST with the HK2 reverse transcriptase as a query. The phylogeny of nucleic acid sequences identified in representative Repbase consensa by BLAST using the HK2 reverse transcriptase as a query as inferred by RAxML under the GTR + G model of evolution and aligned by MACSE.

FILE 37 in Newick (nwk) format | Phylogeny inferred from amino acid sequences from representative Repbase consensa searched using HMMER for RVP. The phylogeny of amino acid sequences identified in representative Repbase consensa by HMMER using the protein family HMM RVP as a query as inferred by RAxML under the WAG + G + F model of evolution and aligned by MACSE.

FILE 38 in Newick (nwk) format | Phylogeny inferred from nucleic acid sequences from representative Repbase consensa searched using HMMER for RVP. The phylogeny of nucleic acid sequences identified in representative Repbase consensa by HMMER using the protein family HMM RVP as a query as inferred by RAxML under the GTR + G model of evolution and aligned by MACSE.

FILE 39 in Newick (nwk) format | Phylogeny inferred from amino acid sequences from representative Repbase consensa searched using HMMER for RVT_1. The phylogeny of amino acid sequences identified in representative Repbase consensa by HMMER using the protein family HMM RVT_1 as a query as inferred by RAxML under the RtREV + G + I model of evolution and aligned by MACSE.

FILE 40 in Newick (nwk) format | Phylogeny inferred from nucleic acid sequences from representative Repbase consensa searched using HMMER for RVT_1. The phylogeny, inferred by RAxML under the GTR + G model of evolution, of nucleic acid sequences identified in representative Repbase consensa by HMMER using the protein family HMM RVT_1 as a query aligned by MACSE.

FILE 41 in FASTA (fa) format | Alignment used to generate **Figure 4**.

FILE 42 in PNG (png) format | Expression of non-HML2/HML3 ERVs in human tissues. The expression of ERVK loci in GRCh38 containing an uninterrupted

protease was measured with bowtie2 using RNA-Seq data from different conditions (SRA accession numbers for each condition are ALS – SRP064478, Breast Cancer – SRP058722, Prostate Cancer – ERP000550). Normalization to fragments per kilobase of exon per million mapped reads (FPKM) is relative to the entire locus. This graphic was made using R with the tidyverse library, and GIMP.

FILE 43 in PNG (png) format | Expression of ERVK in cervical spinal cord RNA-Seq of male and female donors with ALS. The expression of ERVK loci in GRCh38 containing an uninterrupted protease was measured with bowtie2 using RNA-Seq data from different conditions (SRA accession numbers for each condition are ALS – SRP064478, Breast Cancer – SRP058722, Prostate Cancer – ERP000550). Normalization to fragments per kilobase of exon per million mapped reads (FPKM) is relative to the entire locus. Samples were sexed using expression of XIST. This graphic was made using R with the tidyverse library, and GIMP.

REFERENCES

- 1000 Genomes Project (2018). Available at: <http://www.internationalgenome.org/>
- Abba, M. C., Gong, T., Lu, Y., Lee, J., Zhong, Y., Lacunza, E., et al. (2015). A molecular portrait of high-grade Ductal carcinoma *in situ*. *Cancer Res.* 75, 3980–3990. doi: 10.1158/0008-5472.CAN-15-0506
- Abudu, A., Takaori-Kondo, A., Izumi, T., Shirakawa, K., Kobayashi, M., Sasada, A., et al. (2006). Murine retrovirus escapes from murine APOBEC3 via two distinct novel mechanisms. *Curr. Biol.* 16, 1565–1570. doi: 10.1016/j.cub.2006.06.055
- Agoni, L., Guha, C., and Lenz, J. (2013). Detection of human endogenous retrovirus K (HERV-K) transcripts in human prostate cancer cell lines. *Front. Oncol.* 3:180. doi: 10.3389/fonc.2013.00180
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Ananworanich, J., Nuesch, R., Teeratakulpisarn, S., Srasuebkul, P., Chuenyam, T., Siangphoe, U., et al. (2003). In vivo cell-mediated immunity in subjects with undetectable viral load on protease inhibitor-based versus non-protease inhibitor-based highly active antiretroviral therapy. *J. Acquir. Immune Defic. Syndr.* 32, 570–572. doi: 10.1097/00126334-200304150-00016
- Bannert, N., and Kurth, R. (2004). Retroelements and the human genome: new perspectives on an old relation. *Proc. Natl. Acad. Sci. U.S.A.* 101(Suppl. 2), 14572–14579. doi: 10.1073/pnas.0404838101
- Bannert, N., and Kurth, R. (2006). The evolutionary dynamics of human endogenous retroviral families. *Annu. Rev. Genomics Hum. Genet.* 7, 149–173. doi: 10.1146/annurev.genom.7.080505.115700
- Barbeau, B., and Mesnard, J. M. (2015). Does chronic infection in retroviruses have a sense? *Trends Microbiol.* 23, 367–375. doi: 10.1016/j.tim.2015.01.009
- Bauerova-Zabranska, H., Stokrova, J., Strisovsky, K., Hunter, E., Ruml, T., and Pichova, I. (2005). The RNA binding G-patch domain in retroviral protease is important for infectivity and D-type morphogenesis of Mason-Pfizer monkey virus. *J. Biol. Chem.* 280, 42106–42112. doi: 10.1074/jbc.M508031200
- Becker, J., Perot, P., Cheynet, V., Oriol, G., Mugnier, N., Mommert, M., et al. (2017). A comprehensive hybridization model allows whole HERV transcriptome profiling using high density microarray. *BMC Genomics* 18:286. doi: 10.1186/s12864-017-3669-7
- Belshaw, R., Dawson, A. L., Woolven-Allen, J., Redding, J., Burt, A., and Tristram, M. (2005). Genomewide screening reveals high levels of insertional polymorphism in the human endogenous retrovirus family HERV-K(HML2): implications for present-day activity. *J. Virol.* 79, 12507–12514. doi: 10.1128/JVI.79.19.12507-12514.2005
- Bogerd, H. P., Wiegand, H. L., Hulme, A. E., Garcia-Perez, J. L., O'Shea, K. S., Moran, J. V., et al. (2006). Cellular inhibitors of long interspersed element 1 and Alu retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8780–8785. doi: 10.1073/pnas.0603313103
- Boissinot, S., Chevet, P., and Furano, A. V. (2000). L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol. Biol. Evol.* 17, 915–928. doi: 10.1093/oxfordjournals.molbev.a026372
- Boller, K., Schonfeld, K., Lischer, S., Fischer, N., Hoffmann, A., Kurth, R., et al. (2008). Human endogenous retrovirus HERV-K113 is capable of producing intact viral particles. *J. Gen. Virol.* 89(Pt 2), 567–572. doi: 10.1099/vir.0.83534-0
- Bray, S., Turnbull, M., Hebert, S., and Douville, R. N. (2016). Insight into the ERVK Integrase - Propensity for DNA Damage. *Front. Microbiol.* 7:1941. doi: 10.3389/fmicb.2016.01941
- Brohawn, D. G., O'Brien, L. C., and Bennett, J. P. Jr. (2016). RNAseq analyses identify tumor necrosis factor-mediated inflammation as a major abnormality in ALS spinal cord. *PLoS One* 11:e0160520. doi: 10.1371/journal.pone.0160520
- Chiodi, F. (2006). A link between immune hyperactivation of T cells during HIV-1 infection and the virus protease? *AIDS* 20, 769–771. doi: 10.1097/01.aids.0000216378.84313.b6
- Christensen, T. (2016). Human endogenous retroviruses in neurologic disease. *APMIS* 124, 116–126. doi: 10.1111/apm.12486
- Contreras-Galindo, R., Kaplan, M. H., Contreras-Galindo, A. C., Gonzalez-Hernandez, M. J., Ferlenghi, I., Giusti, F., et al. (2012). Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *J. Virol.* 86, 262–276. doi: 10.1128/JVI.00602-11
- Cowley, M., and Oakey, R. J. (2013). Transposable elements re-wire and fine-tune the transcriptome. *PLoS Genet.* 9:e1003234. doi: 10.1371/journal.pgen.1003234
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088
- Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat. Methods* 9:772. doi: 10.1038/nmeth.2109
- Denner, J., and Young, P. R. (2013). Koala retroviruses: characterization and impact on the life of koalas. *Retrovirology* 10:108. doi: 10.1186/1742-4690-10-108
- Dewannieux, M., Harper, F., Richaud, A., Letzelter, C., Ribet, D., Pierron, G., et al. (2006). Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res.* 16, 1548–1556. doi: 10.1101/gr.5565706
- Douville, R., Liu, J., Rothstein, J., and Nath, A. (2011). Identification of active loci of a human endogenous retrovirus in neurons of patients with amyotrophic lateral sclerosis. *Ann. Neurol.* 69, 141–151. doi: 10.1002/ana.22149
- Dunn, B. M., Goodenow, M. M., Gustchina, A., and Wlodawer, A. (2002). Retroviral proteases. *Genome Biol.* 3:REVIEWS3006. doi: 10.1186/gb-2002-3-4-reviews3006
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics* 14, 755–763. doi: 10.1093/bioinformatics/14.9.755
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Frank, O., Giehl, M., Zheng, C., Hehlmann, R., Leib-Mosch, C., and Seifarth, W. (2005). Human endogenous retrovirus expression profiles in samples from brains of patients with schizophrenia and bipolar disorders. *J. Virol.* 79, 10890–10901. doi: 10.1128/JVI.79.17.10890-10901.2005
- Franklin, G. C., Chretien, S., Hanson, I. M., Rochefort, H., May, F. E., and Westley, B. R. (1988). Expression of human sequences related to those of mouse mammary tumor virus. *J. Virol.* 62, 1203–1210.

- Gifford, R., and Tristem, M. (2003). The evolution, distribution and diversity of endogenous retroviruses. *Virus Genes* 26, 291–315. doi: 10.1023/A:1024455415443
- Gremme, G., Steinbiss, S., and Kurtz, S. (2013). GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 10, 645–656. doi: 10.1109/TCBB.2013.68
- Grow, E. J., Flynn, R. A., Chavez, S. L., Bayless, N. L., Wossidlo, M., Wesche, D. J., et al. (2015). Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 522, 221–225. doi: 10.1038/nature14308
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hartl, M. J., Wohrl, B. M., Rosch, P., and Schweimer, K. (2008). The solution structure of the simian foamy virus protease reveals a monomeric protein. *J. Mol. Biol.* 381, 141–149. doi: 10.1016/j.jmb.2008.05.064
- Hidaka, K., Kimura, T., Abdel-Rahman, H. M., Nguyen, J. T., McDaniel, K. F., Kohlbrenner, W. E., et al. (2009). Small-sized human immunodeficiency virus type-1 protease inhibitors containing allophenylnorstatine to explore the S2' pocket. *J. Med. Chem.* 52, 7604–7617. doi: 10.1021/jm9005115
- Holmes, E. C. (2011). The evolution of endogenous viral elements. *Cell Host Microbe* 10, 368–377. doi: 10.1016/j.chom.2011.09.002
- Hong, L., Hartsock, J. A., Foundling, S., Ermoloeff, J., and Tang, J. (1998). Active-site mobility in human immunodeficiency virus, type 1, protease as demonstrated by crystal structure of A28S mutant. *Protein Sci.* 7, 300–305. doi: 10.1002/pro.5560070209
- Ido, E., Han, H. P., Kezdy, F. J., and Tang, J. (1991). Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S. *J. Biol. Chem.* 266, 24359–24366.
- Impens, F., Timmerman, E., Staes, A., Moens, K., Arien, K. K., Verhasselt, B., et al. (2012). A catalogue of putative HIV-1 protease host cell substrates. *Biol. Chem.* 393, 915–931. doi: 10.1515/hsz-2012-0168
- Jaskolski, M., Miller, M., Rao, J. K., Leis, J., and Wlodawer, A. (1990). Structure of the aspartic protease from Rous sarcoma retrovirus refined at 2-Å resolution. *Biochemistry* 29, 5889–5898. doi: 10.1021/bi00477a002
- Jha, A. R., Nixon, D. F., Rosenberg, M. G., Martin, J. N., Deeks, S. G., Hudson, R. A., et al. (2011). Human endogenous retrovirus K106 (HERV-K106) was infectious after the emergence of anatomically modern humans. *PLoS One* 6:e20234. doi: 10.1371/journal.pone.0020234
- Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* 110, 462–467. doi: 10.1159/000084979
- Kaneko-Ishino, T., and Ishino, F. (2012). The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. *Front. Microbiol.* 3:262. doi: 10.3389/fmicb.2012.00262
- Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., et al. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. U.S.A.* 85, 4686–4690. doi: 10.1073/pnas.85.13.4686
- Konvalinka, J., Litterst, M. A., Welker, R., Kottler, H., Rippmann, F., Heuser, A. M., et al. (1995). An active-site mutation in the human immunodeficiency virus type 1 proteinase (PR) causes reduced PR activity and loss of PR-mediated cytotoxicity without apparent effect on virus maturation and infectivity. *J. Virol.* 69, 7180–7186.
- Kourjian, G., Rucevic, M., Berberich, M. J., Dinter, J., Wambua, D., Boucau, J., et al. (2016). HIV protease inhibitor-induced cathepsin modulation alters antigen processing and cross-presentation. *J. Immunol.* 196, 3595–3607. doi: 10.4049/jimmunol.1600055
- Kourjian, G., Xu, Y., Mondesire-Crump, I., Shimada, M., Gourdain, P., and Le Gall, S. (2014). Sequence-specific alterations of epitope production by HIV protease inhibitors. *J. Immunol.* 192, 3496–3506. doi: 10.4049/jimmunol.1302805
- Kozak, C. A. (2014). Origins of the endogenous and infectious laboratory mouse gammaretroviruses. *Viruses* 7, 1–26. doi: 10.3390/v7010001
- Kraus, B., Boller, K., Reuter, A., and Schnierle, B. S. (2011). Characterization of the human endogenous retrovirus K Gag protein: identification of protease cleavage sites. *Retrovirology* 8:21. doi: 10.1186/1742-4690-8-21
- Kuhelj, R., Rizzo, C. J., Chang, C. H., Jadhav, P. K., Towler, E. M., and Korant, B. D. (2001). Inhibition of human endogenous retrovirus-K10 protease in cell-free and cell-based assays. *J. Biol. Chem.* 276, 16674–16682. doi: 10.1074/jbc.M008763200M008763200
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Lee, Y. N., and Bieniasz, P. D. (2007). Reconstitution of an infectious human endogenous retrovirus. *PLoS Pathog.* 3:e10. doi: 10.1371/journal.ppat.0030010
- Li, W., Lee, M. H., Henderson, L., Tyagi, R., Bachani, M., Steiner, J., et al. (2015). Human endogenous retrovirus-K contributes to motor neuron disease. *Sci. Transl. Med.* 7:307ra153. doi: 10.1126/scitranslmed.aac8201
- Lin, Y. C., Beck, Z., Morris, G. M., Olson, A. J., and Elder, J. H. (2003). Structural basis for distinctions between substrate and inhibitor specificities for feline immunodeficiency virus and human immunodeficiency virus proteases. *J. Virol.* 77, 6589–6600. doi: 10.1128/JVI.77.12.6589-6600.2003
- Louis, J. M., Ishima, R., Torchia, D. A., and Weber, I. T. (2007). HIV-1 protease: structure, dynamics, and inhibition. *Adv. Pharmacol.* 55, 261–298. doi: 10.1016/S1054-3589(07)55008-8
- Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., et al. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63. doi: 10.1038/nature11244
- Manghera, M., Ferguson, J., and Douville, R. (2014). Endogenous retrovirus-K and nervous system diseases. *Curr. Neurol. Neurosci. Rep.* 14:488. doi: 10.1007/s11910-014-0488-y
- Manghera, M., Ferguson, J., and Douville, R. (2015). ERVK polyprotein processing and reverse transcriptase expression in human cell line models of neurological disease. *Viruses* 7, 320–332. doi: 10.3390/v7010320
- Mayer, J., Blomberg, J., and Seal, R. L. (2011). A revised nomenclature for transcribed human endogenous retroviral loci. *Mob. DNA* 2:7. doi: 10.1186/1759-8753-2-7
- Mayer, J., and Meese, E. U. (2002). The human endogenous retrovirus family HERV-K(HML-3). *Genomics* 80, 331–343. doi: 10.1006/geno.2002.6839
- Menendez-Arias, L., Weber, I. T., Soss, J., Harrison, R. W., Gotte, D., and Oroszlan, S. (1994). Kinetic and modeling studies of subsites S4-S3' of Moloney murine leukemia virus protease. *J. Biol. Chem.* 269, 16795–16801.
- Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racie, L., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789. doi: 10.1038/35001608
- Michaud, H. A., de Mulder, M., SenGupta, D., Deeks, S. G., Martin, J. N., Pilcher, C. D., et al. (2014). Trans-activation, post-transcriptional maturation, and induction of antibodies to HERV-K (HML-2) envelope transmembrane protein in HIV-1 infection. *Retrovirology* 11:10. doi: 10.1186/1742-4690-11-10
- Morandi, E., Tarlinton, R. E., and Gran, B. (2015). Multiple sclerosis between genetics and infections: human endogenous retroviruses in monocytes and macrophages. *Front. Immunol.* 6:647. doi: 10.3389/fimmu.2015.00647
- Mueller-Lantzsch, N., Sauter, M., Weiskircher, A., Kramer, K., Best, B., Buck, M., et al. (1993). Human endogenous retroviral element K10 (HERV-K10) encodes a full-length gag homologous 73-kDa protein and a functional protease. *AIDS Res. Hum. Retroviruses* 9, 343–350. doi: 10.1089/aid.1993.9.343
- Naveira, H., Bello, X., Abal-Fabeiro, J. L., and Maside, X. (2014). Evidence for the persistence of an active endogenous retrovirus (ERVE) in humans. *Genetica* 142, 451–460. doi: 10.1007/s10709-014-9789-y
- Ono, M. (1986). Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. *J. Virol.* 58, 937–944.
- Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. J., et al. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 52, 1–8. doi: 10.1006/geno.1998.5423
- Prudencio, M., Gonzales, P. K., Cook, C. N., Gendron, T. F., Daugherty, L. M., Song, Y., et al. (2017). Repetitive element transcripts are elevated in the brain of C9orf72 ALS/FTLD patients. *Hum. Mol. Genet.* 26, 3421–3431. doi: 10.1093/hmg/ddx233
- Ranwez, V., Harispe, S., Delsuc, F., and Douzery, E. J. (2011). MACSE: multiple alignment of coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6:e22594. doi: 10.1371/journal.pone.0022594
- Ren, S., Peng, Z., Mao, J. H., Yu, Y., Yin, C., Gao, X., et al. (2012). RNA-seq analysis of prostate cancer in the Chinese population identifies recurrent gene fusions,

- cancer-associated long noncoding RNAs and aberrant alternative splicings. *Cell Res.* 22, 806–821. doi: 10.1038/cr.2012.30
- Reynier, F., Verjat, T., Turrel, F., Imbert, P. E., Marotte, H., Mougin, B., et al. (2009). Increase in human endogenous retrovirus HERV-K (HML-2) viral load in active rheumatoid arthritis. *Scand. J. Immunol.* 70, 295–299. doi: 10.1111/j.1365-3083.2009.02271.x
- Ruda, V. M., Akopov, S. B., Trubetskoy, D. O., Manuylov, N. L., Vetchinova, A. S., Zavalova, L. L., et al. (2004). Tissue specificity of enhancer and promoter activities of a HERV-K(HML-2) LTR. *Virus Res.* 104, 11–16. doi: 10.1016/j.virusres.2004.02.036
- Rumlova, M., Krizova, I., Keprova, A., Hadravova, R., Dolezal, M., Strohalmova, K., et al. (2014). HIV-1 protease-induced apoptosis. *Retrovirology* 11:37. doi: 10.1186/1742-4690-11-37
- Sauter, M., Schommer, S., Kremmer, E., Remberger, K., Dolken, G., Lemm, I., et al. (1995). Human endogenous retrovirus K10: expression of Gag protein and detection of antibodies in patients with seminomas. *J. Virol.* 69, 414–421.
- Schlesinger, S., and Goff, S. P. (2013). Silencing of proviruses in embryonic cells: efficiency, stability and chromatin modifications. *EMBO Rep.* 14, 73–79. doi: 10.1038/embor.2012.182
- Schlesinger, S., Meshorer, E., and Goff, S. P. (2014). Asynchronous transcriptional silencing of individual retroviral genomes in embryonic cells. *Retrovirology* 11:31. doi: 10.1186/1742-4690-11-31
- Seifarth, W., Frank, O., Zeilfelder, U., Spiess, B., Greenwood, A. D., Hehlmann, R., et al. (2005). Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. *J. Virol.* 79, 341–352. doi: 10.1128/JVI.79.1.341-352.2005
- Shoeman, R. L., Kesselmier, C., Mothes, E., Honer, B., and Traub, P. (1991). Non-viral cellular substrates for human immunodeficiency virus type 1 protease. *FEBS Lett.* 278, 199–203. doi: 10.1016/0014-5793(91)80116-K
- Snasel, J., Shoeman, R., Horejsi, M., Hruskova-Heidingsfeldova, O., Sedlacek, J., Ruml, T., et al. (2000). Cleavage of vimentin by different retroviral proteases. *Arch. Biochem. Biophys.* 377, 241–245. doi: 10.1006/abbi.2000.1776
- Solis, M., Nakhaei, P., Jalalirad, M., Lacoste, J., Douville, R., Arguello, M., et al. (2011). RIG-I-mediated antiviral signaling is inhibited in HIV-1 infection by a protease-mediated sequestration of RIG-I. *J. Virol.* 85, 1224–1236. doi: 10.1128/JVI.01635-10
- Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Stewart, C., Kural, D., Stromberg, M. P., Walker, J. A., Konkel, M. K., Stutz, A. M., et al. (2011). A comprehensive map of mobile element insertion polymorphisms in humans. *PLoS Genet.* 7:e1002236. doi: 10.1371/journal.pgen.1002236
- Tien, C., Huang, L., Watanabe, S. M., Speidel, J. T., Carter, C. A., and Chen, C. (2018). Context-dependent autoprocessing of human immunodeficiency virus type 1 protease precursors. *PLoS One* 13:e0191372. doi: 10.1371/journal.pone.0191372
- Towler, E. M., Gulnik, S. V., Bhat, T. N., Xie, D., Gustschina, E., Sumpter, T. R., et al. (1998). Functional characterization of the protease of human endogenous retrovirus, K10: can it complement HIV-1 protease? *Biochemistry* 37, 17137–17144. doi: 10.1021/bi9818927
- Turner, G., Barbulescu, M., Su, M., Jensen-Seaman, M. I., Kidd, K. K., and Lenz, J. (2001). Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr. Biol.* 11, 1531–1535. doi: 10.1016/S0960-9822(01)00455-9
- Tyagi, R., Li, W., Parades, D., Bianchet, M. A., and Nath, A. (2017). Inhibition of human endogenous retrovirus-K by antiretroviral drugs. *Retrovirology* 14:21. doi: 10.1186/s12977-017-0347-4
- Vargiu, L., Rodriguez-Tome, P., Sperber, G. O., Cadeddu, M., Grandi, N., Blikstad, V., et al. (2016). Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* 13:7. doi: 10.1186/s12977-015-0232-y
- Wagner, R. N., Reed, J. C., and Chanda, S. K. (2015). HIV-1 protease cleaves the serine-threonine kinases RIPK1 and RIPK2. *Retrovirology* 12:74. doi: 10.1186/s12977-015-0200-6
- Wheeler, T. J., Clements, J., Eddy, S. R., Hubley, R., Jones, T. A., Jurka, J., et al. (2013). Dfam: a database of repetitive DNA based on profile hidden Markov models. *Nucleic Acids Res.* 41, D70–D82. doi: 10.1093/nar/gks1265
- Wildschutte, J. H., Ram, D., Subramanian, R., Stevens, V. L., and Coffin, J. M. (2014). The distribution of insertionally polymorphic endogenous retroviruses in breast cancer patients and cancer-free controls. *Retrovirology* 11:62. doi: 10.1186/s12977-014-0062-310.1186/PREACCEPT-1720768941312026
- Wildschutte, J. H., Williams, Z. H., Montesion, M., Subramanian, R. P., Kidd, J. M., and Coffin, J. M. (2016). Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc. Natl. Acad. Sci. U.S.A.* 113, E2326–E2334. doi: 10.1073/pnas.1602336113
- Yoshikawa, R., Takeuchi, J. S., Yamada, E., Nakano, Y., Misawa, N., Kimura, Y., et al. (2017). Feline immunodeficiency virus evolutionarily acquires two proteins, Vif and protease, capable of antagonizing feline APOBEC3. *J. Virol.* 91:e00250-17. doi: 10.1128/JVI.00250-17

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Turnbull and Douville. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Transcriptional Modulation of Human Endogenous Retroviruses in Primary CD4⁺ T Cells Following Vorinostat Treatment

OPEN ACCESS

Edited by:

Hao Shen,
University of Pennsylvania,
United States

Reviewed by:

Dirk Dittmer,
University of North Carolina at Chapel
Hill, United States
Wenhui Hu,
Temple University, United States

*Correspondence:

Christopher H. Woelk
christopher.woelk@merck.com

[†]Present address:

Cory H. White and
Christopher H. Woelk,
Merck Exploratory Science Center,
Merck Research Laboratories,
Cambridge, MA, United States;
Gkikas Magiorkinis
Department of Hygiene,
Epidemiology and Medical Statistics,
Medical School, National and
Kapodistrian University of Athens,
Athens, Greece

Specialty section:

This article was submitted to
Microbial Immunology,
a section of the journal
Frontiers in Immunology

Received: 24 October 2017

Accepted: 09 March 2018

Published: 12 April 2018

Citation:

White CH, Beliakova-Bethell N,
Lada SM, Breen MS, Hurst TP,
Spina CA, Richman DD, Frater J,
Magiorkinis G and Woelk CH (2018)
Transcriptional Modulation of Human
Endogenous Retroviruses in Primary
CD4⁺ T Cells Following Vorinostat
Treatment.
Front. Immunol. 9:603.
doi: 10.3389/fimmu.2018.00603

Cory H. White^{1†}, Nadejda Beliakova-Bethell^{2,3}, Steven M. Lada², Michael S. Breen⁴,
Tara P. Hurst⁵, Celsa A. Spina^{2,6}, Douglas D. Richman^{2,3,6}, John Frater⁷,
Gkikas Magiorkinis^{5†} and Christopher H. Woelk^{1*†}

¹ Faculty of Medicine, University of Southampton, Southampton, Hants, United Kingdom, ² San Diego VA Medical Center and Veterans Medical Research Foundation, San Diego, CA, United States, ³ Department of Medicine, University of California San Diego, La Jolla, CA, United States, ⁴ Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States, ⁵ Department of Zoology, University of Oxford, Oxford, United Kingdom, ⁶ Department of Pathology, University of California San Diego, La Jolla, CA, United States, ⁷ Nuffield Department of Clinical Medicine, Peter Medawar Building for Pathogen Research, South Parks Road, Oxford, United Kingdom

The greatest obstacle to a cure for HIV is the provirus that integrates into the genome of the infected cell and persists despite antiretroviral therapy. A “shock and kill” approach has been proposed as a strategy for an HIV cure whereby drugs and compounds referred to as latency-reversing agents (LRAs) are used to “shock” the silent provirus into active replication to permit “killing” by virus-induced pathology or immune recognition. The LRA most utilized to date in clinical trials has been the histone deacetylase (HDAC) inhibitor—vorinostat. Potentially, pathological off-target effects of vorinostat may result from the activation of human endogenous retroviruses (HERVs), which share common ancestry with exogenous retroviruses including HIV. To explore the effects of HDAC inhibition on HERV transcription, an unbiased pharmacogenomics approach (total RNA-Seq) was used to evaluate HERV expression following the exposure of primary CD4⁺ T cells to a high dose of vorinostat. Over 2,000 individual HERV elements were found to be significantly modulated by vorinostat, whereby elements belonging to the ERVL family (e.g., LTR16C and LTR33) were predominantly downregulated, in contrast to LTR12 elements of the HERV-9 family, which exhibited the greatest signal, with the upregulation of 140 distinct elements. The modulation of three different LTR12 elements by vorinostat was confirmed by droplet digital PCR along a dose–response curve. The monitoring of LTR12 expression during clinical trials with vorinostat may be indicated to assess the impact of this HERV on the human genome and host immunity.

Keywords: human endogenous retroviruses, histone deacetylase inhibitor, primary CD4⁺ T cells, total RNA-Seq, long terminal repeat

INTRODUCTION

Vorinostat is a histone deacetylase (HDAC) inhibitor also known as suberoylanilide hydroxamic acid. HDAC inhibitors act on HDAC enzymes and block the removal of acetyl groups from histones resulting in a relaxed chromatin state (1) and the modulation of the expression of large numbers of genes (2, 3). In addition, HDAC inhibitors appear to affect the acetylation states of transcription

factors at the protein level, which alters their activity and leads to further transcriptional changes (4). HDAC inhibitors have wide ranging therapeutic value and have been considered for the treatment of cancer (5) and neurodegenerative disorders (6), as well as in “shock and kill” strategies to facilitate an HIV cure (7). The therapeutic efficacy of HDAC inhibitors against cancer is thought to stem from their ability to induce tumor cell apoptosis (5). Vorinostat is approved by the Federal Drug Administration (FDA) for the treatment of refractory cutaneous T-cell lymphoma (8). In an HIV cure setting, HDAC inhibitors may provide the “shock” capable of flushing HIV out of the persistent reservoir, while antiretroviral therapy is used to prevent new infections so that the cell lysis mediated by viral replication or the immune system may then “kill” actively replicating cells (7). Due to the pre-existing FDA approvals for human use, vorinostat has already been used in a number of completed (9, 10) and ongoing (11) clinical trials assessing shock and kill strategies for an HIV cure.

Human endogenous retroviruses (HERVs), which constitute approximately 8% of the human genome, are themselves descended from ancient exogenous retroviruses (12) and thus share common ancestry with HIV. HERV structure reflects that of retroviruses with two long terminal repeat (LTR) elements flanking *gag*, *pol*, and *env* genes, although HERVs most frequently exist in the genome as solitary LTR elements due to the loss of genes through recombination (13). Since vorinostat activates the expression of HIV, there have been concerns that this drug may also upregulate HERVs with potentially pathological consequences (14). For example, HERV pathology could result from the modulation of the expression of protein coding genes or the formation of chimeric proteins with aberrant function leading to oncogenesis (15), as well as the dysregulation of inflammatory immune responses through the expression of HERV encoded proteins (e.g., *gag* and *env*) (16). Indeed, HERV expression has previously been associated with a wide repertoire of diseases including diabetes, schizophrenia, autoimmune diseases (e.g., multiple sclerosis and rheumatoid arthritis), and cancer (17). However, the difficulties in associating HERVs with disease should be stressed due to their ubiquitous nature in human populations although polymorphisms between individuals could explain disease specificity (17). Finally, it was previously shown that HIV capsids could be successfully pseudotyped *in vitro* with HERV-W Env resulting in infectious virus particles (18). This raises the possibility that coexpression of HERVs and HIV might lead to novel retroviral strains with new properties through transcomplementation or recombination, although the latter may be unlikely due to the large evolutionary distance between HERV elements and HIV (19).

To explore the ability of vorinostat to modulate the expression of HERV elements in the human genome, our previous analysis utilized a targeted approach [i.e., real-time reverse transcription polymerase chain reaction (RT-qPCR)], to assess the expression of the *env* and *pol* genes of specific HERV families (i.e., HERV-K, HERV-W, and HERV-FRD) following HDAC inhibitor treatment (14). This study showed that cell line model systems of chronic HIV infection (i.e., J-LAT-8.4 and U1 cells) treated with different concentrations of vorinostat (i.e., 1 μ M and 1 mM) for 24 h did not significantly alter the expression

of these HERV elements. Furthermore, treatment of uninfected and HIV-infected primary CD4⁺ T cells with another HDAC inhibitor, panobinostat (20 nM), for 24 h did not result in the upregulation of these HERV genes. In contrast, Kronung et al. (20) previously applied another targeted RT-qPCR approach to study the expression of transcripts of the *TP63* and *TNFRSF10B* genes that are under control of an LTR12 promoter derived from the HERV-9 family. Treatment with vorinostat (1 or 5 μ M) for 18 h upregulated these genes *via* the LTR12 promoter across various cells lines (i.e., GH, H1299, K562, U2OS, HeLa, Ovar-3, and HuT-78) suggesting that this drug may indeed modulate HERV elements. However, discrepancies have been noted between cell lines and primary cells with respect to the host gene transcriptional profile induced by vorinostat (2). The main motivation for the current study was to resolve these discrepancies and determine if vorinostat can modulate HERV elements in primary CD4⁺ T cells using an unbiased approach (i.e., total RNA-Seq). Uninfected instead of HIV-infected primary CD4⁺ T cells were selected for study to disambiguate the effects of vorinostat on HERV elements since the Tat protein of HIV has also been shown to activate HERV elements, e.g., HERV-K(HML-2) (21, 22).

MATERIALS AND METHODS

Isolation of Primary CD4⁺ T Cells

For subsequent total RNA-Seq analysis, cryopreserved primary CD4⁺ T cells that were viably frozen were obtained from four different healthy donors (AllCells, Inc., Emeryville, CA, USA) and thawed in RPMI with 20% human serum. Dead cells resulting from thawing frozen cells were removed using Viahance magnetic negative selection (Biophysics Assay Laboratory Inc., Worcester, MA, USA). For dose-response analysis, peripheral blood was isolated from two additional healthy donors by venipuncture according to the protocols approved by an institutional review board into polypropylene syringes containing sodium heparin. Primary CD4⁺ T cells were isolated using the RosetteSep CD4⁺ T cell enrichment cocktail (StemCell Technologies Inc., Vancouver, Canada). Aliquots taken from CD4⁺ T cell samples were subjected to flow cytometry to assess purity (i.e., >95% cells expressing CD4).

Treatment of Primary CD4⁺ T Cells With Vorinostat

Primary CD4⁺ T cells (2.5 million cells per milliliter) were plated into six-well tissue culture plates at 2 ml per well. For the four donors subjected to total RNA-Seq analysis, wells were either treated with a high dose of vorinostat (10 μ M) dissolved in dimethyl sulfoxide (DMSO) or left untreated (i.e., DMSO solvent only). For the two donors subjected to dose-response analysis by digital droplet PCR, the wells were treated with 0.34, 1, 3, and 10 μ M of vorinostat dissolved in DMSO or left untreated (i.e., DMSO solvent only). In all cases, after 24 h of vorinostat exposure, the samples were washed twice with 10 ml of phosphate buffered saline and resuspended in RLT Plus buffer (Qiagen, Valencia, CA, USA) containing β -mercaptoethanol for RNA extraction.

RNA Isolation

Total RNA was extracted from primary CD4⁺ T cells using the RNeasy Plus Kit (Qiagen, Valencia, CA, USA) and genomic DNA removed using an on-column DNase treatment. RNA integrity was assessed using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA) and RNA integrity numbers of samples were on average 8.9 (SD \pm 0.29).

Total RNA-Seq Data Generation

Cytoplasmic and mitochondrial ribosomal RNAs (rRNAs) were removed from total RNA extractions using the Ribo-Zero Gold (Human/Mouse/Rat) rRNA Removal Kit (Epicentre, Madison, WI, USA). RNA-Seq libraries were prepared using the TruSeq™ Stranded Total RNA Library Prep Kit (Illumina, San Diego, CA, USA) and sequenced to a depth of 100 million reads using the HiSeq2000 (Illumina) to generate 50 bp paired-end reads.

Total RNA-Seq Data Analysis

Sequence data in FASTA format for greater than 92,000 distinct HERV elements from the human genome were downloaded from the Endogenous Retrovirus Database (HERVd, 2012 release) (23, 24). This FASTA file of HERV sequence data was converted into a Bowtie index using the *bowtie-build* command (25) and also used to manually construct a gene transfer format (GTF) file. Duplication has led to the expansion of HERV elements throughout the human genome that are often fragmented due to insertions. To enable accurate quantification of HERV elements, reads were mapped with the “-m 1” option using the Bowtie index to ensure that only reads uniquely mapping to a single HERV element with no mismatches were retained. To maximize the alignment of reads to fragments of HERV elements, paired-end reads were decoupled into single-end reads for mapping purposes. The number of reads mapping to each HERV element was then counted using *htseq-count* with the GTF file (26). Raw counts were converted to counts per million (cpm) mapped reads, HERV elements were removed that did not have at least 1 cpm in at least half of the samples, and subjected to trimmed mean of M-values normalization. Finally, reads from

the total RNA-Seq data were also mapped to the human genome using TopHat (27) (default settings with coverage based search for junctions disabled) in order to visualize read pile up against HERV elements in a genomic context using the UCSC genome browser (28).

Total RNA-Seq Data Access

Metadata, FASTQ files, and a raw HERV expression matrix have been submitted to the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE102187.

Droplet Digital PCR (ddPCR) Analysis

Custom TaqMan assays were used to quantify the upregulation by vorinostat of three LTR12 elements with HERVd designations rv_007357, rv_007420, and rv_010177, using the QX100 Droplet Digital PCR System (Bio-Rad, Hercules, CA, USA) as previously described (29, 30). Briefly, the PrimerQuest Tool (31) (Integrated DNA Technologies, Coralville, IA, USA) was used to design two TaqMan assays against rv_007357 and a single assay against each of the remaining LTR12 elements. Primer and probe sequences for these TaqMan assays were subject to BLAT analysis against the human genome to confirm specificity and are presented in **Table 1**. Five nanograms of RNA in a 20 μ l PCR reaction volume were used for each target in duplicate. The TaqMan Gene Expression Assay (Hs03044961_g1) for the ribosomal protein L27 (*RPL27*) gene was selected as a normalizer (32). LTR12 element expression was assessed between the vorinostat treated (10 μ M) and untreated condition using the samples used to generate the original total RNA-Seq data (Donors 1–4). LTR12 expression was also assessed in a vorinostat dose–response curve (0, 0.34, 1, 3, and 10 μ M) for two additional donors (Donors 5 and 6).

Statistical Analyses

For total RNA-Seq data, HERV elements were identified as differentially expressed between vorinostat treated and untreated samples with a false discovery rate (FDR) corrected *p*-value <0.05 using EdgeR (33). EdgeR adopts the negative binomial model as the main approach to model RNA-Seq data. This model

TABLE 1 | Custom TaqMan assays used to confirm the upregulation of LTR12 elements by vorinostat treatment.^a

LTR12 element	Primer and probe set	Designation	Sequence	Length	GC%	Amplicon size (bp)
rv_007357	Set 1	Forward (sense)	GAGCGTATGGCGTTATGTAGTT	22	45.5	114
		Probe (sense)	TTGAGCCGATGAGATCGCTAAGCC	24	54.0	
		Reverse (antisense)	AGCGGTATGTCTCCCTTTA	20	50.0	
	Set 2	Forward (sense)	GGAGGAACGAAACACTCATCT	21	47.6	102
		Probe (antisense)	TGCAACTTTCACAGAGTCGTCTCACC	26	50.0	
		Reverse (antisense)	CGTCTACCCACTTCAGAAA	20	50.0	
rv_007420	Set 1	Forward (sense)	GGTAGTGAGAGAGAACGGTATG	22	50.0	124
		Probe (sense)	TCCTCTGCTCATTCTGGTTGTGCT	24	50.0	
		Reverse (antisense)	CTAAGAGCTCCACGGTATAG	22	50.0	
rv_010177	Set 1	Forward (sense)	ACTCCAGACACACCGTCTTA	20	50.0	96
		Probe (sense)	ATTGGTAGCTTTCCCGAGTCAGCG	24	54.0	
		Reverse (antisense)	TCATTCCATTACAGTGGGGTTC	21	47.6	

^aThe “rv” designations from the Human Endogenous Retrovirus Database (HERVd) are listed for each LTR12 element. Two primer and probe sets were used for the LTR12 element with designation rv_007357.

GC%, percentage of guanine and cytosine bases in corresponding primer or probe; bp, base pair.

approach requires an estimate of the true biological coefficient of variation. The square root of this value, the dispersion parameter, was estimated in edgeR by initially measuring a single dispersion parameter using all genes, while taking into account trends or gene abundances (i.e., trended dispersion). Then genewise (i.e., tagwise) dispersion estimates are measured and an empirical Bayes method was used to shrink these genewise dispersion estimates toward the trended dispersion. Gene expression data were then fit to the generalized linear model (GLM), and a GLM likelihood test was used to assess differential gene expression. Model parameters included a sample donor variable to account for the paired structure of the data (paired before and after vorinostat treatment). Significance values were adjusted for multiple testing using the Benjamini and Hochberg (34) method to control the FDR.

Validation of gene expression initially assessed by total RNA-Seq was performed using ddPCR, which is more sensitive than RT-qPCR, since target RNA is distributed across thousands of oil emulsion droplets that each undergoes reverse transcription and a subsequent end point PCR reaction. The number of target RNA molecules present was calculated from the fraction of positive end point reactions using Poisson statistics because some droplets contain no template while others contain one or more copies (35). All ddPCR data were expressed as copies of target RNA molecules (e.g., LTR12 element) per million copies of *RPL27* mRNA molecules and then \log_2 transformed. Prior to \log_2 transformation, a small regularization constant of 0.01 was added to all values used in this calculation to avoid taking a logarithm of zero in some instances. Differential expression of LTR12 elements was assessed between the vorinostat treated (10 μ M) and untreated condition in a paired *t*-test (*p*-value < 0.05) using the samples used to generate the original total RNA-Seq data (Donors 1–4), as well as along the aforementioned dose–response curve for Donors 5 and 6.

RESULTS

Global Modulation of HERVs Upon Treatment With Vorinostat

Primary CD4⁺ T cells were isolated from four seronegative donors and treated with a high dose of vorinostat (10 μ M) or left untreated for 24 h. This high dose of vorinostat was initially used in an exploratory approach to optimize the ability to identify HERV elements modulated by this HDAC inhibitor. These eight samples were subjected to total RNA-Seq analysis, and the resulting data mapped against HERV sequences curated in HERVd (23). The non-parametric equivalent of a paired *t*-test identified 2,101 distinct HERV elements modulated by vorinostat with an FDR corrected *p*-value of less than 0.05 that mapped to 120 different HERV families across the human genome. In a conservative approach, HERV elements with an absolute \log_2 fold change of more than |3| were identified, leaving 451 upregulated HERV elements from 81 distinct HERV families and 363 downregulated elements from 82 families annotated from HERVd (Figure 1). LTR16C and LTR33 elements, which originated from the ERVL family, were predominantly downregulated, whereas LTR12 elements from the HERV-9 family were predominantly upregulated. The upregulation of LTR12 elements was by far the most

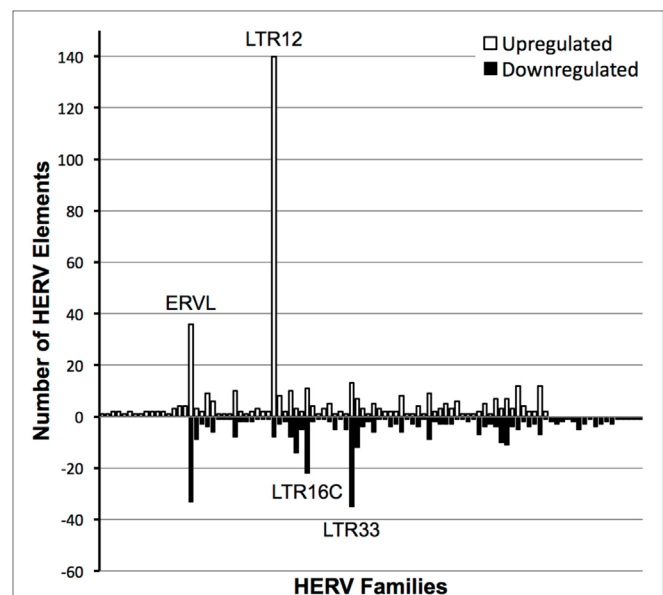


FIGURE 1 | Number of distinct human endogenous retrovirus (HERV) elements per HERV family modulated by vorinostat treatment. HERV elements significantly differentially expressed (false discovery rate-corrected *p*-value < 0.05) with a \log_2 fold change of greater than 3 or less than -3 between vorinostat treated and untreated samples are depicted. HERV elements were grouped into HERV families according to designations in the Human Endogenous Retrovirus Database. White bars represent the number of HERV elements upregulated by vorinostat and black bars the number of downregulated elements. For example, there were 140 distinct HERV elements belonging to the LTR12 family which were significantly upregulated with a \log_2 fold change greater than 3 by vorinostat treatment.

dramatic of which the most upregulated (rv_005487) had a \log_2 fold change of 11.985 (actual fold change 4,054). Furthermore, the top 100 upregulated HERV elements contained 46 that were from the LTR12 HERV family (data not shown). Other ERVL elements (i.e., not belonging to the LTR16C or LTR33 families) had a balance of up- and downregulated elements. In summary, vorinostat clearly modulated HERV elements across the genome, but appears to have specificity for certain elements (e.g., LTR12) and families (e.g., ERVL and HERV-9).

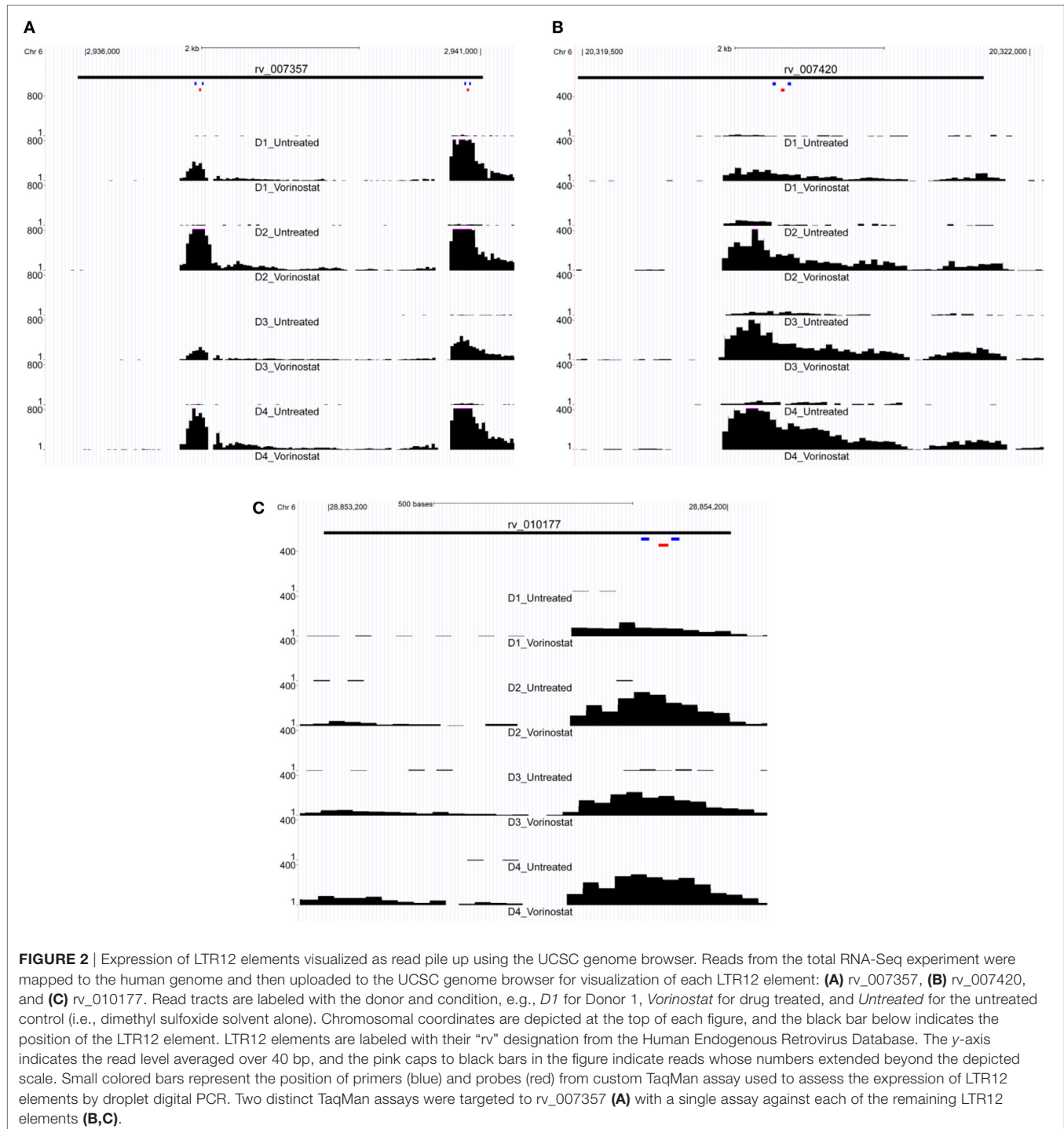
Previously, using targeted RT-qPCR analysis we demonstrated that cell lines chronically infected with HIV (i.e., J-LAT8.4 or U1 cells) exposed to low (1 μ M) and high (1 mM) doses of vorinostat did not lead to the consistent upregulation of the following HERV elements: HK2 *env*, HK2 *pol*, HERV-W *env* (syncytin-1), and HERV-FRD *env* (syncytin-2) (14). To confirm these findings, the expression of these elements in the total RNA-Seq data of the current study was examined following vorinostat treatment of primary CD4⁺ T cells. There was no difference in expression of these elements between vorinostat treated and untreated cells (Figure S1 in Supplementary Material).

ddPCR Confirmation of LTR12 Upregulation by Vorinostat

LTR12 elements were selected for ddPCR validation since these elements were the most upregulated by vorinostat. Three

LTR12 elements on chromosome 6 (rv_007357, rv_007420, and rv_010177) were selected for ddPCR analysis since they were (1) upregulated by at least eightfold (\log_2 fold change of 3), (2) longer than 1,000 bp and upregulated along large segments of the HERV element, (3) not confounded by the presence of a neighboring gene within 5 kb, and (4) consistently upregulated across all donors (Figure 2). The most upregulated LTR12 element (rv_005487) is approximately 1.5 kb in the human genome but was not selected

for ddPCR analysis because only a 250 bp fragment of this element was expressed. Two TaqMan primer and probe sets were designed against rv_007357 to capture the two major peaks of expression in this LTR12 element, whereas a single primer and probe set was used to target rv_007420 and rv_010177 (Table 1; Figure 2). Significant upregulation of all three elements upon vorinostat exposure was confirmed by ddPCR and fold changes reflected those detected by RNA-Seq analysis (Figure 3).



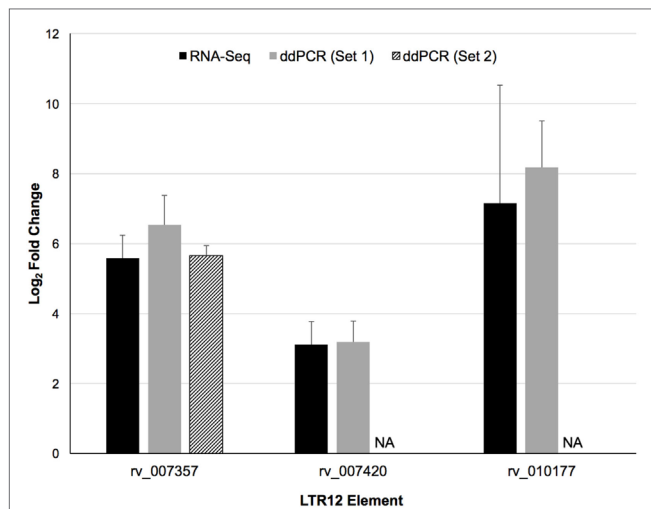


FIGURE 3 | Confirmation of the upregulation of three distinct LTR12 elements by vorinostat in primary CD4⁺ T cells using droplet digital PCR (ddPCR). Log₂ fold changes between vorinostat treated and untreated conditions were averaged across all four donors (Donors 1–4) for the RNA-Seq (black bars) and the ddPCR (gray and hatched bars) data for each LTR12 element. Error bars represent SDs across donors. The labels “Set 1” and “Set 2” indicate the two distinct primer and probe sets used to target the same LTR12 element (i.e., rv_007357). A second primer and probe set was not used to target the other LTR12 elements, and the missing bar is thus labeled “NA” for not applicable. LTR12 elements are labeled with their “rv” designation from the Human Endogenous Retrovirus Database.

Dose Responsive Upregulation of LTR12 by Vorinostat

A high dose of vorinostat (10 μ M) was used to treat primary CD4⁺ T cells subjected to RNA-Seq analysis. This dose is just above what can be achieved with intravenous administration of vorinostat for the treatment of B- and T-cell malignancies (36) and much higher than doses achieved via oral administration in clinical trials to explore “Shock and Kill” strategies for an HIV cure (9, 10). Furthermore, 10 μ M of vorinostat appears cytotoxic for just over 20% of healthy T lymphocytes (37). Therefore, the expression of the three LTR12 elements was examined by ddPCR over a more pharmacologically relevant dose–response curve (0.34, 1, 3, and 10 μ M), which included less cytotoxic doses (e.g., 0.34 and 1 μ M), in primary CD4⁺ T cells isolated from two independent donors. The expression of each LTR12 element was clearly dose dependent for each donor and greater than twofold (log₂ fold change > 1) for rv_007357 at the lowest dose of 0.34 μ M for both donors (Figure 4). In summary, vorinostat upregulated LTR12 elements at doses used to treat blood cell malignancies (36) and at even lower doses (i.e., 0.34 μ M) that are relevant to clinical trials with HIV-infected individuals (9, 10).

DISCUSSION

A total RNA-Seq experiment was used to investigate the modulation of HERVs in primary CD4⁺ T cells exposed to vorinostat. The primary result from this work demonstrated the power

of such an untargeted (i.e., unbiased) approach. Our previous studies, using a targeted RT-qPCR approach, had concluded that vorinostat did not modulate HERV elements of the HERV-K, HERV-W, and HERV-FRD families (14). Although this result was confirmed for elements from these families (Figure S1 in Supplementary Material), the current untargeted approach suggests that vorinostat does indeed modulate HERV elements in primary CD4⁺ T cells, predominantly LTRs (e.g., LTR12, LTR16C, and LTR33) of the ERV9 and ERVL families (Figure 1). Brocks and coworkers (38) recently used Cap Analysis Gene Expression Sequencing (CAGE-Seq) (39) to identify expression from transcription start sites (TSS) in a lung cancer cell line (NCI-1299) exposed to vorinostat. The CAGE-Seq data were used to identify treatment induced non-annotated TSS, which were shown to be enriched for LTR12 elements, thus confirming the expression of these elements by vorinostat in a cell line using a different methodology. Furthermore, Kronung and colleagues (20) noted that vorinostat could activate the expression of LTR12-driven genes (e.g., *TP63* and *TNFRSF10B*) in cell lines (i.e., J-LAT8.4 and U1) but did not modulate the expression of HERV-E, HERV-H or MaLR-driven genes, and thus concluded that vorinostat was specific for ERV9 LTRs (i.e., LTR12). However, the current study suggests that vorinostat modulates the expression of LTRs outside of the ERV9 family (e.g., LTR16C and LTR33). In summary, these results advocate for untargeted approaches (e.g., total RNA-Seq and CAGE-Seq), often referred to as “fishing expeditions,” since targeted approaches may not interrogate all relevant transcripts.

The ERV9 family, which includes the LTR12 element, is one of the most successful inhabitants of the human genome due to its continued proliferation until almost six million years ago, around the time of the human-chimp split (40). In contrast, the ERVL family is probably the oldest family and appears to lack an *env* gene consistent with these elements being ancient retrotransposons that entered genomes before the mammalian radiation (41). It is not clear why the transcription of both of these HERV families is modulated by an HDAC inhibitor (i.e., vorinostat) but suggests that HDACs may be an important epigenetic checkpoint in their transcription. HERV elements have recently been associated with non-coding regulatory RNAs with diverse properties ranging from contributing to the pluripotency of human cells (42) to promoting immunoglobulin M production in B-cell driven immune responses independent of T-cells (43). Therefore, agents such as vorinostat that alter the function of HDACs may need to undergo additional evaluation for HERV upregulation to assess their impact on the function of immune cells.

Modulation of HERV elements by vorinostat may be considered an “off-target” effect with respect to the primary goal of HIV activation for shock and kill strategies to facilitate a cure (7). A limitation of this study is that the pathological consequences of these off-target effects remain unknown (16), but one concern might be the oncogenic effects of LTR-driven genes modulated by vorinostat. Lamprecht et al. (44) have demonstrated that activation of an LTR of the MaLR family drives the expression of a proto-oncogene (i.e., *CSF1R*) that may lead to the development of Hodgkin lymphoma. However, vorinostat

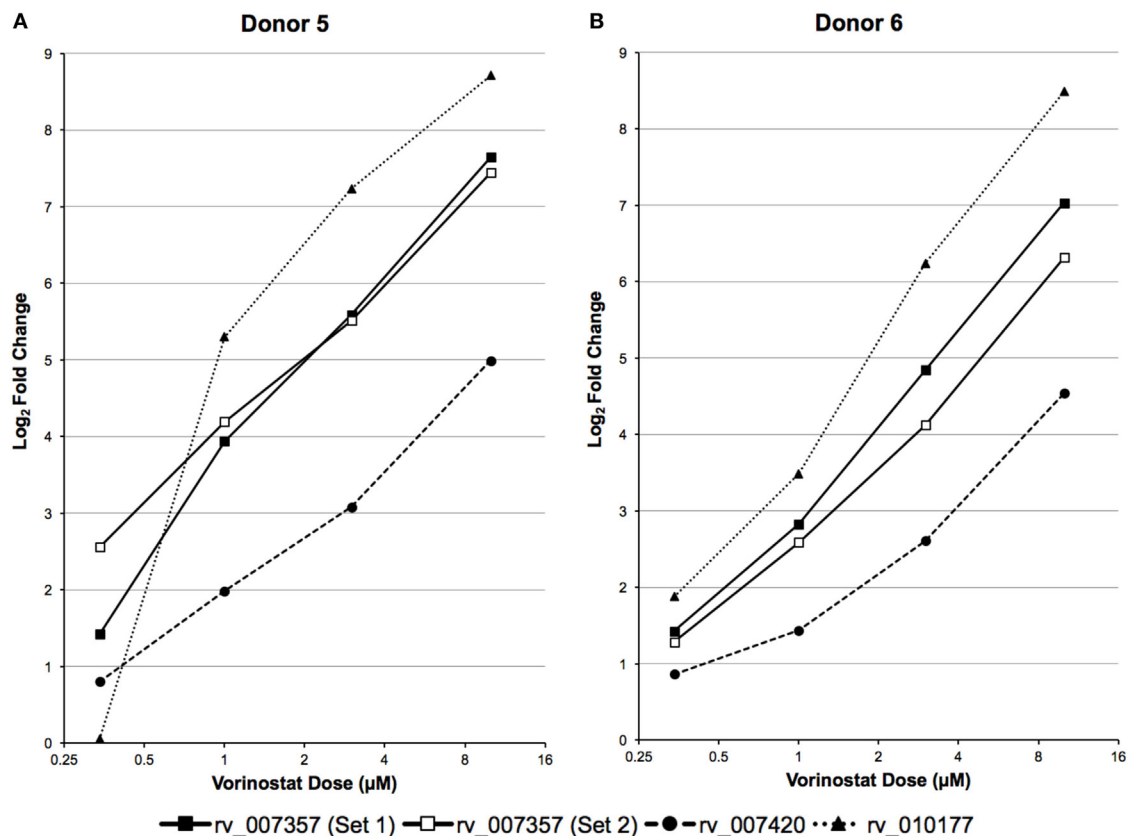


FIGURE 4 | LTR12 elements are upregulated by vorinostat in a dose responsive manner. LTR12 expression was measured by droplet digital PCR using log₂ fold changes between each dose of vorinostat (0.34, 0.1, 0.3, and 10 μM) and the untreated control as calculated for (A) Donor 5 and (B) Donor 6. The labels “Set 1” and “Set 2” indicate the two distinct primer and probe sets used to target the same LTR12 element (i.e., rv_007357). LTR12 elements are labeled with their “rv” designation according to the Human Endogenous Retrovirus Database.

appears to drive the expression of pro-apoptotic genes (i.e., *TP3* and *TNFRSF10B*) in cell lines and thus protect against tumorigenesis (20). An alternate view would be that the HERV elements modulated by vorinostat encode products that might facilitate HIV activation. To explore this more fully, future analyses should better characterize the transcripts expressed from HERV elements upregulated by vorinostat treatment with respect to the relevant RNA products (e.g., messenger RNA, long non-coding RNA or micro RNA) and determine their role, if any, in HIV activation.

Another potential oncogenic concern would be that replication-defective HERV elements could recombine to form replication-competent virus with tumor inducing potential. This phenomenon has been observed *in vivo* in mice (45–48). Specifically, Young et al. (45) demonstrated that a series of recombination events could restore a replication-defective ERV (i.e., *Emv2*) to a replication competent virus in antibody-deficient mice (*Rag1*^{−/−}) that eventually led to thymic and splenic tumors. In humans, the most recently integrated HERV elements belong to the HML-2 family of the HERV-K group, known as HERV-K(HML-2), which have maintained open reading frames encoding functional viral proteins that are expressed but form non-infectious particles (49–52). Dewwanix et al.

(53) demonstrated that the human genome still has the coding potential to resurrect infectious retroviruses from replication defective HERV-K(HML-2) elements. However, this required a three-fragment recombination event *in vitro*, and such a resurrected virus has not been observed to our knowledge *in vivo* in humans. In summary, despite these observations, it is highly unlikely that the HERV elements upregulated by vorinostat in this study, which are predominantly LTR fragments from older HERV families, could recombine to reconstitute the full-length genome required to generate an infectious element.

A final concern is that vorinostat activation of both HIV and HERV elements may lead to recombination and the evolution of novel retroviruses with unknown pathogenicity (19). Acute HIV infection has been shown to lead to the activation of HERV-K(HML-2) elements (22, 54), probably mediated through interactions with the Tat protein. Although subsequent intra-HERV-K recombination has been suggested (55), this remains unconfirmed by other groups (22). Furthermore, to the best of our knowledge, such a recombination event between distantly related retroviruses, such as HERVs and HIV, has never been described. It is unlikely that the HERV elements upregulated in this study by vorinostat exhibit sufficient similarity to facilitate

efficient homologous recombination with the HIV genome (56). For example, the LTR12 elements examined in this study exhibit limited similarity with the HIV LTR from the HXB2 strain (accession number K03455): rv_007357 (38%), rv_007420 (23%), and rv_010177 (10%). A limitation of this study is that HERV element modulation by vorinostat was conducted in the absence of HIV infection. HIV infection was excluded in this pilot investigational study because the virus itself can modulate HERV element expression (21, 22), and this confounder was removed so that the effects of HDAC inhibition on HERV element expression could be unambiguously assessed. Future studies of vorinostat treatment of HIV-infected cells could screen for HIV:HERV recombinants, although these are unlikely to be found for the reasons stated earlier.

Analysis of total RNA-Seq data and validation by ddPCR primarily focused on HERV elements upregulated by vorinostat, i.e., LTR12. A potential limitation is that HERV elements may be embedded or in close proximity to genes that are upregulated, and their signal results from read through transcription. This was not likely for the three LTR12 elements selected for ddPCR confirmation (Figures 3 and 4), since they were at least 5 kbp from protein coding genes in the genome. In addition, a large number of HERV elements were also downregulated (e.g., ERVL, Figure 1). This is reflective of our previous work examining vorinostat-modulated gene expression in which the number of genes downregulated was similar to the number upregulated (2–4). In a similar vein, vorinostat leads to chromatin relaxation and then the milieu of transcription factors present in the nucleus regulates which HERV elements are upregulated and which are downregulated due to the corresponding transcription factor binding sites in these elements. Specifically, we have previously determined that vorinostat upregulated the high mobility group (HMG) AT-hook 1 (*HMGAI*) transcription factor at the transcript, protein, and acetylation level (4).

In summary, the modulation of a large number of HERV elements by vorinostat was demonstrated using an unbiased approach (i.e., RNA-Seq). Evidence for the pathogenic consequences of HERV modulation is not of sufficient strength to limit the use of vorinostat in shock and kill approaches toward an HIV cure. However, HERV elements such as LTR12 could be monitored as off-target biomarkers during shock and kill clinical trials with HDAC inhibitors and trial subjects should be screened to further explore HIV:HERV interactions.

REFERENCES

1. Rasmussen TA, Tolstrup M, Winckelmann A, Ostergaard L, Sogaard OS. Eliminating the latent HIV reservoir by reactivation strategies: advancing to clinical trials. *Hum Vaccin Immunother* (2013) 9(4):790–9. doi:10.4161/hv.23202
2. Beliakova-Bethell N, Zhang JX, Singhania A, Lee V, Terry VH, Richman DD, et al. Suberoylanilide hydroxamic acid induces limited changes in the transcriptome of primary CD4⁺ T cells. *AIDS* (2013) 27(1):29–37. doi:10.1097/QAD.0b013e32835b3e26
3. Reardon B, Beliakova-Bethell N, Spina CA, Singhania A, Margolis DM, Richman DR, et al. Dose-responsive gene expression in suberoylanilide hydroxamic acid-treated resting CD4⁺ T cells. *AIDS* (2015) 29(17):2235–44. doi:10.1097/QAD.0000000000000839

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Institutional Review Board (IRB) at the University of California San Diego (UCSD) with written informed consent from all subjects who donated blood for this study. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the IRB at UCSD.

AUTHOR CONTRIBUTIONS

CWh, NB-B, and CWo contributed to the design of the work. NB-B and SL were responsible for the acquisition of the data which was analyzed by CWh, NB-B, SL, and MB. All the authors contributed to the interpretation of the data. CWh, NB-B, and CWo were primarily responsible for drafting the manuscript with significant inputs from TH and GM with edits and revisions suggested by all other authors.

ACKNOWLEDGMENTS

We acknowledge support from the IRIDIS 4 High Performance Computing Facility and associated support services at the University of Southampton, the Collaboratory of AIDS Researchers for Eradication (CARE), the Bioinformatics Core at the University of Southampton, and the Genomics and Sequencing Core at the UCSD Center for AIDS Research.

FUNDING

This work was performed with the support of the UCSD Center for AIDS Research from the National Institutes of Health (NIH) to DR (P30 AI36214); the U.S. Department of Veterans Affairs grant (K2BX002731) to NBB; the James B. Pendleton Charitable Trust; as well as other NIH grants to DR (the CARE Martin Delaney Collaboratory U19 AI096113) and to CWo (AI104282). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fimmu.2018.00603/full#supplementary-material>.

4. White CH, Johnston HE, Moesker B, Manousopoulou A, Margolis DM, Richman DD, et al. Mixed effects of suberoylanilide hydroxamic acid (SAHA) on the host transcriptome and proteome and their implications for HIV reactivation from latency. *Antiviral Res* (2015) 123:78–85. doi:10.1016/j.antiviral.2015.09.002
5. West AC, Johnstone RW. New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest* (2014) 124(1):30–9. doi:10.1172/JCI69738
6. Chuang DM, Leng Y, Marinova Z, Kim HJ, Chiu CT. Multiple roles of HDAC inhibition in neurodegenerative conditions. *Trends Neurosci* (2009) 32(11):591–601. doi:10.1016/j.tins.2009.06.002
7. Archin NM, Margolis DM. Emerging strategies to deplete the HIV reservoir. *Curr Opin Infect Dis* (2014) 27(1):29–35. doi:10.1097/QCO.0000000000000026
8. Mann BS, Johnson JR, Cohen MH, Justice R, Pazdur R. FDA approval summary: vorinostat for treatment of advanced primary cutaneous T-cell

- lymphoma. *Oncologist* (2007) 12(10):1247–52. doi:10.1634/theoncologist.12-10-1247
9. Archin NM, Liberty AL, Kashuba AD, Choudhary SK, Kuruc JD, Crooks AM, et al. Administration of vorinostat disrupts HIV-1 latency in patients on antiretroviral therapy. *Nature* (2012) 487(7408):482–5. doi:10.1038/nature11286
 10. Elliott JH, Wightman F, Solomon A, Ghneim K, Ahlers J, Cameron MJ, et al. Activation of HIV transcription with short-course vorinostat in HIV-infected patients on suppressive antiretroviral therapy. *PLoS Pathog* (2014) 10(10):e1004473. doi:10.1371/journal.ppat.1004473
 11. Frater J. Thirty years of treating HIV-1 infection: where next? *Trans R Soc Trop Med Hyg* (2015) 109(4):229–30. doi:10.1093/trstmh/trv010
 12. Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet* (2008) 42:709–32. doi:10.1146/annurev.genet.42.110807.091501
 13. Hughes JF, Coffin JM. Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proc Natl Acad Sci U S A* (2004) 101(6):1668–72. doi:10.1073/pnas.0307885100
 14. Hurst T, Pace M, Katzourakis A, Phillips R, Klennerman P, Frater J, et al. Human endogenous retrovirus (HERV) expression is not induced by treatment with the histone deacetylase (HDAC) inhibitors in cellular models of HIV-1 latency. *Retrovirology* (2016) 13:10. doi:10.1186/s12977-016-0242-4
 15. Lock FE, Rebollo R, Miceli-Royer K, Gagnier L, Kuah S, Babaian A, et al. Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A* (2014) 111(34):E3534–43. doi:10.1073/pnas.1405507111
 16. Magiorkinis G, Belshaw R, Katzourakis A. ‘There and back again’: revisiting the pathophysiological roles of human endogenous retroviruses in the post-genomic era. *Philos Trans R Soc Lond B Biol Sci* (2013) 368(1626):20120504. doi:10.1098/rstb.2012.0504
 17. Moyes D, Griffiths DJ, Venables PJ. Insertional polymorphisms: a new lease of life for endogenous retroviruses in human disease. *Trends Genet* (2007) 23(7):326–33. doi:10.1016/j.tig.2007.05.004
 18. An DS, Xie Y, Chen IS. Envelope gene of the human endogenous retrovirus HERV-W encodes a functional retrovirus envelope. *J Virol* (2001) 75(7):3488–9. doi:10.1128/JVI.75.7.3488-3489.2001
 19. Blomberg J, Ushameckis D, Jern P. Evolutionary aspects of human endogenous retroviral sequences (HERVs) and disease. In: Sverdlow E, editor. *Retroviruses and Primate Evolution*. Austin, TX: Landes Biosciences (2005). p. 204–38.
 20. Kronung SK, Beyer U, Chiamonte ML, Dolfini D, Mantovani R, Döbelstein M. LTR12 promoter activation in a broad range of human tumor cells by HDAC inhibition. *Oncotarget* (2016) 7(23):33484–97. doi:10.18632/oncotarget.9255
 21. Gonzalez-Hernandez MJ, Swanson MD, Contreras-Galindo R, Cookinham S, King SR, Noel RJ Jr, et al. Expression of human endogenous retrovirus type K (HML-2) is activated by the Tat protein of HIV-1. *J Virol* (2012) 86(15):7790–805. doi:10.1128/JVI.07215-11
 22. Bhardwaj N, Maldarelli F, Mellors J, Coffin JM. HIV-1 infection leads to increased transcription of human endogenous retrovirus HERV-K (HML-2) proviruses in vivo but not to increased virion production. *J Virol* (2014) 88(19):11108–20. doi:10.1128/JVI.01623-14
 23. Paces J, Pavlicek A, Paces V. HERVd: database of human endogenous retroviruses. *Nucleic Acids Res* (2002) 30(1):205–6. doi:10.1093/nar/30.1.205
 24. Paces J, Pavlicek A, Zika R, Kapitonov VV, Jurka J, Paces V. HERVd: the human endogenous retroviruses database: update. *Nucleic Acids Res* (2004) 32(Database issue):D50. doi:10.1093/nar/gkh075
 25. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* (2009) 10(3):R25. doi:10.1186/gb-2009-10-3-r25
 26. Anders S, Pyl PT, Huber W. HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics* (2015) 31(2):166–9. doi:10.1093/bioinformatics/btu638
 27. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* (2009) 25(9):1105–11. doi:10.1093/bioinformatics/btp120
 28. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res* (2002) 12(6):996–1006. doi:10.1101/gr.229102
 29. Massanella M, Singhania A, Beliakova-Bethell N, Pier R, Lada SM, White CH, et al. Differential gene expression in HIV-infected individuals following ART. *Antiviral Res* (2013) 100(2):420–8. doi:10.1016/j.antiviral.2013.07.017
 30. Beliakova-Bethell N, Massanella M, White C, Lada SM, Du P, Vaida F, et al. The effect of cell subset isolation method on gene expression in leukocytes. *Cytometry A* (2014) 85(1):94–104. doi:10.1002/cyto.a.22352
 31. Owczarzy R, Tataurov AV, Wu Y, Manthey JA, McQuisten KA, Almabrazi HG, et al. IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res* (2008) 36(Web Server issue):W163–9. doi:10.1093/nar/gkn198
 32. de Jonge HJ, Fehrmann RS, de Bont ES, Hofstra RM, Gerbens F, Kamps WA, et al. Evidence based selection of housekeeping genes. *PLoS One* (2007) 2(9):e898. doi:10.1371/journal.pone.0000898
 33. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* (2010) 26(1):139–40. doi:10.1093/bioinformatics/btp616
 34. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B (Methodological)* (1995) 57(1):289–300.
 35. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* (2012) 83(22):8604–10. doi:10.1021/ac202028g
 36. Zain J, O’Connor OA. Targeting histone deacetylases in the treatment of B- and T-cell malignancies. *Invest New Drugs* (2010) 28(Suppl 1):S58–78. doi:10.1007/s10637-010-9591-3
 37. Xargay-Torrent S, Lopez-Guerra M, Saborit-Villarroya I, Rosich L, Campo E, Roue G, et al. Vorinostat-induced apoptosis in mantle cell lymphoma is mediated by acetylation of proapoptotic BH3-only gene promoters. *Clin Cancer Res* (2011) 17(12):3956–68. doi:10.1158/1078-0432.CCR-10-3412
 38. Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat Genet* (2017) 49(7):1052–60. doi:10.1038/ng.3889
 39. Takahashi H, Lassmann T, Murata M, Carninci P. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat Protoc* (2012) 7(3):542–61. doi:10.1038/nprot.2012.005
 40. Lopez-Sanchez P, Costas JC, Naveira HF. Paleogenomic record of the extinction of human endogenous retrovirus ERV9. *J Virol* (2005) 79(11):6997–7004. doi:10.1128/JVI.79.11.6997-7004.2005
 41. Benit L, Lallemand JB, Casella JF, Philippe H, Heidmann T. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J Virol* (1999) 73(4):3301–8.
 42. Santoni FA, Guerra J, Luban J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* (2012) 9:111. doi:10.1186/1742-4690-9-111
 43. Zeng M, Hu Z, Shi X, Li X, Zhan X, Li XD, et al. MAVS, cGAS, and endogenous retroviruses in T-independent B cell responses. *Science* (2014) 346(6216):1486–92. doi:10.1126/science.1250529
 44. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat Med* (2010) 16(5):571–9. doi:10.1038/nm.2129
 45. Young GR, Eksmond U, Salcedo R, Alexopoulou L, Stoye JP, Kassiotis G. Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature* (2012) 491(7426):774–8. doi:10.1038/nature11599
 46. Paprotka T, Delviks-Frankenberry KA, Cingoz O, Martinez A, Kung HJ, Tepper CG, et al. Recombinant origin of the retrovirus XMRV. *Science* (2011) 333(6038):97–101. doi:10.1126/science.1205292
 47. Coffin JM, Stoye JP, Frankel WN. Genetics of endogenous murine leukemia viruses. *Ann N Y Acad Sci* (1989) 567:39–49. doi:10.1111/j.1749-6632.1989.tb16457.x
 48. Stoye JP, Moroni C, Coffin JM. Virological events leading to spontaneous AKR thymomas. *J Virol* (1991) 65(3):1273–85.
 49. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* (2011) 8:90. doi:10.1186/1742-4690-8-90

50. Lower R, Lower J, Frank H, Harzmann R, Kurth R. Human teratocarcinomas cultured in vitro produce unique retrovirus-like viruses. *J Gen Virol* (1984) 65(Pt 5):887–98. doi:10.1099/0022-1317-65-5-887
51. Buscher K, Trefzer U, Hofmann M, Sterry W, Kurth R, Denner J. Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Res* (2005) 65(10):4172–80. doi:10.1158/0008-5472.CAN-04-2983
52. Schmitt K, Reichrath J, Roesch A, Meese E, Mayer J. Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. *Genome Biol Evol* (2013) 5(2):307–28. doi:10.1093/gbe/evt010
53. Dewannieux M, Harper F, Richaud A, Letzelter C, Ribet D, Pierron G, et al. Identification of an infectious progenitor for the multiple-copy HERV-K human endogenous retroelements. *Genome Res* (2006) 16(12):1548–56. doi:10.1101/gr.5565706
54. Contreras-Galindo R, Kaplan MH, He S, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Kappes F, et al. HIV infection reveals widespread expansion of novel centromeric human endogenous retroviruses. *Genome Res* (2013) 23(9):1505–13. doi:10.1101/gr.144303.112
55. Contreras-Galindo R, Kaplan MH, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Ferlenghi I, Giusti F, et al. Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *J Virol* (2012) 86(1):262–76. doi:10.1128/JVI.00602-11
56. An W, Telesnitsky A. Effects of varying sequence similarity on the frequency of repeat deletion during reverse transcription of a human immunodeficiency virus type 1 vector. *J Virol* (2002) 76(15):7897–902. doi:10.1128/JVI.76.15.7897-7902.2002

Conflict of Interest Statement: The senior author, CWo, is currently employed by Merck & Co. who manufactures vorinostat but the research presented in this manuscript was completed prior to this appointment while still a professor at the University of Southampton. The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 White, Beliakova-Bethell, Lada, Breen, Hurst, Spina, Richman, Frater, Magiorkinis and Woelk. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Variable Baseline *Papio cynocephalus* Endogenous Retrovirus (PcEV) Expression Is Upregulated in Acutely SIV-Infected Macaques and Correlated to STAT1 Expression in the Spleen

OPEN ACCESS

Edited by:

Gkikas Magiorkinis,
National and Kapodistrian University
of Athens, Greece

Reviewed by:

Mani Larijani,
Memorial University of Newfoundland,
Canada
Claudia Matteucci,
University of Rome Tor Vergata, Italy

*Correspondence:

Neil Berry
neil.berry@nibsc.org
Robert Belshaw
robert.belshaw@plymouth.ac.uk

† Joint senior authors

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 28 January 2019

Accepted: 08 April 2019

Published: 15 May 2019

Citation:

Maze EA, Ham C, Kelly J, Ussher L,
Almond N, Towers GJ, Berry N and
Belshaw R (2019) Variable Baseline
Papio cynocephalus Endogenous
Retrovirus (PcEV) Expression Is
Upregulated in Acutely SIV-Infected
Macaques and Correlated to STAT1
Expression in the Spleen.
Front. Immunol. 10:901.
doi: 10.3389/fimmu.2019.00901

Emmanuel Atangana Maze^{1,2}, Claire Ham², Jack Kelly¹, Lindsay Ussher¹, Neil Almond²,
Greg J. Towers³, Neil Berry^{2*†} and Robert Belshaw^{1*†}

¹ School of Biomedical Sciences, Faculty of Medicine and Dentistry, University of Plymouth, Plymouth, United Kingdom,

² Division of Infectious Disease Diagnostics, National Institute of Standards and Control (NIBSC), Potters Bar,
United Kingdom, ³ Division of Infection and Immunity, University College London, London, United Kingdom

Retroviral replication leaves a DNA copy in the host cell chromosome, which over millions of years of infection of germline cells has led to 5% of the human genome sequence being comprised of endogenous retroviruses (ERVs), distributed throughout an estimated 100,000 loci. Over time these loci have accrued mutations such as premature stop codons that prevent continued replication. However, many loci remain both transcriptionally and translationally active and ERVs have been implicated in interacting with the host immune system. Using archived plasma and tissue samples from past macaque studies, experimentally infected with simian immunodeficiency virus (SIV), the expression of one macaque ERV in response to acute viral infection was explored together with a measure of the innate immune response. Specifically, RNA levels were determined for (a) *Papio cynocephalus* Endogenous Retrovirus (PcEV), an ERV (b) STAT1, a key gene in the interferon signaling pathway, and (c) SIV, an exogenous pathogen. Bioinformatic analysis of DNA sequences of the PcEV loci within the macaque reference genome revealed the presence of open reading frames (ORFs) consistent with potential protein expression but not ERV replication. Quantitative RT-PCR analysis of DNase-treated RNA extracts from plasma derived from acute SIV-infection detected PcEV RNA at low levels in 7 of 22 macaques. PcEV RNA levels were significantly elevated in PBMC and spleen samples recovered during acute SIV infection, but not in the thymus and lymph nodes. A strong positive correlation was identified between PcEV and STAT1 RNA levels in spleen samples recovered from SIV-positive macaques. One possibility is that SIV infection induces PcEV expression in infected lymphoid tissue that contributes to induction of an antiviral response.

Keywords: endogenous retrovirus, STAT1, macaque, PcEV, SIV, RNA, innate immunity

INTRODUCTION

Endogenous Retroviruses (ERV) are descendants of ancient retroviral infections that have become established in the germline and proliferated to now represent ~5% of the human genome and other mammals, rising to ~8% if the older group of Mammalian Apparent LTR-Retrotransposons (MaLRs) are included. Individual proviruses, termed loci, accumulate mutations over time during their Mendelian transmission that eventually render them replication-defective. In the human reference genome, no locus has retained full-length Open Reading Frames (ORFs) for all genes which precludes the possibility of yielding infectious, cell-free progeny virions (1, 2). However, ERVs and other retroelements have been shown to play a role in epigenetic gene regulation (3–7), with many ERV loci having been co-opted. Examples include HERV-E Long Terminal Repeat (LTR)-driven tissue-specific expression of a human salivary amylase gene (8), ERV-derived syncytins that play a role in placentation (9), and promoter-containing LTRs of ERVs that influence epigenetic gene control and pluripotency (10).

The study of ERVs in non-human primates (NHP) has potential to broaden the wealth of data generated from mouse studies. Smaller mammals tend to have more ERV loci integrating into their genome than larger mammals (11) and, consistent with this general pattern, replication-competent ERV loci have been found in the mouse (12) but not in humans (1, 2, 12–14). In terms of the number of recently integrated ERV loci, the macaque genome is more similar to the human than the mouse (11). Macaques thus represent an opportunity to derive additional relevant data that could augment more conventional mouse studies. To help bridge this gap we have utilized archived materials derived from previous simian immunodeficiency virus (SIV) studies.

Recently, it has been suggested ERVs and other retroelements play a role in innate immune signaling by acting as Pathogen Associated Molecular Patterns (PAMPs) (15–21). Indeed, it has been proposed that retroelement PAMPs, such as cytosolic DNA resulting from reverse transcription, set an activation threshold for triggering the innate immune response (22). These authors speculate that the level at which activation occurs is an evolutionary trade-off between avoidance of tolerance of exogenous pathogens on one hand and constant triggering by endogenous elements on the other. The latter appears to occur in the inherited autoimmune disease Aicardi-Goutieres syndrome where deficiency of TREX1 DNA endonuclease may lead to a build-up of retroelement PAMPs, which triggers inappropriate innate signaling (23). If ERVs were involved in the early immune response, it might be expected that a correlation would exist between ERV expression and measures of innate immune activation in archived macaque samples taken during acute infection.

Our study focuses on *Papio cynocephalus* Endogenous Retrovirus (PcEV), a recently integrated ERV first described in baboons (24). PcEV is one of three ERV lineages copying within the macaque genome during the last 5 million years (25). Since a range of components in the ERV replication cycle can potentially be detected by the innate immune system (26), we argue that the

likelihood of an ERV locus being biologically relevant increases if its sequence is free of inactivating mutations such as premature stop codons. Induction of type-1 interferons (IFN-1), via the JAK/STAT pathway, represents a key part of the innate response to viral infections (27, 28). Interferon-stimulated genes (ISGs) act to limit viral replication and play a role in the control of the viral burden (29). However, the immunomodulatory role of ISG induction in acute-phase and chronic SIV infection is complex and relatively poorly understood, contributing to both dampening and inflammatory immune responses (30, 31). IFN-1 signaling results in the transcription of hundreds of ISGs, an important member being STAT1 (32, 33). Basal IFN-1 levels set the abundance of STAT1, which influences the susceptibility toward infection (34). STAT1 therefore provides a measure of IFN-1 activation induced during acute SIV infection in macaques (35), hence its expression was explored in this study.

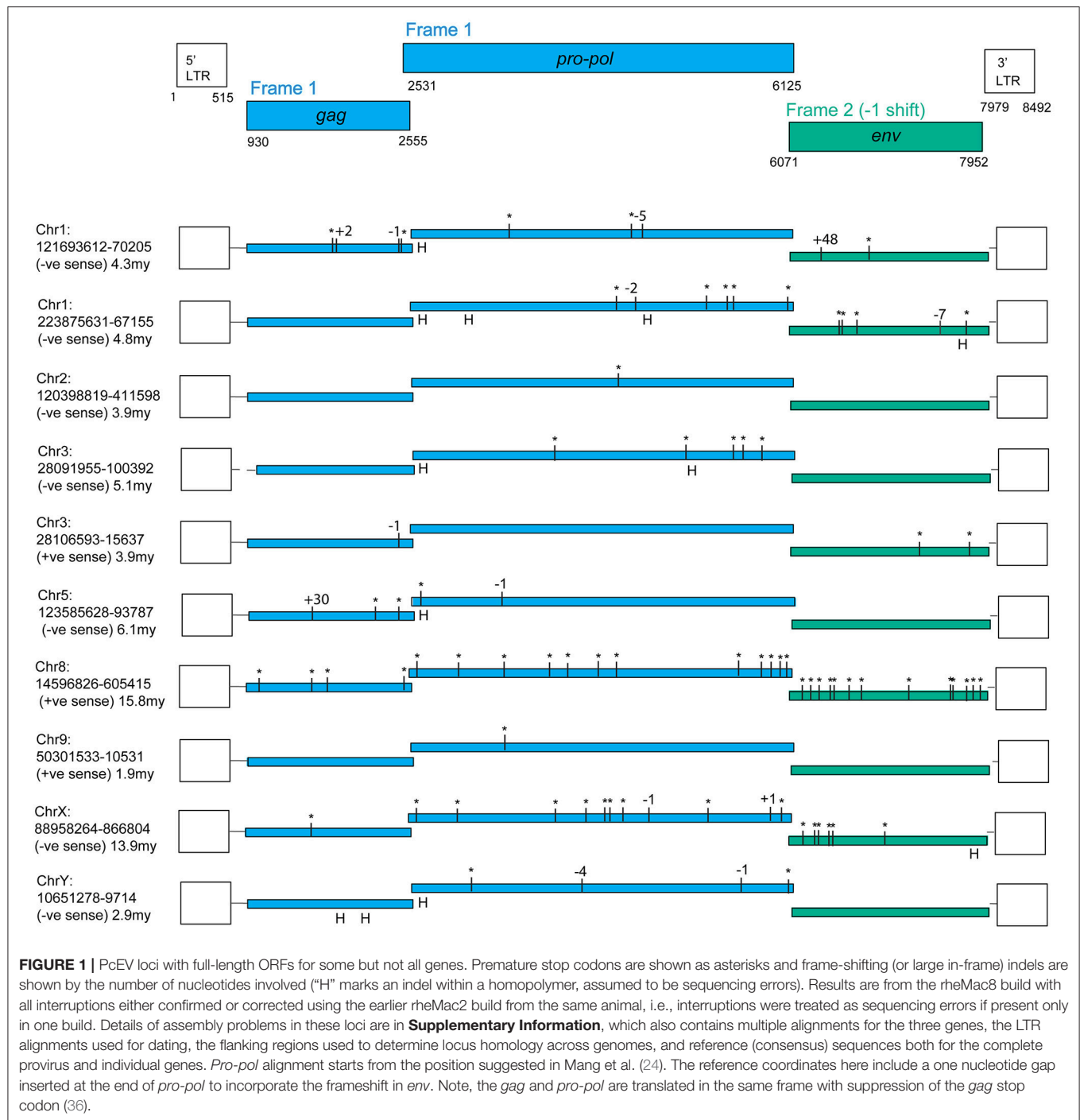
Archived macaque plasma and tissue samples were analyzed to determine PcEV activity during acute SIV infection when *de novo* responses are likely to be highest, particularly lymphoid tissues targeted during primary infection (35). This activity was also compared with STAT1 RNA levels in the same RNA preparations. Our study represents the first analysis of PcEV loci in the macaque reference genome sequences and demonstrates the presence of multiple potentially protein-coding, but not replicating, PcEV loci. In conjunction with identification of low levels of PcEV RNA in plasma of a proportion of acutely SIV-infected macaques, PcEV was actively transcribed in tissues with a level of cell-associated gene expression that is upregulated in response to acute SIV infection. In the spleen, this appeared to be directly correlated to STAT1 RNA expression. The potential significance of these findings linking ERV and innate immunity are discussed.

RESULTS

Bioinformatic Analysis Suggests PcEV Protein Production but Not Replication

An online search of the most recent rhesus macaque (*Macaca mulatta*, RM) genome assembly (rheMac8) using the constructed PcEV reference sequence revealed 72 PcEV loci (plus another 76 matches in unassembled regions), most of which are represented only by fragments. **Figure 1** illustrates the only 10 loci that do not have large regions (more than a few hundred nucleotides) missing from the reference genome. Among these loci are examples of full-length ORFs for all genes but no single locus has full-length ORFs in all genes.

The fragmentation observed is largely the result of genome assembly errors. Four of these 10 most intact loci contain scaffold gaps, where sequenced regions either side have not been joined. One example is locus chr9:50301533-10531, which in the rheMac8 assembly has its *pro-pol* reading frame interrupted by a single premature stop codon and by a long insertion which consists of a tandem duplication at a region marked as a scaffold gap. However, an examination of this locus sequence from an earlier assembly of the RM genome (rheMac2), which was from the same individual animal, and from the homologous



locus in the cynomolgus macaque (*Macaca fascicularis*, CM) genome (macFas5), a species that diverged from the RM ~1–2 million years ago (mya) (37, 38), demonstrates that neither the premature stop codon nor the insertion belong to this locus. These three *pro-pol* sequences are not homologous along their entire length and have a clearly visible transition, a “breakpoint,” near to position 3,937 in the reference genome: the ~40% of the rheMac8 *pro-pol* sequence before this breakpoint is no more

similar to the other two sequences than it is to a range of other PcEV loci, while downstream the sequences of all three loci are very similar.

This can also be demonstrated phylogenetically (**Figure 2**), where a tree of *pro-pol* sequences before the breakpoint shows the rheMac8 sequence is not recovered in the same clade as the rheMac2 and macFas5 sequences, while a tree of the *pro-pol* sequences after this break-point shows all three sequences

recovered together in a well-supported clade (as seen in the other loci). This represents a clear assembly error. Indeed, the locus in rheMac2 and in macFas5 are in the antisense direction while the locus in rheMac8 is in the sense direction. Consistent with widespread assembly problems, only approximately half of the 'intact' loci in rheMac8 are intact in macFas5, and *vice versa*. There is also much more fragmentation in PcEV loci than in the human HERV-K(HML2) loci (39), which belong to an older ERV lineage (40). The human ERVs were only sequenced after being cloned within bacterial artificial chromosomes (41), which avoided the problem with the macaque genomes of trying to assemble simultaneously multiple very similar ERV loci from short next generation sequencing reads.

Evidence was identified of insertionally polymorphic PcEV loci, i.e., loci present in some individuals but not others (42). The intact loci recovered are mostly several millions of years old (Figure 1), the most recently integrated locus being chr9:50301533-10531. The LTRs of this locus differ by two substitutions which, given an estimated rate of nucleotide substitution of $\sim 1 \times 10^{-9}$ /nucleotide/year (25), indicates an age of ~ 1.9 million years. All these loci are also found in the CM reference genome and are therefore expected to be fixed in the RM population. However, two loci with large scaffold gaps were estimated to have integrated within at least the last 700,000 years; both were insertionally polymorphic. Locus chr1:55452680-60247 is heterozygous for presence/absence (=provirus/pre-integration site) in the reference genome and locus chr5:114918771-21894 is represented solely by the pre-integration site in the rheMac3 genome sequence, which is from a different individual (from China rather than India; further details of these loci are included in **Supplementary Information**).

We infer from these observations that PcEV transcription arising from some of these 10 loci may result in PcEV protein production but is unlikely to result in PcEV replication.

PcEV Loci Are Likely to be Transcribed and Lack STAT1 Binding Sites

The LTRs of the above loci, excluding the two very old ones, all have an undisrupted TATA binding protein (TBP) binding site (=TATA box) at position 382 (Figure 3; Table S1). This correlates to the position of the conserved TATA box found in PcEV in the baboon genome (24). Since the TATA box usually has at least minimal promoting activity (43), loci without disruption in the TBP binding site are likely to be transcribed. Another TBP binding site was predicted at position 51 but is unlikely to be real since it is poorly conserved among the loci and not found in other γ -retroviruses, such as HERV-W and MuLV (43, 44). The LTRs also contain many undisrupted transcription factor binding sites. The regions containing these sites correlated with ones already described in the baboon (24), so were named in accordance with that earlier study. Direct-repeat enhancers (DR1, DR2, DR3) contained sites for GR α , GATA1, and GATA2. Three CCAAT box-associated binding regions (CCAAT box 1, 2, and 3) contained sites for GR β , CEBP β , and NFY. A fourth CCAAT box region (box 4) contains these three sites plus the site for CEBP α . Such CCAAT boxes directly

upstream of a TATA box have been shown to efficiently promote transcription activity of the HERV-W LTR (43). Other retroviral LTRs similarly harbor binding sites for many transcription factors (45).

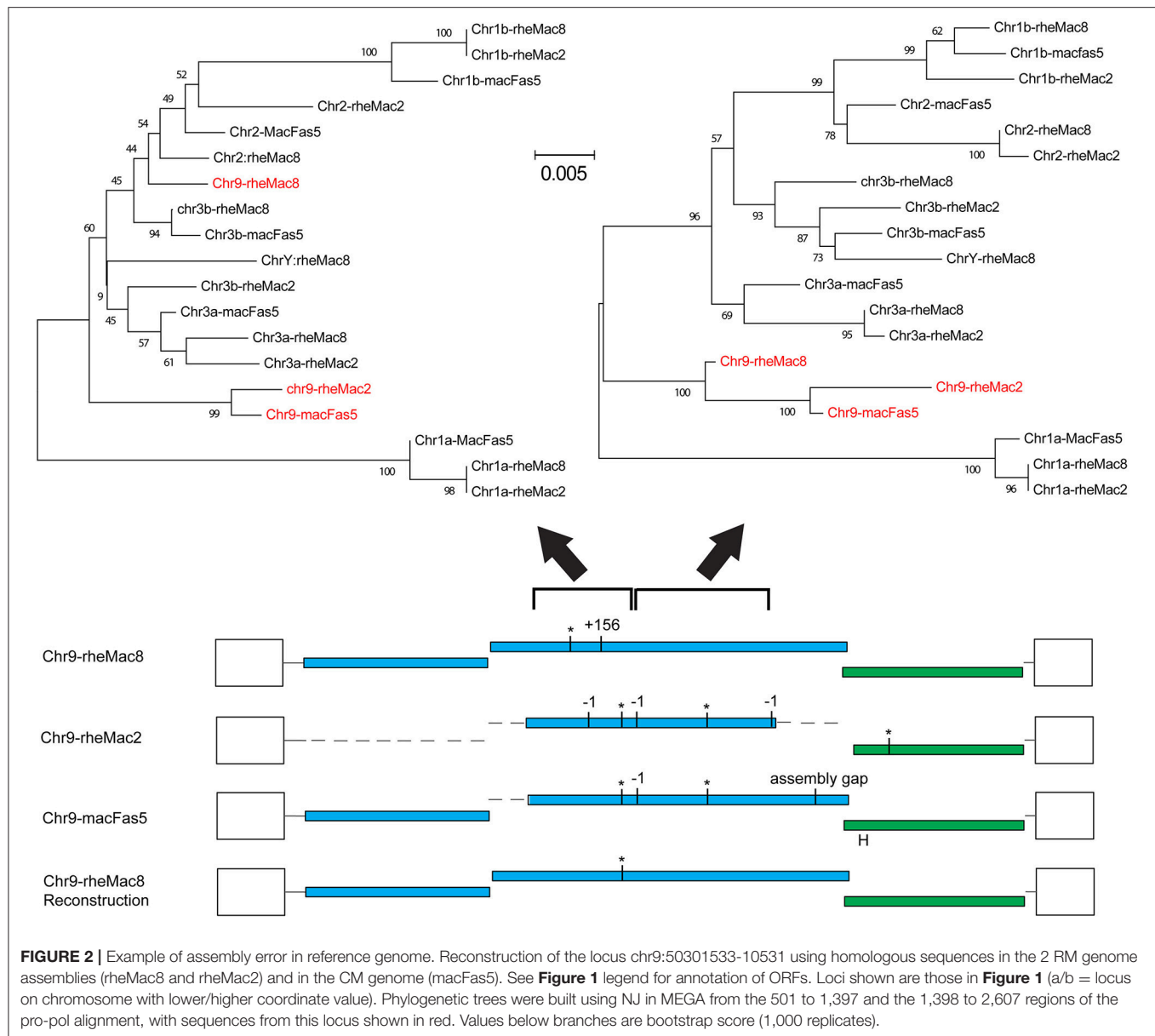
The screening of PcEV loci for putative transcription factor binding sites included key inflammation-related transcription factors: STAT1, NF κ B (46), and STAT3 (47). Predicted binding sites for STAT1 and NF κ B were not identified, so it is therefore unlikely that STAT1 or NF κ B directly regulates PcEV expression through binding to the retroviral promoter. The only inflammation-related transcription factor binding site found was for STAT3 at position 149 (Figure 3).

Low Levels of Plasma PcEV RNA in One-Third of Acutely SIV-infected Macaques

Cell-free PcEV RNA levels were measured by RT-qPCR in plasmas from unchallenged control (naïve) macaques ($n = 5$) or macaques at or around peak SIV viraemia (SIV+) challenged with one of four different SIV strains; SIVmac239 ($n = 5$), SIVmac251 ($n = 5$), SIVmacC8 ($n = 6$), or SIVsmE660 ($n = 6$). SIVmacC8 represents a minimally attenuated *nef*-disrupted variant of wild-type SIVmac251/32H (35), while all others are wild-type SIVs belonging to the SIVmac/SIVsm lineage. All viral challenges were performed in CM, except for SIVmac239 conducted in RM. Three naïve animals were CM, two were RM. As shown in Figure 4, 7 of 22 macaques (32%) exhibited low but detectable levels of PcEV in their plasma (39; 84; 94; 139; 70; 35; 165 PcEV copies/ml). These seven macaques include instances of infection with each of the four SIV strains. PcEV RNA was not detected in the plasma of naïve macaques, although the small sample size does not enable any inference about the relationship between PcEV detection in plasma and SIV status (Fisher's Exact Test comparing presence vs. absence of PcEV in plasma, from naïve and pooled SIV+ macaques, $p=0.27$). Stringent protocols were employed for DNase treatment to remove contaminating genomic DNA, including paired +/- RT reaction steps in each experiment (Figure S1). Neither PcEV nor SIV were amplified in RT minus reactions (Table 1), providing additional confidence that all PcEV molecules originated from an RNA template. Hence, low but detectable PcEV RNA was detectable in around one-third of SIV+ plasmas during acute infection.

Variable Baseline Tissue PcEV Transcriptional Activity Is Upregulated During Acute SIV Infection

Levels of detectable cellular PcEV transcripts in a range of tissues targeted by SIV were determined, isolating RNA from PBMC, spleen, thymus, peripheral and mesenteric lymph nodes (PLN/MLN) from naïve and acutely SIV+ macaques. Data generated for cell-associated PcEV transcriptional activity were compared for each preparation, normalized by co-amplification of the constitutively expressed cellular GAPDH gene with PcEV. Baseline PcEV RNA transcriptional activity was identified in the majority of tissues from all naïve macaques (Figure 5),



though many were at low levels and none exceeded 60 PcEV RNA copies/1,000 copies GAPDH. In these naïve tissues, basal PcEV transcriptional activity for MLN, thymus, PBMC, spleen and PLN was determined as 19, 20, 25, 33, and 52 PcEV RNA copies/ml, respectively. Differences between these tissue means were not statistically significant (Kruskal-Wallis test, $p = 0.07$).

Mean PcEV levels were higher in tissues recovered from SIV+ macaques during acute infection (10–14 days post-infection) compared to naïve macaques (**Figure 5**). PcEV RNA was significantly increased in PBMC (Mann-Whitney test, $p = 0.01$) and thymus ($p = 0.03$), but increases were not significantly different in spleen ($p = 0.23$), MLN ($p = 0.38$), and PLN ($p = 0.20$). Differences between tissue means were not statistically significant (Kruskal-Wallis test, $p = 0.83$). Taken

together, the data indicate that SIV infection *per se* upregulates the expression of PcEV RNA from a variable baseline level of transcriptional activity in a range of tissues targeted by SIV during acute infection.

Cellular STAT1 RNA Levels Are Elevated During Acute SIV Infection

STAT1 represents an independent variable in these studies and its relationship with SIV status and tissue type was determined in the same manner as for PcEV (**Figure 6**). As with PcEV, a global STAT1 induction was observed during acute SIV infection compared to naïve macaques. Considering each tissue separately, STAT1 was significantly increased in PBMC (Mann-Whitney test, $p < 0.001$), thymus ($p < 0.001$), and MLN ($p = 0.03$), but not spleen ($p = 0.15$) or PLN ($p = 0.60$). Hence, during acute SIV

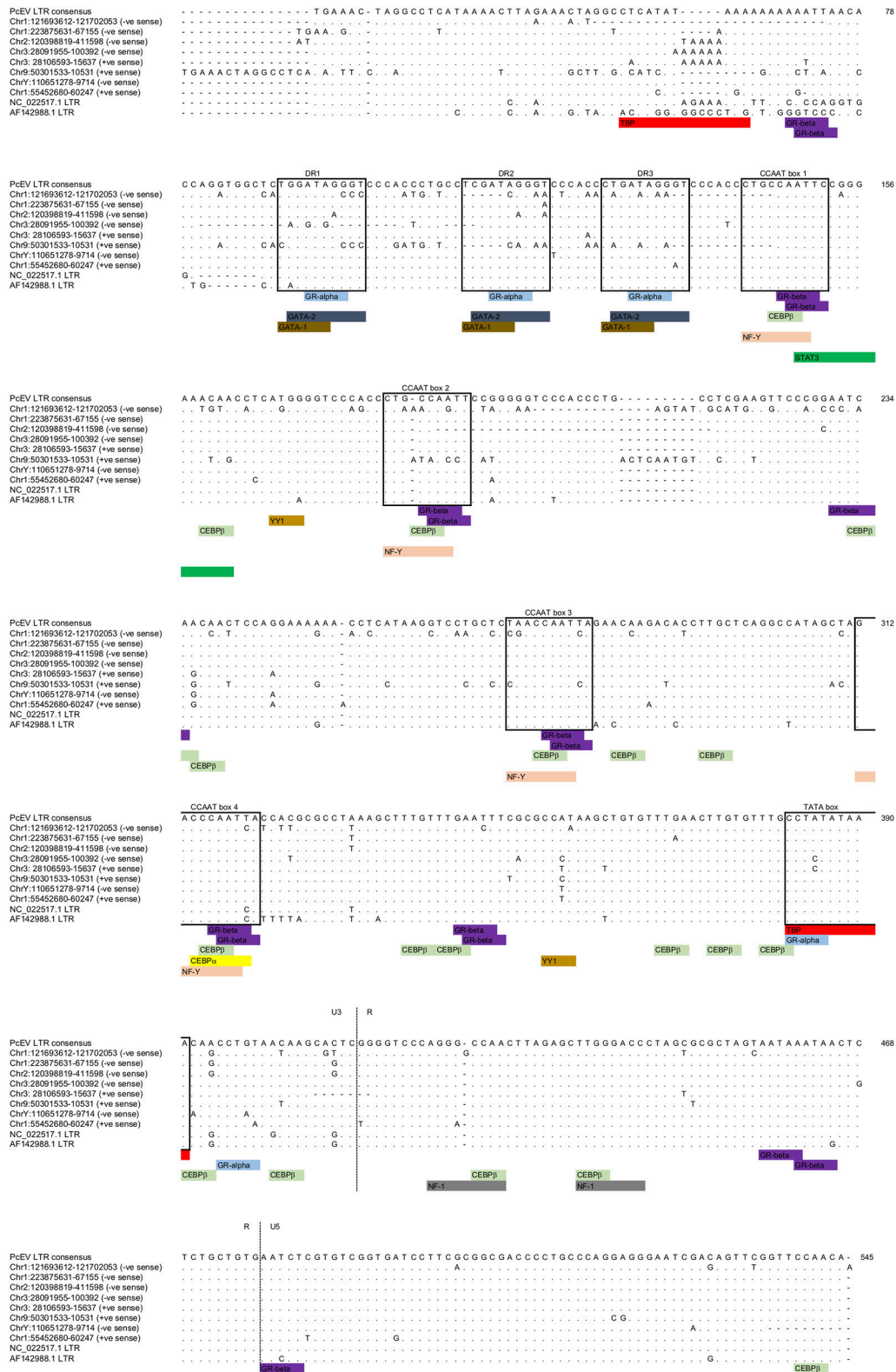
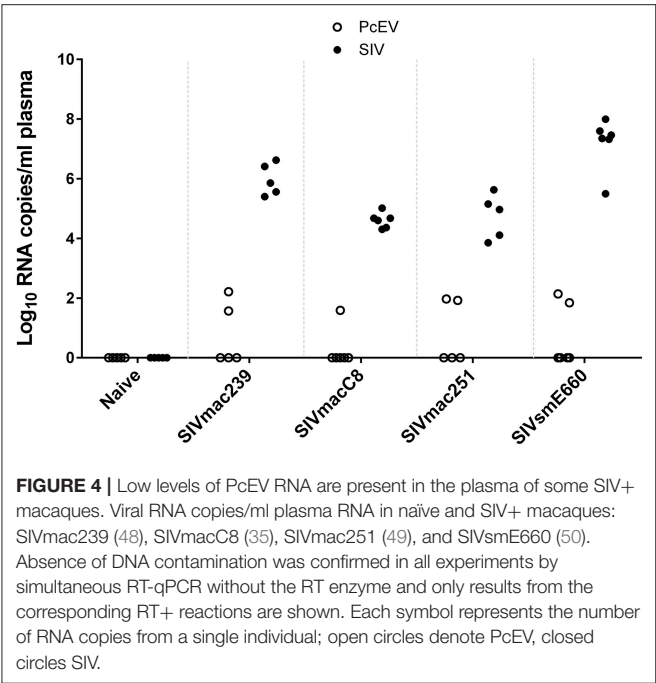


FIGURE 3 | PcEV LTR TATA box and transcription factor binding sites. Prediction of transcription factor binding sites and TATA box in the eight PcEV loci described in the macaque genome in this study. LTR sequences from PcEV loci in the baboon (accessions NC_022517 and AF142988) are also shown.



infection cellular STAT1 levels are elevated, though at variable levels for individual tissues, as SIV infection stimulates a robust immune response.

Cell-Associated PcEV Transcription Levels Are Correlated to STAT1 in SIV+ Spleens

Finally, expression levels of the two genes were compared in the same RNA preparations from each tissue sample. **Figure 7** shows tissue-specific patterns identifying the relationship between localized PcEV and STAT1 expression. Cellular PcEV and STAT1 RNA levels were strongly correlated in the spleen of SIV+ macaques (Spearman test, $r = 0.90$, $p = 0.005$) and showed a positive though statistically non-significant relationship in spleens derived from naïve macaques ($r = 0.68$, $p = 0.11$). Two SIV+ macaques (H19 and H20) with elevated splenic PcEV levels were PcEV+ in plasma, but the macaque with the highest level (G24) was PcEV RNA- in plasma. Four PBMC samples from macaques that were PcEV+ in plasma (including H19 and H20) exhibited a wide range of PcEV transcription levels.

STAT1 and PcEV levels were not significantly correlated in the other tissues but most relationships were positive. Pooling all tissue data from the 22 SIV+ macaques (a total of 30 samples) yielded a strong positive correlation (Spearman test, $r = 0.47$; $p = 0.0091$). Pooled PcEV and STAT1 levels in the 13 naïve macaques were also strongly correlated ($r = 0.61$; $p = 0.0004$), albeit at a lower magnitude of induction for both PcEV and STAT1. Such correlations in pooled data might be artifactual but it is interesting that no such correlation was found between STAT1 and cellular SIV levels in these total RNA preparations ($r = 0.29$, $p = 0.21$); there was also no correlation between STAT1 and plasma SIV RNA levels ($r = 0.25$, $p = 0.22$).

TABLE 1 | PcEV and SIV RNA levels in plasma as determined by qPCR with (+) and without (–) reverse transcriptase (RT).

SIV strain	Individual	PcEV RNA copies/ml		SIV RNA copies/ml	
		RT+	RT-	RT+	RT-
Naïve ^a	E79	No Ct	No Ct	No Ct	No Ct
	E80	No Ct	No Ct	No Ct	No Ct
	E81	No Ct	No Ct	No Ct	No Ct
SIVmac239 ^a	E79	No Ct	No Ct	2.60×10^6	No Ct
	E80	No Ct	No Ct	7.24×10^5	No Ct
	E81	No Ct	No Ct	4.23×10^6	No Ct
SIVmacC8 ^a	E5	No Ct	No Ct	1.04×10^5	No Ct
	E6	No Ct	No Ct	2.03×10^4	No Ct
	G3	39	No Ct	4.01×10^4	No Ct
	G4	No Ct	No Ct	4.75×10^4	No Ct
	G5	No Ct	No Ct	5.14×10^3	No Ct
	G6	No Ct	No Ct	2.31×10^4	No Ct
SIVmac251 ^a	J41	No Ct	No Ct	1.42×10^5	No Ct
	J42	No Ct	No Ct	9.35×10^4	No Ct
	J44	84	No Ct	4.29×10^5	No Ct
	J17	No Ct	No Ct	7.22×10^3	No Ct
	J18	94	No Ct	1.29×10^4	No Ct
SIVsmE660 ^a	G25	No Ct	No Ct	2.25×10^7	No Ct
	G26	139	No Ct	3.13×10^5	No Ct
	G27	No Ct	No Ct	2.12×10^7	No Ct
	G28	No Ct	No Ct	3.97×10^7	No Ct
	G7	70	No Ct	9.91×10^7	No Ct
	G8	No Ct	No Ct	2.89×10^7	No Ct
Naïve ^b	H17	No Ct	No Ct	No Ct	No Ct
	H18	No Ct	No Ct	No Ct	No Ct
SIVmac239 ^b	H19	37	No Ct	3.64×10^5	No Ct
	H20	165	No Ct	2.52×10^5	No Ct

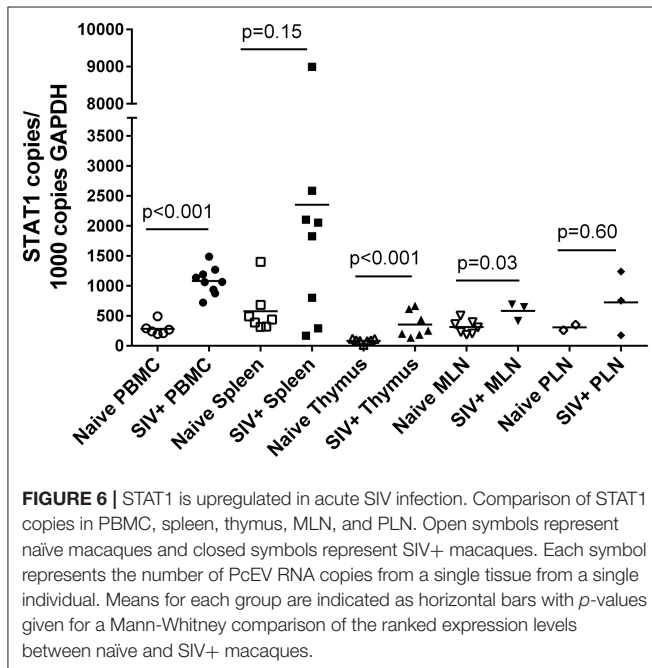
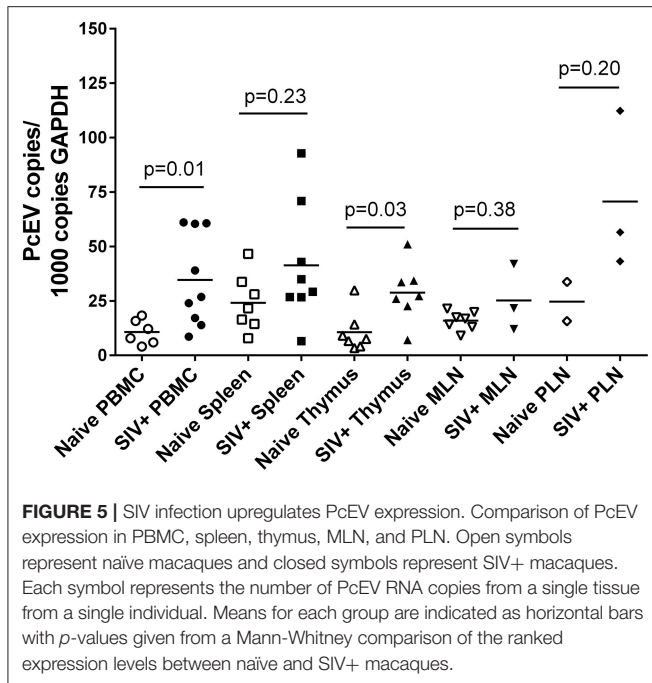
Species of macaque is indicated; CM, cynomolgus macaque; RM, rhesus macaque. PcEV RNA levels are expressed as simple integers; SIV RNA levels use a power of ten notation. Ct, cycle threshold.

^aIn CM.

^bIn RM.

DISCUSSION

The transcriptional activity of an ERV in macaques has been investigated using a combination of bioinformatic prediction analyses and biological measures of *in vivo* activity, exploiting archived biological materials collected from previous macaque challenge studies. PcEV expression was upregulated in several key lymphoid tissues of acutely SIV-infected macaques, when *de novo* immune responses are likely to be highest. In the spleen PcEV expression was positively correlated to cellular STAT1 RNA levels. Analysis of baseline data indicated low, variable cell-associated PcEV RNA levels in the absence of a clear external stimulus, such as SIV. Introduction of SIV induced PcEV in all tissues measured, attaining statistical significance in the thymus and PBMC. While caution needs to be exercised in not over-interpreting our findings given a number of caveats (e.g., inherent variations in SIV strain dynamics, tissues targeted during acute



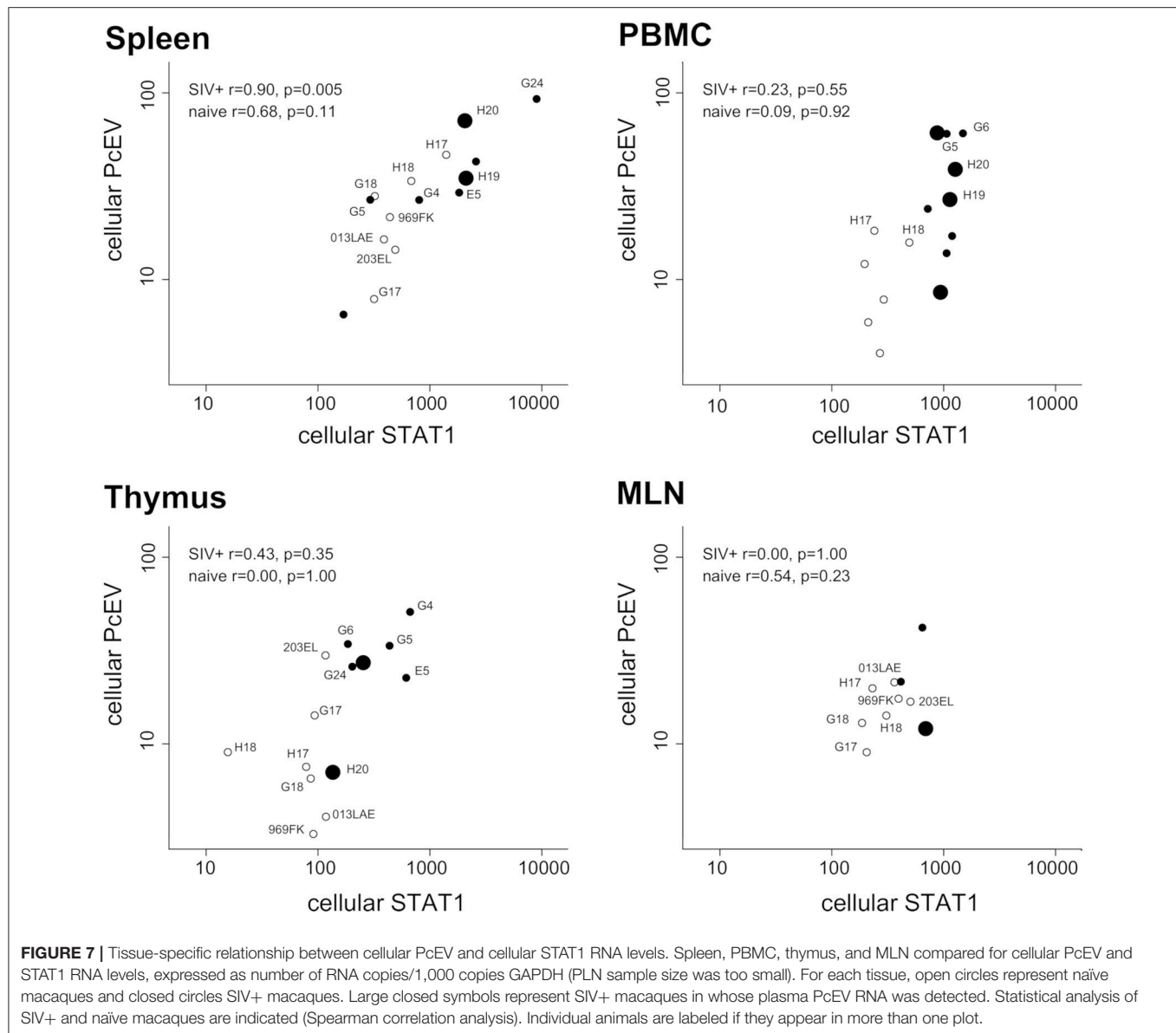
infection, macaque species), these data clearly suggest a global upregulation of PcEV transcriptional activity in response to acute SIV infection.

It is widely recognized that SIV establishes productive infection in multiple lymphoid tissues during the primary viraemic phase. STAT1, part of the IFN-1 cascade, represents one component of the dynamic response to SIV, triggering a range of innate immunity genes during acute infection (35), but this

response fails to clear the virus in the long term. This study confirmed induction of STAT1, part of the IFN-1 cascade in response to SIV, highlighting tissue-specific differences. In the spleen, PcEV and STAT1 RNA levels were positively correlated but this relationship was less clear in PBMC, thymus and MLN and warrants further scrutiny, especially where small sample sizes precluded a comprehensive analysis. With both PcEV and STAT1, exposure to SIV *per se* appears sufficient to trigger a response independent of the level of viral replication.

Although little is known about the macaque, in the mouse model it is well-established that microbial triggering of innate and adaptive sensors leads to increased ERV transcription (51). Treatment with both lipopolysaccharide (LPS), a Toll-like receptor 4 (TLR4) agonist, and polyinosinic-polycytidylic acid [poly(I:C)], a TLR3 agonist, significantly induced expression of two mouse ERV proviruses (52). Examples of human ERVs whose expression is enhanced in the presence of IFN or viral infection include HERV-K(HML2), reported to be increased in PBMC from HIV-1-infected individuals (13, 53, 54), and HERV-W was reported to be increased by influenza virus infection *in vitro* (55, 56). The HERV-K(HML2) LTR has two interferon-stimulated response elements (ISREs) and binding sites for inflammatory transcription factors such as NFκB (57, 58). PcEV lacks such motifs and the absence of clear STAT1 binding sites suggests that STAT1 proteins do not directly drive PcEV transcription. However, prediction of transcription binding sites is complex and lowering the dissimilarity margin of transcription binding site prediction from 5% to 15% allocates the STAT3 binding sites in PcEV also to NFκB and STAT1. STAT3 is activated at the protein level by IFN-1 (59) and binding motifs in PcEV LTR for CEPBβ, whose RNA expression is elevated by type-II IFN (60), were identified in the sequence analyses. Alternatively, PcEV might be upregulated by accessory proteins of exogenous retroviruses as HIV-1 Tat increases HERV-K(HML2) transcription (61, 62). Another possibility is inflammation-induced demethylation of LTRs, where the HIV-1 5'LTR undergoes demethylation following administration of TNFα *in vitro* and LPS to transgenic mice harboring latent HIV genomes in their lymphocytes (63, 64).

With regard to how PcEV might impact on innate immune signaling pathways, two main mechanisms can be envisaged (26). The first, via an exogenous route, implies binding of ERV particles or glycoproteins to innate receptors such as TLRs, as demonstrated *in vitro* where recombinant Env glycoprotein from HERV-W triggers secretion of pro-inflammatory cytokines via TLR4 in PBMC (65). The second, via an endogenous route, is that ERV transcription and reverse transcription may generate intermediate reaction products such as dsRNA, possibly by transcription and hybridization of ERV sense and antisense transcripts, and RNA:DNA hybrids generated during reverse transcription detected by TLRs and cytosolic RIG-I-like receptors, such as MDA5 (18–20). Only low PcEV RNA levels were detected in plasma from around one-third of SIV+ macaques, suggesting PcEV particles are unlikely to be produced in large quantities. It seems plausible, however, that PcEV might trigger induction mechanisms through released viral components generated in a transient nature, supported



by the finding of full-length ORFs for the *env* gene in several loci. In humans, ERV-encoded reverse transcriptase has been suggested to reverse transcribe viral RNA intracellularly (66, 67). In autoimmune TREX1-deficient mice, ERV DNA accumulates in the cytoplasm, associated with inflammatory phenotypes (23) which can be treated using RT inhibitors (68). Hence, the correlation between PcEV and STAT1 expression in this *in vivo* model requires further clarification.

One study limitation is that we restricted our analyses to data from only a single time point. This might not reveal any delayed correlations between PcEV, SIV, and STAT1 levels, especially given differences in cellular and anatomical localization of SIV (35). Cell sorting experiments to derive specific cell sub-sets would help address this issue and refine analyses. Differences in viral kinetics and tropism might also play a confounding role,

although all SIV strains studied, even the minimally attenuated SIVmacC8, exhibit robust early replication kinetics *in vivo* as determined by plasma viral RNA levels 10–14 days post-infection.

To assess the relevance of the macaque as a model system, similarities and differences between human and macaque ERVs require consideration. Humans have a single ERV lineage that has been replicating within the last few million years: HERV-K (HML2). Early studies reported HERV-K(HML2) RNA in plasma from HIV patients (53), but more recent studies failed to reproduce this (13, 14). Evidence for low level PcEV RNA was identified in the plasma of a proportion of SIV+ macaques. Detection of retroviral RNA in plasma presumes some form of protein protection, whether as fully or partially packaged particles, so some macaques in the population may harbor a

replication-competent PcEV locus. Although we did not find any potentially replication-competent PcEV in the reference macaque genome sequences, this may reflect a combination of genome assembly problems with the likelihood that such loci would be unfixed in the macaque population, i.e., present only in the genomes of some individual macaques (42). Evidence for at least two unfixed PcEV loci was found. This phenomenon occurs in HERV-K(HML2), where loci with full-length ORFs in all genes are present only in some non-reference genomes (69, 70). Replication-competent PcEV loci could also be generated by recombination: in the mouse model, recombination between replication-defective loci reconstituted replication-competent loci with resulting infectious virions (71). Although macaques might differ from humans in having as yet unrecovered replication-competent PcEV loci, levels of cell-associated HERV-K(HML2) transcription appear similar: ~30 copies/1,000 copies GAPDH in PBMC of non-acute HIV infection cases and ~15 RNA/1,000 copies GAPDH in PBMC from uninfected individuals (13).

Assuming SIV induces PcEV expression *in vitro*, knock-down of PcEV transcription and individual sensors followed by measurement of the IFN-1 response would determine whether PcEV is involved in the IFN-1 response. The other two recently active ERV lineages in the macaque, Chimpanzee Endogenous Retrovirus (CERV), and Simian Endogenous type D Retrovirus (SERV) (25), also require study since within both lineages single examples of loci with full-length ORFs in all genes were identified. Another important group of retroelements, Long Interspersed Nuclear Elements (LINEs), also require scrutiny because they, unlike ERVs, are known to be copying in the human population (72) and potentially replication-competent loci have been found in the macaque genome (73). LINEs have an entirely intra-cellular replication cycle that would help in distinguishing between endogenous and exogenous routes of innate triggering. Considering several ERVs and LINEs together would provide a wider perspective on whether retroelements are involved in fundamental immune response induction.

In conclusion, the current study demonstrates that PcEV retains transcriptional activity despite the apparent lack of replication. Expression appears to be upregulated by acute SIV infection and in some tissues a correlation exists between the expression levels of PcEV and STAT1, the latter being part of the IFN-1 cascade. The mechanistic processes at play warrant detailed investigation to determine whether ERVs are overlooked components of innate immune responses and activation signaling mechanisms.

MATERIALS AND METHODS

Extraction and Alignment of PcEV Locus Sequences

Bioinformatic analyses were performed using UCSC's Genome Browser website and its BLAT tool (74) on the most recent reference genome sequence of the RM, rheMac8 (Mmul_8.0.1). This assembly derives from short-read Illumina whole genome sequencing (MacaM) (75)

with later closing of some gaps using long-read PacBio sequencing of the same individual animal (comment in GenBank accession GCA_000772875.3).

Sequences of loci were compared with those from an earlier build of the RM genome, rheMac2 (76) (GenBank accession GCA_000002255.2). This build is an earlier short-read whole genome sequencing project largely from the same (female) individual (ID 17573), with some sequencing of an unrelated male used in finishing and to provide a Y chromosome. Sequences were also compared to those of a CM, a closely related species sharing a common ancestor only 0.91 (± 0.11) million years ago, which is more recent than the common ancestor of all rhesus macaques (38). The reference genome sequence of the CM, macFas5, appears to have been built independently of the RM genome using a range of short-read technologies (comment in GenBank accession GCA_000364345.1). The other RM genome assembly available on the UCSC website, rheMac3, is very fragmentary.

PcEV was described initially from the baboon (24) (GenBank reference AF142988), erroneously referred to as BaEV in an earlier analysis of the macaque genome (25) (see **Supplementary Information**). This GenBank sequence was used to locate and download the sequences of several PcEV loci in the macaque using the UCSC Genome Browser. A multiple alignment was then made of the downloaded sequences using MEGA (77, 78) and a new consensus reference sequence built with full-length ORFs in all genes (sequences from different loci differed only by a few percent so alignments were unambiguous). This reference was used to re-search the macaque genome. Multiple alignments were made manually from downloaded sequences and individual loci examined visually for full-length ORFs and other motifs.

Previously published data on recently active macaque ERVs lineages (11, 25) are presented here in more detail (**Figures S2–S4**).

Prediction of Transcription Factor Motifs

A consensus PcEV LTR sequence was built based on an alignment of LTR sequences from the loci displayed in **Figure 1**, excluding the two much older loci, plus NC_022517 (36) and AF142988 (24). This sequence was screened with the ALGGEN-PROMO online tool (79, 80) for the presence of putative binding sites for a range of transcription factors linked to either inflammation or retroviral infection (43, 44). The dissimilarity between query and binding site matrices from the database was fixed at a threshold of 5%. The following transcription factors and binding sites were selected for NHPs only, with the accession number in the TRANSFAC public database given in square brackets: AP1 [T00029]; NF1/CTF [T00094]; C/EBP α [T00105]; NFY [T00150]; CREB [T00163]; cRel [T00168]; CTF [T00174]; GATA1 [T00306]; GATA2 [T00308]; GATA3 [T00311]; GR α [T00337]; NF1 [T00539]; NFAT1 [T00550]; C/EBP β [T00581]; C/EBP Δ [T00583]; NFkB [T00590]; NFkB1 [T00593]; RelA [T00594]; POU2F1 [T00641]; Sp1 [T00759]; TBP [T00794]; YY1 [T00915]; STAT1 α , [T01492]; STAT3 [T01493]; STAT1 β [T01573]; NFYA [T01804];

GR β [T01920]; NFAT2 [T01945]; NFAT1 [T01948]; STAT1 [T04759]; GR [T05076].

Dating the Integration of ERV Loci

Integrations were dated using the nucleotide divergence between the two LTRs of the proviruses. These LTRs form the flanks of the provirus (the complete integrated DNA form of a retrovirus) and are identical at the time of integration, subsequently accumulating substitutions at the host background rate. The following equation was used to calculate locus age (81): integration time = sequence distance between 5' and 3' LTR/(rate in 5' LTR + rate in 3' LTR).

Thus, estimated age = (a/b)/ $r \times 2$ where,

a = number of mismatches between the two LTRs,

b = length of the LTR,

r = estimated rate of nucleotide substitution.

Note, because substitutions will have occurred along the two branches leading to the 5' and the 3' LTRs it is possible to alternatively multiply the LTR length by two to reflect the effect of this. The rate of nucleotide substitution was taken as 1.0×10^{-9} substitutions/nucleotide/year based on analyses of primate ERV sequences (25). This value is about half the rate estimated for mammals generally (82, 83) but is close to estimates for the rate of neutral molecular evolution in Old World Monkeys and Apes (38).

For loci with no substitutions within their LTRs, the probability (P_0) of observing this if the locus had integrated at time "age" can be found from the Poisson distribution with $P_0 = e^{-\lambda}$ where λ is the average number of events in the time period. Thus, using the notation $P_0 = e^{-(2 \times b \times r \times \text{age})}$, the probability of having no substitutions in the LTRs is just <0.5 at age 700,000 years.

Study Population and RNA Preparation

Plasma and cells were derived from acutely SIV-infected Indian RM and CM challenged with different SIV isolates in historical studies of SIV pathogenesis and vaccination (35, 48, 50, 84). Total RNA was extracted from 140 μ L to 1 mL plasmas using QiaAmp (Qiagen Ltd) according to the manufacturer's instruction. Plasmas obtained from CM were extracted from a volume of 140 μ L, the four RM plasmas were extracted from 1 mL plasma. Each RNA preparation underwent a DNase step to remove contaminant DNA and all qPCR samples were run with an RT minus control, an essential step when working with ERVs. Healthy human plasma contains about 1,000 genome copies/mL, in the form of cell-free DNA (85), with many ERV copies in every cell. By contrast, exogenous retroviruses have only a single DNA copy in a small minority of PBMC (86), so the danger of DNA contamination is greatly reduced. For quantification of cell-associated transcripts, RNA was extracted from frozen cells, washed once in PBS and lysed with guanidine isothiocyanate (Sigma Ltd). Chloroform (200 μ L) was added followed by centrifugation at 13,000 rpm for 2 min, aqueous phase collected, 1 volume 100% ethanol added and samples loaded onto an RNeasy silica column (Qiagen Ltd). Columns were washed and RNA eluted according to the manufacturer's instruction to a final volume of 50 μ L and quantitated using

nanodrop. Each RNA preparation was diluted to 10 ng/ μ L, and 50 ng (5 μ L) added to each qPCR reaction.

Primers and Probe Sequences

Primer sequences for qPCR are shown below in 5'-3' orientation. PcEV primers were based on an alignment of PcEV *pro-pol* sequences (Figure S5) designed to anneal to all loci. SIV primers are based on conserved regions in *gag* (49) except for SIVsmE660 based on conserved LTR sequences (50). PcEV, GAPDH, and STAT1 sequences were as follows:

PcEV: CCGTGTCTATCAAGCAATATCC (forward), GGCAGAAGAGGAGTGTCCAGG (reverse) and AACTCGGAGTGTTGCGAC (probe).

GAPDH: GGCTGAGAACGGGAAGCTC (forward), AGG GATCTCGCTCCTGGAA (reverse), and TCATCAATGGAA GCCCATCACCA (probe).

STAT1: CAATACCTCGCACAGTGGTTAGAAAA (forward) and CGGATGGTGGCAAATGAAAC (reverse).

SIVsmE660: CTCCACGCTTGCTTGCTTAA (forward), AGGGTCCTAACAGACCAGGG (reverse), TCCCATATCTCT CCTAGYCGCCGC (probe).

Other SIVs: AGTGCCAACAGGCTCAGAAAA (forward), TGCGTGAATGCACCAGATG (reverse) and TTA AAAAGC CTTTATAATACTGTCTGCG (probe).

Plasmids and *in vitro* Transcription

pBluescript II SK+ (pBS, 3.0 kb) was from Stratagene Ltd (1 μ g/ μ L). pBS-PcEV was derived separately by inserting a PcEV amplicon into the multiple cloning site of a pBS vector using *Apal* (nt 604) restriction sites (GENEWIZ, Inc). pBS-PcEV was linearized by *HindIII* restriction located 16nt downstream of the PcEV sequence. Reaction mixes included 1 μ g of pBS-PcEV plasmid (0.2 μ g/ μ L), 2 U/ μ L final concentration of enzyme and 15 μ L of water/enzyme buffer. Linearization was at 37°C for 1 h. *In vitro* transcription and purification were carried using MAXIscript kit (Ambion Ltd) following manufacturer's instructions.

Standard Curves

PcEV transcripts were DNase-treated and 10-fold serial dilutions made. PcEV copy number was determined by Poisson end-point calculations, with Ct values obtained for each plasma sample compared to the standard curve to estimate number of PcEV copies per reaction (for 5 μ L total RNA input). The number of PcEV copies/mL for each plasma sample was determined taking into account extraction starting volume. SIV standards used were derived from a SIVmac251 plasma RNA reference series (49, 87). For intracellular RT-PCR assessments, standards for GAPDH and STAT1 were as previously described (35).

DNase Treatment of RNA Preparations

Removal of DNA contaminant was conducted using DNA-free kit (Ambion Ltd). 5 μ L of 10X DNase buffer was added to 50 μ L RNA preparation. 2 μ L TURBO DNase was added and incubated for 1 h at 37°C. 10 μ L Inactivation Reagents were added and incubated for 5 min with intermittent flicking of tubes. Inactivation reagents were pelleted by centrifugation at

10,000 g for 2 min. Supernatant containing RNA was harvested, assayed or stored at -80°C for further analysis. In the absence of DNAase treatment, similar copy numbers between RT+/RT- reactions were noted: H17(RT+ = $1.18\text{E}+04$; RT- = $1.27\text{E}+04$); H18(RT+ = $7.06\text{E}+03$; RT- = $9.12\text{E}+03$); H19(RT+ = $2.39\text{E}+04$; RT- = $2.37\text{E}+04$); H20(RT+ = $6.96\text{E}+03$; RT- = $8.99\text{E}+03$). Observed levels of DNA contamination are consistent with theoretical expectations: assuming macaques have similar levels of cell-free DNA in plasma, given at least 50 intact PcEV copies exist in each macaque cell, with an expectation of at least 5×10^4 DNA copies in plasma.

qPCR and qRT-PCR

Quantification of PcEV and SIV transcripts in plasma were determined with 5 μL of DNase-treated RNA per qPCR reaction, assayed in triplicate using RNA Ultrasense One-Step Quantitative RT-PCR kit (Invitrogen Ltd). For PcEV, qPCR reactions consisted of 25.9 μL nuclease-free water, 1 μL forward and reverse primers (5 μM each, final concentration: 100 nM), 2.5 μL PcEV-specific Taqman probe (10 μM), 10 μL of 5X master mix, 0.1 μL ROX reference dye, 2.5 μL Superscript III RT and Platinum Taq polymerase enzyme mix. Cycling parameters were 48°C for 30 min RT step, 95°C for 5 min, 95°C for 15 s (40 cycles); 61°C extension 1 min. For SIV, mix contained 25.9 μL nuclease-free water, 1 μL each forward and reverse primer (5 μM each, final concentration: 100 nM), 2.5 μL SIV-specific probe (10 μM). Cycling for 52°C 60 min RT step, 95°C 10 min, 95°C for 30 s (40 cycles) and 61°C extension for 90 s.

For quantification of cell-associated transcripts, each RNA preparation was diluted to 10 ng/ μL , so that 50 ng (5 μL) was added in each qPCR reaction. For PcEV and SIV, qPCR was conducted as stated above. For GAPDH, the mix contained 25.9 μL nuclease-free water, 1 μL each forward and reverse primer (5 μM each, final concentration: 100 nM), 2.5 μL GAPDH-specific probe (10 μM). Cycling parameters were 52°C 60 min RT step, 95°C , 10 min, 40 cycles of 95°C , 30 s (40 cycles) and 61°C extension for 90 s. For STAT1, the Power SYBR Green RNA-to-Ct 1-step kit was used (Applied Biosystems Ltd). Mix consisted of 15.6 μL nuclease-free water, 4 μL each forward and reverse primer (10 μM each, final concentration: 400 nM), 25 μL 2X Master Mix (containing SYBR Green and polymerase) and 0.4 μL RT-enzyme mix. Cycling parameters were 48°C 30 min RT, 95°C for 10 min, 95°C for 15 s (40 cycles) and 60°C extension for 1 min. Melt curve was 95°C for 15 s, 60°C for 15 s and 95°C for 15 s. RNA preparations were confirmed to lack detectable DNA by performing a parallel quantitative PCR lacking reverse transcriptase for PcEV, SIV and GAPDH on all RNA extracts. In the RNA Ultrasense system, enzyme mix was replaced by 0.2 μL Platinum Taq polymerase (10 U/ μL) (Invitrogen Ltd).

REFERENCES

1. Stoye JP. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat Rev Microbiol.* (2012) 10:395–406. doi: 10.1038/nrmicro2783
2. Jern P, Coffin JM. Effects of retroviruses on host genome function. *Annu Rev Genet.* (2008) 42:709–32. doi: 10.1146/annurev.genet.42.110807.091501

Statistical Analysis

GraphPad Prism version 6.01 was used for all statistical analyses. The Mann-Whitney test was performed for comparing non-paired samples in naïve vs. SIV+ groups; the Kruskal-Wallis test was used to compare expression levels between macaques infected with different strains of SIV. Since most datasets had distributions that were significantly different from normal, correlation coefficient r and p -values are presented from a non-parametric Spearman test. Fisher's Exact Test was used to assess whether SIV is associated with PcEV in the plasma of SIV+ animals.

ETHICS STATEMENT

All biological samples used in this study were derived from previously conducted experiments, which received the appropriate ethical approval from the local NIBSC ethical committee. Animal procedures were performed in strict accordance with UK Home Office guidelines under a license granted by the Secretary of State for the Home Office which approved the work described. These previous experiments have been published (35, 48, 81–83, 85) along with full details of experimental design, procedures and ethical approval.

AUTHOR CONTRIBUTIONS

EM, NB, and RB conceived and designed the experiments. EM and CH performed the experiments. RB, JK, and LU performed the bioinformatic analyses. EM, NA, GT, NB, and RB wrote the paper. EM and RB carried out the statistical analyses. All authors approved the final draft of the paper.

FUNDING

The project was supported by NIBSC and a University of Plymouth Ph.D. studentship.

ACKNOWLEDGMENTS

We are grateful to Giada Mattiuzzo (NIBSC, Division of Virology) for providing GAPDH and STAT1 standards. Amelia Ellisson and David Camp helped with bioinformatic analysis of PcEV loci.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.00901/full#supplementary-material>

3. Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature.* (2010) 463:237–40. doi: 10.1038/nature08674
4. Fasching L, Kapopoulou A, Sachdeva R, Petri R, Jönsson ME, Männe C, et al. TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *Cell Rep.* (2015) 10:20–8. doi: 10.1016/j.celrep.2014.12.004

5. Brattås PL, Jönsson ME, Fasching L, Nelander Wahlestedt J, Shahsavani M, Falk R, et al. TRIM28 controls a gene regulatory network based on endogenous retroviruses in human neural progenitor cells. *Cell Rep.* (2017) 18:1–11. doi: 10.1016/j.celrep.2016.12.010
6. Lavie L, Kitova M, Maldener E, Meese E, Mayer J. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K (HML-2). *J Virol.* (2005) 79:876. doi: 10.1128/JVI.79.2.876
7. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature.* (2014) 516:242–5. doi: 10.1038/nature13760
8. Ting CN, Rosenberg MP, Snow CM, Samuelson LC, Meisler MH. Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* (1992) 6:1457–65. doi: 10.1101/gad.6.8.1457
9. Lavalie C, Cornelis G, Dupressoir A, Esnault C, Heidmann O, Vernochet C, et al. Paleovirology of “syncytins,” retroviral env genes exapted for a role in placentation. *Philos Trans R Soc B Biol Sci.* (2013) 368:20120507. doi: 10.1098/rstb.2012.0507
10. Lu X, Sachs F, Ramsay L, Jacques P-É, Göke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol.* (2014) 21:423–5. doi: 10.1038/nsmb.2799
11. Katzourakis A, Magiorkinis G, Lim AG, Gupta S, Belshaw R, Gifford R. Larger mammalian body size leads to lower retroviral activity. *PLoS Pathog.* (2014) 10:e1004214. doi: 10.1371/journal.ppat.1004214
12. Stocking C, Kozak CA. Endogenous retroviruses: murine endogenous retroviruses. *Cell Mol Life Sci.* (2008) 65:3383–98. doi: 10.1007/s00018-008-8497-0
13. Bhardwaj N, Maldarelli F, Mellors J, Coffin JM. HIV-1 infection leads to increased transcription of human endogenous retrovirus HERV-K (HML-2) proviruses *in vivo* but not to increased virion production. *J Virol.* (2014) 88:11108–20. doi: 10.1128/JVI.01623-14
14. Karamitros T, Paraskevis D, Hatzakis A, Psychogiou M, Elefsiniotis I, Hurst T, et al. A contaminant-free assessment of Endogenous Retroviral RNA in human plasma. *Sci Rep.* (2016) 6:1–12. doi: 10.1038/srep33598
15. Bannert N, Hofmann H, Block A, Hohn O. HERVs new role in cancer: from accused perpetrators to cheerful protectors. *Front Microbiol.* (2018) 9:178. doi: 10.3389/fmicb.2018.00178
16. Chuong EB, Elde NC, Feschotte C. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science.* (2016) 351:1083–7. doi: 10.1126/science.aad5497
17. Hung T, Pratt GA, Sundaraman B, Townsend MJ, Chaivorapol C, Bhargale T, et al. The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. *Science.* (2015) 350:455–9. doi: 10.1126/science.aac7442
18. Chiappinelli KB, Strissel PL, Desrichard A, Li H, Henke C, Akman B, et al. Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell.* (2015) 162:974–86. doi: 10.1016/j.cell.2015.07.011
19. Strick R, Strissel PL, Baylin SB, Chiappinelli KB. Unraveling the molecular pathways of DNA-methylation inhibitors: human endogenous retroviruses induce the innate immune response in tumors. *Oncoimmunology.* (2016) 5:e1122160. doi: 10.1080/2162402X.2015.1122160
20. Roulois D, Loo Yau H, Singhania R, Wang Y, Danesh A, Shen SY, et al. DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell.* (2015) 162:961–73. doi: 10.1016/j.cell.2015.07.056
21. Grandi N, Tramontano E. Human endogenous retroviruses are ancient acquired elements still shaping innate immune responses. *Front Immunol.* (2018) 9:2039. doi: 10.3389/fimmu.2018.02039
22. Volkman HE, Stetson DB. The enemy within: Endogenous retroelements and autoimmune disease. *Nat Immunol.* (2014) 15:415–22. doi: 10.1038/ni.2872
23. Stetson DB, Ko JS, Heidmann T, Medzhitov R. Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell.* (2008) 134:587–98. doi: 10.1016/j.cell.2008.06.032
24. Mang R, Goudsmit J, van der Kuyl AC. Novel endogenous type C retrovirus in baboons: complete sequence, providing evidence for baboon endogenous virus gag-pol ancestry. *J Virol.* (1999) 73:7021–6.
25. Magiorkinis G, Blanco-Melo D, Belshaw R. The decline of human endogenous retroviruses: Extinction and survival. *Retrovirology.* (2015) 12:8. doi: 10.1186/s12977-015-0136-x
26. Hurst TP, Magiorkinis G. Activation of the innate immune response by endogenous retroviruses. *J Gen Virol.* (2015) 96:1207–18. doi: 10.1099/jgv.0.000017
27. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. *Nat Rev Immunol.* (2014) 14:36–49. doi: 10.1038/nri3581
28. Platanias LC. Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat Rev Immunol.* (2005) 5:375–86. doi: 10.1038/nri1604
29. Schoggins JW, Wilson SJ, Panis M, Murphy MY, Jones CT, Bieniasz P, et al. A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature.* (2011) 472:481–5. doi: 10.1038/nature09907
30. Sandler NG, Bosinger SE, Estes JD, Zhu RTR, Tharp GK, Boritz E, et al. Type I interferon responses in rhesus macaques prevent SIV infection and slow disease progression. *Nature.* (2014) 511:601–5. doi: 10.1038/nature13554
31. Uday NS, Douek DC. Interferons and HIV Infection: the good, the bad, and the ugly. *Pathog Immun.* (2016) 1:107. doi: 10.20411/pai.v1i1.125
32. Der SD, Zhou A, Williams BRG, Silverman RH. Identification of genes differentially regulated by interferon alpha, beta, or gamma using oligonucleotide arrays. *Proc Natl Acad Sci USA.* (1998) 95:15623–8. doi: 10.1073/pnas.95.26.15623
33. Lanford RE, Guerra B, Lee H, Chavez D, Brasky KM, Bigger CB. Genomic response to interferon- α in chimpanzees: implications of rapid downregulation for hepatitis C kinetics. *Hepatology.* (2006) 43:961–72. doi: 10.1002/hep.21167
34. Gough DJ, Messina NL, Clarke CJP, Johnstone RW, Levy DE. Constitutive type I interferon modulates homeostatic balance through tonic signaling. *Immunity.* (2012) 36:166–74. doi: 10.1016/j.immuni.2012.01.011
35. Ferguson D, Mattiuzzo G, Ham C, Stebbings R, Li B, Rose NJ, et al. Early biodistribution and persistence of a protective live attenuated SIV vaccine elicits localised innate responses in multiple lymphoid tissues. *PLoS ONE.* (2014) 9:e104390. doi: 10.1371/journal.pone.0104390
36. Kato S, Matsuo K, Nishimura N, Takahashi N, Takano T. The entire nucleotide sequence of baboon endogenous virus DNA: a chimeric genome structure of murine type C and simian type D retroviruses. *Japanese J Genet.* (1987) 62:127–37. doi: 10.1266/jjg.62.127
37. Jiang J, Yu J, Li J, Li P, Fan Z, Niu L, et al. Mitochondrial genome and nuclear markers provide new insight into the evolutionary history of macaques. *PLoS ONE.* (2016) 11:e0154665. doi: 10.1371/journal.pone.0154665
38. Osada N, Hashimoto K, Kameoka Y, Hirata M, Tanuma R, Uno Y, et al. Large-scale analysis of Macaca fascicularis transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*. *BMC Genomics.* (2008) 9:90. doi: 10.1186/1471-2164-9-90
39. Subramanian RP, Wildschutte JH, Russo C, Coffin JM. Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology.* (2011) 8:90. doi: 10.1186/1742-4690-8-90
40. Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, et al. Long-term reinfection of the human genome by endogenous retroviruses. *Proc Natl Acad Sci USA.* (2004) 101:4894–9. doi: 10.1073/pnas.0307800101
41. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. *Nature.* (2001) 409:860–921. doi: 10.1038/35057062
42. Marchi E, Kanapin A, Magiorkinis G, Belshaw R. Unfixed endogenous retroviral insertions in the human population. *J Virol.* (2014) 88:9529–37. doi: 10.1128/JVI.00919-14
43. Lee WJ. Analysis of transcriptional regulatory sequences in the human endogenous retrovirus W long terminal repeat. *J Gen Virol.* (2003) 84:2229–35. doi: 10.1099/vir.0.19076-0
44. Speck N, Baltimore D. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol Cell Biol.* (1987) 7:1101–10. doi: 10.1128/MCB.7.3.1101
45. Luciw PA, Leung NJ. Mechanisms of retrovirus replication. In: Levy JA, editor. *The Retroviridae*. Boston, MA: The Viruses. Springer (1992) 159–298. doi: 10.1007/978-1-4615-3372-6_5
46. Uleri E, Mei A, Mameli G, Poddige L, Serra C, Dolei A. HIV Tat acts on endogenous retroviruses of the W family and this occurs via

- Toll-like receptor 4: inference for neuroAIDS. *AIDS*. (2014) 28:2659–70. doi: 10.1097/QAD.0000000000000477
47. Fan Y, Timani K, He J. STAT3 and its phosphorylation are involved in HIV-1 tat-induced transactivation of glial fibrillary acidic protein. *Curr HIV Res*. (2015) 13:55–63. doi: 10.2174/1570162X13666150121115804
 48. Berry N, Manoussaka M, Ham C, Ferguson D, Tudor H, Mattiuzzo G, et al. Role of occult and post-acute phase replication in protective immunity induced with a novel live attenuated SIV vaccine. *PLoS Pathog*. (2016) 12:e1006083. doi: 10.1371/journal.ppat.1006083
 49. Berry N, Stebbings R, Ferguson D, Ham C, Alden J, Brown S, et al. Resistance to superinfection by a vigorously replicating, uncloned stock of simian immunodeficiency virus (SIVmac251) stimulates replication of a live attenuated virus vaccine (SIVmacC8). *J Gen Virol*. (2008) 89:2240–51. doi: 10.1099/vir.0.2008/001693-0
 50. Berry N, Ham C, Mee ET, Rose NJ, Mattiuzzo G, Jenkins A, et al. Early potent protection against heterologous SIVsmE660 challenge following live attenuated SIV vaccination in mauritian cynomolgus macaques. *PLoS ONE*. (2011) 6:e23092. doi: 10.1371/journal.pone.0023092
 51. Kassiotis G, Stoye JP. Immune responses to endogenous retroelements: taking the bad with the good. *Nat Rev Immunol*. (2016) 16:207–19. doi: 10.1038/nri.2016.27
 52. Young GR, Mavrommatis B, Kassiotis G. Microarray analysis reveals global modulation of endogenous retroelement transcription by microbes. *Retrovirology*. (2014) 11:59. doi: 10.1186/1742-4690-11-59
 53. Contreras-Galindo R, Kaplan MH, Contreras-Galindo AC, Gonzalez-Hernandez MJ, Ferlenghi I, Giusti F, et al. Characterization of human endogenous retroviral elements in the blood of HIV-1-infected individuals. *J Virol*. (2012) 86:262–76. doi: 10.1128/JVI.00602-11
 54. Contreras-Galindo R, López P, Vélez R, Yamamura Y. HIV-1 infection increases the expression of human endogenous retroviruses type K (HERV-K) *in vitro*. *AIDS Res Hum Retroviruses*. (2007) 23:116–22. doi: 10.1089/aid.2006.0117
 55. Nellaker C, Yao Y, Jones-Brando L, Mallet F, Yolken RH, Karlsson H. Transactivation of elements in the human endogenous retrovirus W family by viral infection. *Retrovirology*. (2006) 3:44. doi: 10.1186/1742-4690-3-44
 56. Li F, Nellaker C, Sabunciyar S, Yolken RH, Jones-Brando L, Johansson A-S, et al. Transcriptional derepression of the ERVWE1 locus following influenza A virus infection. *J Virol*. (2014) 88:4328–37. doi: 10.1128/JVI.03628-13
 57. Manghera M, Douville RN. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology*. (2013) 10:16. doi: 10.1186/1742-4690-10-16
 58. Manghera M, Ferguson-Parry J, Lin R, Douville RN. NF- κ B and IRF1 induce endogenous retrovirus K expression via interferon-stimulated response elements in its 5' long terminal repeat. *J Virol*. (2016) 90:9338–49. doi: 10.1128/JVI.01503-16
 59. Ho HH, Ivashkin LB. Role of STAT3 in type I interferon responses: Negative regulation of STAT1-dependent inflammatory gene activation. *J Biol Chem*. (2006) 281:14111–8. doi: 10.1074/jbc.M511797200
 60. Roy SK, Wachira SJ, Weihua X, Hu J, Kalvakolanu DV. CCAAT/enhancer-binding protein- β regulates interferon-induced transcription through a novel element. *J Biol Chem*. (2000) 275:12626–32. doi: 10.1074/jbc.275.17.12626
 61. Gonzalez-Hernandez MJ, Swanson MD, Contreras-Galindo R, Cookinham S, King SR, Noel RJ, et al. Expression of human endogenous retrovirus type K (HML-2) is activated by the tat protein of HIV-1. *J Virol*. (2012) 86:7790–805. doi: 10.1128/JVI.07215-11
 62. Gonzalez-Hernandez MJ, Cavalcoli JD, Sartor MA, Contreras-Galindo R, Meng F, Dai M, et al. Regulation of the human endogenous retrovirus K (HML-2) transcriptome by the HIV-1 tat protein. *J Virol*. (2014) 88:8924–35. doi: 10.1128/JVI.00556-14
 63. Ishida T, Hamano A, Koiwa T, Watanabe T. 5' long terminal repeat (LTR)-selective methylation of latently infected HIV-1 provirus that is demethylated by reactivation signals. *Retrovirology*. (2006) 3:69. doi: 10.1186/1742-4690-3-69
 64. Tanaka J, Ishida T, Choi B-I, Yasuda J, Watanabe T, Iwakura Y. Latent HIV-1 reactivation in transgenic mice requires cell cycle -dependent demethylation of CREB/ATF sites in the LTR. *AIDS*. (2003) 17:167–75. doi: 10.1097/e01.aids.0000042971.95433.b0
 65. Rolland A, Jouvin-Marche E, Viret C, Faure M, Perron H, Marche PN. The envelope protein of a human endogenous retrovirus-w family activates innate immunity through CD14/TLR4 and promotes Th1-like responses. *J Immunol*. (2006) 176:7636–44. doi: 10.4049/jimmunol.176.12.7636
 66. Dube D, Contreras-Galindo R, He S, King SR, Gonzalez-Hernandez MJ, Gitlin SD, et al. Genomic flexibility of human endogenous retrovirus type K. *J Virol*. (2014) 88:9673–82. doi: 10.1128/JVI.01147-14
 67. Contreras-Galindo RA, Dube D, Fujinaga K, Kaplan MH, Markovitz DM. Susceptibility of human endogenous retrovirus type-K to reverse transcriptase inhibitors. *J Virol*. (2017) 91:e01309-17. doi: 10.1128/JVI.01309-17
 68. Beck-Engeser GB, Eilat D, Wabl M. An autoimmune disease prevented by anti-retroviral drugs. *Retrovirology*. (2011) 8:91. doi: 10.1186/1742-4690-8-91
 69. Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM. Discovery of unfixed endogenous retrovirus insertions in diverse human populations. *Proc Natl Acad Sci USA*. (2016) 113:E2326–34. doi: 10.1073/pnas.1602336113
 70. Turner G, Barbulescu M, Su M, Jensen-Seaman MI, Kidd KK, Lenz J. Insertional polymorphisms of full-length endogenous retroviruses in humans. *Curr Biol*. (2001) 11:1531–5. doi: 10.1016/S0960-9822(01)00455-9
 71. Young GR, Eksmond U, Salcedo R, Alexopoulou L, Stoye JP, Kassiotis G. Resurrection of endogenous retroviruses in antibody-deficient mice. *Nature*. (2012) 491:774–8. doi: 10.1038/nature11599
 72. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, et al. LINE-1 retrotransposition activity in human genomes. *Cell*. (2010) 141:1159–70. doi: 10.1016/j.cell.2010.05.021
 73. Han K, Konkel MK, Xing J, Wang H, Lee J, Meyer TJ, et al. Mobile DNA in Old World monkeys: a glimpse through the rhesus macaque genome. *Science*. (2007) 316:238–40. doi: 10.1126/science.1139462
 74. Kent WJ. BLAT — The BLAST-like alignment tool. *Genome Res*. (2002) 12:656–64. doi: 10.1101/gr.229202
 75. Zimin AV, Cornish AS, Maudhoo MD, Gibbs RM, Zhang X, Pandey S, et al. A new rhesus macaque assembly and annotation for next-generation sequencing analyses. *Biol Direct*. (2014) 9:20. doi: 10.1186/1745-6150-9-20
 76. Gibbs RA, Rogers J, Katze MG, Bumgarner R, Weinstock GM, Mardis ER, et al. Evolutionary and biomedical insights from the rhesus macaque genome. *Science*. (2007) 316:222–34. doi: 10.1126/science.1139247
 77. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. (2013) 30:2725–9. doi: 10.1093/molbev/mst197
 78. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol*. (2011) 28:2731–9. doi: 10.1093/molbev/msr121
 79. Messeguer X, Escudero R, Farré D, Núñez O, Martínez J, Albà MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. *Bioinformatics*. (2002) 18:333–4. doi: 10.1093/bioinformatics/18.2.333
 80. Farré D, Roset R, Huerta M, Adsua JE, Roselló L, Albà MM, et al. Identification of patterns in biological sequences at the ALGEN server: PROMO and MALGEN. *Nucleic Acids Res*. (2003) 31:3651–3. doi: 10.1093/nar/gkg605
 81. Martins H, Villesen P. Improved integration time estimation of endogenous retroviruses with phylogenetic data. *PLoS ONE*. (2011) 6:e14745. doi: 10.1371/journal.pone.0014745
 82. Subramanian S, Kumar S. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. *Genome Res*. (2003) 13:838–44. doi: 10.1101/gr.1152803
 83. Yi S, Ellsworth DL, Li WH. Slow molecular clocks in old world monkeys, apes, and humans. *Mol Biol Evol*. (2002) 19:2191–8. doi: 10.1093/oxfordjournals.molbev.a004043
 84. Mattiuzzo G, Rose NJ, Almond N, Towers GJ, Berry N. Upregulation of TRIM5 α gene expression after liveattenuated simian immunodeficiency virus vaccination in Mauritian cynomolgus macaques, but TRIM5 α genotype has no impact on virus acquisition or vaccination outcome. *J Gen Virol*. (2013) 94:606–11. doi: 10.1099/vir.0.047795-0

85. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, et al. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Anal Bioanal Chem.* (2014) 406:6499–512. doi: 10.1007/s00216-014-7835-3
86. Parkin NT. Measurement of HIV-1 viral load for drug resistance surveillance using dried blood spots: literature review and modeling of contribution of DNA and RNA. *Aids Rev.* (2014) 16:160–71.
87. Ham C, Srinivasan P, Thorstensson R, Verschoor E, Fagrouche Z, Sernicola L, et al. International multicenter study to assess a panel of reference materials for quantification of Simian immunodeficiency virus RNA in plasma. *J Clin Microbiol.* (2010) 48:2582–5. doi: 10.1128/JCM.00082-10

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Maze, Ham, Kelly, Ussher, Almond, Towers, Berry and Belshaw. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Pockets of HIV Non-infection Within Highly-Infected Risk Networks in Athens, Greece

Leslie D. Williams¹, Evangelia-Georgia Kostaki², Eirini Pavlitina³, Dimitrios Paraskevis², Angelos Hatzakis², John Schneider⁴, Pavlo Smyrnov⁵, Andria Hadjickou⁶, Georgios K. Nikolopoulos⁷, Mina Psychogiou⁸ and Samuel R. Friedman^{1*}

¹ Institute for Infectious Disease Research, National Development and Research Institutes, New York, NY, United States,

² Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece, ³ Transmission Reduction Intervention Project, Athens, Greece, ⁴ Departments of Medicine and Public Health Sciences, University of Chicago Medical Center, Center for AIDS Elimination, Chicago, IL, United States,

⁵ Alliance for Public Health, Kyiv, Ukraine, ⁶ Medical School, University of Cyprus, Nicosia, Cyprus and European University Cyprus, Nicosia, Cyprus, ⁷ Medical School, University of Cyprus, Nicosia, Cyprus, ⁸ First Department of Internal Medicine, Laikon General Hospital, Medical School, National and Kapodistrian University of Athens, Athens, Greece

OPEN ACCESS

Edited by:

Tara Patricia Hurst,
Abcam, United Kingdom

Reviewed by:

Ellsworth Marvin Campbell,
Centers for Disease Control and
Prevention (CDC), United States

Taisuke Izumi,
Henry M. Jackson Foundation,
United States

*Correspondence:

Samuel R. Friedman
friedman@ndri.org

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 06 May 2018

Accepted: 23 July 2018

Published: 24 August 2018

Citation:

Williams LD, Kostaki E-G, Pavlitina E, Paraskevis D, Hatzakis A, Schneider J, Smyrnov P, Hadjickou A, Nikolopoulos GK, Psychogiou M and Friedman SR (2018) Pockets of HIV Non-infection Within Highly-Infected Risk Networks in Athens, Greece. *Front. Microbiol.* 9:1825. doi: 10.3389/fmicb.2018.01825

As part of a network study of HIV infection among people who inject drugs (PWID) and their contacts, we discovered a connected subcomponent of 29 uninfected PWID. In the context of a just-declining large epidemic outbreak, this raised a question: What explains the existence of large pockets of uninfected people? Possible explanations include “firewall effects” (Friedman et al., 2000; Dombrowski et al., 2017) wherein the only HIV+ people that the uninfected take risks with have low viral loads; “bottleneck effects” wherein few network paths into the pocket of non-infection exist; low levels of risk behavior; and an impending outbreak. We considered each of these. Participants provided information on their enhanced sexual and injection networks and assisted us in recruiting network members. The largest connected component had 241 members. Data on risk behaviors in the last 6 months were collected at the individual level. Recent infection was determined by LAg (Sedia™ Biosciences Corporation), data on recent seronegative tests, and viral load. HIV RNA was quantified using Artus HI Virus-1 RG RT-PCR (Qiagen). The 29 members of the connected subcomponent of uninfected participants were connected (network distance = 1) to 17 recently-infected and 24 long-term infected participants. Fourteen (48%) of these 29 uninfected were classified as “extremely high risk” because they self-reported syringe sharing and had at least one injection partner with viral load >100,000 copies/mL who also reported syringe sharing. Seventeen of the 29 uninfected were re-interviewed after 6 months, but none had seroconverted. These findings show the power of network research in discovering infection patterns that standard individual-level studies cannot. Theoretical development and exploratory network research studies may be needed to understand these findings and deepen our understanding of how HIV does and does not spread through communities. Finally, the methods developed here provide practical tools to study “bottleneck” and “firewall” network hypotheses in practice.

Keywords: networks, HIV transmission, non-infection, HIV risk, firewall effects, bottleneck effects

INTRODUCTION

We present a case study of a large sub-network of *non-infection* that we encountered during the Transmission Reduction Intervention Project (TRIP). TRIP traced the injection and sexual networks of recently-infected people in a successful attempt to recruit and intervene with additional recently-infected people to get them into treatment both to protect their health and to reduce their transmitting HIV to others during the early infection period of high viral load. In the course of this project, we discovered a large, connected sub-component of 29 *uninfected people* within a larger network that contained many recently-infected members.

This paper explores how such a large connected “pocket” of the uninfected could exist. It considers three possible explanations for the existence of such a sub-network:

1. The “bottleneck effect” (Klov Dahl, 1985). A network bottleneck can be said to exist if there are very few risk network links between a subnetwork of the uninfected and a subnetwork that contains infected people.
2. The “firewall effect” (Friedman et al., 2000; Dombrowski et al., 2013a,b; Khan et al., 2013). In the firewall effect, long-term infected people with low viral loads who link highly-infectious newly infected people to uninfected people essentially greatly reduce potential transmission between the two groups. In practice, a firewall effect would be observed if the only links uninfected people have with infected people are with those who have low viral loads and have been infected for a year or more.
3. Low levels of risk behavior. Usually, this would be studied by determining if the uninfected participants in a study reported engaging rarely or never in high-risk behavior. In a network study in which their partners are also studied, it is also possible to determine if these partners engage in low levels of risk behavior.

The paper also describes the network location and risk links among members of the sub-network of uninfected participants.

MATERIALS AND METHODS

Research methods have previously been described (Nikolopoulos et al., 2016, 2017), so we do so only briefly here.

Setting: The study took place (6/2013–7/2015) in Athens, Greece, where an HIV outbreak among people who inject drugs (PWID) began in 2011 (Paraskevis et al., 2011, 2013, 2015; Nikolopoulos et al., 2015b).

Laboratory Methods: HIV testing used a microparticle anti-HIV-1/2 EIA (AxSYM HIV-1/2 gO, Abbott) confirmed by Western Blot (MP Diagnostics). All HIV+ participants were tested by Limiting Antigen Avidity Assay (LAG; Sedia™ Biosciences Corporation) (Duong et al., 2012, 2015; Nikolopoulos et al., 2017). This test is based on antibody maturation to categorize HIV infection as “recent” or “long-standing.” An Optical Density (ODn) score of 1.5 was used as a cut-off for recent infection, with a median of three ODn values ≤ 1.5 indicating recent infection. This corresponds to a window period of 130 days (Duong et al., 2015). HIV RNA was quantified

for all HIV-positive samples with Artus HI Virus-1 RG RT-PCR (Qiagen). Antibody-negative samples in social networks were tested for viremia (and thus acute infection) in pools of 10.

Questionnaire

Participants were interviewed using a questionnaire containing items on demographics, sexual and injecting behaviors, drug treatment, and antiretroviral treatment. A main focus of this interview was to collect information on participants’ network members and the venues where they interacted with risk network members to enable network and venue recruitment. Participants were asked to name people they had injected or had sex with in the prior 6 months; people who injected or had sex in their presence in the prior 6 months; and people who injected, used drugs, or had sex with people the participants had injected or sex with. They were also asked about places they usually visit to use drugs, to have sex, or to meet new sex partners. We worked with respondents to make a list of sex or drug injecting venues; and staff visited venues to recruit participants for the appropriate arms of the study.

Recruitment

TRIP used social network tracing and venue recruitment methods to locate those who had recently been infected. These methods have been shown to be able to locate infections downstream, upstream, and sideways across infection chains (Friedman et al., 2014). To be eligible for the study, all participants had to be 18 years or older and able to answer the questionnaire.

Recently-infected participants in TRIP were people who were very likely to have acquired HIV in the past 6 months, and included both original participants who were first enrolled in the project and whose networks were subsequently traced (“seeds”), and their network contacts. Long-term infected individuals were TRIP participants (both seeds and their contacts) who had probably been infected more than 9 months ago. The classification of participants as either recently- or longer-term-infected was based on HIV testing histories, LAG ODn values, and viral load levels.

Seeds

More specifically, seeds were newly-HIV-diagnosed PWID referred to the study by the allied ARISTOTLE project (Sypsa et al., 2014, 2017; Hatzakis et al., 2015) or other testing facilities. Seeds with LAG ODn ≤ 1.5 and no indication of advanced disease were classified as recently infected. Most recently-infected seeds also had documented seroconversion in the prior 6 months. Seeds with LAG ODn > 1.5 , and without documented seroconversion in the last 6 months, were classified as long-term infected. They were matched to recently infected seeds for age (± 5 years), risk group, and gender. Many had tested positive for HIV > 3 months before their participation in TRIP but learned about their infection shortly before their TRIP baseline interview.

Network Tracing

The named network and venue members of recently- and longer-term infected seeds were recruited as follows. We recruited

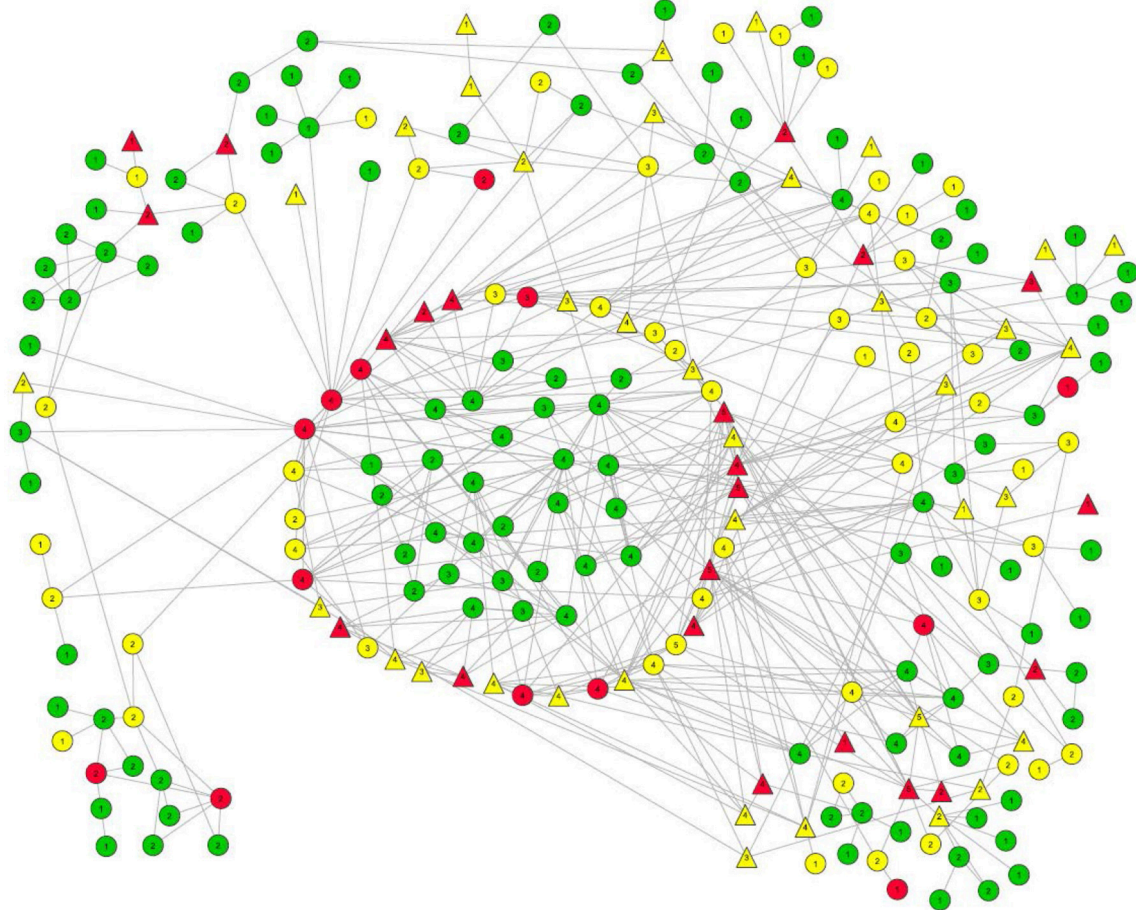


FIGURE 1 | Largest risk network component in the TRIP Athens data. Members of the pocket of non-infection that is the focus of this paper are located in the center. The ring directly around them is comprised of their direct alters. Other members of this large component are located around the perimeter.

HIV infection status:

Long term positives are marked in yellow.

Recently-infected are in red.

Uninfected persons are in green.

Numbers in the node symbols represent what K-level of the Seidman k-core the participant belongs to. Thus, if there is a “3” inside the node symbol, that participant is linked with at least two other members of the 3-core, and each member of that 3-core is linked to at least two other members of that 3-core.

Viral load:

Triangles represent those with viral load >100,000 copies/mL.

Everyone else is represented by a circle.

Link types:

Solid line = Injection Link

Dash line = Sex Link

Small-Dotted Line = Both Sex and Injection Link.

injection and sex partners, and other risk environment contacts, for two steps (i.e., the Step 1 network members recruited directly by the seed, and the Step 2 network members recruited by the Step 1 network members). We tested them for HIV. If they tested positive, we conducted LAg tests and measured HIV viral load. People with recent HIV infection in networks were defined as newly diagnosed individuals with documented testing history of negative serology in the last 6 months and/or LAg ODn ≤ 1.5 , without any indication of advanced disease. Antibody-negative samples were tested for HIV RNA in pools

of 10 to identify acute infections. To maximize the number of potential highly infectious people recruited, we also recruited the network members of people with “borderline-recent infection” found in networks. People with borderline-recent infection were defined as newly diagnosed individuals with LAg ODn ≥ 1.5 but with documented (or reliable, self-reported) history of testing HIV-negative within the last 9 months and/or high viral load (>100,000 copies/ml). For analytic purposes, we included people with borderline-recent infection as part of the recently-infected group in the analyses in this paper.

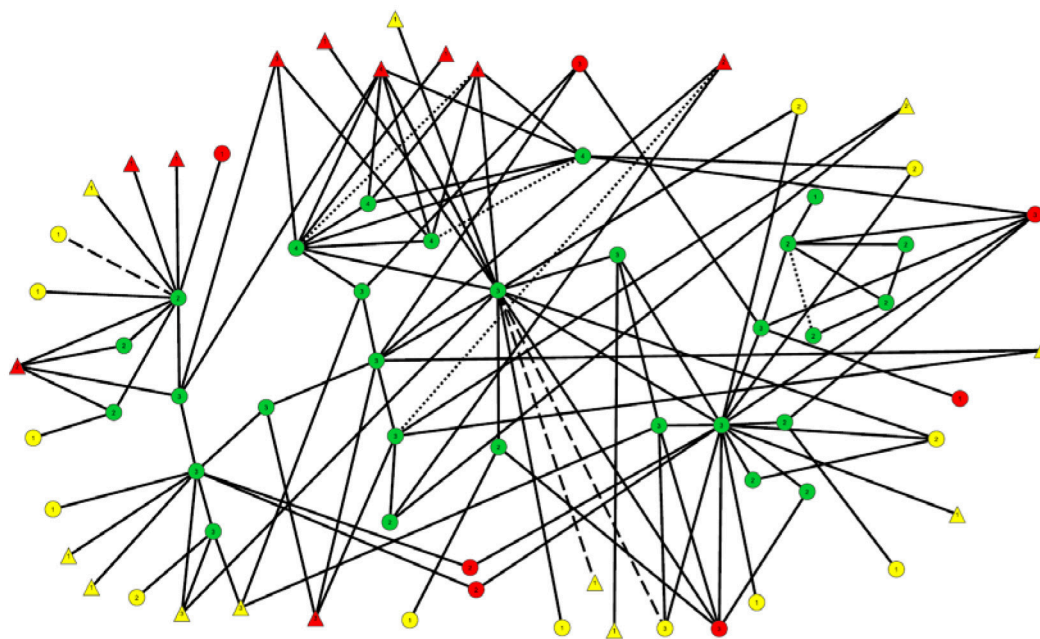


FIGURE 2 | Subnetwork of 29 uninfected participants and the recently- and long-term infected to whom they are linked. The nodes and links in this diagram are the same as those in the center of **Figure 1**, although their locations have been modified to make this diagram easier to interpret.

HIV infection status:

- Long term positives are marked in yellow.
- Recently-infected are in red.
- Uninfected persons are in green.

Viral load:

- Triangles represent those with viral load >100,000 copies/mL.
- Everyone else is represented by a circle.

Link types:

- Solid line = Injection Link
- Dash line = Sex Link
- Small-Dotted Line = Both Sex and Injection Link.

Based on the logic that infection spreads among members of social networks, and that people often find new sexual and injection partners within their social networks, TRIP did not stop when it encountered an uninfected network member but traced the network for at least one additional step (i.e., at least 2 steps from each seed). When recently or borderline-recently infected participants were located in network tracing, their risk and social contacts were recruited for 2 additional steps. For example, if a network member who was 2 steps away from his/her seed was classified as recently infected, we recruited his/her social network members and then the social network partners of those partners.

Incentives

Participants received 10 euros for baseline interviews and 5 euros for each network contact they named who participated in TRIP. As part of HIV testing, we provided participants with standard counseling and appropriate referrals to care. Recently and acutely infected participants received expedited assistance.

Follow-Up

Participants were followed up approximately 6 months later. Those who were uninfected at their first interview were offered the chance to be re-tested for HIV infection (and for recent infection and viral load if infected).

Informed Consent

The project was approved by the Institutional Review Boards of the Hellenic Scientific Society for the Study of AIDS and Sexually Transmitted Diseases in Athens and National Development and Research Institutes (NDRI) in New York. All participants provided written informed consent.

Analyses

Participants recruited in one of four classifications were included in the present analyses: recently-infected seeds, network/venue members of recently-infected seeds; longer-term-infected seeds (LT seeds); and network/venue members of LT seeds. (In these analyses, people with borderline-recent infections were categorized as recently-infected.) Statistical analyses were

TABLE 1 | Numbers and kinds of links of members of the 29 member subcomponent of uninfected participants with each other and with the 17 recently infected participants and long-term infected participants.

	Number of injection-only network connections	Number of sexual-only network connections	Number of network connections where both participants both have sex and inject drugs with each other	Number of venue-based connections
Subcomponent of 29 linked HIV-uninfected	33	0	2	0
Those 17 recently-infected participants with whom a subcomponent member has a direct risk network connection	45	0	2	2
Those 24 long-term-infected participants with whom a subcomponent member has a direct risk network connection	33	3	0	2

For the two venue-based links between uninfected and recently-infected participants, there were no “risky links” since at least one of the dyad members reported not engaging in any sex without a condom and also no syringe sharing. The two venue links in the last row are both cases in which long-term infected participants were recruited from the venues of negatives in the uninfected pocket. In both of these links, each of the dyad members reported syringe sharing, although we do not know if they ever shared with each other.

conducted with SPSS Statistics 21. **Table 2** compares descriptive statistics for members of the subcomponent of negatives with descriptive statistics for the recently-infected and long-term infected participants to whom they are linked. Although not shown in the table, one-way ANOVAs (for continuous variables) and Chi-square tests of independence (for binary variables) were used to compare these groups on all characteristics presented in the table. These tests produce approximate *p*-values that can only be used as heuristic guides because these three subsets of participants were recruited through chain-referral. As such, the sample violates the assumptions of sampling independence that underlie statistical inference.

Network Analyses and visualizations were conducted using Visone 2.16. Calculations of Seidman *k*-core specify subsets of a component whose members are all linked to *k* or more members of that same subset. In any given component, there can be only one 2-core. There can be multiple *k*-cores with *k* > 2; all of their members, by definition, are members of the 2-core. Participants who are not members of a core with *k* > 1 are only weakly tied to the network and thus to patterns of viral transmission. Thus, *k*-core analysis lets us understand how the uninfected component members “fit into” the large connected component, and the extent to which they are linked to denser parts of the network.

RESULTS

Forty-five recently-infected, 105 long-term infected, and 181 uninfected participants were recruited. The largest connected component had 241 members, and is shown in **Figure 1**. Within this large connected component there was a subcomponent (i.e., “pocket”) of 29 connected uninfected PWID (located in the center of **Figure 1**). These 29 participants and the participants with whom they had a direct risk network link are the focus of this paper. (A direct network link usually means that at least one of two participants named the other as a network member during the interview. However, we also considered participants to be directly linked in cases where our field staff saw them together at

injection venues and therefore categorized them as people who probably injected together, even if they did not report this on their questionnaires. Only 4 such links were identified among our 29 negative pocket members and any of their direct network connections.)

All but one of the members of this 29-member subcomponent are members of the Seidman 2-core of the large component, as are all of the infected participants to whom they are directly linked. Indeed, most of the 29 are members of a 3-core as well.

Figure 2 shows the 29 members of the connected subcomponent of uninfected participants and their risk ties to each other and to the 17 recently-infected and 24 long-term infected participants with whom they have direct risk-network connections. **Table 1** shows that the uninfected had many links with each other (35 total links in Row 1) and with members of the recently-infected (47 total links in Row 2) and longer-term infected (36 total links in Row 3) participants, and that almost all of these risk links were injection links rather than sexual links.

Table 2 presents sociodemographic and behavioral characteristics, HIV prevalence rate, and selected other variables for each of these 3 groups and for the total sample. As mentioned above, statistical comparisons only produce approximate *p*-values due to violations of sampling assumptions. Only one comparison was significant at *p* < 0.05: that the long-term infected were more likely to be unemployed.

Twenty-one (72%) of the 29 uninfected “pocket” members were directly linked (network distance = 1) to at least one recently-infected participant, and 16 (55%) to at least one long-term infected participant. We classified 14 out of 29 (48%) uninfected “pocket” members as being at “extremely high risk” because they self-reported syringe sharing and had at least one direct link to at least one injection partner who self-reported sharing syringes and had a viral load > 100,000 copies per mL. These 14 extremely high risk uninfected participants said they used a syringe someone else had already used a mean of approximately 45 times in the last 6 months. Another six of the 29 were linked to someone who shared syringes

TABLE 2 | Sociodemographic and other characteristics of study participants by major analytic category*.

	Total sample	Subcomponent of 29 linked HIV-uninfected	Recently-infected participants with whom a subcomponent member has a direct risk network connection	Those long-term-infected participants with whom a subcomponent member has a direct risk network connection	Total of the 29 member subcomponent of uninfected participants and of the infected participants to whom they are linked
Total	331	29	17	24	70
Males	260 (78.6%)	26 (89.7%)	13 (76.5%)	17 (70.8%)	56 (80.0%)
Median age in years (IQR)	34 (30–40)	35 (30–39)	35 (32.5–49)	31 (28.25–41)	34 (29–42.5)
Education—at least high school (11 years) completed	129 (40.0%)	9 (31.0%)	6 (35.3%)	10 (41.7%)	25 (35.7%)
Homeless	81 (24.5%)	10 (34.5%)	9 (52.9%)	14 (58.3%)	33 (47.1%)
PWID (injecting over the last 6 months)	304 (91.8%)	29 (100%)	17 (100%)	24 (100%)	70 (100%)
Median duration of injection in years (IQR)	13 (7–18)	10 (5–16)	10.5 (5.75–15.5)	9.5 (4.25–15.75)	10 (5–15.5)
On drug/alcohol treatment at enrollment	124 (37.5%)	5 (17.2%)	3 (17.6%)	6 (25.0%)	14 (20.0%)
Unemployed/unable to work	229 (69.2%)	17 (58.6%)	14 (82.4%)	21 (87.5%)	52 (74.3%)
Sex workers	30 (9.1%)	2 (6.9%)	1 (5.9%)	3 (12.5%)	6 (8.6%)
Male sex workers (% of males)	8 (3.1%)	1 (3.8%)	0 (0.0%)	0 (0.0%)	1 (1.8%)
Female sex workers (% of females)	22 (31.0%)	1 (33.3%)	1 (25.0%)	3 (42.9%)	5 (35.7%)
HIV prevalence rate	45.3%	0%	100%	100%	58.6%
Recently HIV infected	45 (13.6%)	0 (0%)	17 (100%)	0 (0%)	17 (24.3%)
Mean number sex partners (S.D.)	4.3 (17.8)	2.0 (2.5)	3.2 (7.1)	11.4 (41.5)	5.5 (24.6)
Reported condomless sex with at least some partners	160 (48.3%)	11 (37.9%)	10 (58.8%)	5 (20.8%)	26 (37.1%)
Mean proportion of partners with whom participants reported always using condoms (S.D.)	0.46 (0.45)	0.61 (0.44)	0.41 (0.45)	0.72 (0.46)	0.58 (0.45)
Mean number injection partners (S.D.)	10.3 (30.2)	10.3 (14.6)	4.7 (3.1)	12.8 (21.2)	9.8 (15.8)
Reported receptive syringe sharing with at least some partners	112/297 (37.8%)	14 (48.3%)	10 (58.8%)	7 (29.2%)	31 (44.3%)
Reporting giving personally used syringes to at least some partners	107/297 (36.0%)	17 (58.6%)	12 (70.6%)	7 (29.2%)	36 (51.4%)
Median reported categorical frequency of injection	Once per day	2–3 times per day	2–3 times per day	2–3 times per day	2–3 times per day
Mean estimated proportion of injection partners with whom participants engaged in receptive syringe sharing	0.07 (0.15)	0.08 (0.17)	0.12 (0.21)	0.05 (0.10)	0.08 (0.16)
Viral load > 100,000 copies/mL	74 (22.4%)	0 (0%)	10 (58.8%)	11 (45.8%)	21 (30.0%)

* Cell entries take the form of n (%).

and had a viral load > 100,000 copies but self-reported that they themselves had not shared syringes in the last 6 months.

Seventeen of the 29 members of the uninfected subcomponent were re-interviewed and had blood taken in a 6-month follow-up. None of these 17 tested positive for HIV. At study intake, 12 (70.6%) of the 17 were in at least one partnership defined as having “extremely high risk.”

DISCUSSION

The research in this paper shows both the power of risk network research and the limitations of current theories about the spread of HIV (and perhaps other agents) through networks and communities. Unlike phylogenetic research or behavioral epidemiology, the network design used in this study can investigate the ties among people who are infected and uninfected, and thus can pose questions about why groups of people who are uninfected remain that way despite having risk network links to people who both have high viral loads and engage in risky behavior.

Of note, neither of the two existing network-level theories can explain why the 29-member subcomponent remains uninfected. These 29 members have many sexual and/or injection ties both to recently-infected and to longer-term-infected participants, which shows that the networks do not create a bottle-neck that is preventing transmission to the 29 member subcomponent. Similarly, the large number of sexual and/or injection ties to participants who have high viral loads and/or are recently-infected shows that something besides the firewall effect is protecting the subcomponent members.

In a study of HIV in New York in the early 1990s (Friedman et al., 1997), we showed (1) that membership in the 2-core of the large component was associated with being HIV-infected and also with higher levels of risk behavior, and (2) that 3-core membership was also associated with additional risk. Thus, it is particularly puzzling to find a large subcomponent of the non-infected with most of its members in the 2-core (and, indeed, many are in a 3-core), and thus not peripheral to the risk network.

Insofar as we can test them, behavioral theories also do not explain why the uninfected subcomponent remains uninfected. Syringe sharing on the part of both the uninfected and their infected injection partners is widespread, and many of these infected partners have high viral loads and thus should be quite infectious. Nonetheless, since we lack relationship-specific data about how often a given participant engaged in a risk behavior with a specific other participant, it remains possible that the uninfected people whom we designated as “extremely high risk” may not have engaged in syringe sharing with their participant partners who had high viral loads and who also engage in syringe sharing (with unknown persons). A related limitation is that four of the 48 links of uninfected participants with recently-infected participants were “venue links,” which means that we cannot be certain that they are directly linked as friends or partners.

One possible explanation for the fact that the “pocket” members remained uninfected is that the epidemic outbreak among Athens PWID is fairly new—it started in 2011 (Paraskevis et al., 2013; Nikolopoulos et al., 2015a; Sypsa et al., 2017), and that therefore, HIV simply had not reached them yet. We cannot rule this out, but the fact that none of the 17 uninfected participants for whom we have follow-up testing data seroconverted by the 6 month follow-up provides a (low statistical power) piece of evidence that suggests that something more is going on here.

Thus, we are left with a conundrum: none of the existing theories can explain our observations. It is, of course, possible that our data are an anomaly, which suggests that replication research is sorely needed. It is also possible that some members of the pocket of non-infection could have a degree of genetic immunity to HIV (Tsiara et al., 2018). On the other hand, these data are sufficiently strong to suggest that the theoretical development of the field is incomplete and that some deep thinking is required. The focus of this deep thinking should go beyond the question of why individuals with high-risk connections are not infected, and should instead consider the question of how such a large, at-risk connected cluster remains uninfected. Relatedly, if this phenomenon turns out to be common, future efforts should seek to understand the contradiction between this phenomenon and the fact that large-scale epidemic outbreaks do happen.

Future replication research should seek to obtain detailed data on the risk and protective behaviors each member of each dyad engages in with each of their specific network members. It should also collect and analyze specimens for measuring possible individual resistance to infection (e.g., via human leukocyte antigen typing).

Future theory and research should not only seek to understand how such a large “pocket” of uninfected network members can remain so, given the observed risks, but should also seek to explore some additional questions posed by the present findings: (1) Given the large number of longer-term infected participants with viral loads >100,000 copies/mL in **Table 2**, as these people with extremely high viral load develop more effective antibody responses and their viral loads decrease, will this establish effective firewalls to reduce further viral transmission? And (2) as a corollary question, in the context of an epidemic among Athens PWID that began in 2011 and had just passed its period of highest incidence at the time TRIP began recruiting, why were these high viral loads so prevalent?

Finally, the straightforward methods used here to study subnetworks of non-infection provide a template for studying “bottleneck” and “firewall” network hypotheses in practice. This template should be useful as additional theories are developed.

AUTHOR CONTRIBUTIONS

All authors contributed to design and data collection, and critiqued and approved the final text. LW had primary responsibility for analyses. LW and SF had primary responsibility for writing the article. SF had primary responsibility for

the project as a whole (all sites), while GN took primary responsibility for the conduct of the research at the Athens site.

ACKNOWLEDGMENTS

This research was supported by US National Institute on Drug Abuse grants DP1 DA034989 Preventing HIV Transmission

by Recently-infected Drug Users (Transmission Reduction Intervention Project) and P30 DA11041 Center for Drug Use and HIV Research. The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

REFERENCES

- Dombrowski, K., Curtis, R., Friedman, and Khan, B. S. (2013a). Topological and historical considerations for infectious disease transmission among injecting drug users in Bushwick, Brooklyn (USA). *World J. AIDS* 3, 1–9. doi: 10.4236/wja.2013.31001
- Dombrowski, K., Khan, B., Habecker, P., Hagan, H., Friedman, S. R., Saad, M., et al. (2017). The interaction 14 of risk network structures and virus natural history in the non-spreading of HIV 15 among People who inject drugs in the early stages of the epidemic. *AIDS Behav.* 21, 1004–1015. doi: 10.1007/s10461-016-1568-6
- Dombrowski, K., Khan, B., McLean, K., Curtis, R., Wendel, T., Misshula, E., et al. (2013b). A re-examination of connectivity trends via exponential random graph modeling in two idu risk networks. *Subst. Use Misuse* 48, 1485–1497. doi: 10.3109/10826084.2013.796987
- Duong, Y. T., Kassanjee, R., Welte, A., Morgan, M., De, A., Dobbs, T., et al. (2015). Recalibration of the Limiting Antigen Avidity EIA to Determine Mean Duration of Recent Infection in Divergent HIV-1 Subtypes. *PLoS ONE* 10:e0114947. doi: 10.1371/journal.pone.0114947
- Duong, Y. T., Qiu, M., De, A. K., Jackson, K., Dobbs, T., Kim, A. A., et al. (2012). Detection of recent HIV-1 infection using a new limiting-antigen avidity assay: potential for HIV-1 incidence estimates and avidity maturation studies. *PLoS ONE* 7:e33328. doi: 10.1371/journal.pone.0033328
- Friedman, S. R., Downing, M. J. Jr., Smyrnov, P., Nikolopoulos, G., Schneider, J. A., Livak, B., et al. (2014). Socially-integrated transdisciplinary HIV prevention. *AIDS Behav.* 18, 1821–1834. doi: 10.1007/s10461-013-0643-5
- Friedman, S. R., Kottiri, B. J., Neaigus, A., Curtis, R., Vermund, S. H., and Des Jarlais, D. C. (2000). Network-related mechanisms may help explain long-term HIV-1 seroprevalence levels that remain high but do not approach population-group saturation. *Am. J. Epidemiol.* 152, 913–922. doi: 10.1093/aje/152.10.913
- Friedman, S. R., Neaigus, A., Jose, B., Curtis, R., Goldstein, M., Ildefonso, G., et al. (1997). Sociometric risk networks and HIV risk. *Am. J. Public Health* 87, 1289–1296.
- Hatzakis, A., Sypsa, V., Paraskevis, D., Nikolopoulos, G., Tsiara, C., Micha, K., et al. (2015). Design and baseline findings of a large-scale rapid response to an HIV outbreak in people who inject drugs in Athens, Greece: the ARISTOTLE programme. *Addiction* 110, 1453–1467. doi: 10.1111/add.12999
- Khan, B., Dombrowski, K., Saad, M., McLean, K., and Friedman, S. (2013). Network firewall dynamics and the sub-saturation stabilization of HIV. *Discrete Dynam. Nature Soc.* 2013:720818. doi: 10.1155/2013/720818
- Klov Dahl, A. S. (1985). Social networks and the spread of infectious diseases: the AIDS example. *Soc. Sci. Med.* 21, 1203–1216.
- Nikolopoulos, G. K., Fotiou, A., Kanavou, E., Richardson, C., Detsis, M., Pharris, A., et al. (2015a). National income inequality and declining GDP growth rates are associated with increases in HIV diagnoses among people who inject drugs in Europe: a panel data analysis. *PLoS ONE* 10:e0122367. doi: 10.1371/journal.pone.0122367
- Nikolopoulos, G. K., Katsoulidou, A., Kantzanou, M., Rokka, C., Tsiara, C., Sypsa, V., et al. (2017). Evaluation of the limiting antigen avidity EIA (LAG) in people who inject drugs in Greece. *Epidemiol. Infect.* 145, 401–412. doi: 10.1017/S0950268816002417
- Nikolopoulos, G. K., Pavlitina, E., Muth, S. Q., Schneider, J., Psychogiou, M., Williams, L. D., et al. (2016). A network intervention that locates and intervenes with recently HIV-infected persons: the transmission reduction intervention project (TRIP). *Sci. Rep.* 6:38100. doi: 10.1038/srep38100.NIHMSID837702
- Nikolopoulos, G. K., Sypsa, V., Bonovas, S., Paraskevis, D., Malliori-Minerva, M., Hatzakis, A., et al. (2015b). Big events in Greece and HIV infection among people who inject drugs. *Subst. Use Misuse* 50, 825–838. doi: 10.3109/10826084.2015.978659
- Paraskevis, D., Nikolopoulos, G., Fotiou, A., Tsiara, C., Paraskeva, D., Sypsa, V., et al. (2013). Economic recession and emergence of an HIV-1 outbreak among drug injectors in athens metropolitan area: a longitudinal study. *PLoS ONE* 8:e78941. doi: 10.1371/journal.pone.0078941
- Paraskevis, D., Nikolopoulos, G., Tsiara, C., Paraskeva, D., Antoniadou, A., Lazanas, M., et al. (2011). HIV-1 outbreak among injecting drug users in Greece: a preliminary report. *Euro Surveill.* 16:19962. doi: 10.2807/ese.16.36.19962-en
- Paraskevis, D., Paraschiv, S., Sypsa, V., Nikolopoulos, G., Tsiara, C., Magiorkinis, G., et al. (2015). Enhanced HIV-1 surveillance using molecular epidemiology to study and monitor HIV-1 outbreaks among intravenous drug users (IDUs) in Athens and Bucharest. *Infect. Genet. Evol.* 35, 109–121. doi: 10.1016/j.meegid.2015.08.004
- Sypsa, V., Paraskevis, D., Malliori, M., Nikolopoulos, G. K., Panopoulos, A., Kantzanou, M., et al. (2014). Homelessness and other risk factors for HIV infection in the current outbreak among injection drug users in Athens, Greece. *Am. J. Public Health* 105, 196–204. doi: 10.2105/AJPH.2013.301656
- Sypsa, V., Psychogiou, M., Paraskevis, D., Nikolopoulos, G., Tsiara, C., Paraskeva, D., et al. (2017). Rapid decline in HIV incidence among persons who inject drugs during a fast-track combination prevention program after an HIV outbreak in Athens. *J. Infect. Dis.* 215, 1496–1505. doi: 10.1093/infdis/jix100
- Tsiara, C. G., Nikolopoulos, G. K., Dimou, N. L., Pantavou, K. G., Bagos, P. G., Mensah, B., et al. (2018). Interleukin gene polymorphisms and susceptibility to HIV-1 infection: a meta-analysis. *J. Genet.* 97, 235–251. doi: 10.1007/s12041-018-0907-y

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Williams, Kostaki, Pavlitina, Paraskevis, Hatzakis, Schneider, Smyrnov, Hadjikou, Nikolopoulos, Psychogiou and Friedman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Immunomodulatory Function of HBeAg Related to Short-Sighted Evolution, Transmissibility, and Clinical Manifestation of Hepatitis B Virus

Anna Kramvis^{1*}, Evangelia-Georgia Kostaki², Angelos Hatzakis² and Dimitrios Paraskevis²

¹ Hepatitis Virus Diversity Research Unit, Department of Internal Medicine, Faculty of Health Science, University of the Witwatersrand, Johannesburg, South Africa, ² Department of Hygiene, Epidemiology and Medical Statistics, Medical School, National and Kapodistrian University of Athens, Athens, Greece

OPEN ACCESS

Edited by:

Tara Patricia Hurst,
Abcam, United Kingdom

Reviewed by:

Timokratis Karamitros,
University of Oxford, United Kingdom
Masaya Sugiyama,
National Center For Global Health and
Medicine, Japan

*Correspondence:

Anna Kramvis
Anna.Kramvis@wits.ac.za

Specialty section:

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

Received: 19 July 2018

Accepted: 03 October 2018

Published: 24 October 2018

Citation:

Kramvis A, Kostaki E-G, Hatzakis A
and Paraskevis D (2018)
Immunomodulatory Function of
HBeAg Related to Short-Sighted
Evolution, Transmissibility, and Clinical
Manifestation of Hepatitis B Virus.
Front. Microbiol. 9:2521.
doi: 10.3389/fmicb.2018.02521

Hepatitis B virus (HBV) infection, a global public health problem can be asymptomatic, acute or chronic and can lead to serious consequences of infection, including cirrhosis, and hepatocellular carcinoma. HBV, a partially double stranded DNA virus, belongs to the family *Hepadnaviridae*, and replicates via reverse transcription of an RNA intermediate. This reverse transcription is catalyzed by a virus-encoded polymerase that lacks proof reading ability, which leads to sequence heterogeneity. HBV is classified into nine genotypes and at least 35 subgenotypes, which may be characterized by distinct geographical distributions. This HBV diversification and distinct geographical distribution has been proposed to be the result of the co-expansion of HBV with modern humans, after their out-of-Africa migration. HBeAg is a non-particulate protein of HBV that has immunomodulatory properties as a tolerogen that allows the virus to establish HBV infection *in vivo*. During the natural course of infection, there is seroconversion from a HBeAg-positive phase to a HBeAg-negative, anti-HBe-positive phase. During this seroconversion, there is loss of tolerance to infection and immune escape-HBeAg-negative mutants can be selected in response to the host immune response. The different genotypes and, in some cases, subgenotypes develop different mutations that can affect HBeAg expression at the transcriptional, translational and post-translational levels. The ability to develop mutations, affecting HBeAg expression, can influence the length of the HBeAg-positive phase, which is important in determining both the mode of transmission and the clinical course of HBV infection. Thus, the different genotypes/subgenotypes have evolved in such a way that they exhibit different modes of transmission and clinical manifestation of infection. Loss of HBeAg may be a sign of short-sighted evolution because there is loss of tolerogenic ability of HBeAg and HBeAg-negative virions are less transmissible. Depending on their ability to lead to HBeAg seroconversion, the genotype/subgenotypes exhibit varying degrees

of short-sighted evolution. The “arms race” between HBV and the immune response to HBeAg is multifaceted and its elucidation intricate, with transmissibility and persistence being important for the survival of the virus. We attempt to shed some light on this complex interplay between host and virus.

Keywords: genotypes, subgenotypes, tolerogen, transmission, HBeAg seroconversion

Hepatitis B virus (HBV), which belongs to the family *Hepadnaviridae* is the smallest DNA virus infecting man. HBV has a partially double-stranded, circular DNA genome of ~3,200 base pairs, with four overlapping reading frames (ORFs). HBV replicates via an RNA intermediate, through the process of reverse transcription catalyzed by the viral polymerase that lacks proof reading ability. Thus, sequence heterogeneity is a feature of the virus and within the host HBV exists as a quasispecies of mixed viral strains. Based on an intergroup divergence of >7.5% across the complete genome, HBV has been classified phylogenetically into 9 genotypes, A–I (Norder et al., 2004; Kramvis et al., 2005; Yu et al., 2010; Kramvis, 2014) (Figure 1), with a putative 10th genotype, “J,” isolated from a single individual (Tatematsu et al., 2009). Genotypes A–D, F, H, and I are classified further into at least 35 subgenotypes, using between ~4 and 8% intergroup nucleotide divergence across the complete genome and good bootstrap support, (Norder et al., 2004; Kramvis et al., 2005, 2008; Kramvis, 2014). The genotypes, and in some cases the subgenotypes, have distinct global and local geographical distributions, with the genotypes, prevailing in the two regions where HBV is endemic, south East Asia, and Africa, being different (Figure 2) (Kramvis et al., 2005; Kramvis, 2014). This HBV diversification and distinct geographical distribution has been proposed to be the result of the co-expansion of HBV with modern humans, after their out-of-Africa migration (Paraskevis et al., 2013).

In addition to encoding for the structural proteins, HBcAg (core or capsid protein) and the viral envelope proteins [three forms of HBsAg, small (S), middle (M), and large (L)] and the polymerase/reverse transcriptase, the compact genome of HBV encodes for two non-particulate proteins, the X protein (a transcriptional transactivator) and HBeAg (Tiollais et al., 1981). Antibodies are directed against both structural and non-structural proteins of HBV. Anti-HBc is non-neutralizing and a sign of exposure to HBV. Anti-HBs is the neutralizing antibody, directed against “a” dominant epitope of the viral envelope and anti-HBs levels >10 IU/L, following either natural infection or vaccination, are indicators of immunity. Anti-HBe is directed against B cell epitopes shared by both HBeAg and HBcAg, with HBeAg acting as a decoy for HBcAg. HBcAg and HBeAg T cell epitopes are also cross-reactive in both humans (Ferrari et al., 1991) and mice (Milich et al., 1987). Anti-HBe seroconversion can occur up to 6 years before the actual loss of HBeAg or the onset of liver damage (Thompson et al., 2007).

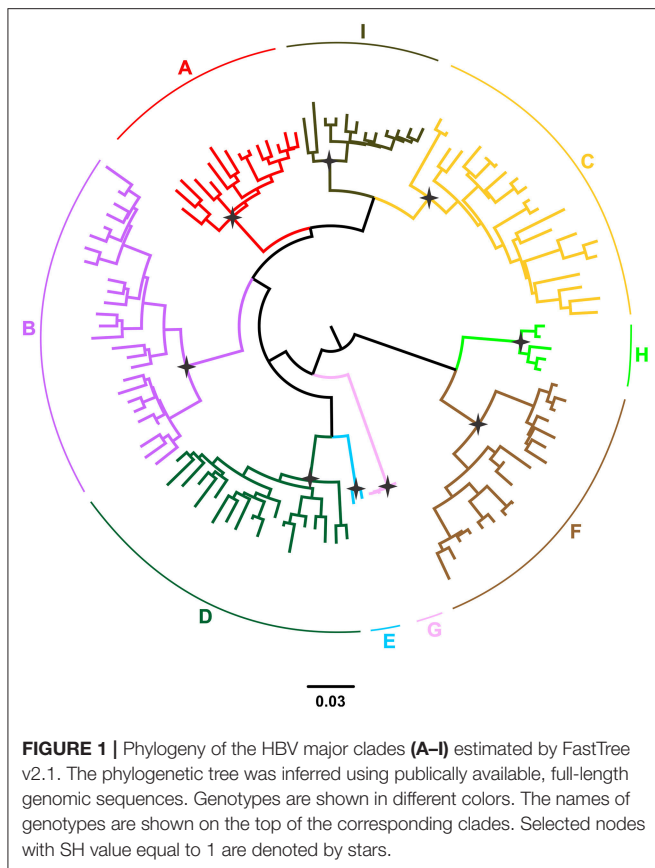
The complex interplay between the host and viral factors (including genotype/subgenotype, viral load and HBeAg status) play an important role in determining the clinical outcomes of HBV infection. HBV infection can be asymptomatic, acute,

chronic (HBsAg-positive for longer than 6 months), which can lead to serious consequences of infection, including cirrhosis, and hepatocellular carcinoma (HCC; liver cancer). HBV is generally non-cytopathic. The liver damage associated with either acute or chronic hepatitis B is as a result of the immune response attack on hepatocytes, in a bid to eliminate HBV during the immune clearance or reactive phase, which leads to necroinflammation. In order to overcome the effects of the immune response viruses can code for immunomodulatory proteins, such as HBeAg in the case of HBV, and also evolve genetically in order to escape the immune response. Immune escape mutants can be selected during the course of natural infection in response to the host immune response. In HBV the complex patterns of purifying selection are as a result of the overlapping ORFs (Mizokami et al., 1997) and the high frequency of recombination (Simmonds and Midgley, 2005; Zhou and Holmes, 2007). The “arms race” between HBV and the immune response is multifaceted and its elucidation intricate, with transmissibility and persistence being important for the survival of the virus. Here an attempt will be made to shed some light on this complex interplay.

HBeAG AS AN IMMUNOMODULATOR

The 25 kDa precursor of HBeAg, with an additional 29 amino acids on its amino terminus relative to HBcAg, is translated from the *preC/C* ORF on the precore mRNA [1901–2452/2488 from the *EcoRI* site, (Kramvis et al., 2005)] (Summers and Mason, 1982; Messageot et al., 2003). HBcAg is translated from the pregenomic RNA from position 1901. The basic core promoter (BCP) of HBV controls the transcription of the *preC/C* region from both transcripts (Yuh et al., 1992; Yu and Mertz, 1996) (reviewed in Kramvis and Kew, 1999). The amino terminal signal peptide directs the precursor to the endoplasmic reticulum (ER), where it is post-translationally modified by cleavage on the amino and carboxyl termini to give rise to the mature HBeAg that is expressed in the hepatocyte cytosol and also secreted in its soluble form in the serum (Revill et al., 2010).

Even though its exact function has not been determined, the conservation of HBeAg in all hepadnaviruses signifies an important role of this non-particulate secreted protein (Revill et al., 2010). Moreover, HBeAg must impart an evolutionary advantage to HBV because even though HBeAg-negative mutant strains of HBV exist (see below), they have not replaced the wild-type virus. From various animal studies it is evident that HBeAg is not involved in viral infection, replication, and assembly (Chang et al., 1987; Schlicht et al., 1987; Chen et al., 1992; Milich and Liang, 2003), but is important for natural infection *in vivo* (Milich and Liang, 2003). It is thought that the virus has retained the



secretory form of HBeAg because it has immunomodulatory functions. HBeAg downregulates the immune response to HBcAg by deletional, nondeletional, central, and peripheral immune tolerance (Milich et al., 1990, 1998; Milich, 1997; Milich and Liang, 2003; Chen et al., 2004, 2005). Thus, HBeAg-mediated immune regulation may predispose to chronicity and persistence following *in utero* or perinatal transmission and prevent severe liver injury during adult infections (Milich and Liang, 2003). Clinically, HBeAg is an index of viral replication, infectivity, inflammation, severity of disease and response to antiviral therapy.

HBeAg is expressed by both human and non-human hepadnaviruses, with all mammalian-infecting sequences showing high conservation within the amino terminus of the precursor and key immunomodulatory epitopes (Revill et al., 2010), including B-cell epitopes and T-cell recognition sites in the region that overlaps with HBcAg (Milich et al., 1987; Belnap et al., 2003; Billaud et al., 2005). Thus, HBeAg:

- *elicits both humoral and cell-mediated immunity, which differ from that directed to HBcAg* (Huang et al., 2006)

HBcAg and HBeAg share extensive amino acid homology but differ in their structure and localization. HBcAg is particulate and found in the cell whereas HBeAg is non-particulate/monomeric and can be found in the cell but is also secreted extracellularly (Milich and Liang, 2003). Because of these differences they are targeted by different antibodies (Imai et al., 1982). The antibody

to HBcAg can either be T-cell dependent or independent and the non-cross-reactive antibody to HBeAg is greater than that to that directed against HBcAg (Milich and McLachlan, 1986). HBeAg exhibits low T-cell dependent antibody response (Milich and Liang, 2003). The CD8⁺ T-cell (CTL) response against HBcAg is important in HBV elimination in humans (Bertoletti et al., 1991; Chisari and Ferrari, 1995). HBcAg and HBeAg are highly cross-reactive at the CD4⁺ and CD8⁺ T-cell levels (Milich et al., 1988; Bertoletti et al., 1991; Kuhrober et al., 1997; Townsend et al., 1997) and are indistinguishable in terms of CTL priming and CTL target recognition *in vitro* (Kuhrober et al., 1997; Townsend et al., 1997; Frelin et al., 2009). HBeAg expressed in hepatocytes *in vivo* is a superior target for HBcAg/HBeAg CTLs compared to HBcAg alone (Frelin et al., 2009). In order to be recognized by the CTLs, cytosolic HBeAg and/or its precursors must be processed and presented in the context of major histocompatibility complex (MHC) class I molecules (Frelin et al., 2009), whereas HBeAg presentation for CD4⁺ cells is via the MHC class II pathway (Milich and Liang, 2003).

- *functions as a T cell tolerogen* (Chen et al., 2005) *and regulates the immune response against the intracellular nucleocapsid* (Chen et al., 2004)

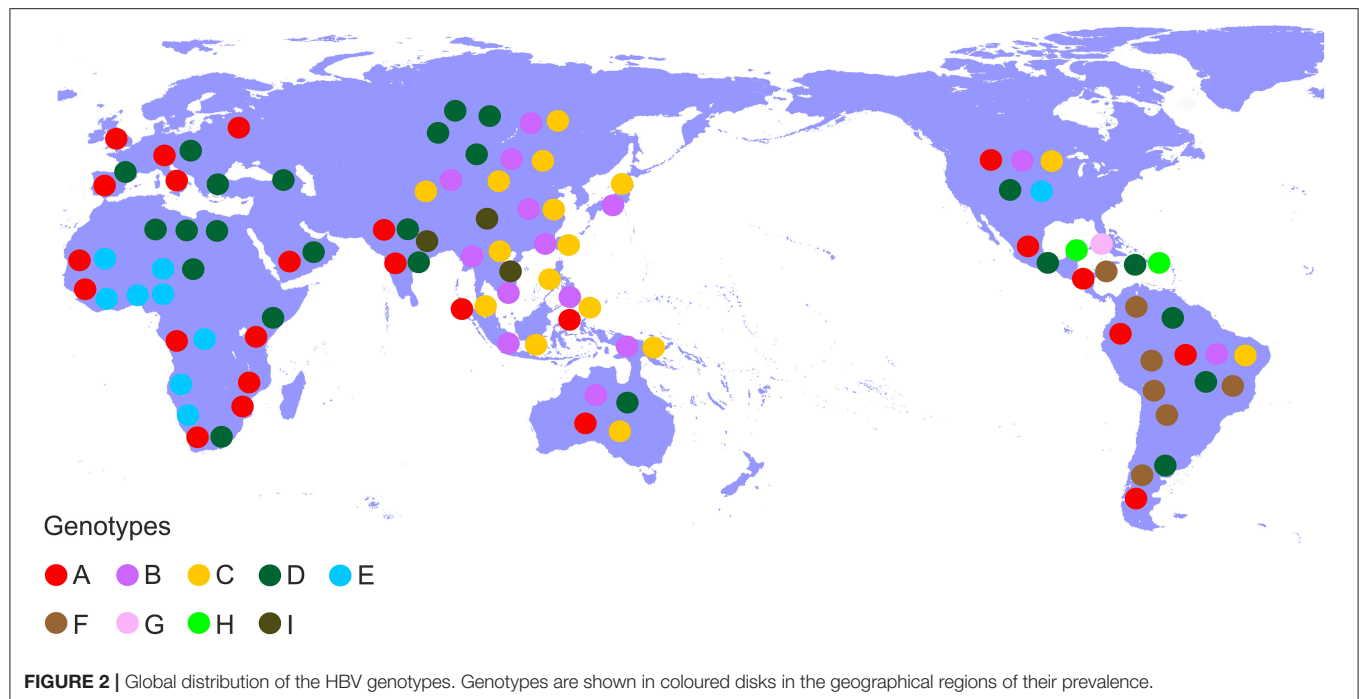
An important function of the non-particulate HBeAg, which is the only HBV protein to cross the placenta (Milich and Liang, 2003), is to establish neonatal T cell tolerance to both HBeAg and HBcAg (Milich et al., 1990; Hsu et al., 1992; Chen et al., 2004). Therefore, it is essential that HBeAg remains non-particulate so that it can cross the placenta (Schodel et al., 1993).

- *regulates the innate immune response to viral infection within hepatocytes.*

The adaptive immune response is attenuated, by the preferential activation of Th2 cells over Th1 cells, following NF- κ B activation by HBeAg (Yang et al., 2006). Th2 cells induce a non-protective humoral response whereas Th1 cells stimulate macrophages, which eliminate virions (Forsthuber et al., 1996; Huang et al., 2006). HBeAg in the serum leads to a switch from the normal Th1-mediated anti-HBc antibody response to Th2 phenotype (Milich et al., 1998). The difference in the activation of the different Th cell subsets is probably because the primary antigen-presenting cells are different for HBcAg and HBeAg, being B cells for the former and dendritic/macrophage cells for the latter (Milich et al., 1997). This preferential activation by HBeAg of Th2 cells, together with the depletion of Th1 cells, which are required for viral clearance, can lead to the persistence of HBV. HBeAg also down-regulates the expression of the Toll-like receptor (TLR) family within hepatocytes (Riordan et al., 2003; Visvanathan et al., 2007). When peripheral blood monocyte cells, were pre-treated with HBeAg, they displayed an impaired TLR signaling response (Visvanathan et al., 2007).

HBeAg TO ANTI-HBe SEROCONVERSION

Broadly, the natural history of HBV can be divided into at least four phases, which are differentiated by the level of viral replication and the host immune response:



- The **high replicative, low inflammatory phase** (previously immune tolerant phase) (Gish et al., 2015) is HBsAg/HBeAg-positive, with high HBV DNA levels ($>2 \times 10^5$ IU/ml) (Ni et al., 2007), with minimal liver inflammation, normal liver enzyme levels and asymptomatic (Chang et al., 1988). HBV does not select mutations frequently.
- The **immune clearance or reactive phase** is characterized by fluctuating alanine amino transferase (ALT) and HBV DNA levels and ending with spontaneous HBeAg loss (Ni et al., 2007; Liaw and Chu, 2009). HBeAg seroconversion is accompanied by elevated ALT and decreased HBV DNA levels (Ni et al., 2007). Core and precore mutations emerge in HBV.
- The **HBeAg-negative chronic hepatitis phase** (Gish et al., 2015), characterized by necroinflammation with high or fluctuating ALT levels and unsuccessful immune clearance, moderate to high viral loads and progressive liver disease.
- The **low replicative phase** (previously “inactive HBsAg-positive carrier” phase) (Gish et al., 2015) (post-HBeAg seroconversion). Although like the HBeAg-negative chronic hepatitis phase, this phase is also characterized by the absence of HBeAg, anti-HBe positivity, ALT levels are normal and HBV DNA levels are low or undetectable ($<2 \times 10^3$ IU/ml).

The length of the HBeAg-positive phase is important in determining both the mode of transmission and the clinical course of HBV infection. Mother-to-child transmission and subsequent chronic infection in the infant are favoured when mothers are HBeAg-positive (Chen et al., 2012), whereas children born to HBeAg-negative mothers are more likely to develop acute hepatitis B or may be infected horizontally later in life, if not immunized (Hadziyannis and Vassilopoulos,

2001; Kramvis, 2016). Following acute HBV infection, the development of chronic infection requires the expression of HBeAg (Hadziyannis, 2011) with a longer duration of HBeAg-positivity, increasing the risk of liver disease progression (Chu and Liaw, 2007). Although HBeAg loss is universally considered to be a favourable outcome (Hsu et al., 2002; Chu and Liaw, 2007), usually accompanied by decreased viral load and clinical remission, active hepatitis, cirrhosis, and hepatocellular carcinoma can develop in a minority of patients even after seroconversion (Hsu et al., 2002).

HBeAg seroconversion is a dynamic process lasting a number of years and involves the interaction between the host immune response and changes in the quasispecies and viral loads of HBV (Lim et al., 2007).

VIRAL EVOLUTION AND EMERGENCE OF MUTANTS AFFECTING HBeAG EXPRESSION

The loss of HBeAg between the high replicative, low inflammatory phase and HBeAg-negative chronic hepatitis phase is accompanied by:

- *Decrease in viral sequence diversity*

Viral sequence diversity was found to be 2.4-fold higher in both spontaneous and interferon-induced HBeAg seroconverters compared with non-seroconverters (Lim et al., 2007). Mutations, in both the HBeAg and HBcAg, were positively selected in 70% of the seroconverters compared to 5% of the non-seroconverters (Lim et al., 2007).

TABLE 1 | Mutations affecting HBeAg expression.

Level of effect	Genomic region	Mutation	Regression coefficient* (Kramvis et al., 2008)		References
			Positive	Negative	
Genotype/subgenotype variant	<i>preC</i>	1858C	A (3.04) C2 (2.59) F (1.49) F2 (1.95) H (1.48)	B (−1.73) C (−0.55) C1 (−2.34) D (−2.39) E (−1.68) F1 (−1.95)	Kramvis et al., 2008
Transcriptional	BCP	1762T/1764A	C (0.89) G (1.80)	B (−0.92) F (−0.78)	Buckwold et al., 1996
Translational	BCP	Kozak 1809–1812	A1		Ahn et al., 2003
	<i>preC</i>	Precore start codon 1817TAA	G (3.64)	D (−2.12)	
	<i>preC</i>	G1896A	B (0.89) D (0.97) G	C (−0.86) E (−1.42)	Carman et al., 1989
Post-translational	<i>preC</i>	G1862T	A1		Chen et al., 2008; Bhoola and Kramvis, 2017

*The logistic regression method was used to identify HBV mutations associated with specific genotypes/ subgenotypes (Afifi et al., 2004). Larger values show tighter correlations, with positive values related to a positive association, and negative values to a negative association. All associations were statistically significant using 2-sided p-values (Kramvis et al., 2008).

● Increase in viral substitution rates

Viral substitution rates have been shown to increase between the high replicative, low inflammatory phase and the immune clearance/reactive phase (Hannoun et al., 2000). Following seroconversion, immune pressure continues to drive viral mutagenesis even though the viral loads may be low (Lim et al., 2007) and HBV replicates less efficiently in the HBeAg-negative phase of disease (Volz et al., 2007).

● Emergence of HBeAg-negative mutants

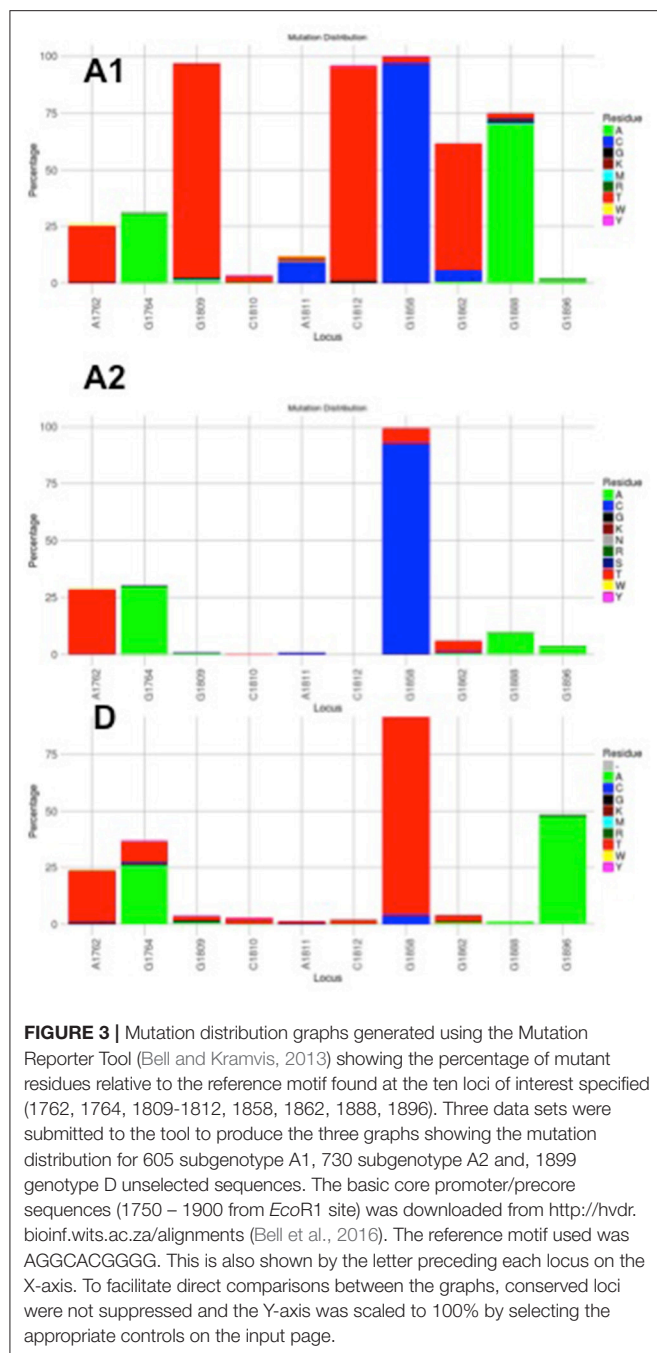
The viral quasispecies following seroconversion are defective in HBeAg production as a result of mutations in the BCP and *preC/C* ORF. Mutations in the BCP and *preC* can influence the expression of HBeAg at the transcriptional, translational and post-translational levels (Kramvis, 2016) (Table 1).

Position 1858 in the *preC/C* can be either a C or T (Figure 3) and is differentially associated with the different genotypes/subgenotypes (Table 1). This variation at position 1858 can influence, which of the mutations, affecting HBeAg expression, will develop (Kramvis et al., 2008) and their frequency (Revill et al., 2010).

Although the HBV sequences deposited in the public databases, such as GenBank, may not be entirely representative of the strains circulating globally, they provide the best available datasets of genotyped strains of HBV (Bell et al., 2016). From this data set it is evident that the genotype can affect the frequency of the BCP/PC mutations. The BCP double mutation 1762T/1764A occurs in ~25% of downloaded sequences (Bell et al., 2016) without difference between subgenotypes A1, A2, and genotype D (Figure 3). In two case control studies, independently of HBeAg status, 1762T/1764A was significantly more frequent in genotypes A and C compared to D and B, respectively (Orito

et al., 2001; Tanaka et al., 2004). A mutation, which leads to a stop codon that truncates HBeAg during translation, G1896A, does not develop in the genotypes/subgenotypes in which 1858C occurs frequently (Li et al., 1993; Lok et al., 1994) and in which 1858C is positively associated with, namely, A, C2, F2, and H (Kramvis et al., 2008). On the other hand, G1896A is frequent in genotypes/subgenotypes, which have 1858T, C1, D, E, and F (Kramvis et al., 2008; Revill et al., 2010). As illustrated in Figure 3 more than 40% of genotype D sequences downloaded from Genbank had G1896A whereas it was found in very few genotype A sequences. Overall G1896A occurs in up to 30% of all sequences, most frequently detected in genotypes G (100%), and B (50%), followed by genotypes D (40%) and C (23%) (Revill et al., 2010). All genotype G and the majority of subgenotype B6 sequences encode this mutation and it is least frequently observed in genotype A (1.5%) and genotype F (5%) (Revill et al., 2010). Thus, because of these differences the estimated annual rate of HBeAg to anti-HBeAg seroconversion can differ between the genotypes. Genotype B has higher annual rate of seroconversion of 15.5% compared to genotype C (7.9%) (Kao et al., 2004) and E (7.4%) (Shimakawa et al., 2016). In a Taiwanese study, it was shown that HBeAg seroconversion of patients infected with genotype C occurred 10 years later than those infected with genotype B (Kao et al., 2004).

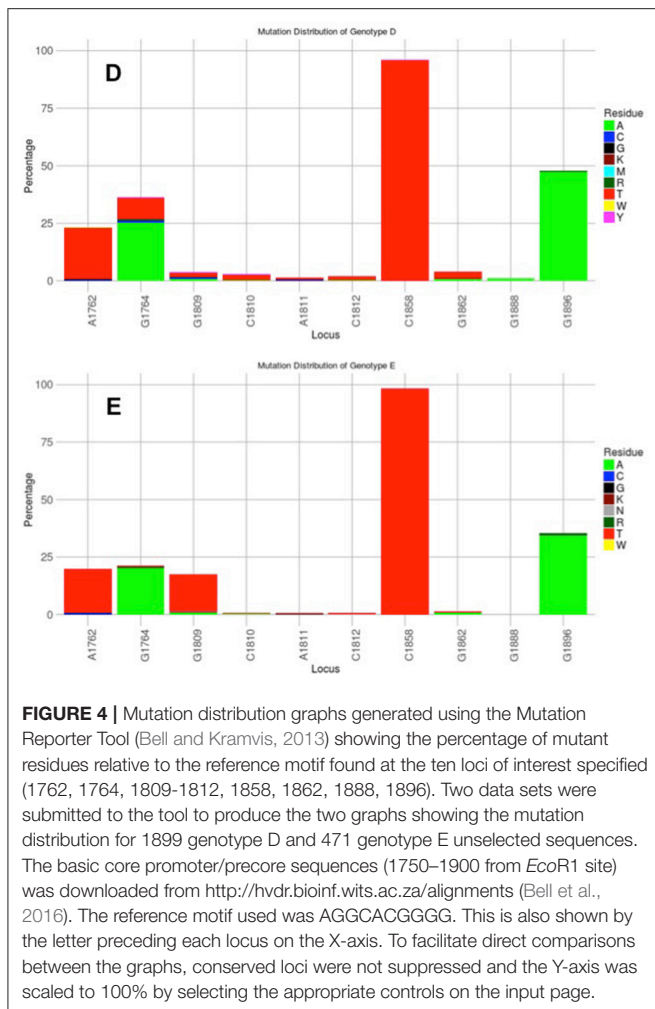
Subgenotype A2 does not select the G1896A mutation, nor the other mutations found in subgenotype A1 that contribute to the HBeAg-negativity of this unique strain (Table 1). In particular, the mutations in the Kozak sequence upstream from the precore start codon (1809–1812), characteristic of subgenotype A1 (Baptista et al., 1999; Kramvis and Kew, 2007; Revill et al., 2010), interfere with translation of HBeAg by a leaky scanning mechanism (Ahn et al., 2003), do not occur in subgenotype A2. Therefore, generally subgenotype A2 has a high



frequency of HBeAg-positivity compared to subgenotype A1 and this is statistically significant in individuals younger than 30 years (Tanaka et al., 2004). The reason for the high frequency of HBeAg-positivity in subgenotype A2 is possibly because the only mutation that is selected in A2 that can affect HBeAg expression is 1762T/1764A. This double mutation is found more frequently in HBeAg-negative carriers of subgenotype A2 HBV, whereas, the frequency of 1762T/1764A is not affected by HBeAg status in subgenotype A1 (Tanaka et al., 2004). In genotype D isolates from HBeAg-negative individuals, both 1762T/1764A and 1896A were

more frequent than from HBeAg-positive individuals (Tanaka et al., 2004). The BCP 1762T/1764A mutations are not selected in subgenotype B6 strains, prevalent in people living in Canadian Arctic and who are frequently HBeAg-negative, as a result of 1896A and precore start codon mutations (Osiowy et al., 2011). Even though genotypes D and E cannot be differentiated in the precore region and both have 1858T, they do not develop G1896A at the same frequency. In Sudan, where genotypes D and E co-circulate viral loads were significantly higher in genotype E-infected patients compared to genotype D-infected, with patients infected with genotype E, showing a significantly higher frequency of HBeAg-positivity in blood donors (Mahgoub et al., 2011), asymptomatic carriers and liver disease patients (Yousif et al., 2013). The high frequency of HBeAg-negativity in genotype D was as a result of G1896A (Mahgoub et al., 2011; Yousif et al., 2013), which correlates with the statistically positive and negative association of G1896A, with genotype D and E, respectively (Kramvis et al., 2008) (Table 1). This is confirmed by a comparison of the frequency of G1896A in sequences downloaded from the public databases (Bell et al., 2016), where G1896A occurred 47.2% of genotype D compared to 34.2% genotype E sequences ($p < 0.0001$) (Figure 4). Genotype G is unique in that all sequences have a premature stop codon at position 2 of the precore precursor protein and therefore HBeAg is not expressed (Stuyver et al., 2000).

Subgenotype A1 is the most “sophisticated” in terms of its control of HBeAg expression (Kramvis and Kew, 2007). Firstly, like all other genotypes and subgenotype A2, it can develop 1762T/1764A. Secondly, instead of Kozak 1809–1812 GCAC, which is present in subgenotypes A2 and D3, subgenotype A1 has TCAT (Kramvis and Kew, 2007). These variations are characteristic of subgenotype A1 and affect HBeAg expression at the translational level (Ahn et al., 2003), by converting the Kozak region from an optimal to a suboptimal translation context (Kimbi et al., 2004) and, causing decreased translation of HBeAg by a ribosomal leaky scanning mechanism (Ahn et al., 2003). Compared to subgenotypes A2 and D3, transfection with subgenotype A1 led to a lower expression of the precore/core precursor in the secretory pathway and a higher co-localization in the nucleus (Bhoola and Kramvis, 2016). This reduction in HBeAg levels is comparable to that observed in the presence of 1762T/1764A and when occurring together, the Kozak and BCP 1762T/1764A mutations reduce HBeAg expression in an additive manner (Ahn et al., 2003). Thirdly, a G to T transversion at position 1862 in the precore region of HBV occurs more frequently in subgenotype A1 isolates from HBeAg-negative than in HBeAg-positive South Africans (Kramvis et al., 1997, 1998) and affects HBeAg expression at the post-translational level (Sugauchi et al., 2004; Chen et al., 2008; Inoue et al., 2009; Bhoola and Kramvis, 2017). This mutation results in a valine to phenylalanine substitution at the –3 position of the signal peptide cleavage site at position 19 of the precursor protein. Phenylalanine interferes with signal peptide cleavage (Nielsen et al., 1997). This leads to decreased HBeAg expression as a result of the increased retention of the precursor in the cytoplasm of the hepatocyte (Chen et al., 2008; Inoue et al., 2009; Bhoola and Kramvis, 2017). When this mutation was introduced into



a genotype D plasmid driven by a cytomegalovirus promoter, it resulted in a 54% reduction in the secretion of HBeAg relative to the wild-type and to the formation of aggresomes (Chen et al., 2008). When this mutation was introduced into a subgenotype A1 backbone, the decrease in expression of secreted HBeAg was less than that in the genotype D context (22%). The mutant was found to lead to the accumulation of the HBeAg precursor protein in the ER and ER-Golgi intermediate compartment (ERGIC). This accumulation resulted in an earlier activation of the three UPR pathways, but not to an increase in apoptosis (Bhoola and Kramvis, 2016).

Thus, in the various genotypes and subgenotypes, HBeAg loss, and immune response escape are as a result of a number of mutations. In order to survive, HBV has to balance its ability to establish an infection (HBeAg-positive, immunotolerogenic phase) and escape the immune response (HBeAg-negative phase) in order to persist.

INTERACTION OF THE VIRUS AND HOST

Previous studies have shown that a few years before seroconversion and as soon as the levels of HBeAg begin

declining, the genetic diversity of the virus increases (Hannoun et al., 2000; Lim et al., 2007; Cheng et al., 2013). Notably viral diversity does not follow this trend in non-seroconverters but remains at the low-level seen over the course of the immunotolerant phase (Lim et al., 2007; Wu et al., 2011; Cheng et al., 2013). Given the stable rate at which mutations accumulate in the HBV genome during the different stages of HBV replication, the differences in viral diversity observed, firstly between seroconverters and non-seroconverters, during the course of the infection and close to the time of seroconversion are because of the strong selective pressure of the immune response (Cheng et al., 2013). Specifically, as soon as the levels of HBeAg are decreasing, tolerance diminishes, leading to increased immune response and the selection of escape variants, which are infrequently selected during the immunotolerant phase. The increased genetic diversity is inversely related to HBV-DNA levels, intimating that the variants selected before and during the HBeAg negative phase have a lower replicative capacity than the virions circulating in the absence of immune selective pressure (Cheng et al., 2013). Thus, the intra-host viral evolution also follows different stages according to the HBeAg and immune-status of the host. During the immunotolerant phase, the virus replicates without constraint from the host and thus there is no need to select multiple variants; in contrast, during the immune clearance and seroconversion phase, different variants are selected because of the strong selective pressure (Lim et al., 2007; Cheng et al., 2013). Interestingly, during the latter phase, multiple variants with higher genetic variability in the non-overlapping genomic regions are selected and circulate.

Therefore, there is a complex interaction of the virus and the host. As it has been previously found, HBV has co-expanded with modern humans for at least 28,000 years suggesting a long period of interaction (Paraskevis et al., 2013). HBV major clades (genotypes) have been generated as a result of founder effects of different strains spreading in distinct geographic regions. Similarly, subgenotypes, belonging within HBV major clades, are the result of more recent events of strains. The time of the most common recent origin (tMRCA) of different genotypes has been previously estimated to occur several 1,000 years ago (Paraskevis et al., 2013), while the tMRCA of the subgenotypes is more recent (Paraskevis et al., 2013). Given that HBV has co-expanded with humans for a long time and also that some genotypes have been more frequently found in specific areas and populations (i.e., F/H in indigenous populations in America; B and C in Asia; many different subgenotypes have been infecting mostly indigenous populations across the globe) suggests that some of the genomic characteristics related to the HBeAg expression probably have been shaped over a long time period. This is presently a hypothesis and concrete evidence will come from the analysis of HBV from ancient samples.

Stochastic mutations lead to the break of immune tolerance and/or increased immune reactivity, which drives viral evolution from a low- to a high-level positive selection stage that in turn leads to decreased HBV DNA levels because of increased immune pressure and less efficient viral replication. This increased sequence diversity in the intra-host quasispecies in seroconverters prior to the loss of HBeAg was found to be

largely as a result of sequence variation in the precore/core open reading frame and its regulatory elements (Harrison et al., 2011). This sequence variation in the region encoding HBeAg can influence the immunomodulatory role of HBeAg, the ability of HBV to persist and be transmitted, as well as its virulence. The driving forces of viral evolution during the different stages of HBV infection are complex and difficult to present in a linear fashion. The balance and trade-offs of the immunomodulation and immune escape of HBeAg can be influenced by the different genotypes/subgenotypes.

SHORT SIGHTED EVOLUTION, TRANSMISSIBILITY AND CLINICAL MANIFESTATION RELATED TO HBeAg EXPRESSION

Broadly, HBeAg-negativity is preferable at the individual host level whereas HBeAg-positivity is more favourable at the population level (Milich and Liang, 2003). As the natural course of HBV infection progresses from the HBeAg-positive (high replicative/low inflammatory) phase to the HBeAg-negative (chronic hepatitis) phase, T cell tolerance is ending as HBeAg expression changes from its secreted soluble to the cytosolic form. As an immunotolerogen and a decoy, secreted HBeAg is beneficial during the early phases of the infection, whereas the cytosolic form represents a liability considering that this form is processed and presented to MHC class I. Furthermore, HBeAg expressed *in vivo* is superior to HBcAg as a target, directly, or indirectly, of HBcAg/HBeAg-specific CTLs (Frelin et al., 2009). This differential CTL recognition of HBeAg, expressed by wild-type virus, may result in the preferential elimination of wild-type relative to HBeAg-negative HBV (Frelin et al., 2009). As a consequence, viral diversity increases during the course of infection, with the emergence of strains defective in HBeAg expression, which can escape the immune response and persist albeit at the cost, in some cases, of reduced replication, decreased transmissibility and increased virulence (Figure 5). Thus, this adaptive evolution of HBV during the course of HBeAg-seroconversion can be considered *short-sighted* when it limits the ability of the virus to transmit to new hosts, either by reducing per contact transmissibility or reducing contact rate of infected hosts because of increased virulence that can result in death (Lythgoe et al., 2017). The increased virulence can also be a result of the loss of immune tolerance. The ability of the different genotypes/subgenotypes to lead to HBeAg seroconversion can influence the degree of short-sighted evolution of HBV.

RELATIONSHIP OF GENOTYPES AND SUBGENOTYPES TO SHORT-SIGHTED EVOLUTION

As discussed above, mutations in the BCP/precore region, can affect HBeAg expression at the transcriptional, translational and post-translational level, and these mutations do not develop in all genotypes/subgenotypes at the same frequency (Table 1,

Figures 3, 4). Subgenotype A2 and genotype H, develop only mutations 1762T/1764A, which affect transcription of precore mRNA and result in decreased levels of HBeAg (Buckwold et al., 1996; Baptista et al., 1999), without switching off HBeAg expression. Thus, these (sub)genotypes, which have been shown to have a relatively high frequency of HBeAg-positivity (Tanaka et al., 2004) are expected to have a long HBeAg high replicative, low inflammatory phase. This allows HBV to be transmitted sexually since it can remain asymptomatic allowing maturity for sexual activity and a healthy host (Araujo et al., 2011) before the development of cirrhosis and HCC. In fact, in a study carried out in the USA, genotype A (subgenotype A2 is predominates over A1 in the USA) was predominantly transmitted via sexual (82%) or parenteral (54%) routes compared to genotype B and C, which were transmitted vertically (Chu et al., 2003). On the other hand, the persistence of 1762T/1764A mutant expressing HBeAg, even at relatively low levels, means that the hepatocytes expressing HBeAg will be targeted, leading to liver injury and consequent development of cirrhosis and HCC much later in life. A high prevalence of 1762T/1764A has been detected in HBV strains from HCC patients compared to asymptomatic carriers of HBV (Baptista et al., 1999).

G1896A is the classical HBeAg-negative mutation that affects the translation of HBeAg by introducing a stop codon that truncates the HBeAg precursor and HBeAg is not expressed. Genotypes/subgenotypes with 1858T can develop the G1896A HBeAg-negative mutants (Table 1) and develop HBeAg-negative chronic hepatitis. Loss of HBeAg may be a sign of short-sighted evolution because there is loss of tolerogenic ability of HBeAg, HBeAg-negative virions are less transmissible and cannot establish chronic hepatitis. Establishment of chronic HBV infection following perinatal transmission requires wild-type HBV, whereas transmission of mixed wild-type/G1896A quasiespecies leads to early immune elimination and resolution of the acute infection (Raimondo et al., 1993). G1896A occurs more frequently in genotype D than in genotype E strains (Figure 4), despite both having 1858T. This means that women of childbearing age infected with genotype E will remain HBeAg-positive for longer, allowing for mother-to-child transmission, which has resulted in high prevalence and geographical restriction of genotype E to Africa and among African emigrants to other regions (Mulders et al., 2004). Vertical transmission is advantageous in restricted human populations (Li et al., 2017). The higher HBeAg-positivity seen in individuals infected with genotype E compared to those infected with genotype D could confer tolerance and less serious clinical manifestations than genotype D (Yousif et al., 2013). Genotype E was found to prevail in Sudanese blood donors (Mahgoub et al., 2011), whereas the liver disease patients were infected with genotype D (Yousif et al., 2013). High ratios (>50%) of G1896A in the HBeAg-positive phase in individuals infected with genotype D are accompanied by persistently high viraemia and ALT elevation after anti-HBe seroconversion and a higher risk of cirrhosis (Chu et al., 2002).

Genotype G is a HBeAg-negative strain because of a premature stop codon at position 2 of the precore precursor protein and G1896A. Although genotype G is replication competent on its own (Li et al., 2007), it is always found as a

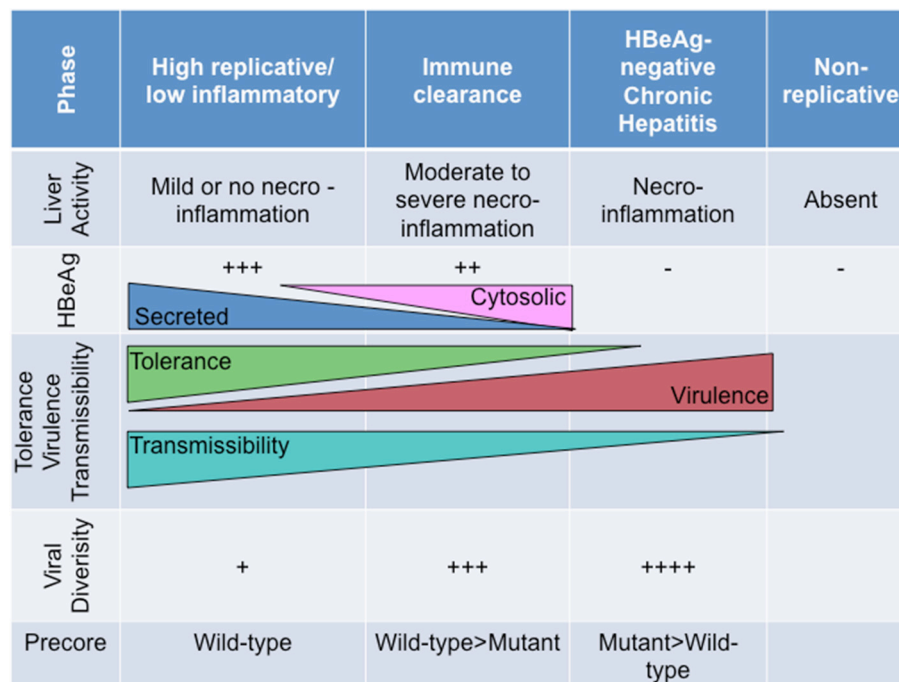


FIGURE 5 | The relationship of HBeAg expression and precore region diversity to the natural history of HBV infection and its effect on tolerance, virulence and transmissibility of HBV.

co-infection with either subgenotype A2 or genotype H because the absence of the immunotolerogenic properties of HBeAg does not allow it to establish a persistent infection (Li et al., 2007). Like subgenotype A2 (Figure 3), genotype H develops 1896A infrequently and therefore can supply HBeAg *in trans* following co-infection. Genotype G can be transmitted and propagated without the helper virus and has been shown to replace genotype A in the HBeAg-negative phase of infection (Li et al., 2007) but requires the presence of helper virus expressing HBeAg in the early phases of infection in immunocompetent hosts (Kato et al., 2002). The absence of immunomodulatory HBeAg means that this genotype G alone can only infect immunocompromised individuals (Li et al., 2007). This is the reason why this genotype is relatively rare and is transmitted by high risk transmission chains including males-who-have-sex-with-men (MSMs), intravenous drug users (IVDUs) and blood transfusions.

Subgenotype A1 is the only strain that develops G1862T and HBV strains with G1862T have been isolated from tumorous but not from adjacent non-tumorous liver tissue (Kramvis et al., 1998). Thus, patients, infected with subgenotype A1 with the characteristic suboptimal Kozak sequence preceding the precore start codon, together with 1762T/1764A and 1862T mutant, would have severely decreased levels of HBeAg. The reduction or absence of HBeAg in the serum would result in the loss of immunotolerance and in the immune response being directed to the hepatocytes because of the reduction of soluble HBeAg, which acts as a decoy. This together with the increased ER stress, can result in liver damage, thus contributing to the higher hepatocarcinogenic potential of this subgenotype (Kramvis and

Kew, 2007). In individuals infected with subgenotype A1, HCC develops 6.5 years earlier than in individuals infected with other (sub)genotypes (Kramvis and Kew, 2007). The early HBeAg seroconversion and the development of HCC before the age of 30, means that this subgenotype is characterized by the highest degree of short-sighted evolution as its mode of transmission is limited to early horizontal transmission and a short HBeAg-positive phase. This could be the reason why it is the only subgenotype, which can develop the unique mutations that affect HBeAg expression, Kozak 1809-1812 and G1862T at the translational and post-translational levels, respectively.

KNOWLEDGE GAPS AND FUTURE STUDIES

There is a paucity of studies on the natural history of subgenotype A1 of HBV, which prevails in Africa as opposed to subgenotype A2 found outside Africa. Moreover, there are no studies on the immune response to this particular strain of HBV. Further studies on the infection by this subgenotype in ethnic groups of non-African descent are important to determine whether the host genetic background influences its natural history. A relatively high prevalence of subgenotype A1 was found in HCC patients in southern India, and similarly to African studies, there was an association of subgenotype A1 with HCC and its development at a younger age (Gopalakrishnan et al., 2013). These studies are important and relevant considering the recent

human migrations from Africa, which can lead to the dispersal of the strain globally.

CONCLUSION

HBV is a very successful, highly evolved virus for a number of reasons. It has evolved in such a way that it can be transmitted by various modes, *in utero*, perinatally, horizontally early in life and sexually. Various (sub)genotypes have preferable mode/s of transmission. HBV can establish chronic infection and even though this may have serious consequences, including the development of cirrhosis, and HCC that may limit mobility and hence sexual transmission, these arise later in life long after the perinatal and neonatal transmission of the virus. Loss of HBeAg may be a sign of short-sighted evolution because there is loss of tolerogenic ability of HBeAg and HBeAg-negative virions are less transmissible. The different genotype/subgenotypes exhibit different degrees of short-sighted evolution.

REFERENCES

- Affi, A., Clark, V. A., and May, S. (2004). *Computer-Aided Multivariate Analysis*. 4th Edn. Boca Raton, FL: Chapman & Hall/CRC Press.
- Ahn, S. H., Kramvis, A., Kawai, S., Spangenberg, H. C., Li, J., Kimbi, G., et al. (2003). Sequence variation upstream of precore translation initiation codon reduces hepatitis B virus e antigen production. *Gastroenterology* 125, 1370–1378. doi: 10.1016/j.gastro.2003.07.016
- Araujo, N. M., Waizbort, R., and Kay, A. (2011). Hepatitis B virus infection from an evolutionary point of view: how viral, host, and environmental factors shape genotypes and subgenotypes. *Infect. Genet. Evol.* 11, 1199–1207. doi: 10.1016/j.meegid.2011.04.017
- Baptista, M., Kramvis, A., and Kew, M. C. (1999). High prevalence of 1762(T) 1764(A) mutations in the basic core promoter of hepatitis B virus isolated from black Africans with hepatocellular carcinoma compared with asymptomatic carriers. *Hepatology* 29, 946–953. doi: 10.1002/hep.510290336
- Bell, T. G., and Kramvis, A. (2013). Mutation reporter tool: an online tool to interrogate loci of interest, with its utility demonstrated using hepatitis B virus. *Virol. J.* 10:62. doi: 10.1186/1743-422X-10-62
- Bell, T. G., Yousif, M., and Kramvis, A. (2016). Bioinformatic curation and alignment of genotyped hepatitis B virus (HBV) sequence data from the GenBank public database. *Springerplus* 5:1896. doi: 10.1186/s40064-016-3312-0
- Belnap, D. M., Watts, N. R., Conway, J. F., Cheng, N., Stahl, S. J., Wingfield, P. T., et al. (2003). Diversity of core antigen epitopes of hepatitis B virus. *Proc. Natl. Acad. Sci. U.S.A.* 100, 10884–10889. doi: 10.1073/pnas.1834404100
- Bertoletti, A., Ferrari, C., Fiaccadori, F., Penna, A., Margolskee, R., Schlicht, H. J., et al. (1991). HLA class I-restricted human cytotoxic T cells recognize endogenously synthesized hepatitis B virus nucleocapsid antigen. *Proc. Natl. Acad. Sci. U.S.A.* 88, 10445–10449. doi: 10.1073/pnas.88.23.10445
- Bhoola, N. H., and Kramvis, A. (2016). Hepatitis B e Antigen Expression by Hepatitis B Virus Subgenotype A1 Relative to Subgenotypes A2 and D3 in Cultured Hepatocellular Carcinoma (Huh7) Cells. *Intervirology* 59, 48–59. doi: 10.1159/000446240
- Bhoola, N. H., and Kramvis, A. (2017). Expression of wild-type or G1862T mutant HBe antigen of subgenotype A1 of hepatitis B virus and the unfolded protein response in Huh7 cells. *J. Gen. Virol.* 98, 1422–1433. doi: 10.1099/jgv.0.000793
- Billaud, J. N., Peterson, D., Schodel, F., Chen, A., Sallberg, M., Garduno, F., et al. (2005). Comparative antigenicity and immunogenicity of hepatitis B virus core proteins. *J. Virol.* 79, 13641–13655. doi: 10.1128/JVI.79.21.13641-13655.2005

AUTHOR CONTRIBUTIONS

AK conceptualized and wrote the first draft of the paper, carried out mutation analysis. DP contributed to the writing and editing of the paper. E-GK carried out the phylogenetic analyses and prepared **Figures 1, 2**. AH contributed to the writing and editing of the paper. All authors listed have made a substantial, and intellectual contribution to the work, read and approved the final manuscript for submission.

ACKNOWLEDGMENTS

Research was supported by the Cancer Association of South Africa, Deutsche Forschungsgemeinschaft (German Research Foundation), Japan Society for the Promotion of Science, National Research Foundation of South Africa, Poliomyelitis Research Foundation, South African Medical Research Council, University of the Witwatersrand and the Hellenic Scientific Society for the Study of AIDS and STDs.

- Buckwold, V. E., Xu, Z., Chen, M., Yen, T. S., and Ou, J. H. (1996). Effects of a naturally occurring mutation in the hepatitis B virus basal core promoter on precore gene expression and viral replication. *J. Virol.* 70, 5845–5851.
- Carman, W. F., Jacyna, M. R., Hadziyannis, S., Karayiannis, P., McGarvey, M. J., Makris, A., et al. (1989). Mutation preventing formation of hepatitis B e antigen in patients with chronic hepatitis B infection. *Lancet* 2, 588–591. doi: 10.1016/S0140-6736(89)90713-7
- Chang, C., Enders, G., Sprengel, R., Peters, N., Varmus, H. E., and Ganem, D. (1987). Expression of the precore region of an avian hepatitis B virus is not required for viral replication. *J. Virol.* 61, 3322–3325.
- Chang, M. H., Hwang, L. Y., Hsu, H. C., Lee, C. Y., and Beasley, R. P. (1988). Prospective study of asymptomatic HBsAg carrier children infected in the perinatal period: clinical and liver histologic studies. *Hepatology* 8, 374–377. doi: 10.1002/hep.1840080231
- Chen, C. Y., Crowther, C., Kew, M. C., and Kramvis, A. (2008). A valine to phenylalanine mutation in the precore region of hepatitis B virus causes intracellular retention and impaired secretion of HBe-antigen. *Hepatol. Res.* 38, 580–592. doi: 10.1111/j.1872-034X.2007.00315.x
- Chen, H. L., Lin, L. H., Hu, F. C., Lee, J. T., Lin, W. T., Yang, Y. J., et al. (2012). Effects of maternal screening and universal immunization to prevent mother-to-infant transmission of HBV. *Gastroenterology* 142, 773–81.e2. doi: 10.1053/j.gastro.2011.12.035
- Chen, H. S., Kew, M. C., Hornbuckle, W. E., Tennant, B. C., Cote, P. J., Gerin, J. L., et al. (1992). The precore gene of the woodchuck hepatitis virus genome is not essential for viral replication in the natural host. *J. Virol.* 66, 5682–5684.
- Chen, M., Sallberg, M., Hughes, J., Jones, J., Guidotti, L. G., Chisari, F. V., et al. (2005). Immune tolerance split between hepatitis B virus precore and core proteins. *J. Virol.* 79, 3016–3027. doi: 10.1128/JVI.79.5.3016-3027.2005
- Chen, M. T., Billaud, J. N., Sallberg, M., Guidotti, L. G., Chisari, F. V., Jones, J., et al. (2004). A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proc. Natl. Acad. Sci. U.S.A.* 101, 14913–14918. doi: 10.1073/pnas.0406282101
- Cheng, Y., Guindon, S., Rodrigo, A., Wee, L. Y., Inoue, M., Thompson, A. J., et al. (2013). Cumulative viral evolutionary changes in chronic hepatitis B virus infection precedes hepatitis B e antigen seroconversion. *Gut* 62, 1347–1355. doi: 10.1136/gutjnl-2012-302408
- Chisari, F. V., and Ferrari, C. (1995). Hepatitis B virus immunopathogenesis. *Annu. Rev. Immunol.* 13, 29–60. doi: 10.1146/annurev.iy.13.040195.000333
- Chu, C. J., Keeffe, E. B., Han, S. H., Perrillo, R. P., Min, A. D., Soldevila-Pico, C., et al. (2003). Hepatitis B virus genotypes in the United States: results of a nationwide study. *Gastroenterology* 125, 444–451. doi: 10.1016/S0016-5085(03)00895-3

- Chu, C. M., and Liaw, Y. F. (2007). Chronic hepatitis B virus infection acquired in childhood: special emphasis on prognostic and therapeutic implication of delayed HBeAg seroconversion. *J. Viral Hepat.* 14, 147–152. doi: 10.1111/j.1365-2893.2006.00810.x
- Chu, C. M., Yeh, C. T., Lee, C. S., Sheen, I. S., and Liaw, Y. F. (2002). Precore stop mutant in HBeAg-positive patients with chronic hepatitis B: clinical characteristics and correlation with the course of HBeAg-to-anti-HBe seroconversion. *J. Clin. Microbiol.* 40, 16–21. doi: 10.1128/JCM.40.1.16-21.2002
- Ferrari, C., Bertoletti, A., Penna, A., Cavalli, A., Valli, A., Missale, G., et al. (1991). Identification of immunodominant T cell epitopes of the hepatitis B virus nucleocapsid antigen. *J. Clin. Invest.* 88, 214–222. doi: 10.1172/JCI115280
- Forsthuber, T., Yip, H. C., and Lehmann, P. V. (1996). Induction of TH1 and TH2 immunity in neonatal mice. *Science* 271, 1728–1730. doi: 10.1126/science.271.5256.1728
- Frelin, L., Wahlstrom, T., Tucker, A. E., Jones, J., Hughes, J., Lee, B. O., et al. (2009). A mechanism to explain the selection of the hepatitis e antigen-negative mutant during chronic hepatitis B virus infection. *J. Virol.* 83, 1379–1392. doi: 10.1128/JVI.01902-08
- Gish, R. G., Given, B. D., Lai, C. L., Locarnini, S. A., Lau, J. Y., Lewis, D. L., et al. (2015). Chronic hepatitis B: Virology, natural history, current management and a glimpse at future opportunities. *Antiviral Res.* 121, 47–58. doi: 10.1016/j.antiviral.2015.06.008
- Gopalakrishnan, D., Keyter, M., Shenoy, K. T., Leena, K. B., Thayumanavan, L., Thomas, V., et al. (2013). Hepatitis B virus subgenotype A1 predominates in liver disease patients from Kerala, India. *World J. Gastroenterol.* 19, 9294–9306. doi: 10.3748/wjg.v19.i48.9294
- Hadziyannis, S. J. (2011). Natural history of chronic hepatitis B in Euro-Mediterranean and African countries. *J. Hepatol.* 55, 183–191. doi: 10.1016/j.jhep.2010.12.030
- Hadziyannis, S. J., and Vassilopoulos, D. (2001). Hepatitis B e antigen-negative chronic hepatitis B. *Hepatology* 34(4 Pt 1), 617–624. doi: 10.1053/jhep.2001.27834
- Hannoun, C., Horal, P., and Lindh, M. (2000). Long-term mutation rates in the hepatitis B virus genome. *J. Gen. Virol.* 81(Pt 1), 75–83. doi: 10.1099/0022-1317-81-1-75
- Harrison, A., Lemey, P., Hurler, M., Moyes, C., Horn, S., Pryor, J., et al. (2011). Genomic analysis of hepatitis B virus reveals antigen state and genotype as sources of evolutionary rate variation. *Viruses* 3, 83–101. doi: 10.3390/v3020083
- Hsu, H. Y., Chang, M. H., Hsieh, K. H., Lee, C. Y., Lin, H. H., Hwang, L. H., et al. (1992). Cellular immune response to HBeAg in mother-to-infant transmission of hepatitis B virus. *Hepatology* 15, 770–776. doi: 10.1002/hep.1840150505
- Hsu, Y. S., Chien, R. N., Yeh, C. T., Sheen, I. S., Chiou, H. Y., Chu, C. M., et al. (2002). Long-term outcome after spontaneous HBeAg seroconversion in patients with chronic hepatitis B. *Hepatology* 35, 1522–1527. doi: 10.1053/jhep.2002.33638
- Huang, C. F., Lin, S. S., Ho, Y. C., Chen, F. L., and Yang, C. C. (2006). The immune response induced by hepatitis B virus principal antigens. *Cell. Mol. Immunol.* 3, 97–106. Available online at: <http://www.cmi.ustc.edu.cn/3/2/97.pdf> (Accessed October 11, 2018).
- Imai, M., Nomura, M., Gotanda, T., Sano, T., Tachibana, K., Miyamoto, H., et al. (1982). Demonstration of two distinct antigenic determinants on hepatitis B e antigen by monoclonal antibodies. *J. Immunol.* 128, 69–72.
- Inoue, J., Ueno, Y., Nagasaki, F., Wakui, Y., Kondo, Y., Fukushima, K., et al. (2009). Enhanced intracellular retention of a hepatitis B virus strain associated with fulminant hepatitis. *Virology* 395, 202–209. doi: 10.1016/j.virol.2009.09.028
- Kao, J. H., Chen, P. J., Lai, M. Y., and Chen, D. S. (2004). Hepatitis B virus genotypes and spontaneous hepatitis B e antigen seroconversion in Taiwanese hepatitis B carriers. *J. Med. Virol.* 72, 363–369. doi: 10.1002/jmv.10534
- Kato, H., Orito, E., Gish, R. G., Bzowej, N., Newsom, M., Sugauchi, F., et al. (2002). Hepatitis B e antigen in sera from individuals infected with hepatitis B virus of genotype G. *Hepatology* 35, 922–929. doi: 10.1053/jhep.2002.32096
- Kimbi, G. C., Kramvis, A., and Kew, M. C. (2004). Distinctive sequence characteristics of subgenotype A1 isolates of hepatitis B virus from South Africa. *J. Gen. Virol.* 85(Pt 5), 1211–1220. doi: 10.1099/vir.0.19749-0
- Kramvis, A. (2014). Genotypes and genetic variability of hepatitis B virus. *Intervirology* 57, 141–150. doi: 10.1159/000360947
- Kramvis, A. (2016). The clinical implications of hepatitis B virus genotypes and HBeAg in pediatrics. *Rev. Med. Virol.* 26, 285–303. doi: 10.1002/rmv.1885
- Kramvis, A., Arakawa, K., Yu, M. C., Nogueira, R., Stram, D. O., and Kew, M. C. (2008). Relationship of serological subtype, basic core promoter and precore mutations to genotypes/subgenotypes of hepatitis B virus. *J. Med. Virol.* 80, 27–46. doi: 10.1002/jmv.21049
- Kramvis, A., Bukofzer, S., Kew, M. C., and Song, E. (1997). Nucleic acid sequence analysis of the precore region of hepatitis B virus from sera of southern African black adult carriers of the virus. *Hepatology* 25, 235–240. doi: 10.1002/hep.510250143
- Kramvis, A., Kew, M., and Francois, G. (2005). Hepatitis B virus genotypes. *Vaccine* 23, 2409–2423. doi: 10.1016/j.vaccine.2004.10.045
- Kramvis, A., and Kew, M. C. (1999). The core promoter of hepatitis B virus. *J. Viral Hepat.* 6, 415–427. doi: 10.1046/j.1365-2893.1999.00189.x
- Kramvis, A., and Kew, M. C. (2007). Molecular characterization of subgenotype A1 (subgroup Aa) of hepatitis B virus. *Hepatol. Res.* 37, S27–S32. doi: 10.1111/j.1872-034X.2007.00100.x
- Kramvis, A., Kew, M. C., and Bukofzer, S. (1998). Hepatitis B virus precore mutants in serum and liver of Southern African Blacks with hepatocellular carcinoma. *J. Hepatol.* 28, 132–141. doi: 10.1016/S0168-8278(98)80212-2
- Kuhrober, A., Wild, J., Pudollek, H. P., Chisari, F. V., and Reimann, J. (1997). DNA vaccination with plasmids encoding the intracellular (HBeAg) or secreted (HBeAg) form of the core protein of hepatitis B virus primes T cell responses to two overlapping Kb- and Kd-restricted epitopes. *Int. Immunol.* 9, 1203–1212. doi: 10.1093/intimm/9.8.1203
- Li, J. S., Tong, S. P., Wen, Y. M., Vitvitski, L., Zhang, Q., and Treppe, C. (1993). Hepatitis B virus genotype A rarely circulates as an HBe-minus mutant: possible contribution of a single nucleotide in the precore region. *J. Virol.* 67, 5402–5410.
- Li, K., Zoulim, F., Pichoud, C., Kwei, K., Villet, S., Wands, J., et al. (2007). Critical role of the 36-nucleotide insertion in hepatitis B virus genotype G in core protein expression, genome replication, and virion secretion. *J. Virol.* 81, 9202–9215. doi: 10.1128/JVI.00390-07
- Li, S., Wang, Z., Li, Y., and Ding, G. (2017). Adaptive evolution of proteins in hepatitis B virus during divergence of genotypes. *Sci. Rep.* 7:1990. doi: 10.1038/s41598-017-02012-8
- Liaw, Y. F., and Chu, C. M. (2009). Hepatitis B virus infection. *Lancet* 373, 582–592. doi: 10.1016/S0140-6736(09)60207-5
- Lim, S. G., Cheng, Y., Guindon, S., Seet, B. L., Lee, L. Y., Hu, P., et al. (2007). Viral quasi-species evolution during hepatitis Be antigen seroconversion. *Gastroenterology* 133, 951–958. doi: 10.1053/j.gastro.2007.06.011
- Lok, A. S., Akarca, U., and Greene, S. (1994). Mutations in the pre-core region of hepatitis B virus serve to enhance the stability of the secondary structure of the pre-genome encapsidation signal. *Proc. Natl. Acad. Sci. U.S.A.* 91, 4077–4081. doi: 10.1073/pnas.91.9.4077
- Lythgoe, K. A., Gardner, A., Pybus, O. G., and Grove, J. (2017). Short-sighted virus evolution and a germline hypothesis for chronic viral infections. *Trends Microbiol.* 25, 336–348. doi: 10.1016/j.tim.2017.03.003
- Mahgoub, S., Candotti, D., El Ekiaby, M., and Allain, J. P. (2011). Hepatitis B virus (HBV) infection and recombination between HBV genotypes D and E in asymptomatic blood donors from Khartoum, Sudan. *J. Clin. Microbiol.* 49, 298–306. doi: 10.1128/JCM.00867-10
- Messageot, F., Salhi, S., Eon, P., and Rossignol, J. M. (2003). Proteolytic processing of the hepatitis B virus e antigen precursor. Cleavage at two furin consensus sequences. *J. Biol. Chem.* 278, 891–895. doi: 10.1074/jbc.M207634200
- Milich, D., and Liang, T. J. (2003). Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection. *Hepatology* 38, 1075–1086. doi: 10.1053/jhep.2003.50453
- Milich, D. R. (1997). Influence of T-helper cell subsets and crossregulation in hepatitis B virus infection. *J. Viral. Hepat.* 4 (Suppl. 2), 48–59. doi: 10.1111/j.1365-2893.1997.tb00180.x
- Milich, D. R., Chen, M., Schodel, F., Peterson, D. L., Jones, J. E., and Hughes, J. L. (1997). Role of B cells in antigen presentation of the hepatitis B core. *Proc. Natl. Acad. Sci. U.S.A.* 94, 14648–14653. doi: 10.1073/pnas.94.26.14648
- Milich, D. R., Chen, M. K., Hughes, J. L., and Jones, J. E. (1998). The secreted hepatitis B precore antigen can modulate the immune response to the nucleocapsid: a mechanism for persistence. *J. Immunol.* 160, 2013–2021.
- Milich, D. R., Jones, J. E., Hughes, J. L., Price, J., Raney, A. K., and McLachlan, A. (1990). Is a function of the secreted hepatitis B e antigen to induce immunologic tolerance in utero? *Proc. Natl. Acad. Sci. U.S.A.* 87, 6599–6603. doi: 10.1073/pnas.87.17.6599

- Milich, D. R., and McLachlan, A. (1986). The nucleocapsid of hepatitis B virus is both a T-cell-independent and a T-cell-dependent antigen. *Science* 234, 1398–1401. doi: 10.1126/science.3491425
- Milich, D. R., McLachlan, A., Moriarty, A., and Thornton, G. B. (1987). Immune response to hepatitis B virus core antigen (HBcAg), localization of T cell recognition sites within HBcAg/HBcAg. *J. Immunol.* 139, 1223–1231.
- Milich, D. R., McLachlan, A., Stahl, S., Wingfield, P., Thornton, G. B., Hughes, J. L., et al. (1988). Comparative immunogenicity of hepatitis B virus core and E antigens. *J. Immunol.* 141, 3617–3624.
- Mizokami, M., Orito, E., Ohba, K., Ikey, K., Lau, J. Y., and Gojobori, T. (1997). Constrained evolution with respect to gene overlap of hepatitis B virus. *J. Mol. Evol.* 44 (Suppl. 1), S83–S90. doi: 10.1007/PL00000061
- Mulders, M. N., Venard, V., Njaye, M., Edorh, A. P., Bola Oyefolu, A. O., Kehinde, M. O., et al. (2004). Low genetic diversity despite hyperendemicity of hepatitis B virus genotype E throughout West Africa. *J. Infect. Dis.* 190, 400–408. doi: 10.1086/421502
- Ni, Y. H., Chang, M. H., Chen, P. J., Tsai, K. S., Hsu, H. Y., Chen, H. L., et al. (2007). Viremia profiles in children with chronic hepatitis B virus infection and spontaneous e antigen seroconversion. *Gastroenterology* 132, 2340–2345. doi: 10.1053/j.gastro.2007.03.111
- Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6. doi: 10.1093/protein/10.1.1
- Norder, H., Courouce, A. M., Coursaget, P., Echevarria, J. M., Lee, S. D., Mushahwar, I. K., et al. (2004). Genetic diversity of hepatitis B virus strains derived worldwide: genotypes, subgenotypes, and HBsAg subtypes. *Intervirology* 47, 289–309. doi: 10.1159/000080872
- Orito, E., Mizokami, M., Sakugawa, H., Michitaka, K., Ishikawa, K., Ichida, T., et al. (2001). A case-control study for clinical and molecular biological differences between hepatitis B viruses of genotypes B and C. Japan HBV Genotype Research Group. *Hepatology* 33, 218–223. doi: 10.1053/jhep.2001.20532
- Osiowy, C., Larke, B., and Giles, E. (2011). Distinct geographical and demographic distribution of hepatitis B virus genotypes in the Canadian Arctic as revealed through an extensive molecular epidemiological survey. *J. Viral Hepat.* 18, e11–e19. doi: 10.1111/j.1365-2893.2010.01356.x
- Paraskevis, D., Magiorkinis, G., Magiorkinis, E., Ho, S. Y., Belshaw, R., Allain, J. P., et al. (2013). Dating the origin and dispersal of hepatitis B virus infection in humans and primates. *Hepatology* 57, 908–916. doi: 10.1002/hep.26079
- Raimondo, G., Tanzi, E., Brancatelli, S., Campo, S., Sardo, M. A., Rodino, G., et al. (1993). Is the course of perinatal hepatitis B virus infection influenced by genetic heterogeneity of the virus? *J. Med. Virol.* 40, 87–90. doi: 10.1002/jmv.1890400202
- Revill, P., Yuen, L., Walsh, R., Perrault, M., Locarnini, S., and Kramvis, A. (2010). Bioinformatic analysis of the hepadnavirus e-antigen and its precursor identifies remarkable sequence conservation in all orthohepadnaviruses. *J. Med. Virol.* 82, 104–115. doi: 10.1002/jmv.21645
- Riordan, S. M., Skinner, N., Nagree, A., McCallum, H., McIver, C. J., Kurtovic, J., et al. (2003). Peripheral blood mononuclear cell expression of toll-like receptors and relation to cytokine levels in cirrhosis. *Hepatology* 37, 1154–1164. doi: 10.1053/jhep.2003.50180
- Schlicht, H. J., Salfeld, J., and Schaller, H. (1987). The duck hepatitis B virus pre-C region encodes a signal sequence which is essential for synthesis and secretion of processed core proteins but not for virus formation. *J. Virol.* 61, 3701–3709.
- Schodel, F., Peterson, D., Zheng, J., Jones, J. E., Hughes, J. L., and Milich, D. R. (1993). Structure of hepatitis B virus core and e-antigen. A single precore amino acid prevents nucleocapsid assembly. *J. Biol. Chem.* 268, 1332–1337.
- Shimakawa, Y., Lemoine, M., Njai, H. F., Bottomley, C., Ndw, G., Goldin, R. D., et al. (2016). Natural history of chronic HBV infection in West Africa: a longitudinal population-based study from The Gambia. *Gut* 65, 2007–2016. doi: 10.1136/gutjnl-2015-309892
- Simmonds, P., and Midgley, S. (2005). Recombination in the genesis and evolution of hepatitis B virus genotypes. *J. Virol.* 79, 15467–15476. doi: 10.1128/JVI.79.24.15467-15476.2005
- Stuyver, L., De Gendt, S., Van Geyt, C., Zoulim, F., Fried, M., Schinazi, R. F., et al. (2000). A new genotype of hepatitis B virus: complete genome and phylogenetic relatedness. *J. Gen. Virol.* 81(Pt 1), 67–74. doi: 10.1099/0022-1317-81-1-67
- Sugauchi, F., Kumada, H., Acharya, S. A., Shrestha, S. M., Gamutan, M. T., Khan, M., et al. (2004). Epidemiological and sequence differences between two subtypes (Ae and Aa) of hepatitis B virus genotype A. *J. Gen. Virol.* 85(Pt 4), 811–820. doi: 10.1099/vir.0.79811-0
- Summers, J., and Mason, W. S. (1982). Replication of the genome of a hepatitis B-like virus by reverse transcription of an RNA intermediate. *Cell* 29, 403–415. doi: 10.1016/0092-8674(82)90157-X
- Tanaka, Y., Hasegawa, I., Kato, T., Orito, E., Hirashima, N., Acharya, S. K., et al. (2004). A case-control study for differences among hepatitis B virus infections of genotypes A (subtypes Aa and Ae) and D. *Hepatology* 40, 747–755. doi: 10.1002/hep.20365
- Tatematsu, K., Tanaka, Y., Kurbanov, F., Sugauchi, F., Mano, S., Maeshiro, T., et al. (2009). A genetic variant of hepatitis B virus divergent from known human and ape genotypes isolated from a Japanese patient and provisionally assigned to new genotype J. *J. Virol.* 83, 10538–10547. doi: 10.1128/JVI.00462-09
- Thompson, A., Locarnini, S., and Visvanathan, K. (2007). The natural history and the staging of chronic hepatitis B: time for reevaluation of the virus-host relationship based on molecular virology and immunopathogenesis considerations? *Gastroenterology* 133, 1031–1035. doi: 10.1053/j.gastro.2007.07.038
- Tiollais, P., Charnay, P., and Vyas, G. N. (1981). Biology of hepatitis B virus. *Science* 213, 406–411. doi: 10.1126/science.6264599
- Townsend, K., Sallberg, M., O'Dea, J., Banks, T., Driver, D., Sauter, S., et al. (1997). Characterization of CD8+ cytotoxic T-lymphocyte responses after genetic immunization with retrovirus vectors expressing different forms of the hepatitis B virus core and e antigens. *J. Virol.* 71, 3365–3374.
- Visvanathan, K., Skinner, N. A., Thompson, A. J., Riordan, S. M., Sozzi, V., Edwards, R., et al. (2007). Regulation of Toll-like receptor-2 expression in chronic hepatitis B by the precore protein. *Hepatology* 45, 102–110. doi: 10.1002/hep.21482
- Volz, T., Lutgehetmann, M., Wachtler, P., Jacob, A., Quaas, A., Murray, J. M., et al. (2007). Impaired intrahepatic hepatitis B virus productivity contributes to low viremia in most HBcAg-negative patients. *Gastroenterology* 133, 843–852. doi: 10.1053/j.gastro.2007.06.057
- Wu, S., Imazeki, F., Kurbanov, F., Fukai, K., Arai, M., Kanda, T., et al. (2011). Evolution of hepatitis B genotype C viral quasi-species during hepatitis B e antigen seroconversion. *J. Hepatol.* 54, 19–25. doi: 10.1016/j.jhep.2010.06.018
- Yang, C. Y., Kuo, T. H., and Ting, L. P. (2006). Human hepatitis B viral e antigen interacts with cellular interleukin-1 receptor accessory protein and triggers interleukin-1 response. *J. Biol. Chem.* 281, 34525–34536. doi: 10.1074/jbc.M510981200
- Yousif, M., Mudawi, H., Bakhiet, S., Glebe, D., and Kramvis, A. (2013). Molecular characterization of hepatitis B virus in liver disease patients and asymptomatic carriers of the virus in Sudan. *BMC Infect. Dis.* 13:328. doi: 10.1186/1471-2334-13-328
- Yu, H., Yuan, Q., Ge, S. X., Wang, H. Y., Zhang, Y. L., Chen, Q. R., et al. (2010). Molecular and phylogenetic analyses suggest an additional hepatitis B virus genotype “I”. *PLoS ONE* 5:e9297. doi: 10.1371/journal.pone.009297
- Yu, X., and Mertz, J. E. (1996). Promoters for synthesis of the pre-C and pregenomic mRNAs of human hepatitis B virus are genetically distinct and differentially regulated. *J. Virol.* 70, 8719–8726.
- Yuh, C. H., Chang, Y. L., and Ting, L. P. (1992). Transcriptional regulation of precore and pregenomic RNAs of hepatitis B virus. *J. Virol.* 66, 4073–4084.
- Zhou, Y., and Holmes, E. C. (2007). Bayesian estimates of the evolutionary rate and age of hepatitis B virus. *J. Mol. Evol.* 65, 197–205. doi: 10.1007/s00239-007-0054-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kramvis, Kostaki, Hatzakis and Paraskevis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Hepatitis B Virus Adaptation to the CD8+ T Cell Response: Consequences for Host and Pathogen

Sheila F. Lumley^{1,2}, Anna L. McNaughton¹, Paul Klennerman^{1,2,3}, Katrina A. Lythgoe⁴ and Philippa C. Matthews^{1,2,3*}

¹ Medawar Building for Pathogen Research, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, ² Department of Infectious Diseases and Microbiology, Oxford University Hospitals NHS Foundation Trust, John Radcliffe Hospital, Oxford, United Kingdom, ³ Oxford BRC, John Radcliffe Hospital, Oxford, United Kingdom, ⁴ Nuffield Department of Medicine, Big Data Institute, University of Oxford, Oxford, United Kingdom

OPEN ACCESS

Edited by:

Gkikas Magiorkinis,
National and Kapodistrian University
of Athens, Greece

Reviewed by:

Antonio Bertoletti,
Duke-NUS Medical School,
Singapore

Joerg Timm,

Heinrich-Heine Universität
Düsseldorf, Germany

Robert Thimme,
Albert Ludwigs Universität
Freiburg, Germany

*Correspondence:

Philippa C. Matthews
philippa.matthews@ndm.ox.ac.uk

Specialty section:

This article was submitted
to Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 12 April 2018

Accepted: 25 June 2018

Published: 16 July 2018

Citation:

Lumley SF, McNaughton AL,
Klennerman P, Lythgoe KA and
Matthews PC (2018) Hepatitis
B Virus Adaptation to the
CD8+ T Cell Response:
Consequences for Host
and Pathogen.
Front. Immunol. 9:1561.
doi: 10.3389/fimmu.2018.01561

Chronic viral hepatitis infections are a major public health concern, with an estimated 290 million individuals infected with hepatitis B virus (HBV) globally. This virus has been a passenger in human populations for >30,000 years, and remains highly prevalent in some settings. In order for this endemic pathogen to persist, viral adaptation to host immune responses is pre-requisite. Here, we focus on the interplay between HBV infection and the CD8+ T cell response. We present the evidence that CD8+ T cells play an important role in control of chronic HBV infection and that the selective pressure imposed on HBV through evasion of these immune responses can potentially influence viral diversity, chronicity, and the outcome of infection, and highlight where there are gaps in current knowledge. Understanding the nature and mechanisms of HBV evolution and persistence could shed light on differential disease outcomes, including cirrhosis and hepatocellular carcinoma, and help reach the goal of global HBV elimination by guiding the design of new strategies, including vaccines and therapeutics.

Keywords: hepatitis B virus, evolution, adaptation, diversity, CD8+ T cells, adaptive immunity, human leukocyte antigen

INTRODUCTION

Within hosts, viruses with high mutation rates can rapidly adapt to the selection pressures placed upon them, including natural and vaccine induced immune responses, and antiviral therapy. Hepatitis B virus (HBV) represents a substantial international public health challenge, with an estimated 290 million people chronically infected globally (1). In this review, we explore the evidence for HBV escape from the CD8+ T cell response and examine the influence this process could have on infection outcomes.

Abbreviations: C HBV, core gene; CARs, chimeric antigen receptors; cccDNA, covalently closed circular DNA; GWAS, genome wide association studies; HBV, hepatitis B virus; HBcAg, hepatitis B core antigen; HBeAg, hepatitis B e antigen; HBsAg, hepatitis B surface antigen; HCC, hepatocellular carcinoma; HLA, human leukocyte antigen; KIR, killer-cell immunoglobulin-like receptors; NLG, N-linked glycosylation; ORF, open reading frame; P HBV, polymerase gene; PBL, peripheral blood lymphocyte; RT, reverse transcriptase; S HBV, surface gene; SNP, single-nucleotide polymorphism; TCR, T cell receptor; MHC, major histocompatibility complex; HIV, Human immunodeficiency virus; HCV, Hepatitis C virus.

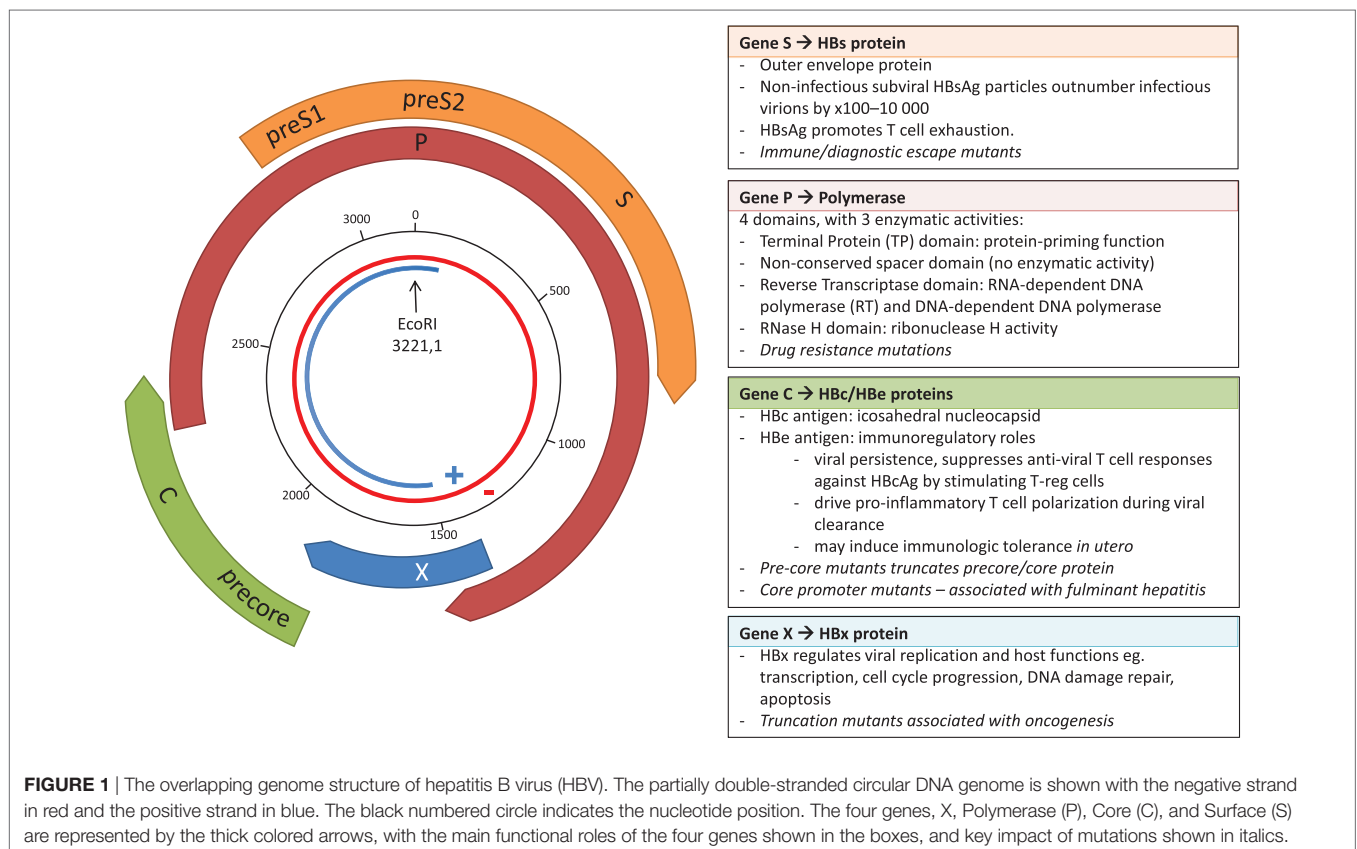
Hepatitis B virus belongs to the Hepadnaviridae family of small, enveloped, primarily hepatotropic viruses. At only 3,200 bp, HBV has one of the smallest genomes of all known pathogenic viruses. The partially double-stranded DNA (dsDNA) circular genome consists of four genes, X, Polymerase (P), Core (C), and Surface (S), and a high proportion of the genome is encoded on overlapping open reading frames (**Figure 1**). During transcription, the partially dsDNA genome is “completed” to form a fully dsDNA molecule, which is subsequently supercoiled to form covalently closed circular DNA (cccDNA). This cccDNA is reverse transcribed by HBV reverse transcriptase (RT), an enzyme lacking 3′–5′ exonuclease proof-reading capacity, and therefore introducing mutations into the HBV genome during each round of replication [in duck hepadnavirus, the mutation rate is estimated at between 0.8×10^{-5} and 4.5×10^{-5} substitutions per nucleotide per replication (2)]. The mutations generated result in a viral quasispecies, comprised of dominant genotype(s) surrounded by clouds of closely related HBV variants.

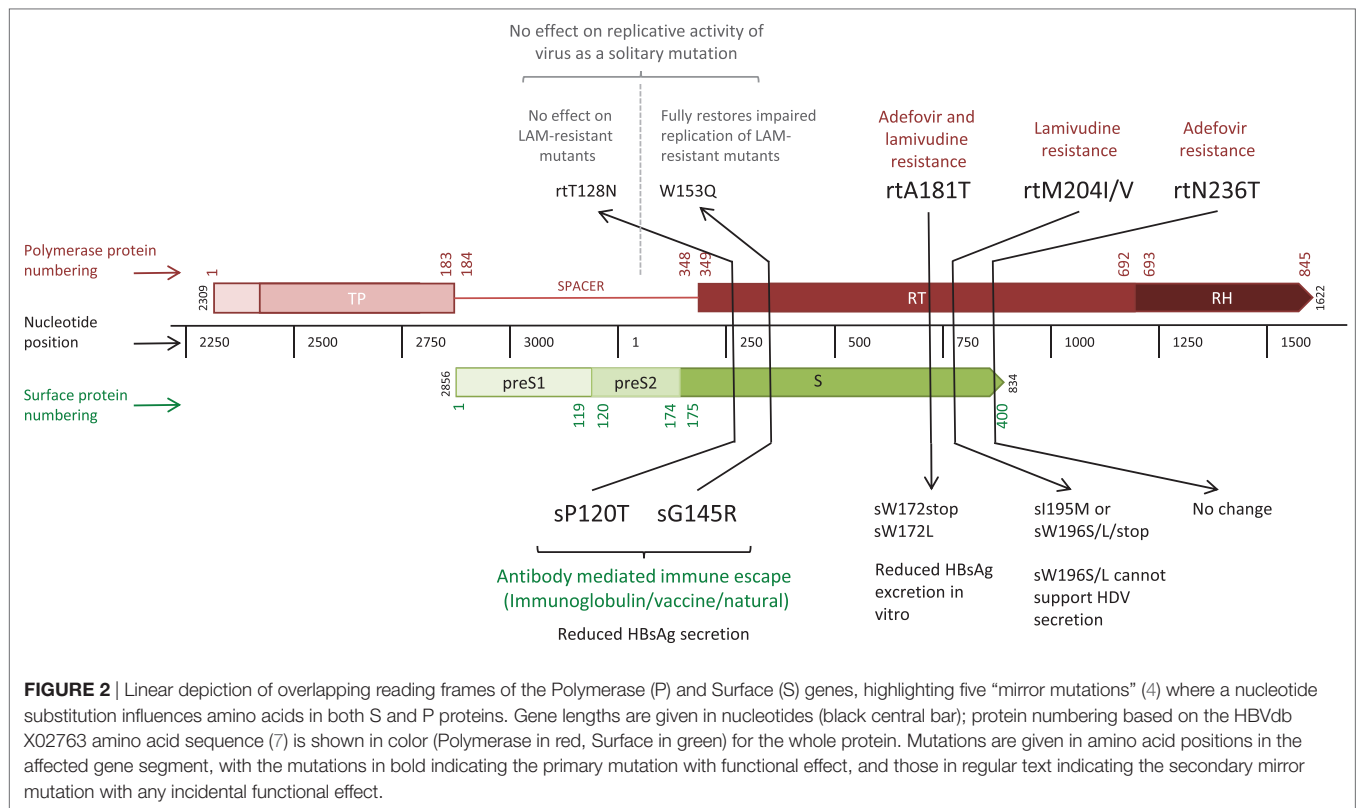
The error prone RT, coupled with high rates of HBV replication [estimated at between 200 and 1,000 virions/hepatocyte/day at the peak of infection (3)] results in the production of a large number of virions harboring mutations. The vast majority of mutations are likely to be deleterious, some are neutral, and a minority provide the virus with a potential selective advantage, such as escape from CD8+ T cell-mediated responses. However, HBV polymorphisms are constrained by the overlapping reading frame structure of the genome, since the majority

of mutations can simultaneously affect multiple genes [these have been described as “mirror” mutations (4), **Figure 2**]. Mutations that are neutral or beneficial for one protein might be detrimental for another. Accordingly, overlapping regions of the HBV genome generally have less diversity compared to non-overlapping regions (5) and the within-host rate of evolution at overlapping regions is about half that of non-overlapping regions (6).

Current vaccination and treatment approaches are hindered by poor diagnosis and access to treatment, drug and vaccine escape mutants, viral rebound on treatment cessation or immunosuppression, and lack of curative therapy (8). To make a significant impact on HBV prevalence, parallel improvements in diagnostics, treatment, and prevention are required; ultimately, new immunotherapeutic strategies may be key to the success of elimination. Developing a more robust picture of the extent, nature, and significance of the interplay between the virus and the host CD8+ T cell response is an important avenue of enquiry, enabling us to predict and tailor therapeutic interventions that may be beneficial in mediating control or clearance of chronic infection.

A robust body of data has been assimilated over the past few decades for HIV and HCV, informing significant understanding of the nature and impact of CD8+ T cell-mediated immune control and escape (**Table 1**). For HBV, there is a relative paucity of such evidence but the field could be advanced by similar approaches. We have therefore set out to assimilate the evidence





for viral adaptation to the host CD8+ T cell response in HBV infection, and to consider the significance of this adaptation both to viral fitness and function, and to host outcomes. Finally, we highlight gaps in our current understanding and knowledge, in order to provide foundations for ongoing research efforts.

THE IMMUNOLOGICAL BASIS FOR ESCAPE

Acute and chronic HBV infections are associated with functionally different CD8+ T cell responses (Table 1). Acute, self-resolving infections are characterized by functionally efficient, multi-specific antiviral CD8+ T cell responses which are sustained after viral clearance (9). Both non-cytolytic and cytolytic mechanisms have been implicated (22). In contrast, chronic infection is typically characterized by a lack of protective T cell memory maturation and exhausted HBV-specific CD8+ T cell responses (22–24).

Th1-polarized CD4+ T cells regulate and maintain CD8+ T cell responses and contribute to HBV clearance (80). Genome wide association studies (GWAS) have linked a range of human leukocyte antigen (HLA) class II alleles with disease outcomes. CD4+ responses are associated with vaccine responses (81) and clearance of acute infection (82, 83). Host HLA class II genotype has also been linked to treatment response (84) and to risk of developing hepatocellular carcinoma (HCC) (85). CD4+ CD25+ regulatory T cells suppress the activation, proliferation, and interferon- γ production of both CD4+ and CD8+ T cells in chronic HBV infection (86, 87).

The highly polymorphic HLA class I genes are thought to be an important host factor for viral control, contributing to differences in HBV outcome observed globally. Host HLA polymorphisms and different HBV genotypes have been demonstrated to influence the rate of disease progression and the long-term outcome of HBV infection (66, 88, 89). However, HBV can subvert various multiple steps of the CD8+ T cell antigen processing and presentation pathway to evade detection by the host (Figure 3, boxes 1–5). Thus, while all individuals with chronic HBV infection are at risk of increased progression to cirrhosis and HCC, individual outcomes depend on the interplay between host, viral, and environmental factors. In addition to HLA genes, other factors are implicated in disease outcome including age and duration of infection, other host genetic factors (90), and exposure to hepatotoxins (91).

MECHANISMS OF HBV ESCAPE FROM CD8+ T CELL RESPONSES

Antigen-Processing Escape Mutants

The amino acids flanking viral epitopes are important for effective antigen processing; mutations in these regions may impair proteasomal processing of the epitopes and are recognized in both HCV and HIV as a mechanism of CD8+ T cell escape (100, 101). Likewise, mutations altering the processing of HBV epitopes could be relevant for HBV escape from the CD8+ T cell-mediated immune response (Figure 3, box 1), however, none have been identified at present, potentially

TABLE 1 | Strands of evidence for the significance of the CD8+ T cell response in control/clearance of infection with blood-borne viruses.

Evidence for role of CD8+ T cell response in control of infection	HBV (details and citations ^a)	HCV (details and citations ^a)	HIV (details and citations ^a)
CD8+ T cell responses in acute infection associated with control and/or clearance	Functionally efficient, multi-specific antiviral CD8+ T cell responses (9)	CD8+ T cell responses are associated with clearance (10)	Inverse relationship between magnitude of acute CD8+ T cell response in acute infection and subsequent viral setpoint (11, 12)
<i>In vitro</i> killing assays demonstrate efficacy/potency of CD8+ T cell response	HBV-specific CD8+ T cells reduce viral loads in HBV-infected HepG2 (hNTCP) cells (13) Strong cytolytic response in woodchuck model (14)	<i>In vitro</i> system using replicons demonstrates HCV-specific CD8+ T cells strongly inhibit viral replication through cytolytic and non-cytolytic mechanisms in a dose-dependent manner (15) Rodent pegiviruses <i>in vitro</i> system shows promise (16)	CD8+ T cells targeting Gag show superior <i>in vitro</i> suppression of HIV replication (17)
Depletion of CD8+ T cell subsets impacts viremic control <i>in vitro</i> and <i>in vivo</i>	CD8+ depletion prolongs infection and delays viral clearance in chimpanzees (18)	Viral infection is prolonged in chimpanzees that previously cleared the infection after CD8+ T cells depletion (19)	Depletion of CD8+ T cell populations is associated with outgrowth of virus <i>in vitro</i> (20) and in the simian model <i>in vivo</i> (21)
Disease progression is associated with T cell exhaustion in chronic infection	Lack of protective T cell memory maturation and exhausted HBV-specific CD8+ T cell responses are seen in chronic infection (22–25) Blockade of inhibitory receptors PD-1 (26), CTLA-4 (27, 28), and Tim-3 (29) partly improve HBV-specific CD8+ T cell function <i>in vitro</i>	High PD-1 expression is associated with chronic infection (30)	PD-1 expression on HIV-specific T cells is associated with T cell exhaustion and disease progression (31)
CD8+ T cell responses may be required to maintain latency; viral mutations are associated with rebound or reactivation	Mutations identified in both neutralizing antibody targets and T cell epitopes are associated with reactivation (32–34)	Latency and reactivation are not typically associated with HCV infection	CD8+ T cell responses are associated with maintenance of latency, including during therapy, (35) and reviewed in Ref. (12) Persistence of escape mutations in viral reservoir can lead to reactivation (36)
Modulation of cell surface HLA expression as a mechanism of immune evasion	Variable surface expression of HLA is seen in human (37–39) and woodchuck hepatitis virus model (40)	Variable surface expression of class I and class II (41, 42)	HIV Nef is associated with downregulation of cell surface HLA expression (43, 44)
Evidence of efficacy of T cell vaccines	Heterologous prime-boost vaccines are promising (45, 46) Vaccine strategies assessed in woodchuck model (47)	Evidence from animal models (48) and human vaccine constructs are in development (49)	Vaccines expand the cellular immune response in rhesus monkeys (50) Reviewed in Ref. (51, 52)
HLA class I “footprints”—specific HLA alleles associated with viral polymorphisms identified, CD8+ T cell epitope escape and reversion occurs	Footprints identified in all four HBV genes (53–58) Further data are required to develop appropriate animal models (59)	Viral footprints in HLA epitopes display escape and reversion mutants identified (60–62)	CD8+ T cell escape mutations are associated with loss of viremic control; reversion to wild type is observed on transmission to an HLA-mismatched recipient [literature reviewed in Ref. (63)] SIV escape mutations selected in macaque model can influence disease outcome (64, 65)
Particular HLA class I alleles associated with disease control	HLA-A genotype is associated with HBeAg status (66) A database of HBV epitopes, “hepitopes,” highlights instances in which outcome is associated with specific T cell responses (53)	HLA-A*03 and HLA-B*27 alleles are protective in HCV infection (61, 67)	Strong association with HLA-B genotype (68); e.g., control epitomized by CD8+ responses restricted by HLA-B*57 and HLA-B*27 (69–71)
GWAS highlighting importance of HLA class I genes in control of chronic infection	Associations identified but mechanisms lacking (72–74). HLA-A*0301 is associated with clearance (75)	HLA-A*03, HLA-B*27 is associated with control (61, 76, 77)	Disease control is associated with SNPs in MHC region of human chromosome (78, 79), including specific association with HLA-B*57 expression

^aThe citations within this table aim to provide a robust overview of the evidence, using a combination of strong examples from the primary literature together with selected review articles that summarize specific aspects of this topic.

HBV, hepatitis B virus; HBeAg, hepatitis B e antigen; HLA, human leukocyte antigen; GWAS, Genome wide association studies; SNPs, single-nucleotide polymorphisms; MHC, major histocompatibility complex; HCV, Hepatitis C virus; HIV, Human immunodeficiency virus.

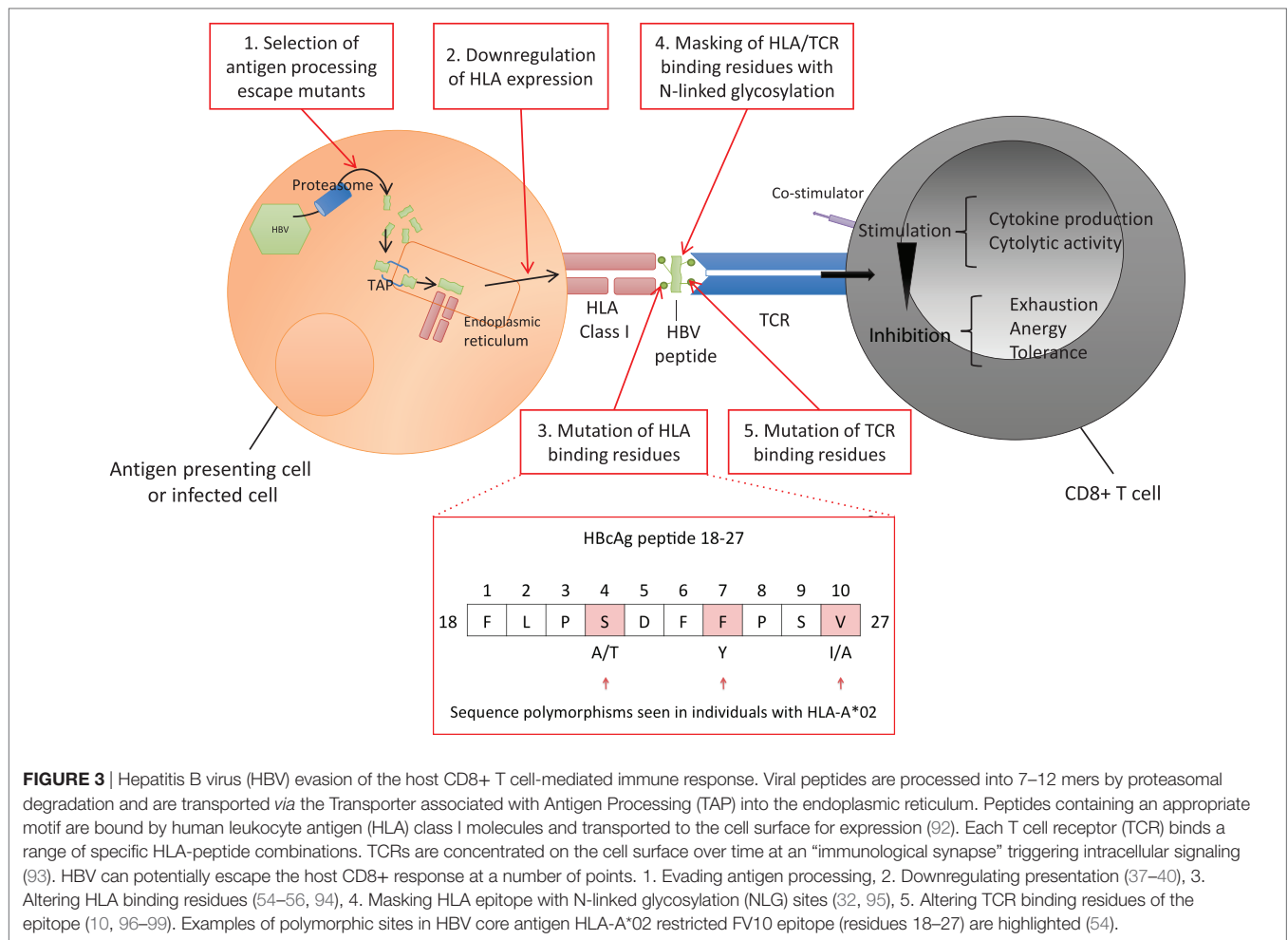


FIGURE 3 | Hepatitis B virus (HBV) evasion of the host CD8+ T cell-mediated immune response. Viral peptides are processed into 7–12 mers by proteasomal degradation and are transported via the Transporter associated with Antigen Processing (TAP) into the endoplasmic reticulum. Peptides containing an appropriate motif are bound by human leukocyte antigen (HLA) class I molecules and transported to the cell surface for expression (92). Each T cell receptor (TCR) binds a range of specific HLA-peptide combinations. TCRs are concentrated on the cell surface over time at an “immunological synapse” triggering intracellular signaling (93). HBV can potentially escape the host CD8+ response at a number of points. 1. Evading antigen processing, 2. Downregulating presentation (37–40), 3. Altering HLA binding residues (54–56, 94), 4. Masking HLA epitope with N-linked glycosylation (NLG) sites (32, 95), 5. Altering TCR binding residues of the epitope (10, 96–99). Examples of polymorphic sites in HBV core antigen HLA-A*02 restricted FV10 epitope (residues 18–27) are highlighted (54).

due to the focus on mutations lying within HLA-restricted epitopes rather than in the flanking regions.

Virus Peptides Regulate Surface HLA Expression

Virus-induced changes in HLA class I surface expression play an important role in viral pathogenesis and persistence (Figure 3, box 2). CD8+ T cells recognize HBV-infected hepatocytes through presentation of HLA class I HBV epitopes on the cell surface (13), however, this expression can be upregulated or downregulated. Decreased presentation of class I MHC molecules on hepatocytes and lymphoid cells is described in the woodchuck hepatitis virus model (40). Changes in surface HLA have also been described in human HBV infection (37), for example, lower HLA class I has been associated with hepatitis B e antigen (HBeAg)-positive vs HBeAg-negative status (38, 39). Interestingly, these studies are three decades old and have not been replicated in the more recent literature. Downregulation of HLA class II molecules by pre-core mutants has also been described during chronic HBV infection (102). Mutations altering the processing or presentation of HBV HLA class I epitopes, although not conclusively demonstrated, could hypothetically

be relevant for escape from the CD8+ T cell-mediated immune response in human infection (Figure 3, box 2).

Selective Mutation of HLA-Binding Residues

Immune escape by selective mutation of HLA-binding residues within HBV CD8+ epitopes is one of the most commonly identified mechanisms for viral CD8+ immune escape (Table 1; Figure 3, box 3). Evidence of this escape mechanism in HBV has emerged through the identification of HLA class I “footprints” (94), mutations that are significantly enriched in patients with certain HLA class I alleles. Older literature was conflicting regarding the frequency and significance of such footprints in HBV (96, 103), but HLA footprints have subsequently been identified in all four HBV genes and mapped to known or predicted HLA epitopes [Table 1 (53)]. In some cases, these mutations result in altered peptide-HLA binding scores, providing a plausible mechanism for HBV immune escape, but have so far only been identified using cross-sectional data (54, 55, 96). However, the pattern of escape is consistent across populations with divergent HLA haplotypes and different HBV genotypes, for example, genotypes B and C in a cohort of Chinese-origin

patients (56), New Zealand-resident Tongans with chronic HBV genotype C3 infections (57) and Iranian patients with genotype D infection (58). Importantly, core mutants in patients with chronic genotype A or D infection have been confirmed to impair CD8+ T cell IFN- γ secretion *in vitro* (54), indicating that these mutants could play an *in vivo* role in immune escape. These studies are limited by the sequencing methods (in which a pre-defined number of variants is typically selected for cloning and sequencing) reducing sensitivity for detection of low-abundance variant detection compared to newer ultra-deep sequencing methods.

Further studies have focused on identifying regions of the HBV genome with high within-host nucleotide diversity, and high rates of nonsynonymous substitutions, to determine which regions may be under strong HLA-mediated selection pressure. In a longitudinal study of eight HBeAg-negative asymptomatic HBV carriers, followed over 25 years, the ratio of synonymous to nonsynonymous mutations (dS/dN) in the core gene was low, suggesting high rates of positive selection, although no specific HLA-restricted epitopes were identified (104). Another study demonstrated low dS/dN ratios in patients with sustained vs unsustained viral control following treatment, with specific surface antigen polymorphisms lying within HLA class I epitopes identified [sV14G, sF20S, sT45I, sI213L (105)], although this was not confirmed either with HLA genotyping or through demonstration of a functional impact on T cell recognition.

Epitope Masking With N-Linked Glycosylation (NLG)

N-linked glycosylation is a post-translational modification that plays an established role in the antigenicity and infectivity of viruses (106, 107). NLG can mask immunogenic epitopes, interfering with antibody recognition of hepatitis B surface antigen (HBsAg), leading to immune and diagnostic escape (32, 95). It can also impact on HBV virion secretion, likely by altering the ability of envelope proteins to interact with the capsid surface (108–110). The number of NLG sites correlates with disease state, with increased NLG reported in patients with reactivated HBV vs chronic infection, and in those with sustained vs unsustained response off treatment (32, 111). Although the role of NLG in HBV evasion of CD8+-mediated T cell immunity is yet to be determined, it could potentially provide another strategy for immune escape by interfering with the binding of an HBV epitope to an HLA molecule, or the binding of an HLA-antigen complex to a cognate T cell receptor (TCR) (Figure 3, box 4).

Alteration of TCR Recognition

Engagement of the TCR with HLA class I/peptide complexes on antigen-presenting cells is key to activating CD8+ T cells; therefore, mutations in the TCR contact residues of an epitope can lead to immune escape. Immunodominance of viral epitopes is not simply determined by the amino acid sequence of the peptide and its binding affinity, but also depends on the peptide concentration and T cell clone, with the same HBV peptide able to induce different signaling cascades in different CD8+ T cell clones (97). Antagonist functions may provide HBV with a

means of immune escape (Figure 3, box 5). Specifically, certain CD8+ T cell epitopes in hepatitis B core antigen (HBcAg) (97) and HBsAg (98) act as TCR antagonists, binding the TCR and inhibiting the CD8+ T cell response. HBeAg may promote HBV chronicity by inducing CD8+ T cell tolerance. However, the underlying mechanisms driving this immune state in humans remain to be elucidated. Indeed, the mechanism may not involve presentation of an HLA class I-restricted epitope, as currently no epitopes have been identified that are unique to the pre-core sequence of HBcAg (a ~29 amino acid stretch not shared with HBcAg) (112).

Chronic HBV infection is characterized by an exhausted CD8+ T cell phenotype associated with reduced cytotoxic activity and enhanced expression of inhibitory markers. TCR binding in the presence of high HBsAg levels induces T cell exhaustion, characterized by poor effector cytotoxic activity, impaired cytokine production and sustained expression of multiple inhibitory receptors. A hierarchy of co-inhibitory receptors, dominated by PD-1, act synergistically to promote CD8+ tolerance. The degree of T cell impairment also depends on suppressive cytokines, interaction with other T cell subsets, and stage of T cell differentiation (113–117). T cell exhaustion is (at least partly) reversible; blockade of inhibitory receptors including PD-1 (26, 117), CTLA-4 (27), and Tim-3 (29) partly improve HBV-specific CD8+ T cell function *in vitro*. In addition, therapy with nucleot(s)ide analogs may lead to a modest reconstitution of HBV-specific T cell function (118). Although this restoration is transient (119), these CD8+ T cells can be associated with viral control upon therapy cessation (120).

ESCAPE OVER SPACE AND TIME

Kinetics of Escape

Evidence of immune-mediated selection has been found in HBV infection (53–56), although the kinetics of immune escape are yet to be robustly delineated. Longitudinal samples from the same individuals form the ideal dataset to address questions about the changes in viral sequence and diversity over time, but this has rarely been undertaken for HBV infection. One longitudinal study of HBV evolution following acquisition from a single source demonstrated an expansion and contraction of HBV diversity, with maximum diversity coinciding with peak viremia, and a predominance of nonsynonymous mutations with greatest diversity in the core gene (3). Further longitudinal data are required to unpick the timing and kinetics of viral evolution.

An area of HBV kinetics that has received some attention is the scenario of HBeAg loss. It is hypothesized that the change from HBeAg-positive to HBeAg-negative occurs by one of two mechanisms:

- (i) Antibody-mediated control (121, 122) usually associated with low HBV DNA levels. This situation is most likely to be characterized by low viral sequence diversity, although the low viral loads make this difficult to study given the limits of sensitivity of next generation sequencing approaches.

- (ii) Selection of pre-core and promotor mutations (123), reducing or eliminating HBeAg production. In this case, HBeAg-negative status is associated with an increase in evolutionary rate and therefore with increased sequence diversity (6, 54, 56, 57, 104, 111, 124–126). The cause/effect relationship between the increased evolutionary rate and the shift in immune activity is unclear. The higher viral mutation rate could lead to the occurrence of stochastic mutations, generating new T cell epitopes that disrupt immune tolerance, or could be the consequence of increased immune reactivity driving escape mutants.

Compartment-Specific Evolution

Compartment-specific evolution has been described for chronic viruses, including HIV (127, 128) and HCV (129, 130), although the evidence for HBV is very limited to date. The practical barriers to sampling tissue compartments longitudinally from the same patient make it difficult to assess the co-evolution of genetically distinct subpopulations over time [as is likely the case for HIV in the genital tract (131)].

Although *hepadnaviruses* are characteristically hepatotropic, HBV DNA is also found in a range of other tissues, including lymphatic cells. In the woodchuck model, life-long replication- and transmission-competent viruses persist in lymphocytes (132). However, it is difficult to demonstrate that hepatitis B virions isolated from different compartments in humans are replication and transmission competent, without a viable method of culturing autologous virus. There are some data to suggest that peripheral blood lymphocytes (PBLs) can support viral replication (133), but secretion of HBeAg and HBsAg from liver macrophages has not been detected [Lucifora, unpublished data, referred to in (134)].

HBV may undergo independent evolution in different tissue compartments, leading to compartmentalization of viral subpopulations (135–137); for example, HBV variants isolated from PBLs may be specifically adapted to this environment (137), potentially harboring relevant immune-escape mutants (135, 137). It has been hypothesized that compartment-specific mutants may serve as a source of reactivation or transmission and have been implicated in reinfection post liver transplant (138), mother to child transmission (137, 139), fulminant hepatic failure in the context of HIV co-infection (140), and antiviral escape (141).

Further work is required to confirm whether HBV does harbor replication and transmission competent viruses in cells other than hepatocytes. If this is confirmed, understanding host-virus dynamics at the compartmental level, studying the emergence of immune and antiviral escape mutants and the factors contributing to persistence and transmission will be crucial for developing improved therapeutics for HBV control.

FUNCTIONAL IMPACT OF ESCAPE MUTATIONS ON HBV

The primary functional impact of mutations within HLA class I-restricted T cell epitopes is to alter the frequency and/or

functionality of the CD8+ T cell immune response. These mutations may have additional impact on the viral replication cycle and treatment response in the following ways:

- (i) Altered structure/function of the viral protein containing the mutated immune epitope (**Figure 2**). This is seen in HBV escape from B cell immunity in which mutations within the S gene are associated with diagnostic failure (HBsAg mutants are not detected by the immunological assay) and treatment failure due to changes in assembly and secretion, virion formation and HBV infectivity (109, 142, 143).
- (ii) Effects of a “mirror” escape mutation on the overlapping gene (**Figure 2**). A mutation leading to amino acid substitutions in both the P and S genes, can simultaneously affect replicative capacity, drug resistance, and immunogenicity (142–145). Furthermore, deletions in regions such as the spacer region of Pol, neutral to Pol function, may lead to loss of immune epitopes in the overlapping preS1–preS2 region.
- (iii) Impact of compensatory mutations mitigating for (i) and (ii). This is seen in replication deficient CD8+ immune escape mutants in HCV (146) and HIV (147, 148) but has not yet been identified in HBV.

The full range of functional impacts of HBV CD8+ immune escape mutants has not been comprehensively explored. Understanding the functional impact of mirror and compensatory mutations that are associated with CD8+ T cell-mediated selection may lead to further insights into the host–virus interaction.

CLINICAL IMPACT OF VIRUS AND HOST POLYMORPHISMS ON HOST OUTCOME

Impact of HBV Mutations on Reactivation

Hepatitis B virus reactivation as a consequence of immunosuppression has emerged as an important issue across a wide range of clinical settings [as previously reviewed (149, 150)]. Reactivation is seen secondary to immunosuppressive therapy for cancer, in particular in the context of therapy with rituximab and fludarabine (33, 151), solid organ transplantation (150), bone marrow transplantation (152), and autoimmune disease [especially with infliximab treatment (153–155)], highlighting that HBV reactivation is associated with a general defect of HBV-specific T cell control. Reactivation has also been documented in immunocompetent patients despite the presence of neutralizing antibodies (156).

Specific mutations associated with HBV reactivation have been identified in both neutralizing antibody targets and T cell epitopes (32–34). In a study of 29 patients with HBV reactivation, 75% of HBV-reactivated patients (vs 3% of chronic HBV controls) carried HBsAg mutations localized in immune-active HBsAg regions, and 5 of 13 identified HBsAg mutations were localized in HLA-restricted T cell epitopes [either class I (sC48G, sV96A, sL175S, and sG185E) or class II (sS171F)] (32). This suggests that in addition to an iatrogenic trigger for reactivation during immunosuppressive therapy, viral sequence can be a contributory factor as a result of CD8+ immune escape mutants.

Impact of Host HLA Class I Haplotype on HBV Infection Outcome

GWAS approaches have linked various single-nucleotide polymorphisms in the HLA class II region with a range of infection outcomes, but there is a lack of such robust evidence for the involvement of HLA class I genes (72, 73, 75). One study identified a relationship between class I HLA-A genotype and HBeAg status (66), suggesting a role for genes at this locus in control of infection. However, confirmation of HLA associations can be difficult due to the variability in study design and methodologies and the small, heterogeneous populations sampled. Furthermore, the mechanisms for these HLA class I associations with disease outcome are poorly understood. Differences in antigen presentation, TCR binding leading to changes in T cell activation, and altered cytokine production may be responsible, either individually or in combination. Effects of linkage disequilibrium with other important neighboring loci, such as HLA class II or killer-cell immunoglobulin-like receptors genes, cannot be excluded. Functional studies are required to determine the basis for these associations.

Impact of HBV Adaptation on Control Strategies

It is likely that to achieve elimination in line with global public health goals (157), new therapies targeting either the host immune system or the HBV replication cycle will be needed. Specific immunotherapies are under development, targeting both the innate and adaptive immune system, which aim to eliminate (or stably suppress) HBV replication (158).

T cell-based immune therapies are attractive options for HBV control (159). Strategies broadly take two approaches, either

aiming to restore functionality and increase the quantity of existing defective host T cells with vaccines and checkpoint inhibitors, or to mimic the T cell response mounted during naturally resolving acute HBV infection by the adoptive transfer of HBV-specific T cells. Adoptive T cell therapy renders T cells HBV-specific by expression of natural HLA-restricted TCRs or HLA-independent chimeric antigen receptors on the T cell surface. Although natural TCRs have the advantage of activating the T cell response in a physiological way, therapy is potentially complicated by the need to match TCR to host HLA alleles, although some cross-reactivity may occur (160).

Given that HBV is able to evade natural immunity (32, 149, 156, 161, 162), vaccine-induced immunity (163, 164), and antiviral therapy (145, 165–169), it should be anticipated that HBV has the potential to mutate and escape from immunotherapeutic control. This is a vital consideration in the development of new HBV control strategies, and good knowledge of the full range of escape strategies should allow us to predict and potentially mitigate this. Care must be taken when developing T cell immunotherapies and polyepitope vaccines as immunodominance is a complex function of the nature and context of the epitope within the peptide, the TCR, the T cell clone, and the environment (98). The HBV literature is skewed toward the investigation of certain populations with specific HBV genotypes and HLA haplotypes, as highlighted in the “Hepitopes” database, a catalog of HLA class I epitopes in HBV, in which a disproportionate 44% of reported CD8+ T cell epitopes are HLA-A*02 restricted (53). The effect of using a polyepitope vaccine or re-directing T cells against peptides presented by discordant HLA alleles needs to be considered. This might inadvertently occur by using a vaccine or T cell-based immunotherapy based on key epitopes from a different genotype to that prevalent in the population to which it is

TABLE 2 | Areas for future focus in determining the nature and characteristics of the CD8+ T cell response to hepatitis B virus (HBV).

Approach	Rationale
Comprehensive case-finding and diagnostic strategy	To build a more complete picture of global HBV prevalence and distribution
Matched host genetic, clinical outcome and viral sequencing data, from populations with varying human leukocyte antigen (HLA) alleles and different infecting HBV genotypes, supported by improved case-finding strategy	To study the differential impact of viral and host genetics on host outcome. An unbiased approach is required to determine HLA genes that may be associated with specific clinical outcomes To investigate mechanisms behind reactivation in the immunosuppressed
Next generation sequencing of full-length HBV genome, including longitudinal deep sequencing data	To study the kinetics of viral transmission, evolution and escape, and the role played by viral quasispecies—higher sensitivity for low-abundance variant detection To identify novel mechanisms for CD8+ immune escape, e.g., antigen-processing escape mutants, regulation of HLA expression, N-linked glycosylation (Figure 3)
Culture systems for autologous HBV	To allow functional impact of patient-isolated HBV mutant strains to be studied, to determine fitness impact of the primary mutation, “mirror mutations” and compensatory mutations on replication, transmission, drug resistance immunogenicity, and clinical outcomes
Determination of 3D crystal structures for HBV proteins	To allow assessment of structural impact of viral polymorphisms, including consequences of immune and drug-mediated escape mutations
Comprehensive functional T cell studies	To understand how HLA class I escape mutants impact T cell function and the impact of viral and host genotype
Compartment-specific sampling to include liver and lymphoid tissue	To determine the presence of replication and transmission competent compartment-specific mutants and their dynamics including emergence of immune and antiviral escape mutants

delivered, and may produce functionally incompetent T cells, unable to recognize the infectious virus strain when used, or potentially lead to immunopathology. Since knowledge about certain host/virus interactions is under-represented, further studies will be required to define the full range of CD8+ T cell epitopes presented by HLA alleles, the antiviral functions of the corresponding CD8+ T cells in each compartment, the potential for generation of immune escape mutants and the impact these have on the immune response.

CHALLENGES

The understanding of HBV escape from the CD8+ immune response is lagging behind that of HIV and HCV. The field struggles with a lack of comprehensive literature, small datasets that can lead to conflicting results, differences in approaches to classifying patient groups into poor/outdated descriptions of “phases” of infection, over-reliance on serostatus, and lack of longitudinal follow-up and deep sequence data. Establishing the role of compartmentalization in infection is complex, with clinical samples scarce due to the risk associated with liver biopsy. These challenges are exacerbated by an under-resourcing of clinical and research approaches in many of the settings where HBV is endemic (8).

FUTURE FOCUS

There are many unanswered questions in the field of HBV and CD8+ immunity. In **Table 2**, we highlight gaps in our current understanding and knowledge, suggest desirable methods to develop, datasets to collate, and questions to be answered in order to provide foundations for ongoing research efforts.

REFERENCES

1. Polaris Observatory Collaborators. Global prevalence, treatment, and prevention of hepatitis B virus infection in 2016: a modelling study. *Lancet Gastroenterol Hepatol* (2018) 3(6):383–403. doi:10.1016/S2468-1253(18)30056-6
2. Pult I, Abbott N, Zhang YY, Summers J. Frequency of spontaneous mutations in an avian hepadnavirus infection. *J Virol* (2001) 75(20):9623–32. doi:10.1128/JVI.75.20.9623-9632.2001
3. Whalley SA, Murray JM, Brown D, Webster GJ, Emery VC, Dusheiko GM, et al. Kinetics of acute hepatitis B virus infection in humans. *J Exp Med* (2001) 193:847–54. doi:10.1084/jem.193.7.847
4. Sheldon J, Soriano V. Hepatitis B virus escape mutants induced by antiviral therapy. *J Antimicrob Chemother* (2008) 61(4):766–8. doi:10.1093/jac/dkn014
5. Mizokami M, Orito E, Ohba K, Ikeo K, Lau JYN, Gojobori T. Constrained evolution with respect to gene overlap of hepatitis B virus. *J Mol Evol* (1997) 44(1 Suppl):83–90. doi:10.1007/PL00000061
6. Harrison A, Lemey P, Hurler M, Moyes C, Horn S, Pryor J, et al. Genomic analysis of hepatitis B virus reveals antigen state and genotype as sources of evolutionary rate variation. *Viruses* (2011) 3(2):83–101. doi:10.3390/v3020083
7. Hayer J, Jadeau F, Deléage G, Kay A, Zoulim F, Combet C. *The Hepatitis B Virus Database (HBVdb) – X02763 [Internet]*. Available from: <https://hbvdb.ibcp.fr> (Accessed: April, 2018).
8. O'Hara GA, McNaughton AL, Maponga T, Jooste P, Ocamo P, Chilengi R, et al. Hepatitis B virus infection as a neglected tropical disease. *PLoS Negl Trop Dis* (2017) 11(10):e0005842. doi:10.1371/journal.pntd.0005842
9. Rehmann B, Fowler P, Sidney J, Person J, Redeker A, Brown M, et al. The cytotoxic T lymphocyte response to multiple hepatitis B virus polymerase epitopes during and after acute viral hepatitis. *J Exp Med* (1995) 181(3):1047–58. doi:10.1084/jem.181.3.1047
10. Chen MT, Billaud J-N, Sällberg M, Guidotti LG, Chisari FV, Jones J, et al. A function of the hepatitis B virus precore protein is to regulate the immune response to the core antigen. *Proc Natl Acad Sci U S A* (2004) 101(41):14913–8. doi:10.1073/pnas.0406282101
11. Streeck H, Nixon DF. T cell immunity in acute HIV-1 infection. *J Infect Dis* (2010) 202(Suppl 2):S302–8. doi:10.1086/655652
12. McBrien JB, Kumar NA, Silvestri G. Mechanisms of CD8(+) T cell-mediated suppression of HIV/SIV replication. *Eur J Immunol* (2018) 48(6):898–914. doi:10.1002/eji.201747172
13. Hoh A, Heeg M, Ni Y, Schuch A, Binder B, Hennecke N, et al. HBV-infected HepG2hNTCP cells serve as a novel immunological tool to analyze the antiviral efficacy of CD8+ T cells in vitro. *J Virol* (2015) 89(14):7433–8. doi:10.1128/JVI.00605-15
14. Frank I, Budde C, Fiedler M, Dahmen U, Viazov S, Lu M, et al. Acute resolving woodchuck hepatitis virus (WHV) infection is associated with a strong cytotoxic T-lymphocyte response to a single WHV core peptide. *J Virol* (2007) 81(13):7156–63. doi:10.1128/JVI.02711-06
15. Jo J, Aichele U, Kersting N, Klein R, Aichele P, Bisse E, et al. Analysis of CD8+ T-cell-mediated inhibition of hepatitis C virus replication using a novel immunological model. *Gastroenterology* (2009) 136(4):1391–401. doi:10.1053/j.gastro.2008.12.034
16. Billerbeck E, Wolfisberg R, Fahnøe U, Xiao JW, Quirk C, Luna JM, et al. Mouse models of acute and chronic hepatitis B virus infection. *Science* (2017) 357(6347):204–8. doi:10.1126/science.aal1962

SUMMARY

Hepatitis B virus is a complicated, unique virus, which has evolved together with *Homo sapiens* over millennia; it has evolved a range of mechanisms that favor transmission and persistence which include the capacity to evade the CD8+ T cell response. By focusing on understanding the evolutionary interplay between host and virus, we can develop better insights into areas where we can target viral “Achilles heels.” The need for novel anti-HBV strategies should drive a deeper exploration of this host–pathogen interaction. Future research will be strengthened by comprehensive cross-sectional and longitudinal studies on HLA-typed hosts with clinical details, across a range of host ethnicities and HBV genotypes, with high quality serological and whole genome HBV deep sequencing data. This will provide a more comprehensive understanding of the nature and mechanisms of HBV evolution and persistence, helping us to reach the goal of global HBV eradication by guiding the design of new strategies, including vaccines and therapeutics.

AUTHOR CONTRIBUTIONS

SL undertook the primary literature review and drafted the manuscript; all authors had substantial input into revisions.

FUNDING

SL is funded by the National Institute for Health Research. PM is funded by the Wellcome Trust (grant number 110110). KL is funded by the Wellcome Trust (grant number 107652/Z15/Z) and The Royal Society. PK is funded by the Wellcome Trust (WT109965MA) and National Institute for Health Research Biomedical Research Centre, Oxford.

17. Julg B, Williams KL, Reddy S, Bishop K, Qi Y, Carrington M, et al. Enhanced anti-HIV functional activity associated with gag-specific CD8 T-cell responses. *J Virol* (2010) 84(11):5540–9. doi:10.1128/JVI.02031-09
18. Thimme R, Wieland S, Steiger C, Ghayeb J, Reimann KA, Purcell RH, et al. CD8(+) T cells mediate viral clearance and disease pathogenesis during acute hepatitis B virus infection. *J Virol* (2003) 77(1):68–76. doi:10.1128/JVI.77.1.68-76.2003
19. Shoukry NH, Grakoui A, Houghton M, Chien DY, Ghayeb J, Reimann KA, et al. Memory CD8+ T cells are required for protection from persistent hepatitis C virus infection. *J Exp Med* (2003) 197(12):1645–55. doi:10.1084/jem.20030239
20. Leitman EM, Palmer CD, Buus S, Chen F, Riddell L, Sims S, et al. Saporin-conjugated tetramers identify efficacious anti-HIV CD8+ T-cell specificities. *PLoS One* (2017) 12(10):e0184496. doi:10.1371/journal.pone.0184496
21. Jin X, Bauer DE, Tuttleton SE, Lewin S, Gettie A, Blanchard J, et al. Dramatic rise in plasma viremia after CD8(+) T cell depletion in simian immunodeficiency virus-infected macaques. *J Exp Med* (1999) 189(6):991–8. doi:10.1084/jem.189.6.991
22. Bertolotti A, Ferrari C. Adaptive immunity in HBV infection. *J Hepatol* (2016) 64(1):S71–83. doi:10.1016/j.jhep.2016.01.026
23. Boni C, Fiscaro P, Valdatta C, Amadei B, Di Vincenzo P, Giuberti T, et al. Characterization of hepatitis B virus (HBV)-specific T-cell dysfunction in chronic HBV infection. *J Virol* (2007) 81(8):4215–25. doi:10.1128/JVI.02844-06
24. Chang JJ, Thompson AJV, Visvanathan K, Kent SJ, Cameron PU, Wightman F, et al. The phenotype of hepatitis B virus-specific T cells differ in the liver and blood in chronic hepatitis B virus infection. *Hepatology* (2007) 46(5):1332–40. doi:10.1002/hep.21844
25. Phillips S, Chokshi S, Riva A, Evans A, Williams R, Naoumov NV. CD8(+) T cell control of hepatitis B virus replication: direct comparison between cytolytic and noncytolytic functions. *J Immunol* (2010) 184(1):287–95. doi:10.4049/jimmunol.0902761
26. Fiscaro P, Valdatta C, Massari M, Loggi E, Biasini E, Sacchelli L, et al. Antiviral intrahepatic T-cell responses can be restored by blocking programmed death-1 pathway in chronic hepatitis B. *Gastroenterology* (2010) 138(2):682–93. doi:10.1053/j.gastro.2009.09.052
27. Cao H, Zhang R, Zhang W. CTLA4 interferes with the HBV specific T cell immune response (review). *Int J Mol Med* (2018) 42(2):703–12. doi:10.3892/ijmm.2018.3688
28. Schurich A, Khanna P, Lopes AR, Han KJ, Peppas D, Micco L, et al. Role of the coinhibitory receptor cytotoxic T lymphocyte antigen-4 on apoptosis-prone CD8 T cells in persistent hepatitis B virus infection. *Hepatology* (2011) 53(5):1494–503. doi:10.1002/hep.24249
29. Wu W, Shi Y, Li S, Zhang Y, Liu Y, Wu Y, et al. Blockade of Tim-3 signaling restores the virus-specific CD8 + T-cell response in patients with chronic hepatitis B. *Eur J Immunol* (2012) 42(5):1180–91. doi:10.1002/eji.201141852
30. Rutebemberwa A, Ray SC, Astemborski J, Levine J, Liu L, Dowd KA, et al. High-programmed death-1 levels on hepatitis C virus-specific T cells during acute infection are associated with viral persistence and require preservation of cognate antigen during chronic infection. *J Immunol* (2008) 181(12):8215–25. doi:10.4049/jimmunol.181.12.8215
31. Day CL, Kaufmann DE, Kiepiela P, Brown JA, Moodley ES, Reddy S, et al. PD-1 expression on HIV-specific T cells is associated with T-cell exhaustion and disease progression. *Nature* (2006) 443(7109):350–4. doi:10.1038/nature05115
32. Salpini R, Colagrossi L, Bellocchi MC, Surdo M, Becker C, Alteri C, et al. Hepatitis B surface antigen genetic elements critical for immune escape correlate with hepatitis B virus reactivation upon immunosuppression. *Hepatology* (2015) 61(3):823–33. doi:10.1002/hep.27604
33. Hayashi K, Ishigami M, Ishizu Y, Kuzuya T, Honda T, Tachi Y, et al. Clinical characteristics and molecular analysis of hepatitis B virus reactivation in hepatitis B surface antigen-negative patients during or after immunosuppressive or cytotoxic chemotherapy. *J Gastroenterol* (2016) 51(11):1081–9. doi:10.1007/s00535-016-1187-z
34. Ando T, Kojima K, Isoda H, Eguchi Y, Honda T, Ishigami M, et al. Reactivation of resolved infection with the hepatitis B virus immune escape mutant G145R during dasatinib treatment for chronic myeloid leukemia. *Int J Hematol* (2015) 102(3):379–82. doi:10.1007/s12185-015-1788-y
35. Cartwright EK, Spicer L, Smith SA, Lee D, Fast R, Paganini S, et al. CD8(+) lymphocytes are required for maintaining viral suppression in SIV-infected macaques treated with short-term antiretroviral therapy. *Immunity* (2016) 45(3):656–68. doi:10.1016/j.immuni.2016.08.018
36. Deng K, Perteau M, Rongvaux A, Wang L, Durand CM, Ghiaur G, et al. Broad CTL response is required to clear latent HIV-1 due to dominance of escape mutations. *Nature* (2015) 517(7534):381–5. doi:10.1038/nature14053
37. Pignatelli M, Waters J, Brown D, Lever A, Iwarson S, Schaff Z, et al. HLA class I antigens on the hepatocyte membrane during recovery from acute hepatitis B virus infection and during interferon therapy in chronic hepatitis B virus infection. *Hepatology* (1986) 6(3):349–53. doi:10.1002/hep.1840060303
38. Chu CM, Shyu WC, Kuo RW, Liaw YF. HLA class I antigen display on hepatocyte membrane in chronic hepatitis B virus infection: its role in the pathogenesis of chronic type B hepatitis. *Hepatology* (1987) 8(3):712–7. doi:10.1002/hep.1840080358
39. Montano L, Miescher GC, Goodall AH, Wiedmann KH, Janossy G, Thomas HC. Hepatitis B virus and HLA antigen display in the liver during chronic hepatitis B virus infection. *Hepatology* (1982) 2(5):557S–61S. doi:10.1002/hep.1840020508
40. Michalak TI, Hodgson PD, Churchill ND. Posttranscriptional inhibition of class I major histocompatibility complex presentation on hepatocytes and lymphoid cells in chronic woodchuck hepatitis virus infection. *J Virol* (2000) 74(10):4483–94. doi:10.1128/JVI.74.10.4483-4494.2000
41. Saito K, Ait-Goughoulte M, Truscott SM, Meyer K, Blazevic A, Abate G, et al. Hepatitis C virus inhibits cell surface expression of HLA-DR, prevents dendritic cell maturation, and induces interleukin-10 production. *J Virol* (2008) 82(7):3320–8. doi:10.1128/JVI.02547-07
42. Kang W, Sung PS, Park S-H, Yoon S, Chang D-Y, Kim S, et al. Hepatitis C virus attenuates interferon-induced major histocompatibility complex class I expression and decreases CD8+ T cell effector functions. *Gastroenterology* (2014) 146(5):1351–4. doi:10.1053/j.gastro.2014.01.054
43. Schwartz O, Maréchal V, Le Gall S, Lemonnier F, Heard JM. Endocytosis of major histocompatibility complex class I molecules is induced by the HIV-1 Nef protein. *Nat Med* (1996) 2(3):338–42. doi:10.1038/nm0396-338
44. Mwimanzu P, Markle TJ, Ueno T, Brockman MA. Human leukocyte antigen (HLA) class I down-regulation by human immunodeficiency virus type 1 negative factor (HIV-1 Nef): what might we learn from natural sequence variants? *Viruses* (2012) 4(9):1711–30. doi:10.3390/v4102014
45. Backes S, Jäger C, Dembek CJ, Kosinska AD, Bauer T, Stephan AS, et al. Protein-prime/modified vaccinia virus Ankara vector-boost vaccination overcomes tolerance in high-antigenemic HBV-transgenic mice. *Vaccine* (2016) 34(7):923–32. doi:10.1016/j.vaccine.2015.12.060
46. Buchmann P, Dembek C, Kuklick L, Jäger C, Tedjukusumo R, von Freyend MJ, et al. A novel therapeutic hepatitis B vaccine induces cellular and humoral immune responses and breaks tolerance in hepatitis B virus (HBV) transgenic mice. *Vaccine* (2013) 31(8):1197–203. doi:10.1016/j.vaccine.2012.12.074
47. Kosinska AD, Zhang E, Johrden L, Liu J, Seiz PL, Zhang X, et al. Combination of DNA prime – adenovirus boost immunization with entecavir elicits sustained control of chronic hepatitis B in the woodchuck model. *PLoS Pathog* (2013) 9(6):e1003391. doi:10.1371/journal.ppat.1003391
48. Wijesundara DK, Gummow J, Li Y, Yu W, Quah BJ, Ranasinghe C, et al. Induction of genotype cross-reactive, hepatitis C virus-specific cell-mediated immunity in DNA-vaccinated mice. *J Virol* (2018) 92(8):e02133-17. doi:10.1128/JVI.02133-17
49. von Delft A, Donnison TA, Lourenço J, Hutchings C, Mullarkey CE, Brown A, et al. The generation of a simian adenoviral vectored HCV vaccine encoding genetically conserved gene segments to target multiple HCV genotypes. *Vaccine* (2018) 36(2):313–21. doi:10.1016/j.vaccine.2017.10.079
50. Barouch DH, O'Brien KL, Simmons NL, King SL, Abbink P, Maxfield LF, et al. Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. *Nat Med* (2010) 16(3):319–23. doi:10.1038/nm.2089
51. McMichael AJ, Koff WC. Vaccines that stimulate T cell immunity to HIV-1: the next step. *Nat Immunol* (2014) 15(4):319–22. doi:10.1038/ni.2844
52. Streeck H. Designing optimal HIV-vaccine T-cell responses. *Curr Opin HIV AIDS* (2016) 11(6):593–600. doi:10.1097/COH.0000000000000313
53. Lumley S, Noble H, Hadley MJ, Callow L, Malik A, Chua YY, et al. Hepitopes: a live interactive database of HLA class I epitopes in hepatitis B virus. *Wellcome Open Res* (2016) 1:9. doi:10.12688/wellcomeopenres.9952.1

54. Kefalakes H, Budeus B, Walker A, Jochum C, Hilgard G, Heinold A, et al. Adaptation of the hepatitis B virus core protein to CD8 + T-cell selection pressure. *Hepatology* (2015) 62(1):47–56. doi:10.1002/hep.27771
55. Nitschke K, Luxenburger H, Kiraithe MM, Thimme R, Neumann-Haefelin C. CD8+ T-cell responses in hepatitis B and C: the (HLA-) A, B, and C of hepatitis B and C. *Dig Dis* (2016) 34(4):396–409. doi:10.1159/000444555
56. Desmond CP, Gaudieri S, James IR, Pfafferoth K, Chopra A, Lau GK, et al. Viral adaptation to host immune responses occurs in chronic hepatitis B virus (HBV) infection, and adaptation is greatest in HBV e antigen-negative disease. *J Virol* (2012) 86(2):1181–92. doi:10.1128/JVI.05308-11
57. Abbott WGH, Tsai P, Leung E, Trevarton A, Ofanoa M, Hornell J, et al. Associations between HLA class I alleles and escape mutations in the hepatitis B virus core gene in New Zealand-resident Tongans. *J Virol* (2010) 84(1):621–9. doi:10.1128/JVI.01471-09
58. Khedive A, Norouzi M, Ramezani F, Karimzadeh H, Alavian SM, Malekzadeh R, et al. Hepatitis B virus surface protein mutations clustered mainly in CTL immune epitopes in chronic carriers: results of an Iranian nationwide study. *J Viral Hepat* (2013) 20(7):494–501. doi:10.1111/jvh.12045
59. Nassal M, Dallmeier K, Schultz U, Sun D. Phenotyping hepatitis B virus variants: from transfection towards a small animal in vivo infection model. *J Clin Virol* (2005) 34(Suppl 1):S89–95. doi:10.1016/S1386-6532(05)80017-4
60. Gaudieri S, Rauch A, Park LP, Freitas E, Herrmann S, Jeffrey G, et al. Evidence of viral adaptation to HLA class I-restricted immune pressure in chronic hepatitis C virus infection. *J Virol* (2006) 80(22):11094–104. doi:10.1128/JVI.00912-06
61. Fitzmaurice K, Petrovic D, Ramamurthy N, Simmons R, Merani S, Gaudieri S, et al. Molecular footprints reveal the impact of the protective HLA-A*03 allele in hepatitis C virus infection. *Gut* (2011) 60(11):1563–71. doi:10.1136/gut.2010.228403
62. Ruhl M, Knuschke T, Schewior K, Glavinic L, Neumann-Haefelin C, Chang DI, et al. CD8+ T-cell response promotes evolution of hepatitis c virus non-structural proteins. *Gastroenterology* (2011) 140(7):2064–73. doi:10.1053/j.gastro.2011.02.060
63. Kloverpris HN, Leslie A, Goulder P. Role of HLA adaptation in HIV evolution. *Front Immunol* (2015) 6:665. doi:10.3389/fimmu.2015.00665
64. Martins MA, Tully DC, Cruz MA, Power KA, Veloso de Santana MG, Bean DJ, et al. Vaccine-induced simian immunodeficiency virus-specific CD8+ T-cell responses focused on a single Nef epitope select for escape variants shortly after infection. *J Virol* (2015) 89(21):10802–20. doi:10.1128/JVI.01440-15
65. Harris M, Burns CM, Becker EA, Braasch AT, Gostick E, Johnson RC, et al. Acute-phase CD8 T cell responses that select for escape variants are needed to control live attenuated simian immunodeficiency virus. *J Virol* (2013) 87(16):9353–64. doi:10.1128/JVI.00909-13
66. Matthews PC, Carlson JM, Beloukas A, Malik A, Jooste P, Ogbu A, et al. HLA-A is a predictor of hepatitis B e antigen status in HIV-positive African adults. *J Infect Dis* (2016) 213(8):1248–52. doi:10.1093/infdis/jiv592
67. McKiernan SM, Hagan R, Curry M, McDonald GSA, Kelly A, Nolan N, et al. Distinct MHC class I and II alleles are associated with hepatitis C viral clearance, originating from a single source. *Hepatology* (2004) 40(1):108–14. doi:10.1002/hep.20261
68. Kiepiela P, Leslie AJ, Honeyborne I, Ramduth D, Thobakgale C, Chetty S, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. *Nature* (2004) 432(7018):769–75. doi:10.1038/nature03113
69. Martinez-Picado J, Prado JG, Fry EE, Pfafferoth K, Leslie A, Chetty S, et al. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *J Virol* (2006) 80(7):3617–23. doi:10.1128/JVI.80.7.3617-3623.2006
70. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, et al. CD8+T-cell responses to different HIV proteins have discordant associations with viral load. *Nat Med* (2007) 13(1):46–53. doi:10.1038/nm1520
71. Frater AJ, Brown H, Oxenius A, Gunthard HF, Hirschel B, Robinson N, et al. Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression. *J Virol* (2007) 81(12):6742–51. doi:10.1128/JVI.00022-07
72. Singh R, Kaul R, Kaul A, Khan K. A comparative review of HLA associations with hepatitis B and C viral infections across global populations. *World J Gastroenterol* (2007) 13(12):1770–87. doi:10.3748/wjg.v13.i12.1770
73. Wang L, Zou ZQ, Wang K. Clinical relevance of HLA gene variants in HBV infection. *J Immunol Res* (2016) 2016:9069375. doi:10.1155/2016/9069375
74. Li X, Liu W, Wang H, Jin X, Fang S, Shi Y, et al. The influence of HLA alleles and HBV subgenotypes on the outcomes of HBV infections in Northeast China. *Virus Res* (2012) 163(1):328–33. doi:10.1016/j.virusres.2011.10.020
75. Thio CL, Thomas DL, Karacki P, Gao X, Marti D, Kaslow RA, et al. Comprehensive analysis of class I and class II HLA antigens and chronic hepatitis B virus infection. *J Virol* (2003) 77(22):12083–7. doi:10.1128/JVI.77.22.12083-12087.2003
76. Ansari MA, Pedergrana V, Ip C LC, Magri A, Von Delft A, Bonsall D, et al. Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. *Nat Genet* (2017) 49(5):666–73. doi:10.1038/ng.3835
77. Neumann-Haefelin C, Oniangue-Ndza C, Kuntzen T, Schmidt J, Nitschke K, Sidney J, et al. Human leukocyte antigen B27 selects for rare escape mutations that significantly impair hepatitis C virus replication and require compensatory mutations. *Hepatology* (2011) 54(4):1157–66. doi:10.1002/hep.24541
78. Fellay J, Shianna KV, Ge D, Colombo S, Ledergerber B, Weale M, et al. A whole-genome association study of major determinants for host control of HIV-1. *Science* (2007) 317(5840):944–7. doi:10.1126/science.1143767
79. International HIV Controllers Study, Pereyra F, Jia X, McLaren PJ, Telenti A, De Bakker PIW, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* (2010) 330(6010):1551–7. doi:10.1126/science.1195271
80. Penna A, Del Prete G, Cavalli A, Bertoletti A, D'Elios MM, Sorrentino R, et al. Predominant T-helper 1 cytokine profile of hepatitis B virus nucleocapsid-specific T cells in acute self-limited hepatitis B. *Hepatology* (1997) 25(4):1022–7. doi:10.1002/hep.510250438
81. Nishida N, Sugiyama M, Sawai H, Nishina S, Sakai A, Ohashi J, et al. Key HLA-DRB1-DQB1 haplotypes and role of the BTNL2 gene for response to a hepatitis B vaccine. *Hepatology* (2018). doi:10.1002/hep.29876
82. Hu Z, Liu Y, Zhai X, Dai J, Jin G, Wang L, et al. New loci associated with chronic hepatitis B virus infection in Han Chinese. *Nat Genet* (2013) 45(12):1499–503. doi:10.1038/ng.2809
83. Kamatani Y, Wattanakayakit S, Ochi H, Kawaguchi T, Takahashi A, Hosono N, et al. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. *Nat Genet* (2009) 41(5):591–5. doi:10.1038/ng.348
84. Hosaka T, Suzuki F, Kobayashi M, Fukushima T, Kawamura Y, Sezaki H, et al. HLA-DP genes polymorphisms associate with hepatitis B surface antigen kinetics and seroclearance during nucleot(s)ide analogue therapy. *Liver Int* (2015) 35(4):1290–302. doi:10.1111/liv.12652
85. Jiang D-K, Sun J, Cao G, Liu Y, Lin D, Gao Y-Z, et al. Genetic variants in STAT4 and HLA-DQ genes confer risk of hepatitis B virus-related hepatocellular carcinoma. *Nat Genet* (2013) 45(1):72–5. doi:10.1038/ng.2483
86. Stoop JN, Van Der Molen RG, Baan CC, Van Der Laan LJW, Kuipers EJ, Kusters JG, et al. Regulatory T cells contribute to the impaired immune response in patients with chronic hepatitis B virus infection. *Hepatology* (2005) 41(4):771–8. doi:10.1002/hep.20649
87. Feng I-C, Koay L-B, Sheu M-J, Kuo H-T, Sun C-S, Lee C, et al. HBcAg-specific CD4+CD25+ regulatory T cells modulate immune tolerance and acute exacerbation on the natural history of chronic hepatitis B virus infection. *J Biomed Sci* (2007) 14(1):43–57. doi:10.1007/s11373-006-9129-z
88. Lin C-L, Kao J-H. Natural history of acute and chronic hepatitis B: the role of HBV genotypes and mutants. *Best Pract Res Clin Gastroenterol* (2017) 31(3):249–55. doi:10.1016/j.bpg.2017.04.010
89. Fung SK, Lok ASF. Hepatitis B virus genotypes: do they play a role in the outcome of HBV infection? *Hepatology* (2004) 40(4):790–2. doi:10.1002/hep.20455
90. Matsuura K, Isogawa M, Tanaka Y. Host genetic variants influencing the clinical course of hepatitis B virus infection. *J Med Virol* (2016) 88(3):371–9. doi:10.1002/jmv.24350
91. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. *Gastroenterology* (2012) 142(6):1264–73. doi:10.1053/j.gastro.2011.12.061
92. Cresswell P, Ackerman AL, Giodini A, Peaper DR, Wearsch PA. Mechanisms of MHC class I-restricted antigen processing and cross-presentation. *Immunol Rev* (2005) 207:145–57. doi:10.1111/j.0105-2896.2005.00316.x

93. Bromley SK, Burack WR, Kenneth G, Somersalo K, Sims TN, Sumen C, et al. The immunological synapse. *Annu Rev Immunol* (2001) 19:375–96. doi:10.1146/annurev.immunol.19.1.375
94. Klenerman P, McMichael A. AIDS/HIV. Finding footprints among the trees. *Science* (2007) 315(5818):1505–7. doi:10.1126/science.1140768
95. Yu DM, Li XH, Mom V, Lu ZH, Liao XW, Han Y, et al. N-glycosylation mutations within hepatitis B virus surface major hydrophilic region contribute mostly to immune escape. *J Hepatol* (2014) 60(3):515–22. doi:10.1016/j.jhep.2013.11.004
96. Bertolotti A, Costanzo A, Chisari FV, Levrero M, Artini M, Sette A, et al. Cytotoxic T lymphocyte response to a wild type hepatitis B virus epitope in patients chronically infected by variant viruses carrying substitutions within the epitope. *J Exp Med* (1994) 180(3):933–43. doi:10.1084/jem.180.3.933
97. Bertolotti A, Sette A, Chisari FV, Penna A, Levrero M, De Carli M, et al. Natural variants of cytotoxic epitopes are T-cell receptor antagonists for antiviral cytotoxic T cells. *Nature* (1994) 369(6479):407–10. doi:10.1038/369407a0
98. Nayersina R, Fowler P, Guilhot S, Missale G, Cerny A, Schlicht HJ, et al. HLA A2 restricted cytotoxic T lymphocyte responses to multiple hepatitis B surface antigen epitopes during hepatitis B virus infection. *J Immunol* (1993) 150(10):4659–71.
99. Chen M, Sällberg M, Hughes J, Jones J, Guidotti LG, Chisari FV, et al. Immune tolerance split between hepatitis B virus precore and core proteins. *J Virol* (2005) 79(5):3016–27. doi:10.1128/JVI.79.5.3016-3027.2005
100. Kimura Y, Gushima T, Rawale S, Kaumaya P, Walker CM. Escape mutations alter proteasome processing of major histocompatibility complex class I-restricted epitopes in persistent hepatitis C virus infection. *J Virol* (2005) 79(8):4870–6. doi:10.1128/JVI.79.8.4870-4876.2005
101. Allen TM, Altfeld M, Yu XG, O'Sullivan KM, Lichterfeld M, Le Gall S, et al. Selection, transmission, and reversion of an antigen-processing cytotoxic T-lymphocyte escape mutation in human immunodeficiency virus type 1 infection. *J Virol* (2004) 78(13):7069–78. doi:10.1128/JVI.78.13.7069-7078.2004
102. Mohamadkhani A, Sotoudeh M, Bowden S, Poustchi H, Jazii FR, Sayehmiri K, et al. Downregulation of HLA class II molecules by G1896A pre-core mutation in chronic hepatitis B virus infection. *Viral Immunol* (2009) 22(5):295–300. doi:10.1089/vim.2009.0031
103. Reherrmann B, Pasquinelli C, Mosier SM, Chisari FV. Hepatitis B virus (HBV) sequence variation of cytotoxic T lymphocyte epitopes is not common in patients with chronic HBV infection. *J Clin Invest* (1995) 96(3):1527–34. doi:10.1172/JCI118191
104. Osioy C, Giles E, Tanaka Y, Mizokami M, Minuk GY. Molecular evolution of hepatitis B virus over 25 years. *J Virol* (2006) 80(21):10307–14. doi:10.1128/JVI.00996-06
105. Chen L, Gan QR, Zhang DQ, Yao LF, Lin RS, Li Q, et al. Increased intrahepatic quasispecies heterogeneity correlates with off-treatment sustained response to nucleos(t)ide analogues in e antigen-positive chronic hepatitis B patients. *Clin Microbiol Infect* (2016) 22(2):201–7. doi:10.1016/j.cmi.2015.10.007
106. Vigerust DJ, Shepherd VL. Virus glycosylation: role in virulence and immune interactions. *Trends Microbiol* (2007) 15(5):211–8. doi:10.1016/j.tim.2007.03.003
107. Anjum S, Wahid A, Afzal MS, Albecka A, Alsaleh K, Ahmad T, et al. Additional glycosylation within a specific hypervariable region of subtype 3a of hepatitis C virus protects against virus neutralization. *J Infect Dis* (2013) 208(11):1888–97. doi:10.1093/infdis/jit376
108. Kwei K, Tang X, Lok AS, Sureau C, Garcia T, Li J, et al. Impaired virion secretion by hepatitis B virus immune escape mutants and its rescue by wild-type envelope proteins or a second-site mutation. *J Virol* (2013) 87(4):2352–7. doi:10.1128/JVI.02701-12
109. Ito K, Qin Y, Guarnieri M, Garcia T, Kwei K, Mizokami M, et al. Impairment of hepatitis B virus virion secretion by single-amino-acid substitutions in the small envelope protein and rescue by a novel glycosylation site. *J Virol* (2010) 84(24):12850–61. doi:10.1128/JVI.01499-10
110. Julithe R, Abou-Jaoude G, Sureau C. Modification of the hepatitis B virus envelope protein glycosylation pattern interferes with secretion of viral particles, infectivity, and susceptibility to neutralizing antibodies. *J Virol* (2014) 88(16):9049–59. doi:10.1128/JVI.01161-14
111. Sede M, Lopez-Ledesma M, Frider B, Pozzati M, Campos RH, Flichman D, et al. Hepatitis B virus depicts a high degree of conservation during the immune-tolerant phase in familiarly transmitted chronic hepatitis B infection: deep-sequencing and phylogenetic analysis. *J Viral Hepat* (2014) 21(9):650–61. doi:10.1111/jvh.12196
112. Milich D, Liang TJ. Exploring the biological basis of hepatitis B e antigen in hepatitis B virus infection. *Hepatology* (2003) 38(5):1075–86. doi:10.1053/jhep.2003.50453
113. Ye B, Liu X, Li X, Kong H, Tian L, Chen Y. T-cell exhaustion in chronic hepatitis B infection: current knowledge and clinical significance. *Cell Death Dis* (2015) 6:e1694. doi:10.1038/cddis.2015.42
114. Webster GJM, Reignat S, Brown D, Ogg GS, Jones L, Seneviratne SL, et al. Longitudinal analysis of CD8+ T cells specific for structural and nonstructural hepatitis B virus proteins in patients with chronic hepatitis B: implications for immunotherapy. *J Virol* (2004) 78(11):5707–19. doi:10.1128/JVI.78.11.5707-5719.2004
115. Kondo Y, Ninomiya M, Kakazu E, Kimura O, Shimosegawa T. Hepatitis B surface antigen could contribute to the immunopathogenesis of hepatitis B virus infection. *ISRN Gastroenterol* (2013) 2013:935295. doi:10.1155/2013/935295
116. Whalley SA, Brown D, Webster GJM, Jacobs R, Reignat S, Bertolotti A, et al. Evolution of hepatitis B virus during primary infection in humans: transient generation of cytotoxic T-cell mutants. *Gastroenterology* (2004) 127(4):1131–8. doi:10.1053/j.gastro.2004.07.004
117. Bengsch B, Martin B, Thimme R. Restoration of HBV-specific CD8+ T cell function by PD-1 blockade in inactive carrier patients is linked to T cell differentiation. *J Hepatol* (2014) 61(6):1212–9. doi:10.1016/j.jhep.2014.07.005
118. Boni C, Laccabue D, Lampertico P, Giuberti T, Viganò M, Schivazappa S, et al. Restored function of HBV-specific T cells after long-term effective therapy with nucleos(t)ide analogues. *Gastroenterology* (2012) 143(4):963–73.e9. doi:10.1053/j.gastro.2012.07.014
119. Boni C, Penna A, Bertolotti A, Lamonaca V, Rapti I, Missale G, et al. Transient restoration of anti-viral T cell responses induced by lamivudine therapy in chronic hepatitis B. *J Hepatol* (2003) 39(4):595–605. doi:10.1016/S0168-8278(03)00292-7
120. Rivino L, Le Bert N, Gill US, Kunasegaran K, Cheng Y, Tan DZM, et al. Hepatitis b virus-specific t cells associate with viral control upon nucleos(t)ide-analogue therapy discontinuation. *J Clin Invest* (2018) 128(2):668–81. doi:10.1172/JCI92812
121. Chu CM, Liaw YF. Chronic hepatitis B virus infection acquired in childhood: special emphasis on prognostic and therapeutic implication of delayed HBeAg seroconversion. *J Viral Hepat* (2007) 14(3):147–52. doi:10.1111/j.1365-2893.2006.00810.x
122. Chen Y-C, Chu C-M, Liaw Y-F. Age-specific prognosis following spontaneous hepatitis B e antigen seroconversion in chronic hepatitis B. *Hepatology* (2010) 51(2):435–44. doi:10.1002/hep.23348
123. Okamoto H, Tsuda F, Akahane Y, Sugai Y, Yoshida M, Moriyama K, et al. Hepatitis B virus with mutations in the core promoter for an e antigen-negative phenotype in carriers with antibody to e antigen. *J Virol* (1994) 68(12):8102–10.
124. Wang H-Y, Chien M-H, Huang H-P, Chang H-C, Wu C-C, Chen P-J, et al. Distinct hepatitis B virus dynamics in the immunotolerant and early immunoclearance phases. *J Virol* (2010) 84(7):3454–63. doi:10.1128/JVI.02164-09
125. Hannoun C, Horal P, Lindh M. Long-term mutation rates in the hepatitis B virus genome. *J Gen Virol* (2000) 81(Pt 1):75–83. doi:10.1099/0022-1317-81-1-75
126. Bozkaya H, Akarca US, Ayola B, Lok ASF. High degree of conservation in the hepatitis B virus core gene during the immune tolerant phase in perinatally acquired chronic hepatitis B virus infection. *J Hepatol* (1997) 26(3):508–16. doi:10.1016/S0168-8278(97)80415-1
127. Philpott S, Burger H, Tsoukas C, Foley B, Anastos K, Kitchen C, et al. Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and intrapatient recombination. *J Virol* (2005) 79(1):353–63. doi:10.1128/JVI.79.1.353-363.2005
128. Schnell G, Price RW, Swanson R, Spudis S. Compartmentalization and clonal amplification of HIV-1 variants in the cerebrospinal fluid during primary infection. *J Virol* (2010) 84(5):2395–407. doi:10.1128/JVI.01863-09
129. Roque-Afonso AM, Ducoulombier D, Di Liberto G, Kara R, Rigou M, Dussaix E, et al. Compartmentalization of hepatitis C virus genotypes

- between plasma and peripheral blood mononuclear cells. *J Virol* (2005) 79(10):6349–57. doi:10.1128/JVI.79.10.6349-6357.2005
130. Russell G, Pizzillo P, Iannolo G, Barbera F, Tuzzolino F, Liotta R, et al. HCV replication in gastrointestinal mucosa: potential extra-hepatic viral reservoir and possible role in HCV infection recurrence after liver transplantation. *PLoS One* (2017) 12(7):e0181683. doi:10.1371/journal.pone.0181683
 131. Bull ME, Heath LM, McKernan-Mullin JL, Kraft KM, Acevedo L, Hitti JE, et al. Human immunodeficiency viruses appear compartmentalized to the female genital tract in cross-sectional analyses but genital lineages do not persist over time. *J Infect Dis* (2013) 207(8):1206–15. doi:10.1093/infdis/jit016
 132. Michalak TI. Occult persistence and lymphotropism of hepadnaviral infection: insights from the woodchuck viral hepatitis model. *Immunol Rev* (2000) 174:98–111. doi:10.1034/j.1600-0528.2002.017406.x
 133. Mason A, Wick M, White H, Perrillo R. Hepatitis B virus replication in diverse cell types during chronic hepatitis B virus infection. *Hepatology* (1993) 18(4):781–9. doi:10.1002/hep.1840180406
 134. Faure-Dupuy S, Durantal D, Lucifora J. Liver macrophages: friend or foe during hepatitis B infection? *Liver Int* (2018). doi:10.1111/liv.13884
 135. Coffin CS, Mulrooney-Cousins PM, Van Marle G, Roberts JP, Michalak TI, Terrault NA. Hepatitis B virus quasiespecies in hepatic and extrahepatic viral reservoirs in liver transplant recipients on prophylactic therapy. *Liver Transpl* (2011) 17(8):955–62. doi:10.1002/lt.22312
 136. Coffin CS, Osiowy C, Gao S, Nishikawa S, van der Meer F, van Marle G. Hepatitis B virus (HBV) variants fluctuate in paired plasma and peripheral blood mononuclear cells among patient cohorts during different chronic hepatitis B (CHB) disease phases. *J Viral Hepat* (2015) 22(4):416–26. doi:10.1111/jvh.12308
 137. Datta S, Panigrahi R, Biswas A, Chandra PK, Banerjee A, Mahapatra PK, et al. Genetic characterization of hepatitis B virus in peripheral blood leukocytes: evidence for selection and compartmentalization of viral variants with the immune escape G145R mutation. *J Virol* (2009) 83(19):9983–92. doi:10.1128/JVI.01905-08
 138. Brind A, Jiang J, Samuel D, Gigou M, Feray C, Bréchet C, et al. Evidence for selection of hepatitis B mutants after liver transplantation through peripheral blood mononuclear cell infection. *J Hepatol* (1997) 26(2):228–35. doi:10.1016/S0168-8278(97)80035-9
 139. Chakravarty R, Neogi M, Roychowdhury S, Panda CK. Presence of hepatitis B surface antigen mutant G145R DNA in the peripheral blood leukocytes of the family members of an asymptomatic carrier and evidence of its horizontal transmission. *Virus Res* (2002) 90(1–2):133–41. doi:10.1016/S0168-1702(02)00147-8
 140. Bagaglio S, Albarello L, Biswas P, Uberti-Foppa C, Fortis C, Morsica G. Virological pattern of hepatitis B infection in an HIV-positive man with fatal fulminant hepatitis B: a case report. *J Med Case Rep* (2009) 3:110. doi:10.1186/1752-1947-3-110
 141. Gao S, Duan Z-P, Chen Y, van der Meer F, Lee SS, Osiowy C, et al. Compartmental HBV evolution and replication in liver and extrahepatic sites after nucleos(tide) analogue therapy in chronic hepatitis B carriers. *J Clin Virol* (2017) 94:8–14. doi:10.1016/j.jcv.2017.06.009
 142. Wu C, Zhang X, Tian Y, Song J, Yang D, Roggendorf M, et al. Biological significance of amino acid substitutions in hepatitis B surface antigen (HBsAg) for glycosylation, secretion, antigenicity and immunogenicity of HBsAg and hepatitis B virus replication. *J Gen Virol* (2010) 91(Pt 2):483–92. doi:10.1099/vir.0.012740-0
 143. Torresi J. The virological and clinical significance of mutations in the overlapping envelope and polymerase genes of hepatitis B virus. *J Clin Virol* (2002) 25:97–106. doi:10.1016/S1386-6532(02)00049-5
 144. Delaney WE, Yang H, Westland CE, Das K, Arnold E, Gibbs CS, et al. The hepatitis B virus polymerase mutation rtV173L is selected during lamivudine therapy and enhances viral replication *in vitro*. *J Virol* (2003) 77(21):11833–41. doi:10.1128/JVI.77.21.11833-11841.2003
 145. Mokaya J, McNaughton AL, Hadley M, Beloukas A, Geretti A-M, Goedhals D, et al. A systematic review of hepatitis B virus (HBV) drug and vaccine escape mutations in Africa: a call for urgent action. *bioRxiv* (2018). doi:10.1101/258350
 146. Oniangue-Ndza C, Kuntzen T, Kemper M, Beral A, Wang YE, Neumann-Haefelin C, et al. Compensatory mutations restore the replication defects caused by cytotoxic T lymphocyte escape mutations in hepatitis C virus polymerase. *J Virol* (2011) 85(22):11883–90. doi:10.1128/JVI.00779-11
 147. Peyerl FW, Bazick HS, Newberg MH, Barouch DH, Sodroski J, Letvin NL. Fitness costs limit viral escape from cytotoxic T lymphocytes at a structurally constrained epitope. *J Virol* (2004) 78(24):13901–10. doi:10.1128/JVI.78.24.13901-13910.2004
 148. Brockman MA, Schneidewind A, Lahaie M, Schmidt A, Miura T, DeSouza I, et al. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 gag alter capsid interactions with cyclophilin A. *J Virol* (2007) 81(22):12608–18. doi:10.1128/JVI.01369-07
 149. Hoofnagle JH. Reactivation of hepatitis B. *Hepatology* (2009) 49(5 Suppl):S156–65. doi:10.1002/hep.22945
 150. Loomba R, Liang TJ. Hepatitis B reactivation associated with immune suppressive and biological modifier therapies: current concepts, management strategies, and future directions. *Gastroenterology* (2017) 152(6):1297–309. doi:10.1053/j.gastro.2017.02.009
 151. Sugauchi F, Tanaka Y, Kusumoto S, Matsuura K, Sugiyama M, Kurbanov F, et al. Virological and clinical characteristics on reactivation of occult hepatitis B in patients with hematological malignancy. *J Med Virol* (2011) 83(3):412–8. doi:10.1002/jmv.21995
 152. Kempinska A, Kwak EJ, Angel JB. Reactivation of hepatitis B infection following allogeneic bone marrow transplantation in a hepatitis B-immune patient: case report and review of the literature. *Clin Infect Dis* (2005) 41(9):1277–82. doi:10.1086/496924
 153. Mok MY, Ng WL, Yuen MF, Wong RW, Lau CS. Safety of disease modifying anti-rheumatic agents in rheumatoid arthritis patients with chronic viral hepatitis. *Clin Exp Rheumatol* (2000) 18(3):363–8.
 154. Ito S, Nakazono K, Murasawa A, Mita Y, Hata K, Saito N, et al. Development of fulminant hepatitis B (precore variant mutant type) after the discontinuation of low-dose methotrexate therapy in a rheumatoid arthritis patient. *Arthritis Rheum* (2001) 44(2):339–42. doi:10.1002/1529-0131(200102)44:2<339::AID-ANR51>3.0.CO;2-Q
 155. Esteve M, Saro C, Gonzalez-Huix F, Suarez F, Forne M, Viver JM. Chronic hepatitis B reactivation following infliximab therapy in Crohn's disease patients: need for primary prophylaxis. *Gut* (2004) 53(9):1363–5. doi:10.1136/gut.2004.040675
 156. Kajiwara E, Tanaka Y, Ohashi T, Uchimura K, Sadoshima S, Kinjo M, et al. Hepatitis B caused by a hepatitis B surface antigen escape mutant. *J Gastroenterol* (2008) 43(3):243–7. doi:10.1007/s00535-007-2150-9
 157. World Health Organization. *Global Hepatitis Report 2017*. Geneva: World Health Organization (2017).
 158. Bertoletti A, Bert NL. Immunotherapy for chronic hepatitis B virus infection. *Gut Liver* (2018). doi:10.5009/gnl17233
 159. Bertoletti A, Tan AT, Koh S. T-cell therapy for chronic viral hepatitis. *Cytotherapy* (2017) 19(11):1317–24. doi:10.1016/j.jcyt.2017.07.011
 160. Wisskirchen K, Metzger K, Schreiber S, Asen T, Weigand L, Dargel C, et al. Isolation and functional characterization of hepatitis B virus-specific T-cell receptors as new tools for experimental and clinical use. *PLoS One* (2017) 12(8):e0182936. doi:10.1371/journal.pone.0182936
 161. Rehmann B, Ferrari C, Pasquinielli C, Chisari FV. The hepatitis B virus persists for decades after patients' recovery from acute viral hepatitis despite active maintenance of a cytotoxic T-lymphocyte response. *Nat Med* (1996) 2(10):1104–8. doi:10.1038/nm1096-1104
 162. Kaymakoglu S, Baran B, Onel D, Badur S, Atamer T, Akyuz F. Acute hepatitis B due to immune-escape mutations in a naturally immune patient. *Acta Gastroenterol Belg* (2014) 77(2):262–5.
 163. Ye Q, Shang S-Q, Li W. A new vaccine escape mutant of hepatitis B virus causes occult infection. *Hum Vaccin Immunother* (2015) 11(2):407–10. doi:10.4161/21645515.2014.994461
 164. Mendy M, D'Mello F, Kanellos T, Oliver S, Whittle H, Howard CR. Envelope protein variability among HBV-Infected asymptomatic carriers and immunized children with breakthrough infections. *J Med Virol* (2008) 80(9):1537–46. doi:10.1002/jmv.21221
 165. Zhang M, Ge G, Yang Y, Cai X, Fu Q, Cai J, et al. Decreased antigenicity profiles of immune-escaped and drug-resistant hepatitis B surface antigen (HBsAg) double mutants. *Virology* (2013) 442:292. doi:10.1016/j.virol.2013.02.022

166. Lacombe K, Boyd A, Gozlan J, Lavocat F, Girard P-M, Zoulim F. Drug-resistant and immune-escape HBV mutants in HIV-infected hosts. *Antivir Ther* (2010) 15(3 Pt B):493–7. doi:10.3851/IMP1495
167. Sloan RD, Ijaz S, Moore PL, Harrison TJ, Teo CG, Tedder RS. Antiviral resistance mutations potentiate hepatitis B virus immune evasion through disruption of its surface antigen a determinant. *Antivir Ther* (2008) 13(3):439–47.
168. Baxa DM, Thekdi AD, Golembieski A, Krishnan PV, Sharif O, Kizy A, et al. Evaluation of anti-HBV drug resistant mutations among patients with acute symptomatic hepatitis B in the United States. *J Hepatol* (2013) 58(2):212–6. doi:10.1016/j.jhep.2012.09.014
169. Coppola N, Tonziello G, Colombaro P, Pisaturo M, Messina V, Moriconi F, et al. Lamivudine-resistant HBV strain rtM204V/I in acute hepatitis B. *J Infect* (2013) 67(4):322–8. doi:10.1016/j.jinf.2013.06.006

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling Editor declared a past co-authorship with one of the authors PK.

Copyright © 2018 Lumley, McNaughton, Klenerman, Lythgoe and Matthews. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Impact of Interferon- α Receptor-1 Promoter Polymorphisms on the Transcriptome of the Hepatitis B Virus-Associated Hepatocellular Carcinoma

Timokratis Karamitros^{1,2*}, George Papatheodoridis³, Dimitrios Paraskevis⁴, Angelos Hatzakis⁴, Jean L. Mbisa⁵, Urania Georgopoulou⁶, Paul Klenerman⁷ and Gkikas Magiorkinis^{1,4*}

OPEN ACCESS

Edited by:

Linda F. Van Dyk,
University of Colorado Denver,
United States

Reviewed by:

Hridayesh Prakash,
All India Institute of Medical
Sciences, India
Sandip Chakraborty,
College of Veterinary Sciences
and Animal Husbandry, India

*Correspondence:

Timokratis Karamitros
tkaram@pasteur.gr;
Gkikas Magiorkinis
gmagi@med.uoa.gr

Specialty section:

This article was submitted
to Microbial Immunology,
a section of the journal
Frontiers in Immunology

Received: 29 September 2017

Accepted: 28 March 2018

Published: 16 April 2018

Citation:

Karamitros T, Papatheodoridis G, Paraskevis D, Hatzakis A, Mbisa JL, Georgopoulou U, Klenerman P and Magiorkinis G (2018) Impact of Interferon- α Receptor-1 Promoter Polymorphisms on the Transcriptome of the Hepatitis B Virus-Associated Hepatocellular Carcinoma. *Front. Immunol.* 9:777. doi: 10.3389/fimmu.2018.00777

¹Department of Zoology, University of Oxford, Oxford, United Kingdom, ²Department of Microbiology, Public Health Laboratories, Hellenic Pasteur Institute, Athens, Greece, ³Academic Department of Gastroenterology, Laiko General Hospital, Medical School of National and Kapodistrian University of Athens, Athens, Greece, ⁴Department of Hygiene and Epidemiology and Medical Statistics, Medical School of National and Kapodistrian University of Athens, Athens, Greece, ⁵Virus Reference Department, Public Health England, London, United Kingdom, ⁶Department of Microbiology, Molecular Virology Laboratory, Hellenic Pasteur Institute, Athens, Greece, ⁷Peter Medawar Building for Pathogen Research and Translational Gastroenterology Unit, University of Oxford, Oxford, United Kingdom

Background and aims: Genetic polymorphisms within the promoter of interferon- α receptor type-1 (IFNAR1) have been associated with the susceptibility to and the outcome of chronic hepatitis B virus (HBV) infection. However, the impact of these polymorphisms in the transcriptome of the HBV-associated hepatocellular carcinoma (HCC) remains largely unexplored.

Methods: Using whole-genome and exome sequencing data from The Cancer Genome Atlas project, we characterized three single-nucleotide polymorphisms (SNPs: –568G/C, –408C/T, –3C/T) and one variable number tandem repeat [VNTR: –77(GT)_n] within the IFNAR1 promoter sequence in 49 HCC patients. RNAseq data from 10 genotyped HCC samples were grouped according to their –77VNTR or –3SNP genotype to evaluate the impact of these polymorphisms on the differential expression on the HCC transcriptome.

Results: There is a fourfold higher impact of the –77VNTR on the HCC transcriptome compared to the –3SNP ($q < 0.1$, $p < 0.001$). The expression of the primary IFNAR1 transcript is not affected by these polymorphisms but a secondary, HCC-specific transcript is expressed only in homozygous –77VNTR $\leq 8/\leq 8$ (GT)_n samples ($p < 0.05$). At the same time, patients carrying at least one –77VNTR > 8 (GT) allele, presented a strong upregulation of the fibronectin-1 (FN-1) gene, which has been associated with the development of HCC. Gene Ontology and pathway enrichment analysis of the differentially expressed genes revealed a strong disruption of the PI3K–AKT signaling pathway, which can be partially triggered by the extracellular matrix FN-1.

Conclusion: The IFNAR-1 promoter polymorphisms are not involved in the expression levels of the main IFNAR-1 transcript. The –77VNTR has a regulatory role on

the expression of a secondary, truncated, HCC-specific transcript, which in turn coincides with disruptions in cancer-associated pathways and in FN-1 expression modifications.

Keywords: hepatitis virus B, interferon- α receptor type-1 promoter, hepatocellular carcinoma, interferon- α receptor, polymorphism, transcriptome, RNAseq, the Cancer Genome Atlas project

INTRODUCTION

Worldwide, more than two billion people have been infected with hepatitis B virus (HBV) and approximately 250 million individuals are chronically infected (1). Infected patients can be inactive chronic HBV carriers (IC) (eAg-negative, eAb-positive, with low levels of HBV DNA and no evidence of liver inflammation) or present with the progressive chronic hepatitis B (CHB) (2–5). HBV is responsible for >50% of hepatocellular carcinomas (HCCs) worldwide (2, 6, 7). On the other hand, HBeAg-negative ICs have a more benign prognosis with very low risk of cirrhosis or HCC, as indicated by long-term follow-up studies (8–10).

The genetic profile of the patient plays a substantial role in the clinical outcome of HBV infection (11–13). The virus also modulates cellular mechanisms and signal pathways during the course of the infection. For example, the persistent expression of the HBV x antigen (HBxAg) is correlated with the development of fibrosis and cirrhosis during CHB, as it can activate fibronectin-1 (FN-1) gene, through the induction of the nuclear factor kappa B (NF-kappa B or NFkB) (14). FN is an omnipresent extracellular matrix glycoprotein. Plasma FN and cellular FN have distinct properties and roles in the strictly regulated mechanism of tissue repair (15). FN is a very important component of ECM and any dysfunction in the fibrinogenesis mechanisms can lead to the development of fibrotic disease (16).

Interferons (IFN)- α/β are cytokines involved in both innate and adaptive immune responses (17, 18), thus play a pivotal role as cancer defense mechanisms (19). Interferon-induced signal transduction pathways represent a fine-tuned network of interactions, which are triggered upon binding of IFN on its receptor (IFNR) (20–22). IFNs are shown to regulate the transcription levels of more than 2,000 genes, which compose the *Interferome*, a gene-network created by integrating information collected from high-throughput experiments (23). As IFNs are anticancer and antiviral cytokines, it is expected that the genetic profiles of the genes involved in these signal transduction pathways (e.g., IFNR) will have an impact on the susceptibility of the patients to cancer and, when HBV is involved, to CHB and HBV-related HCC.

IFN- α is administrated as part of the first-line therapies for CHB (2, 24). The compatible receptor (IFNAR) consists of two subunits IFNAR-1 and IFNAR-2 (25). A number of polymorphisms in the promoter of IFNAR-1 gene and also in its coding sequence have been described (26–29). Those observed in the promoter—three SNPs and one variable number tandem repeat (VNTR) at positions –568, –408, –3 and –77, respectively—are believed to affect the expression of the receptor and are associated with the clinical outcome of HBV infection (27, 30, 31).

We have previously shown that the same polymorphisms are associated with the clinical phase of HBeAg-negative chronic HBV infection in Caucasians. Briefly, patients with genotypes

–568GC/CC, –408CT, and –3CT and patients with less than 8(GT) repeats in the –77VNTR were more frequent among inactive carriers (IC) vs. patients with HBeAg-negative CHB (32). Notably, the (GT)-repeats in the –77VNTR were strongly associated with the clinical outcome of the patients in our study; homozygotes carrying both alleles with ≤ 8 (GT)-repeats, were more likely to be ICs, compared to those carrying both alleles with >8(GT) repeats (OR = 7.14, $p \leq 0.001$).

In this study, we use whole-genome sequencing (WGS), whole-exome sequencing (WXS), and RNAseq data derived from The Cancer Genome Atlas (TCGA) project in combination with a bioinformatics pipeline, which efficiently and accurately genotypes the –77VNTR. We estimate the distribution of the genetic variations of the four IFNAR-1 promoter polymorphisms and we investigate the impact of –77VNTR and –3SNP polymorphisms on the regulation of HCC transcriptome and interferome. IFNAR-1 is transcribed into three transcripts, ENST00000270139 (which is translated into the receptor), ENST00000493503, and ENST00000442071; performing differential expression analysis in patients grouped according to their –77VNTR genotype, we show that –77VNTR modifies the transcriptional profile of IFNAR-1 gene by controlling the expression of the last one, which comprises only a fibronectin III domain and is translated into a truncated form of the receptor. The expression of this transcript coincides with significantly lower expression of FN-1. Finally, the PI3K–AKT signaling pathway, which is partially triggered by FN-1, was significantly enriched with differentially expressed genes. Our results indicate that this secondary transcript could potentially act as a regulatory element, but its functional role remains largely unknown.

PATIENTS AND METHODS

Study Design

We examined four previously described polymorphisms within the IFNAR-1 promoter (27, 30, 31), from now on referred to as interferon- α receptor promoter polymorphisms (IFNARPPs). They include three single-nucleotide polymorphisms (SNPs) at positions –568G/C, –408C/T, and –3C/T and a variable number tandem repeat of the binucleotide GT, –77VNTR(GT)_n.

We used 49 (15 WGS and 34 WXS) TCGA “liver hepatocellular carcinoma (LIHC)” samples (project ID 10464). We filtered all the available (up to March 2016) HCC samples with positive HBV surface antigen. We performed genotyping of the four IFNARPPs, using the pipeline described below. We compared the frequencies of the revealed genotypes with those of a group of 92 chronically infected HBeAg-negative IC as defined in our previous study (32) as they present very low risk of cirrhosis and HCC (8–10). Briefly, IC state was defined for patients with persistently normal ALT

values under strict follow-up and maximum HBV DNA levels $\leq 20,000$ IU/mL. None of these inactive carriers had cirrhosis (33).

To assess the impact of the -77 VNTR and the -3 SNP on the expression profiles in HBV-associated hepatocellular carcinoma, we used tumor RNAseq data available for 10 TCGA HCC genotyped samples after classifying them according to their genotype (for -77 VNTR: $\leq 8/\leq 8$ vs. $\leq 8/>8$ or $>8/>8$ (GT) repeats, and for -3 SNP: CC vs. CT or TT) (Table 1).

Bioinformatics

The .bam files were transformed into paired-end .fastq files using Samtools-bam2fq (34) and were locally mapped against two artificial chromosomes using Bowtie2 (35) in—very sensitive local mode (Figure 1). The artificial reference chromosomes were created after splitting the IFNAR1 promoter sequence (genebank: X60459.1) at the -77 VNTR, leaving 3 GT repeats at each hanging end, to avoid non-specific mapping of reads due to the low complexity of the microsatellite repeats. We extracted the mapped reads and used them to *de novo* assemble the alleles of IFNAR1 promoter for each TCGA patient, using MIRA (36). We further analyzed the *de novo* assembled contigs using R DNASTrings package to count the GT repeats. We used Samtools mpileup and bcftools (34) to call the variations of the three SNPs from the mapping alignments. The genotyping results were confirmed and quality-controlled by visual inspection of the mapping alignments of 15 random samples, using IGV (37).

We performed the RNAseq analysis using Kallisto (38) to map the reads against the Human Transcriptome reference (v.GRCh38.rel79) and to calculate the transcripts abundances. We analyzed the impact of the IFNARPPs on the transcriptional landscape of the interferon-associated genes using the “Interferome” database (23) on a subset our whole-transcriptome results. We used Sleuth (39) and R-base functions to interpret and visualize the RNA-seq analysis results. We performed Gene Ontology (GO) and KEGG pathway enrichment analysis using the differentially expressed genes ($p < 0.001$).

Statistical Analysis

We used the *t*-test, the Fisher’s exact test, and the *z*-test to evaluate the association of the demographic and genetic characteristics of the patients with the disease outcome. We used RStudio v0.99.446 for R v3.2.3 programming language for all statistical computing and graphics.

TABLE 1 | RNAseq samples analyzed in this study.

Sample	ID	-77 VNTR (GTn)	-3 SNP (C/T)	Mapped reads (N)	Reads fraction
A	A4NL	$\leq 8/>8$	CT	52,242,268	0.93
B	A39Y	$\leq 8/>8$	CT	74,221,158	0.95
C	AAV6	$\leq 8/>8$	CT	57,894,784	0.92
D	A7ME	$>8/>8$	CC	63,054,738	0.93
E	A9CZ	$>8/>8$	CC	52,435,977	0.95
F	A1EF	$\leq 8/\leq 8$	CT	35,808,361	0.94
G	A1EK	$\leq 8/\leq 8$	CT	49,188,130	0.90
H	A66X	$\leq 8/\leq 8$	CC	63,083,028	0.93
I	A7PX	$\leq 8/\leq 8$	CC	50,278,517	0.84
J	A115	$\leq 8/\leq 8$	CC	47,977,943	0.95

RESULTS

RNAseq Differential Expression Design

We tested for differentially expressed transcripts in 10 HCC RNAseq samples after defining their -77 VNTR and -3 SNP genotypes. We grouped and compared them according to their -77 VNTR genotype [samples A–E: $>8/\leq 8$ or $>8/>8$ (GT)n vs. F–J: $\leq 8/\leq 8$ (GT)n] and according to their -3 SNP genotype (samples A, B, C, F, G: “CT” vs. D, E, H, I, J: “CC”) (Table 1). The quality-control assessment of the RNAseq analysis is summarized in Figure S1 in Supplementary Material.

The -77 VNTR and -3 SNP Polymorphisms and the IFNAR1 Transcription Profile

IFNAR1 gene generates three different transcripts, but only one of them is translated into the functional receptor protein subunit (transcript “001”: ENST00000270139). Of the two remaining transcripts, one is not translated (transcript “002”: ENST00000493503)

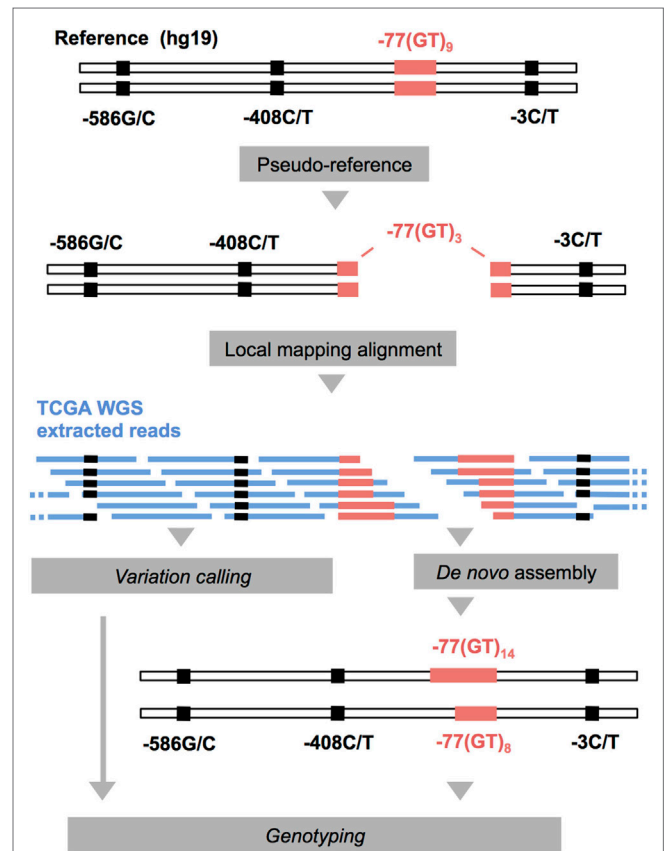


FIGURE 1 | Schematic representation of the computational pipeline applied to whole-genome sequencing (WGS) and WXS The Cancer Genome Atlas (TCGA) data (in blue) for the genotyping of the IFNARPPs. IFNAR1 promoter sequence (genebank: X60459.1) was used for the construction of two pseudo-chromosomes, on which the reads were aligned. The mapped reads were extracted and used for the *de novo* assembly of the two alleles of each sample. This method allows the assembly of alleles varying substantially in length compared to the reference.

and the other (transcript “003”: ENST00000442071) produces a truncated (136 aa long) isoform of the receptor, incorporating only a fibronectin type III domain. We found that –77 VNTR and –3 SNP polymorphisms do not significantly affect the expression levels of the primary transcript “001” and the secondary transcript “002.”

Interestingly, the expression of the secondary transcript “003” was detectable only in homozygous patients carrying both alleles with ≤ 8 (GT) repeats, while it was absent in patients carrying at least one allele with > 8 (GT) repeats (2.47 vs. 0.01 TPM, respectively) (**Figure 2**). We further tested 10 TCGA normal tissue (liver) RNAseq samples for the expression of this transcript. We found them all negative except from one, which presented only basal expression (0.02 TPM). Thus, transcript “003” was HCC-specific.

Impact of –77 VNTR and –3 SNP on the HCC Transcriptome and Interferome

We classified 10 HCC TCGA samples according to their genotype, $\leq 8/\leq 8$ vs. $\leq 8/> 8$ or $> 8/> 8$ (GT) repeats for the –77VNTR, and CC vs. CT or TT for the –3SNP. There were significant changes in the transcription profiles between the groups tested. In detail, for the –77VNTR grouping, there were 246 differentially expressed genes ($p < 0.001$), while for the –3SNP grouping, only 57 genes showed significant changes in their expression levels ($p < 0.001$) (**Figure 3**).

Focusing on the interferon-related genes, we created a subset of the transcriptomics results according to transcript IDs found in the *Interferome* database (23). The –77VNTR

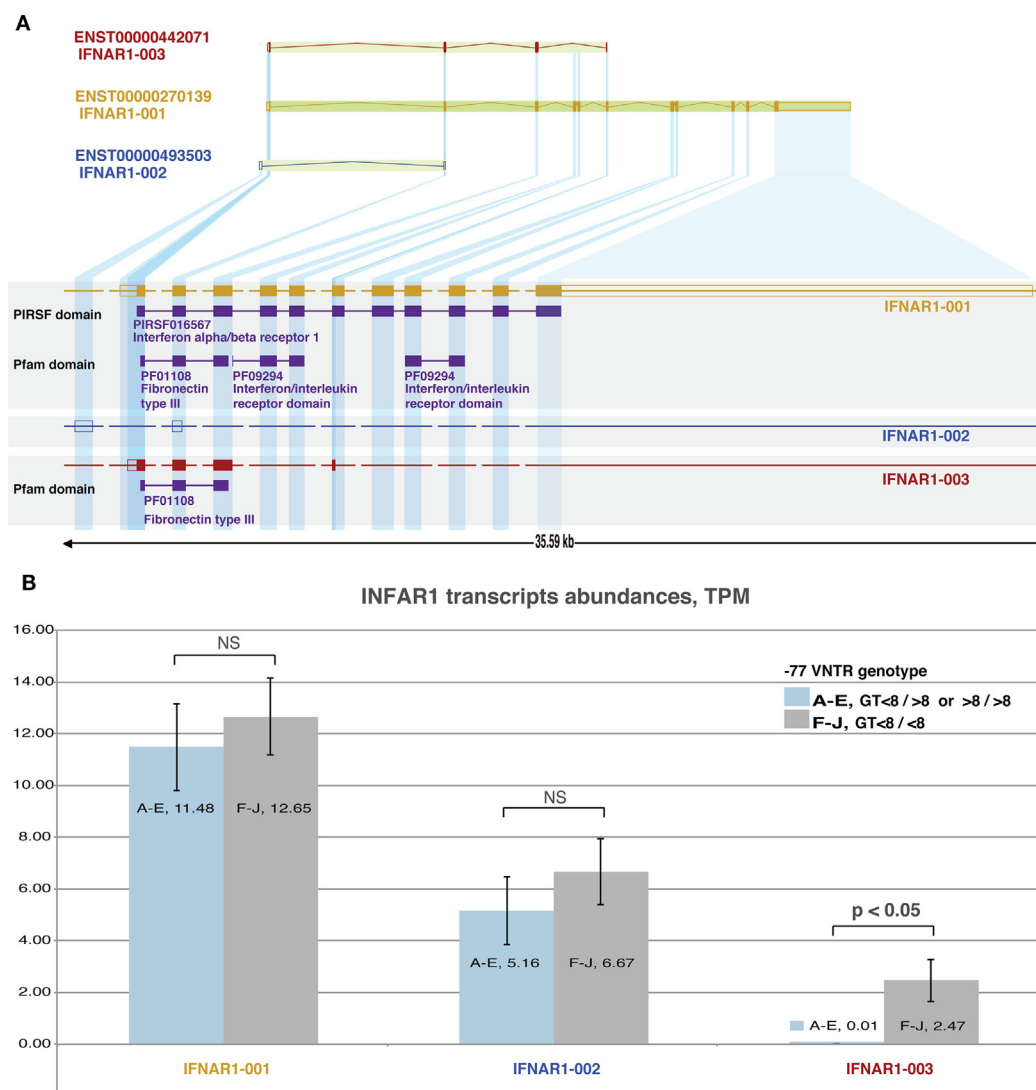


FIGURE 2 | Analysis of IFNAR1 transcripts: **(A)** structure of IFNAR-1 gene. Alternative splicing and transcription initiation sites produce three distinct transcripts that differ in length. Only transcripts 001 and 003 are translated into proteins, which share only the C-terminal part and a fibronectin type III domain. **(B)**: RNAseq-derived expression levels of the three transcripts from 10 hepatocellular carcinoma (HCC) samples grouped based on their –77VNTR(GT)_n genotype. The levels of transcripts 001 (responsible for the production of the receptor) and 002 do not differ significantly between the genotypes. Transcript 003 is expressed only in samples with < 8 / < 8 (GT)_n repeats in the –77VNTR polymorphism. The genotypes of HCC RNA samples (A–J) are described in **Table 1**.

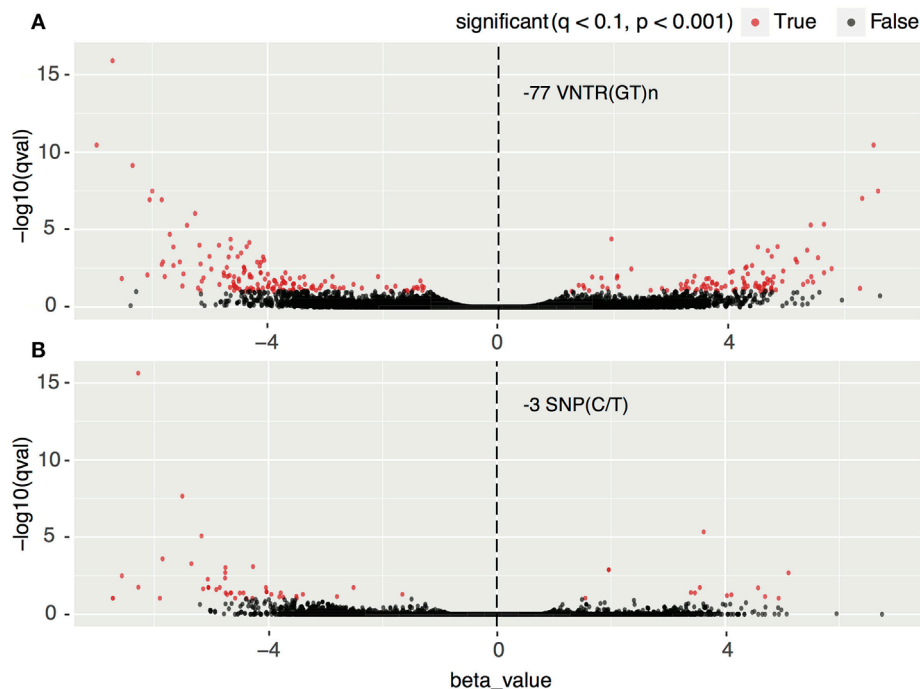


FIGURE 3 | Impact of (A) –77 VNTR (GT)*n* and (B) –3SNP C/T on the hepatocellular carcinoma transcriptome. Transcripts with statistically significant change in their expression are in red. Positive and negative normalized beta values correspond to up- or downregulation, respectively. There are approximately four times more significantly differentially expressed transcripts in the –77 VNTR (GT)*n* compared to the –3SNP C/T genotype group.

polymorphism had a fourfold higher impact on the Interferome compared to the –3SNP, as 46 and 11 transcripts, respectively, are either up- or downregulated ($p < 0.001$). Notably, among the most significantly differentially expressed genes was FN-1, which presented a 4.49-fold lower expression in $\leq 8/\leq 8$ (GT)*n* homozygous patients (Figure 4; Table S2 in Supplementary Material).

Impact of –77 VNTR on Signaling Pathways

Performing GO and KEGG pathway enrichment analysis based on the –77VNTR grouping we found six significantly enriched pathways (Table 2). PI3K–AKT signaling pathway was ranked first, with the higher number and proportion of differentially expressed genes involved (14 genes, $p < 0.05$). The super-family “pathways in cancer” was also found significantly enriched with 15 differentially expressed genes, but since PI3K–AKT pathway was the main contributor, the family was excluded from Table 2. Intriguingly, PI3K–AKT signaling pathway can be triggered by FN-1 (Figure 5).

IFNARPPs and HBV-Associated Hepatocellular Carcinoma

We genotyped the four IFNARPPs in 49 HBV-associated HCC samples, 15 WGS and 34 WXS. We were able to confirm the existence of the –568G allele [previously associated with more pathogenic disease states (26, 32)] in 36 out of 49 HCC

samples, as the read coverage was occasionally reduced over the –568SNP in the WXS dataset. We compared their distributions of the polymorphisms with those described previously for 92 ICs (32). The demographic characteristics of the 49 HCC and the 92 IC patients are summarized in Table S1 in Supplementary Material. There was no statistically significant difference in frequency of the –568SNP genotypes between the IC and HCC groups (Table 3). Given the previously described linkage between alleles –408C/T and –3C/T (–408C to –3C and –408T to –3T) (27) these were analyzed together. There was no statistically significant difference in the prevalence of the –408/–3 SNP polymorphisms in the IC and the HCC groups (Table 3). The –408/–3 TT genotype was not identified in any of the TCGA samples tested. Notably, the number of GT repeats at the –77VNTR of IFNAR-1 promoter was associated with the disease status. Heterozygous $\leq 8/>8$ (GT)*n* patients are less likely to be IC compared to homozygotes carrying both alleles with ≤ 8 GT repeats (OR = 0.41, 95% CI: 0.19, 0.90). The same trend was observed when $>8/>8$ and $\leq 8/>8$ (GT)*n* genotypes were grouped together (OR = 0.40, 95% CI: 0.19, 0.84) (Table 3; Figure S2 in Supplementary Material).

DISCUSSION

Implementing a fully automated computational pipeline, we characterized the genetic profile of the four polymorphisms located in the IFNAR1 promoter region in 49 HBV-associated HCC samples derived from the TCGA project.

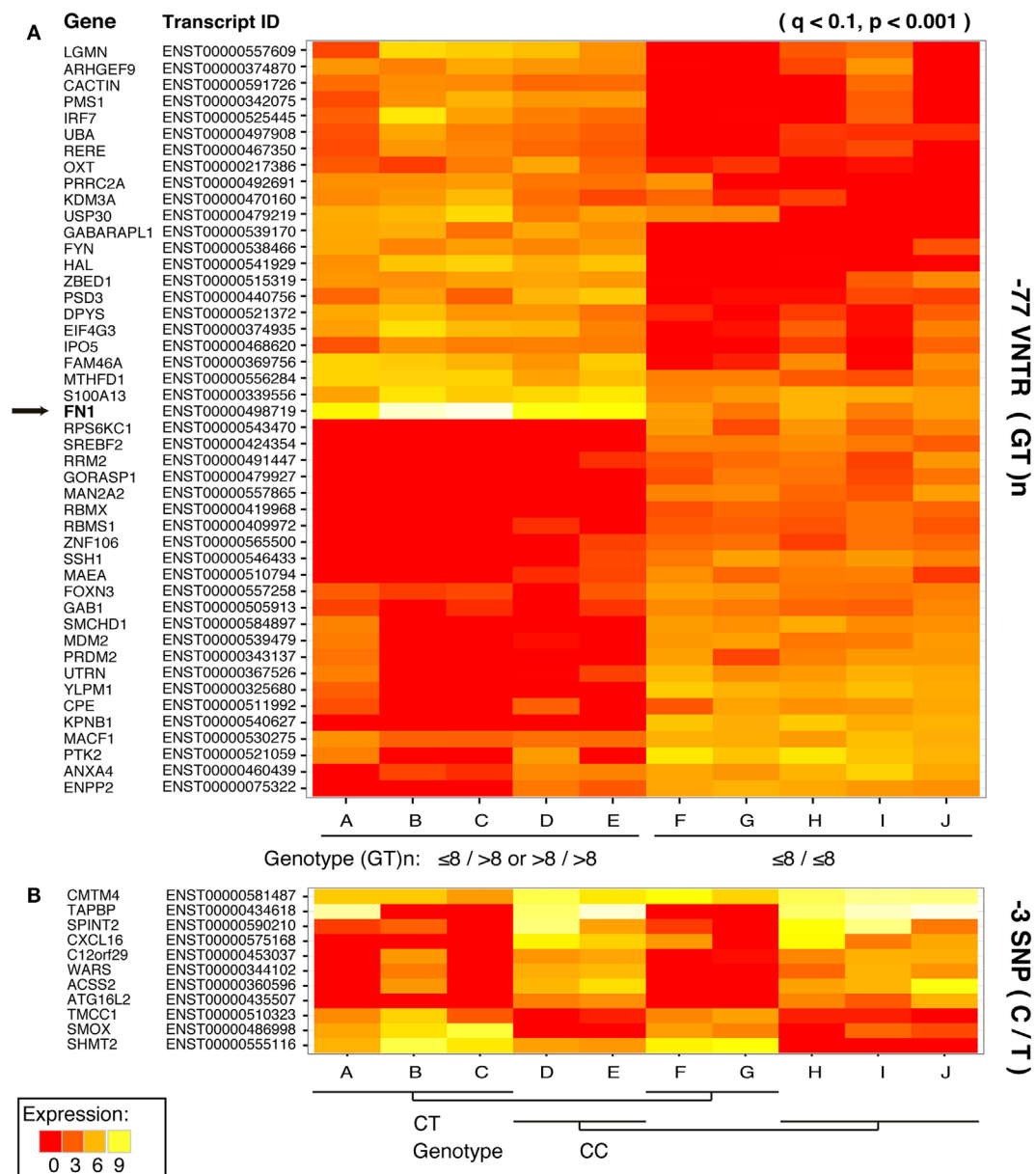


FIGURE 4 | Impact of **(A)** -77 VNTR (GT)n and **(B)** -3SNP C/T on the hepatocellular carcinoma interferome. Only statistically significant differences in the expression levels of the transcripts are presented. Affected transcripts, the respective genes, and the transcripts' product status are on the left.

Whole-genome sequencing data have enormous size and they are usually available as ready-to-use alignments (.bam files), which are binary and compressed. Variation-calling from bam files can be performed with several routine pipelines, each characterized by different biases toward specific types of SNP and in/del genotyping errors (40). For example, reads that differ significantly in length can be ignored and remain unmapped. This is highly likely in VNTRs with heterozygous genotypes in loci where the two alleles differ substantially in length compared to the reference sequence (e.g., -77VNTR(GT)_{5/15}), which in combination with low read coverage and read length (quite common in WGS

data) can lead to inaccurate variant calling. Our pipeline solves this problem, by using an artificial reference of two pseudo-chromosomes, derived from the original sequence, segregated exactly on the VNTR locus, leaving only 3(GT)-repeats at each segregation edge. By using these two pseudo-chromosomes, we were able to specifically mine all the reads corresponding to the IFNAR1 promoter region—originally mapped or unmapped—that in turn precisely reconstructed the two alleles of each sample. Our method provides biologically meaningful results as only one (homozygous) or two (heterozygous) contigs were generated for all the samples tested. There were not any non-specific contigs

generated, indicating that the mining of the reads was highly specific.

Since only 15 TCGA WGS samples met our selection criteria, we used WXS data as well. Although WXS libraries are optimized for mRNA sequencing, non-coding upstream and downstream sequences are randomly captured during the target enrichment

TABLE 2 | Pathway enrichment analysis.

Pathway ID: name	Genes count ^a	% of total Diff. Expr. Genes	p Value	Fold of enrichment
hsa04151: PI3K–Akt signaling pathway	14	4.13	0.0119	2.15
hsa04510: focal adhesion	13	3.83	<0.001	3.35
hsa05205: proteoglycans in cancer	12	3.54	0.0012	3.19
hsa05203: viral carcinogenesis	10	2.95	0.0142	2.59
hsa04010: MAPK signaling pathway	10	2.95	0.0485	2.08
hsa05166: HTLV-I infection	10	2.95	0.0495	2.08

^aA threshold of 10 differentially expressed genes per pathway has been applied.

process (41, 42). Thus, we were able to accurately genotype the proximate to the first IFNAR1 exon polymorphisms, with the exemption of –568SNP, which could only be genotyped in 36 out of the 49 HCC TCGA samples. –568SNP is the most distant one relatively to the start of the transcription, and this resulted in reduced read coverage in 11 WXS samples. The limited number of observations in the HCC group did not allow for statistically significant results with regard to –568 SNP and disease status, although a relative difference in the genotypic distributions was observed (Figure S2 in Supplementary Material). Further studies are needed to confirm the significance of this SNP, as it has been previously associated with the development of CHB and the spontaneous recovery after acute HBV infection (27, 32).

Our results suggest that the number of GT repeats in the –77VNTR is associated with the state of the disease, as alleles with more than 8(GT)_n are more frequent among HCC samples. (Table 3; Figure S2 in Supplementary Material). The finding that more (GT) repeats are associated with a more severe outcome of the disease is in concordance with our previous observations where alleles carrying ≤8(GT) repeats in the –77VNTR, were associated with the IC phase, whereas >8(GT) repeats were more frequent

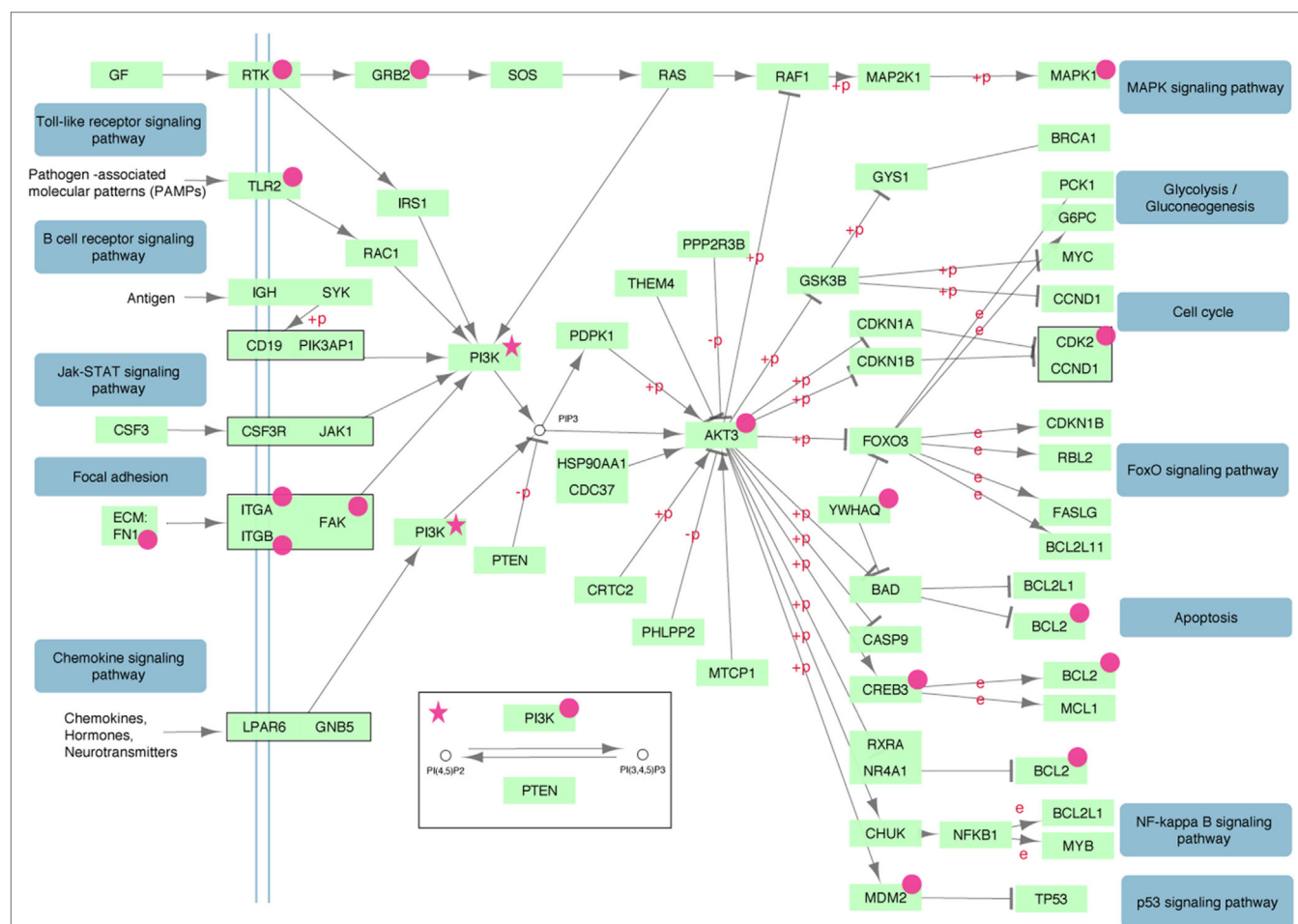


FIGURE 5 | The PI3K–AKT signaling pathway is significantly enriched with differentially expressed genes (magenta). Insulin, ErbB, mTOR, and VEGF signaling sub-pathways triggered by PDK1 have been removed for simplicity.

TABLE 3 | Genotypes of patient samples in relation to clinical status.

	Total (N = 141)	By type of HBV infection		p Value	Crude OR of being IC (95% CI)
		IC (N = 92)	HCC (N = 49)		
IFNARPPs genotypes					
SNP −568*, n (%)				0.313	
CC	54 (42.2)	41 (44.6)	13 (36.1)		1
CG	63 (49.2)	42 (45.7)	21 (58.3)		1.57 (0.65, 3.90)
GG	11 (8.6)	9 (9.8)	2 (5.6)		0.70 (0.06, 4.07)
SNP −408/−3, n (%)				0.094	
CC	62 (44.0)	41 (44.6)	21 (42.9)		1
CT	71 (50.4)	43 (46.7)	28 (57.1)		1.26 (0.59, 2.75)
TT	8 (5.7)	8 (8.7)	0 (0.0)		NA
VNTR −77(GT)n, n (%)					
≤8/≤8	63 (44.7)	48 (52.2)	15 (30.6)		1
≤8/>8	63 (44.7)	36 (39.1)	27 (55.1)	p < 0.05	0.41 (0.19, 0.90)
>8/>8	15 (10.6)	8 (8.7)	7 (14.3)		0.36 (0.11, 1.14)
≤8/≤8	63 (44.7)	48 (52.2)	15 (30.6)		1
≤8/>8 or >8/>8	78 (55.3)	44 (47.8)	34 (69.4)	p < 0.05	0.40 (0.19, 0.84)

Statistically significant associations in bold.

IC, inactive carrier; HCC, hepatocellular carcinoma.

*Genotype available only for 118 samples (92 for IC and 36 for HCC).

in HBeAg-negative CHB patients (32). Zhou et al. studying CHB patients and spontaneously resolved cases after acute HBV infection showed that alleles carrying ≤ 9 (GT) repeats were more frequent in CHB patients, while > 9 (GT) repeats were associated with spontaneous clearance after acute HBV infection (27). Moreover, this result was linked with the rest of the polymorphisms examined in haplotypes, [≤ 9 (GT)n, -568C, -408/-3T] for the CHB patients and > 9 (GT)n, -568G, -408C, -3C in cases of spontaneous clearance after acute HBV infection. In our previous study, alleles -568C and -408/-3T were associated with the IC phase, whereas -568G and -408/-3C were more frequent in CHB patients. In this study, although genotypes -568CG and -408/-3CT show a trend of higher frequency in the HCC samples, compared to genotypes -568CC and -408/-3CC (Figure S2 in Supplementary Material), they were not associated with disease, as the limited availability of HCC samples and the lack of coverage over -568SNP in WXS samples led to lack of statistical power for these comparisons. Notably, genotype -408/-3TT was totally absent in the HCC dataset. Studies in larger datasets may discern whether alleles -568G and -408/-3T and the homozygous genotypes GG and TT are linked to HCC.

The effective expression of IFNAR-1 is essential for the linkage with IFN and the triggering of the downstream signal pathways (43). Even minor modifications in the receptor structure can impair its normal function and alter its antiviral and antiproliferative properties (44, 45). Moreover, the expression levels of both the receptor subunits IFNAR1 and IFNAR2 are associated with the INF- β treatment outcome in multiple sclerosis patients (46). Hepatitis C patients with 5GT repeats in the -77VNTR (genotypes GT_{5/5}/GT_{5/14}) have been reported to be better responders to IFN-based therapy (47). On the other hand, experiments based on luciferase reporter plasmids have suggested that the promoter of the IFNAR-1 is not affected by the -568SNP and the -77VNTR (26, 27), while in another study its activity is significantly affected

by -3SNP (48). In detail, the promoter activity was reduced for -3T plasmids through reduced binding affinity to HMGB1, a factor that was suggested to bind to the -3 element to regulate the transcription levels of IFNAR1. Similarly, reduced levels of IFNAR1 expression in HBV patients with C > T substitution at the -3 position of the IFNAR1 promoter was reported in the same study (48). In another study, an important role of the -3CT SNP was evident regarding the regulation of the transcription factor High Mobility Group B protein 1 (HMGB-1); the C > T transition was shown to reduce the binding affinity of HMGB-1 to the IFNAR-1 promoter sequence, thus lead to reduced expression of IFNAR-1 (48).

In this study, we conducted a thorough RNAseq differential expression analysis on HCC samples grouped according to their -77VNTR and -3SNP genotype. In agreement with our previous findings, where a significant role of the -77VNTR genotypes had arisen with respect to the clinical course of HBV infection, here we report a more significant involvement of -77VNTR in the modification of the HCC transcriptome and interferome, as more genes were found to be significantly differentially expressed compared to those affected by the -3SNP. We also shed light on the controversial question about the impact of these promoter polymorphisms on the expression of IFNAR-1 gene. The expression levels of the receptor are associated with the clinical outcome of chronically HBV-infected patients (49). We conclude that the expression of the major IFNAR-1 transcript, which is responsible for the production of the functional receptor, is not significantly affected by these polymorphisms (Figure 2).

At the same time, a secondary IFNAR1 transcript, which is translated into a truncated form of the receptor, is only produced at detectable levels in samples with $\leq 8/\leq 8$ (GT)n genotype in the -77VNTR, independently from the -3SNP genotype. This transcript is not detectable in normal liver tissue samples,

thus appears to be HCC specific. Secondary IFNAR1 transcripts have been reported in tumor cell lines (50) and here, for the first time, we report that –77VNTR genotype is crucial for their production, which apparently leads to the transcriptional remodeling of IFNAR1 gene.

Grouping the samples according to their –77VNTR genotype revealed that one of the most differentially expressed genes was the extracellular matrix FN-1. Patients carrying at least one –77VNTR >8(GT) allele, presented a strong upregulation of the FN-1 gene, while FN-1 was significantly downregulated in ≤8/≤8(GT)-repeats carrying patients, a phenomenon coinciding with the expression of the secondary IFNAR-1 transcript, which implies its potentially protective role against the development of HCC. This transcript comprises only a fibronectin III domain; FN fragments and modules can inhibit FN-matrix assembly by competing for FN-assembly sites, which could act as a feedback system to regulate FN levels on the cell surface (51). Specifically, the first type III domain of the FN molecule is important for the matrix assembly (52, 53), while even small fragments derived from this module regulate FN polymerization, inducing it at moderate concentrations but inhibiting it at high concentrations (54, 55). FN-1 is involved in HCC as it can be upregulated by HBxAg in an NFκB-dependent way (14), while is generally over-expressed in several cancers [reviewed in Ref. (56)]. NFκB was not found differentially expressed in our dataset ($p = 0.16$) but is linked to PI3K-AKT, which was the most significantly enriched signaling pathway, down to the MDM2 proto-oncogene and the apoptosis regulator Bcl-2 (Figure 5). FN-1 can trigger this pathway and transduce multiple intracellular signals that control cell cycle. Our data suggest that apart from NFκB, complementary mechanisms, partially controlled by FN-1 and involved in cell cycle regulation might exist.

This study is based on data mining and *in silico* analyses, thus our findings here generate a strong hypothesis about links

between the expression of the truncated IFNAR-1 transcript either with the PI3K-AKT signaling pathway and/or with the HCC development. To test this hypothesis, further wet-lab studies will be needed involving knock-down of the transcript, transfection with a plasmid that will produce this transcript and quantitative analysis of the respective transcripts and proteins would confirm these observations and would shed light in these complex regulatory networks.

In conclusion, our results suggest very minimal (if any) involvement of the IFNAR-1 promoter polymorphisms in the expression levels of the IFNAR-1 major transcript but at the same time raise a potential and intriguing role for the –77VNTR regarding the regulation of downstream genes. Our study shows that the majority of modifications of the *Interferome* coincided with the production of the truncated IFNAR-1 transcript. Thus, further study of this truncated transcript could clarify the mechanistic features of the combined antiviral and anticancer roles of IFNAR-1.

AUTHOR CONTRIBUTIONS

TK designed and conducted the analyses, evaluated the results, and wrote the manuscript. GP, DP, AH, JM, UG, PK, and GM wrote and revised the manuscript.

FUNDING

The work was supported the Medical Research Council, UK (Project Reference: MR/K010565/1).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at <https://www.frontiersin.org/articles/10.3389/fimmu.2018.00777/full#supplementary-material>.

REFERENCES

- Schweitzer A, Horn J, Mikolajczyk RT, Krause G, Ott JJ. Estimations of world-wide prevalence of chronic hepatitis B virus infection: a systematic review of data published between 1965 and 2013. *Lancet* (2015) 386(10003):1546–55. doi:10.1016/S0140-6736(15)61412-X
- European Association for the Study of the Liver. EASL clinical practice guidelines: management of chronic hepatitis B virus infection. *J Hepatol* (2012) 57:167–85. doi:10.1016/j.jhep.2012.02.010
- Funk ML, Rosenberg DM, Lok AS. World-wide epidemiology of HBeAg-negative chronic hepatitis B and associated precore and core promoter variants. *J Viral Hepat* (2002) 9(1):52–61. doi:10.1046/j.1365-2893.2002.00304.x
- Papatheodoridis GV, Manolakopoulos S, Liaw YF, Lok A. Follow-up and indications for liver biopsy in HBeAg-negative chronic hepatitis B virus infection with persistently normal ALT: a systematic review. *J Hepatol* (2012) 57(1):196–202. doi:10.1016/j.jhep.2011.11.030
- Hadziyannis SJ, Papatheodoridis GV. Hepatitis B e antigen-negative chronic hepatitis B: natural history and treatment. *Semin Liver Dis* (2006) 26(2):130–41. doi:10.1055/s-2006-939751
- Ribes J, Cleries R, Esteban L, Moreno V, Bosch FX. The influence of alcohol consumption and hepatitis B and C infections on the risk of liver cancer in Europe. *J Hepatol* (2008) 49(2):233–42. doi:10.1016/j.jhep.2008.04.016
- Bosetti C, Levi F, Boffetta P, Lucchini F, Negri E, La Vecchia C. Trends in mortality from hepatocellular carcinoma in Europe, 1980–2004. *Hepatology* (2008) 48(1):137–45. doi:10.1002/hep.22312
- Bellentani S, Scaglioni F, Ciccia S, Bedogni G, Tiribelli C. HCV, HBV and alcohol – the dionysos study. *Dig Dis* (2010) 28(6):799–801. doi:10.1159/000324288
- Hsu YS, Chien RN, Yeh CT, Sheen IS, Chiou HY, Chu CM, et al. Long-term outcome after spontaneous HBeAg seroconversion in patients with chronic hepatitis B. *Hepatology* (2002) 35(6):1522–7. doi:10.1053/jhep.2002.33638
- de Franchis R, Meucci G, Vecchi M, Tatarella M, Colombo M, Del Ninno E, et al. The natural history of asymptomatic hepatitis B surface antigen carriers. *Ann Intern Med* (1993) 118(3):191–4. doi:10.7326/0003-4819-118-3-199302010-00006
- Jung MC, Pape GR. Immunology of hepatitis B infection. *Lancet Infect Dis* (2002) 2(1):43–50. doi:10.1016/S1473-3099(01)00172-4
- Jiang YG, Wang YM, Liu TH, Liu J. Association between HLA class II gene and susceptibility or resistance to chronic hepatitis B. *World J Gastroenterol* (2003) 9(10):2221–5. doi:10.3748/wjg.v9.i10.2221
- Li N, Gao Y-F, Zhang T-C, Chen P, Li X, Su F. Relationship between interleukin 18 polymorphisms and susceptibility to chronic hepatitis B virus infection. *World J Hepatol* (2012) 4(3):105–9. doi:10.4254/wjh.v4.i3.105

14. Norton PA, Reis HM, Prince S, Larkin J, Pan J, Liu J, et al. Activation of fibronectin gene expression by hepatitis B virus x antigen. *J Viral Hepat* (2004) 11(4):332–41. doi:10.1111/j.1365-2893.2004.00555.x
15. To WS, Midwood KS. Plasma and cellular fibronectin: distinct and independent functions during tissue repair. *Fibrogenesis Tissue Repair* (2011) 4:21. doi:10.1186/1755-1536-4-21
16. Liu X-Y, Liu R-X, Hou F, Cui L-J, Li C-Y, Chi C, et al. Fibronectin expression is critical for liver fibrogenesis in vivo and in vitro. *Mol Med Rep* (2016) 14(4):3669–75. doi:10.3892/mmr.2016.5673
17. Shi M, Fu J, Shi F, Zhang B, Tang Z, Jin L, et al. Transfusion of autologous cytokine-induced killer cells inhibits viral replication in patients with chronic hepatitis B virus infection. *Clin Immunol* (2009) 132(1):43–54. doi:10.1016/j.clim.2009.03.001
18. Pham TN, Mercer SE, Michalak TI. Chronic hepatitis C and persistent occult hepatitis C virus infection are characterized by distinct immune cell cytokine expression profiles. *J Viral Hepat* (2009) 16(8):547–56. doi:10.1111/j.1365-2893.2009.01092.x
19. Parker BS, Rautela J, Hertzog PJ. Antitumour actions of interferons: implications for cancer therapy. *Nat Rev Cancer* (2016) 16(3):131–44. doi:10.1038/nrc.2016.14
20. Domanski P, Colamonici OR. The type-I interferon receptor. the long and short of it. *Cytokine Growth Factor Rev* (1996) 7(2):143–51. doi:10.1016/1359-6101(96)00017-2
21. Mogensen KE, Lewerenz M, Reboul J, Lutfalla G, Uze G. The type I interferon receptor: structure, function, and evolution of a family business. *J Interferon Cytokine Res* (1999) 19(10):1069–98. doi:10.1089/107999099313019
22. Langer JA, Cutrone EC, Kotenko S. The Class II cytokine receptor (CRF2) family: overview and patterns of receptor-ligand interactions. *Cytokine Growth Factor Rev* (2004) 15(1):33–48. doi:10.1016/j.cytogfr.2003.10.001
23. Rusinova I, Forster S, Yu S, Kannan A, Masse M, Cumming H, et al. Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Res* (2013) 41(Database issue):D1040–6. doi:10.1093/nar/gks1215
24. Liaw Y-F, Kao J-H, Piratvisuth T, Chan H, Chien R-N, Liu C-J, et al. Asian-Pacific consensus statement on the management of chronic hepatitis B: a 2012 update. *Hepatol Int* (2012) 6(3):531–61. doi:10.1007/s12072-012-9365-4
25. Pestka S, Krause CD, Walter MR. Interferons, interferon-like cytokines, and their receptors. *Immunol Rev* (2004) 202:8–32. doi:10.1111/j.0105-2896.2004.00204.x
26. Zhou J, Smith DK, Lu L, Poon VK, Ng F, Chen DQ, et al. A non-synonymous single nucleotide polymorphism in IFNAR1 affects susceptibility to chronic hepatitis B virus infection. *J Viral Hepat* (2009) 16(1):45–52. doi:10.1111/j.1365-2893.2008.01040.x
27. Zhou J, Lu L, Yuen MF, Lam TW, Chung CP, Lam CL, et al. Polymorphisms of type I interferon receptor 1 promoter and their effects on chronic hepatitis B virus infection. *J Hepatol* (2007) 46(2):198–205. doi:10.1016/j.jhep.2006.08.017
28. He XX, Chang Y, Jiang HJ, Tang F, Meng FY, Xie QH, et al. Persistent effect of IFNAR-1 genetic polymorphism on the long-term pathogenesis of chronic HBV infection. *Viral Immunol* (2010) 23(3):251–7. doi:10.1089/vim.2009.0102
29. Song le H, Xuan NT, Toan NL, Binh VQ, Boldt AB, Kremsner PG, et al. Association of two variants of the interferon-alpha receptor-1 gene with the presentation of hepatitis B virus infection. *Eur Cytokine Netw* (2008) 19(4):204–10. doi:10.1684/ecn.2008.0137
30. Muldoon J, Uriel A, Khoo S, Ollier WE, Hajeer AH. Novel IFN-alpha receptor promoter polymorphisms. *Genes Immun* (2001) 2(3):159–60. doi:10.1038/sj.gene.6363757
31. Aucan C, Walley AJ, Hennig BJ, Fitness J, Frodsham A, Zhang L, et al. Interferon-alpha receptor-1 (IFNAR1) variants are associated with protection against cerebral malaria in the Gambia. *Genes Immun* (2003) 4(4):275–82. doi:10.1038/sj.gene.6363962
32. Karamitros T, Papatheodoridis G, Dimopoulou E, Papageorgiou MV, Paraskevis D, Magiorkinis G, et al. The interferon receptor-1 promoter polymorphisms affect the outcome of caucasians with HBeAg-negative chronic HBV infection. *Liver Int* (2015) 35(12):2506–13. doi:10.1111/liv.12859
33. Papatheodoridis GV, Manolakopoulos S, Margariti A, Papageorgiou MV, Kranidioti H, Katoglou A, et al. The usefulness of transient elastography in the assessment of patients with HBeAg-negative chronic hepatitis B virus infection. *J Viral Hepat* (2014) 21(7):517–24. doi:10.1111/jvh.12176
34. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* (2009) 25(16):2078–9. doi:10.1093/bioinformatics/btp352
35. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* (2012) 9(4):357–9. doi:10.1038/nmeth.1923
36. Chevreur B, Wetter T, Suhai S. Genome sequence assembly using trace signals and additional sequence information. *Computer Science and Biology Proceedings of the German Conference on Bioinformatics*. Hannover (1999).
37. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* (2013) 14(2):178–92. doi:10.1093/bib/bbs017
38. Bray NL, Pimentel H, Melsted P, Pachter L. Erratum: near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* (2016) 34(8):888. doi:10.1038/nbt.3519
39. Pimentel H, Bray NL, Puente S, Melsted P, Pachter L. Differential analysis of RNA-seq incorporating quantification uncertainty. *Nature Methods* (2017) 14: 687.
40. Hwang S, Kim E, Lee I, Marcotte EM. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci Rep* (2015) 5:17875. doi:10.1038/srep17875
41. Karamitros T, Magiorkinis G. A novel method for the multiplexed target enrichment of MinION next generation sequencing libraries using PCR-generated baits. *Nucleic Acids Res* (2015) 43(22):e152. doi:10.1093/nar/gkv773
42. Tsangaras K, Wales N, Sicheritz-Ponten T, Rasmussen S, Michaux J, Ishida Y, et al. Hybridization capture using short PCR products enriches small genomes by capturing flanking sequences (CapFlank). *PLoS One* (2014) 9(10):e109101. doi:10.1371/journal.pone.0109101
43. Massier KB, Hirata MH, Silva AE, Ferraz ML, Nguyen NY, Hirata RD. Interferon-alpha receptor 1 mRNA expression in peripheral blood mononuclear cells is associated with response to interferon-alpha therapy of patients with chronic hepatitis C. *Braz J Med Biol Res* (2004) 37(5):643–7. doi:10.1590/S0100-879X2004000500003
44. de Weerd NA, Matthews AY, Pattie PR, Bourke NM, Lim SS, Vivian JP, et al. A hot spot on interferon alpha/beta receptor subunit 1 (IFNAR1) underpins its interaction with interferon-beta and dictates signaling. *J Biol Chem* (2017) 292(18):7554–65. doi:10.1074/jbc.M116.773788
45. Zhang G, deWeerd NA, Stifter SA, Liu L, Zhou B, Wang W, et al. A proline deletion in IFNAR1 impairs IFN-signaling and underlies increased resistance to tuberculosis in humans. *Nat Commun* (2018) 9(1):85. doi:10.1038/s41467-017-02611-z
46. Karam RA, Rezk NA, Amer MM, Fathy HA. Immune response genes receptors expression and polymorphisms in relation to multiple sclerosis susceptibility and response to INF-beta therapy. *IUBMB Life* (2016) 68(9):727–34. doi:10.1002/iub.1530
47. Matsuyama N, Mishiyo S, Sugimoto M, Furuichi Y, Hashimoto M, Hijikata M, et al. The dinucleotide microsatellite polymorphism of the IFNAR1 gene promoter correlates with responsiveness of hepatitis C patients to interferon. *Hepatol Res* (2003) 25(3):221–5. doi:10.1016/S1386-6346(02)00269-3
48. Zhou J, Huang JD, Poon VK, Chen DQ, Chan CC, Ng F, et al. Functional dissection of an IFN-alpha/beta receptor 1 promoter variant that confers higher risk to chronic hepatitis B virus infection. *J Hepatol* (2009) 51(2):322–32. doi:10.1016/j.jhep.2009.03.020
49. Santos JC, de Deus DMV, de Moura IMF, Lopes EP, Alves MRB, Coelho MRCD. Association between the IFNA1 (-2C→T) polymorphism and increased IFNAR1 gene expression levels in chronic hepatitis B infection. *Intervirology* (2015) 58(6):393–402. doi:10.1159/000444365
50. de Weerd NA, Samarajiva SA, Hertzog PJ. Type I interferon receptors: biochemistry and biological functions. *J Biol Chem* (2007) 282(28):20053–7. doi:10.1074/jbc.R700006200
51. Bourdoulous S, Orend G, MacKenna DA, Pasqualini R, Ruoslahti E. Fibronectin matrix regulates activation of RHO and CDC42 GTPases and cell cycle progression. *J Cell Biol* (1998) 143(1):267–76. doi:10.1083/jcb.143.1.267
52. Aguirre KM, McCormick RJ, Schwarzbauer JE. Fibronectin self-association is mediated by complementary sites within the amino-terminal one-third of the molecule. *J Biol Chem* (1994) 269(45):27863–8.

53. Hocking DC, Sottile J, McKeown-Longo PJ. Fibronectin's III-1 module contains a conformation-dependent binding site for the amino-terminal region of fibronectin. *J Biol Chem* (1994) 269(29):19183–7.
54. Morla A, Ruoslahti E. A fibronectin self-assembly site involved in fibronectin matrix assembly: reconstruction in a synthetic peptide. *J Cell Biol* (1992) 118(2):421–9. doi:10.1083/jcb.118.2.421
55. Morla A, Zhang Z, Ruoslahti E. Superfibronectin is a functionally distinct form of fibronectin. *Nature* (1994) 367(6459):193–6. doi:10.1038/367193a0
56. Wang JP, Hielscher A. Fibronectin: how its aberrant expression in tumors may improve therapeutic targeting. *J Cancer* (2017) 8(4):674–82. doi:10.7150/jca.16901

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Karamitros, Papatheodoridis, Paraskevis, Hatzakis, Mbisa, Georgopoulou, Kleenerman and Magiorkinis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evolution of Two Major Zika Virus Lineages: Implications for Pathology, Immune Response, and Vaccine Development

Jacob T. Beaver, Nadia Lelutiu, Rumi Habib and Ioanna Skountzou*

Department of Microbiology and Immunology, Emory Vaccine Center, Emory University School of Medicine, Atlanta, GA, United States

OPEN ACCESS

Edited by:

Gkikas Magiorkinis,
National and Kapodistrian
University of Athens, Greece

Reviewed by:

Luciana Barros Arruda,
Universidade Federal do Rio de
Janeiro, Brazil
Jose Luiz Proenca-Modena,
Universidade Estadual de
Campinas, Brazil

*Correspondence:

Ioanna Skountzou
iskount@emory.edu

Specialty section:

This article was submitted
to Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 02 May 2018

Accepted: 03 July 2018

Published: 18 July 2018

Citation:

Beaver JT, Lelutiu N, Habib R and
Skountzou I (2018) Evolution of Two
Major Zika Virus Lineages:
Implications for Pathology, Immune
Response, and Vaccine
Development.
Front. Immunol. 9:1640.
doi: 10.3389/fimmu.2018.01640

Zika virus (ZIKV) became a public health emergency of global concern in 2015 due to its rapid expansion from French Polynesia to Brazil, spreading quickly throughout the Americas. Its unexpected correlation to neurological impairments and defects, now known as congenital Zika syndrome, brought on an urgency to characterize the pathology and develop safe, effective vaccines. ZIKV genetic analyses have identified two major lineages, Asian and African, which have undergone substantial changes during the past 50 years. Although ZIKV infections have been circulating throughout Africa and Asia for the later part of the 20th century, the symptoms were mild and not associated with serious pathology until now. ZIKV evolution also took the form of novel modes of transmission, including maternal–fetal transmission, sexual transmission, and transmission through the eye. The African and Asian lineages have demonstrated differential pathogenesis and molecular responses *in vitro* and *in vivo*. The limited number of human infections prior to the 21st century restricted ZIKV research to *in vitro* studies, but current animal studies utilize mice deficient in type I interferon (IFN) signaling in order to invoke enhanced viral pathogenesis. This review examines ZIKV strain differences from an evolutionary perspective, discussing how these differentially impact pathogenesis *via* host immune responses that modulate IFN signaling, and how these differential effects dictate the future of ZIKV vaccine candidates.

Keywords: Zika virus, evolution, transmission, tissue tropism, phylogeny, immunology, vaccines

INTRODUCTION

Zika virus (ZIKV) has garnered international attention due to its rapid worldwide expansion since 2015 when an epidemic struck Brazil, resulting in a newly identified pathology including severe neurological impairments such as microcephaly, which is now part of the congenital ZIKV syndrome, as well as Guillain–Barré syndrome (GBS) afflicting adults (1). The World Health Organization declared ZIKV a Public Health Emergency of International Concern in February 2016, during which time ZIKV was spreading rapidly across South America, the Caribbean, and into the United States (1). This precarious outbreak in Brazil spread rampant across the western continents raising critical questions pertaining to the evolution of this virus. Prior to 2015, ZIKV infections were limited geographically to Africa and Asia and were reported to be asymptomatic, and approximately 20% mildly symptomatic represented as a self-limiting febrile illness with most common symptoms

maculopapular rash, conjunctivitis, and joint pain (2). The mounting evidence that ZIKV is now causing neuropathology and fetal brain disruption, as well as rising concerns over novel modes of ZIKV transmission suggests an evolutionary change in the molecular and genetic structure of ZIKV strains that has contributed to its rapid expansion, severity of pathogenicity, and multiple routes of infections. These increasing adverse effects depicts why an analysis of the phenotypic differences between the African and Asian lineages, as well as between the many strains, which have evolved under each branch, is a vital component of our ongoing effort to develop vaccines or therapeutics and fill major gaps of knowledge regarding ZIKV pathogenesis.

HISTORY OF VIRUS EMERGENCE

Zika virus was discovered in the Zika Forest of Uganda in 1947 by Alexander Haddow and George Dick during a surveillance investigation of yellow fever in rhesus macaques in Uganda (3). The virus was later isolated from the *Aedes africanus* mosquito collected at the same site (4). The first human case occurred in Nigeria in 1954, but it was not until 1966 that ZIKV was first detected in Asia alongside the first evidence of transmission by an urban vector, *Aedes aegypti* mosquitoes from Malaysia (4, 5). We know that two major lineages of ZIKV were formed at this time, African and Asian, which is confirmed by current genetic and phylogenetic analyses (2). ZIKV made no headlines until an outbreak in 2007 on Yap Island, Micronesia, rendering 73% of the residents infected (6). Despite the presence of DENV IgM in all affected individuals, the unique symptomatic presentation was definitively identified as ZIKV-induced. The next outbreak was 6 years later in French Polynesia, spreading to several other islands in Oceania. The most commonly reported symptoms in the Yap Island and French Polynesian outbreaks included rash, fever, arthralgia, and non-purulent conjunctivitis (7, 8). However, the first case with GBS as a complication of ZIKV infection was reported in the 2013 outbreak in French Polynesia (7). It was also in 2013 when it was discovered that ZIKV transmission could occur through blood or other bodily fluids and not just through mosquito bites.

Brazil was the next location to experience an outbreak early in 2015. Phylogenetic and molecular clock analyses revealed that there was a single introduction of ZIKV to the country (9). The virus was likely brought to Brazil by a traveler from French Polynesia after a stop at Easter Island (10–12). A recent article from Passos et al. performed a retrospective blood bank analysis on 210 samples collected from patients during a DENV-4 outbreak that occurred in early 2013 (13). Of these samples, 10% showed a singly positive qRT-PCR result for ZIKV, and only 2% demonstrated consistently positive results across triplicate samples. While the cycle threshold for positive results by this group is less stringent than those of other groups, it regardless provides potential insight that ZIKV may have been present in South American countries as early as April of 2013.

From the confirmed 2015 cases, it took less than 1 year for the virus to spread throughout Brazil, into neighboring South American countries, and into Central and North America.

The increase in GBS cases was reported in Brazil, Colombia, Suriname, and Venezuela and microcephaly cases in NE Brazil, which included neurological disorders and neonatal malformations (12, 14–23). The remarkable rise of infants born with microcephaly in Brazil set off international alarms and garnered global attention (24). Sequence homology studies reveal that of the two ZIKV lineages, the strains responsible for the human outbreaks throughout the Americas were phylogenetically closest to the Asian lineage (25).

MOLECULAR BIOLOGY OF ZIKV

ZIKV Genome Organization

Zika virus is a positive single-stranded RNA virus that belongs in the *Flaviviridae* family. This family includes the human pathogenic viruses, Japanese encephalitis virus, dengue virus type 1–4 (DENV), yellow fever virus (YFV), West Nile virus (WNV), and tick-borne encephalitis virus. To better understand how ZIKV might be evolving, it is important to understand its genomic structure. The genome is a 10.8 kb single-strand, positive-sense RNA molecule that consists of a 5' untranslated region (UTR) (~107 nt), one open reading frame (ORF) (~10.2 kb), and a 3' UTR (~420 nt). The ORF encodes a polyprotein precursor that is processed into three structural proteins; capsid (C), pre-membrane/membrane (prM), and envelope protein (E) as well as seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5). The viral polyprotein is co-translationally or co-post-translationally cleaved by viral NS2/NS3 protease, host signal peptidase (C/prM, prM/E, E/NS1, 2K/NS4B) and a host protease (NS1/NS2A). The pr- fragment of the prM protein is cleaved by furin in the trans-Golgi apparatus to generate mature virions. The major surface glycoprotein involved in host cell binding and membrane fusion is E protein. Viral reproduction is accomplished through the non-structural proteins (NS1–NS5), which serve as self-cleaving peptidases, along with the viral RNA-dependent RNA-polymerase. The genome organization and major protein functions of ZIKV are highly similar to all other members in the *Flavivirus* genus (25–27).

As RNA genome viruses are strategically organized to contain the minimal number of genes required for sufficient replication and host immune evasion, many RNA viruses have evolved innovative methods for manipulating subverting molecules within their host cells (28). Among these are non-coding, subgenomic RNAs. These subgenomic flavivirus RNA components (sfRNA) have been implicated in both the reduction of type I interferon (IFN) transcription, and in mediating resistance to cellular exonucleases that would degrade genomic transcripts, such as Xrn1 (29, 30). While the complete functional role of sfRNAs remains unknown, few key pieces of information have already emerged regarding ZIKV. Of these, work by Donald et al. suggests that sfRNA of ZIKV can not only inhibit the type I IFN response by means of a pseudo-knot tertiary structure, but may do so in a manner that is more broad than those of other flaviviruses, such as DENV (31). Additionally, the difference in ZIKV lineage does not impact the generation of these sfRNAs, and is unlikely to impact the predicted tertiary structure.

Genetic Evolution of the Virus

The MR766 (HQ234498) strain of ZIKV from Uganda is considered the classical strain and is used consistently in both *in vitro* and *in vivo* research studies to model ZIKV infections. While this strain has been passaged 147 times in insect cell cultures and suckling-mouse brain tissues, very few mutations have been detected in its genome. In fact, when genomic sequences are compared between MR766 to two other variations, AY632535 and DQ859059, which were both isolated from Uganda in 1947 from sentinel Rhesus macaques, all three variants were determined to differ in only 0.4% of nucleotide and 0.6% of amino acid sequences (32). The initial low mutation rates may have been responsible for low transmission to humans and possibly subclinical infections. Phylogenetic analysis of available ZIKV genomes reveals that 86.5% of isolates are from humans, 11% are from mosquitoes, 2% are isolated from NHPs. Interestingly, of the available genomes, African lineages are only isolated from mosquitoes and NHPs, while Asian lineages are isolated from both humans and mosquitoes (**Figure 1**).

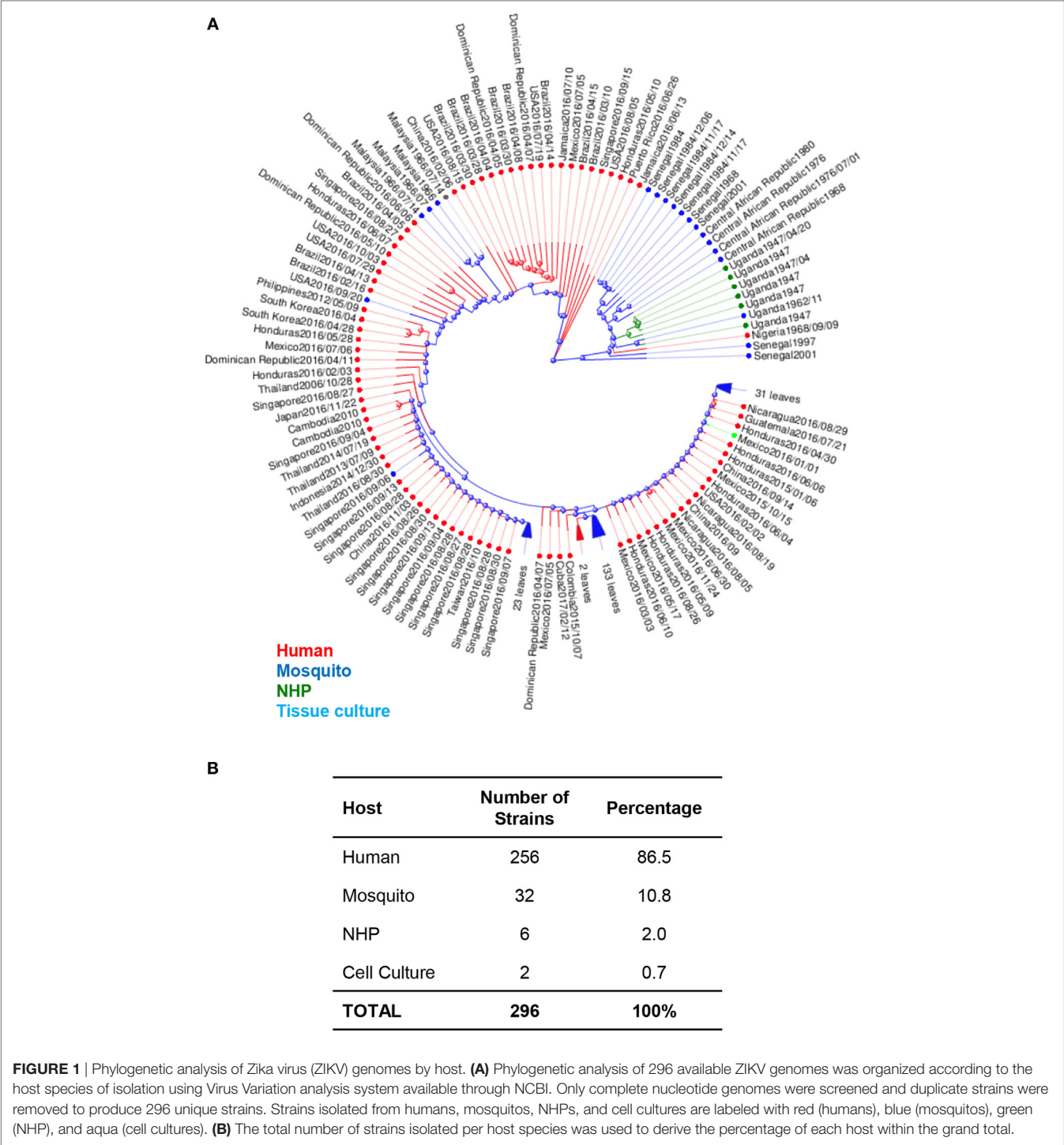
Phylogenetic trees of ZIKV have also been used to study the movement of ZIKV strains across the globe to identify potentially serious mutations that could alter molecular mechanisms, which then lead to enhanced pathology (33). Phylogenetic analysis of available ZIKV genomes reveals approximately 97% of genomes published are from the Asian lineage, and 7% are of African lineage. Among these Asian lineages, 66.9% of all isolates were collected from North, South, and Central America. Of these, 38.8% of isolates were from North and Central America, while only 28.1% were from South America. The remaining 26.1% of Asian lineage isolates were collected from the Asian and Oceania continents. Only 7% of all analyzed strains were from Africa (**Figure 2**).

A phylogenetic analysis juxtaposing MR766 strain to Asian lineage strains, particularly from Suriname and French Polynesia, reveals 50 amino acid lineage-specific differences (2, 34–37). Of these, all variations occur in either the NS1 or NS5 proteins (34, 35). Yet, when Wang et al. compared human to mosquito strains from the French Micronesia (FSM) outbreak in 2007 and the French Polynesia outbreak in 2013 (H/PF/2013), they identified 435 and 446 nucleotide changes in FSM and H/PF/2013, respectively, although 344 nucleotides were identical. Wang et al. considered them as sub-lineages deriving from the same ancestor that arrived in Malaysia in 1966 but had seemingly no clinical impact for 50 years (2, 26). All contemporary human strains within the Americas share higher sequence homology with the Asian lineage P6-740 (Malaysia/1966), which was the sole mosquito isolate (*A. aegypti*) than IbH-30656 (Nigeria/1968) (38). These isolates are most closely related to the H/PF/2013 strain (French Polynesia/2013) than the FSM strain (Micronesia/2007), suggesting that although the two Asian variants evolved from a common ancestor, they further diversified and the genetic distance between the 2007 and 2013 variants increased (2). A third major lineage from Africa is thought to exist based on analysis of only the E and NS5 gene sequences (35, 39). This lineage is designated Africa II, but it is neglected due to incomplete sequencing of the whole genome.

Multiple sequence alignments using 58 complete genome and five envelope sequences of ZIKV as of April 2016, revealed conserved amino acid variations. Nineteen variations were found in the sequence of structural proteins and 47 variations in the non-structural proteins, with the most variations in NS5 although the RNA-dependent polymerase domain had no variations at conserved motifs (26, 40). Eight variation sites were located in E protein; two sites were in the stem region and one site in the transmembrane region of the E protein. Substitutions in stem and transmembrane regions affect virion assembly and membrane fusion, whereas substitutions in Domain III of E protein may affect receptor binding (41, 42).

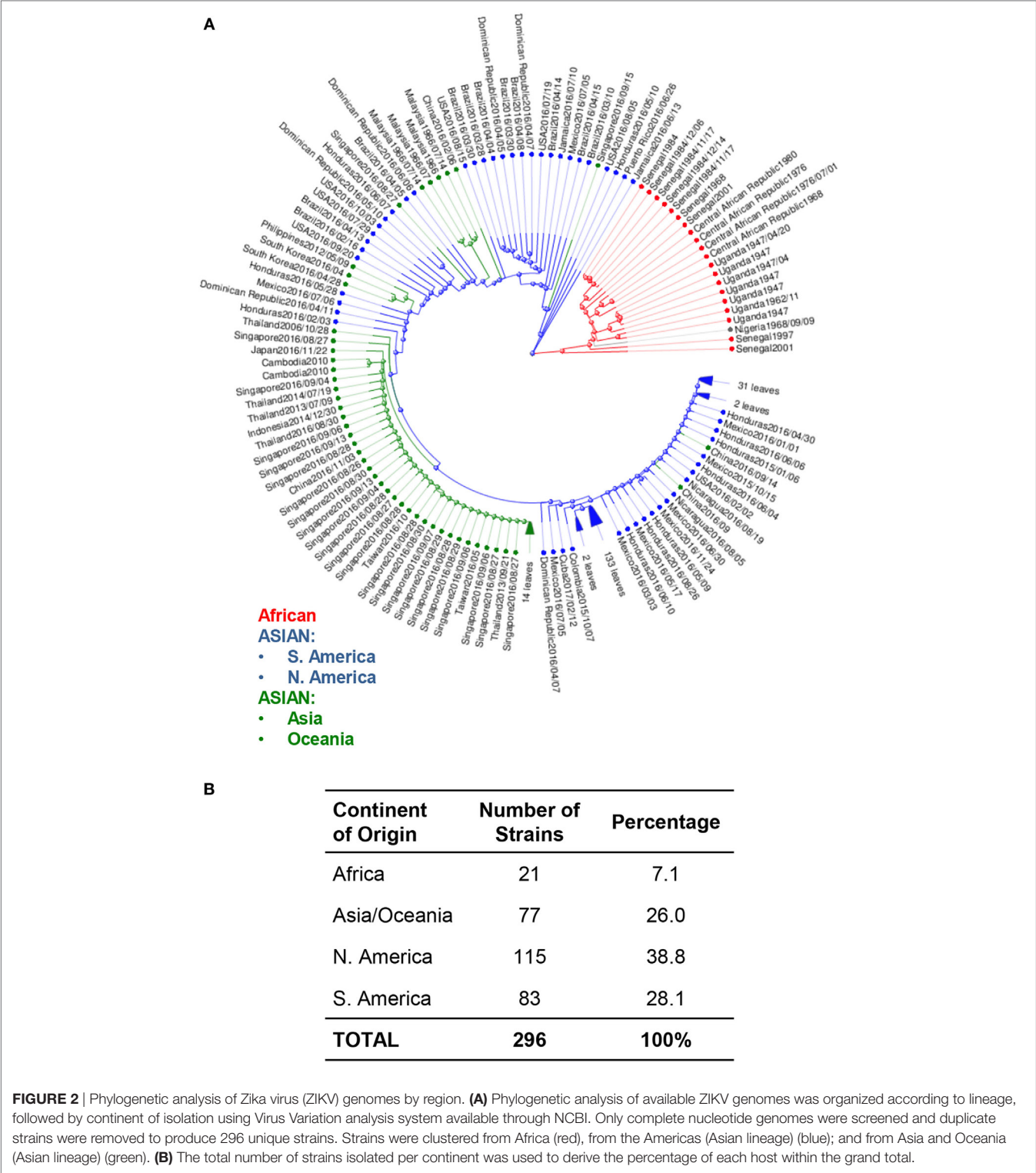
Smith et al. (27) found the difference between African and Asian isolates used in their analysis to be ~75–100 AA residues in the ORF, while the strains within each Asian and American lineage differed by ~10–30 AAs, suggesting that even minimal mutations could have phenotypic impact. A separate analysis by Wang et al. on nucleotide sequences compared 8 African strains (7 from mosquitos, 1 from monkey) with 25 Asian strains (all human) and found 59 amino acid variations that differed between the two major lineages, but were shared within the various strains of either African or Asian ancestry (2). The highest variability (10%) between Asian human and African mosquito strain was in the pr region of prM protein, though the effect of this structural change on viral function is not clear. Yuan et al. demonstrated differences in neurovirulence among Asian lineage strains from Cambodia and Venezuela may be dependent on a single amino acid substitution S139N (43). This substitution occurs in the pr region of prM and Yuan et al. hypothesizes that it may contribute to neurovirulent phenotype, but does not speculate as to a mechanism. These data are important because it demonstrates not only variances among strains of the same lineage but also sheds critical insight on intra-lineage strain-specific evolutionary differences. Comparison of protein sequences using P6-740 as the Asian reference and FSM showed over 400 variations at the nucleotide level and 26 unique substitutions at the protein level. Comparison of FSM, H/PF/2013, and the Brazilian strains from 2015 to 2016 showed that all these strains acquired changes at an additional eight positions for a total of 34 amino acid changes compared to P6-740. All isolates showed identical amino acids at these positions with the exception of T2634M/V in the NS5 protein.

Of note, no known ZIKV mosquito strain has the same nucleotide sequence as the human strains, though this could be due to sampling bias or ZIKV transmission through alternative routes (2). Nonetheless, nucleotide sequence changes can have an impact on viral pathology, replication, transmissibility, and fitness. One such example is the impact on posttranslational modification of the E protein. Faye et al. in 2014 reported a N154-glycosylation site deletion event of E protein in African isolates that did not exist in Asian lineage strains (40). Neurovirulence may depend on glycosylation of the Env protein (Asn 154) (44, 45). Naik and Wu reported that a mutation of putative N-glycosylation sites on DENV NS4B decreased RNA replication suggesting that glycosylation may play important roles in infectivity, maturation, and virulence of flaviviruses (46).



Similar to DENV, ZIKV evolution depends on worldwide spread of the mosquito vector, growing human population size, and increased foreign travel and commerce (37). Sequence analyses demonstrate that the virus originated in Africa within two distinct groups; Uganda and Nigeria, mostly isolated from non-human vectors, and anchored by the MR-766 strain. The Asian cluster was isolated in Malaysia and is anchored by a prototype strain, P6-740, which includes strains from other Southeast Asian countries, such

as Cambodia and French Polynesia. The American clade, which includes strains from Brazil and other American or Caribbean strains, evolved from this Asian cluster and expanded rapidly among naïve populations (37). As ZIKV evolves, it diversifies and creates new interactions with vectors and hosts that impact pathology, which exhibit unique lineage and strain-specific pathological profiles. To this date, 197 fully sequenced African and Asian isolates have been characterized and have been deposited in GenBank.



ZIKV TRANSMISSION AND TISSUE TROPISM

Vector Influence on Viral Evolution

While the African lineage contained eight mosquito isolates, the P6-740 (Malaysia/1966) was the sole mosquito isolate in the Asian

lineage. In 2007, human sera from patients with painful febrile disease and *A. aldopictus* mosquitos were sampled from West Africa and tested positive for ZIKV (47). At the same time, the Micronesia outbreak identified *A. (stregomyia) hensilli* as the likely principal vector (6). In 2013, ZIKV reached French Polynesia, with subsequent spread to Oceanian islands (New Caledonia,

Cook islands, and Easter island), which contained *A. aegypti* and *A. Aldopictus* throughout most of this region (48). Eleven percent of the population was infected causing symptoms such as low-grade fever, rash, conjunctivitis, and arthralgia, as well as GBS (49). *A. aegypti* has not only expanded to Central-South America but is also regarded as the most common vector for DENV (50). The New World strains of *A. aegypti* and *A. albopictus*, which are the most common in USA are poor transmitters of ZIKV (51) suggesting that continuous divergence of the Asian lineage due to genomic evolution can be adapted to direct human to human transmission without the involvement of a vector. Indeed, while *Aedes* is widely accepted as the vector for ZIKV (52–54), work by Guedes et al. has demonstrated that ZIKV can infect and replicate in the midgut, salivary glands, and can be detected in saliva of *Culex* spp. (55). This work suggests, while still a contentious topic requiring further investigation, the transmission vector range for ZIKV may be greater than anticipated.

Non-Vector Transmission

Zika virus and other Flaviviruses (with the exception of Hepatitis C) are transmitted by mosquito bite, but ZIKV has clearly diverged from other flaviviruses. Since the 2015 epidemic, the mirth of published data has made it apparent that ZIKV can be transmitted from human to human through sexual transmission, blood transfusion, ocular transmission, or vertical transmission from mother to fetus (15, 16, 20). While the African strains are better transmitted through mosquitos, the American strains with Asian ancestry may have obtained enhanced transmission capabilities through sexual intercourse. This is supported by the numerous clinical cases of sexual transmission from male to female partners, and the limited data regarding female to male transmission (56–67). Of these clinical cases, originally infected individuals are reported to have traveled to South American countries or Pacific island nations, where they are believed to contracted ZIKV (68). Recent findings from Mead et al. reveal that while ZIKV can be detected in semen from infected men for up to 9 months after infection, sexual transmission of ZIKV typically occurs within 20 days of infection, and the amount of infectious ZIKV in semen decreases rapidly within the first 3 months of infection. Additionally, data from Barreto-Vieira et al. regarding attempts to standardize *in vitro* techniques for and Asian lineage strain of ZIKV demonstrated that mammalian cells associated with mucosal membranes are more susceptible to ZIKV infection than insect cells (69).

Human Infection Studies

Fetoplacental Infections

The new routes of transmission demonstrate a novel tissue tropism for the virus. As such, tropism for the human placenta allowing infection of the fetus is unique for ZIKV, even though it should be noted that other flaviviruses that are not human pathogens do infect placental tissue in their respective hosts. Asian strains of ZIKV attracted global attention for their impact on maternal and fetal health (70, 71). Infectivity studies have shown that South American strains of Asian lineage are capable of infecting human decidua and umbilical cord tissues and are responsible

for apoptosis of chorionic villi, which function in fetal/maternal blood/nutrient exchange (72, 73). In Brazil, mothers giving birth to newborns with microcephaly had reported fever, rash, and conjunctivitis during pregnancy, the most common signs and symptoms of ZIKV infection. However, diagnosis beyond the earliest stage of acute disease is nearly impossible in the dengue-endemic regions and since the majority of exposures to ZIKV may cause asymptomatic infections, large numbers of infected pregnant women may have gone undiagnosed or misdiagnosed (20). Importantly, the severity of these responses varies greatly between patients and indicates that while studies on human immune responses to ZIKV infection are important, there is a large knowledge gap yet to be filled regarding the diversity of ZIKV infections among human demographics. An excellent example of how ZIKV pathogenic severity is dependent on individual genetic backgrounds is the study published by Caires-Junior et al., which compares pairs of twins where one is diagnosed with congenital Zika virus syndrome (CZS) (74). This study demonstrates that upon infection of neural tissues with ZIKV, cells from the CZS twins grew slower and exhibited increased viral replication. Importantly, the transcriptome results of the study reveal a significant difference in the level of an mTOR inhibitor protein, DDIT4L, between CZS and unaffected twins. This finding is significant because it indicates an individual genetic disposition for increased mTOR signaling, and as mTOR signaling pathways are critical for autophagy-mediated viral clearance, means that ZIKV infections are intensified in these individuals.

Fetal Brain Infections

Zika virus's proclivity for neural cells is not a novel finding among flaviviruses, as members of this viral family are known for their neurovirulent qualities. Several studies have sought to investigate the neuroinvasive and pathogenesis of ZIKV within both human-specific cell cultures and in neonatal mice. One example is the work done by van den Pol et al., which investigates ZIKV cellular targeting within the brain (75). Here, *in vivo* mouse studies on neonatal brains reveals ZIKV infection of specific regions of the brain as early as 4 days postinfection, but can be fully determined by 7 days postinfection. Additionally, human cell culture work by van den Pol et al. reveals that human astrocytes demonstrate a 24-h virus incubation period before shedding active virus into culture media. Importantly, a study by Lin et al. using cultured human fetal brain tissues suggests that ZIKV enters the brain *via* the subventricular zone and actively infect and replicate in committed neural cells, and thus as cells propagate to develop the brain, result in an increase of viral presence (76). The authors then speculate as that the immune response to this neural infection may drive microcephaly.

Urogenital Tract Infections

Viral RNA can persist at high levels for months in the sperm of infected men even after resolution of symptoms and persists in the vaginal secretions of infected females for weeks after symptoms resolve (60, 77). This type of persistence in the reproductive tract and sexual transmission is not observed with other flaviviral infections (78).

Ocular Infections

Ocular infections are also unique to ZIKV and likely transmitted through conjunctival fluids, tears, and lacrimal glands. Although the reported cases of African lineage infections were too few to identify similar symptomatology, ocular infections by the Asian lineage have been documented in humans and recapitulated in animal models. Inner retinal vasculopathy (79) or other ocular infections have been reported linked to international travel into South American and Caribbean island nations (79, 80). Ocular abnormalities have been documented in infants with congenital Zika virus syndrome (CZS). In a report by Fernandez et al., postmortem examination of fetuses from terminated pregnancies revealed micro-calcifications of the retina, increased amount of autolysis of tissues at the front of the ocular tract, detachment of retinal and RPE layers, and a distinct lack of neural differentiation of retinal neurons (81). This study by Fernandez et al. corroborates with other clinical cases, such as those reported in Brazil, where 34.5% of all ZIKV microcephalic infants reported in the first 2 months of 2016 had ocular abnormalities, such as retinal mottling, optic nerve degeneration, and a lack of differentiation of retinal neurons (82–90).

Animal Infection Studies

Ocular Infections

To investigate the pathogenic variability of ZIKV lineages as they occur progressively during active infections, many scientific investigations have modeled ZIKV infections in animal models. Noteworthy, among these are murine models, which have been well characterized for use in studying other flaviviruses. A129 and AG129 mice that are deficient in IFN receptor signaling when infected with ZIKV showed specific cellular tropism of ZIKV in retinal cell layers of the eye. A129 mice are deficient in IFN alpha/beta receptors, while AG129 mice are deficient in IFN alpha/beta and gamma receptors. Muller cells and retinal astrocytes infected with the virus resulted in a sustained proinflammatory microenvironment within the ocular tract that contributed to conjunctivitis and uveitis in these mice (91). In a NHP model for ocular and uteroplacental pathogenesis, Mohr et al. demonstrated lack of retinal neuron maturation, anterior segment dysgenesis, and notable chorioretinal lesions in fetal macaques (92). Different ZIKV strains display divergent tropism within host tissues. Miner et al. showed that ZIKV infections induced purulent viral panuveitis in AG129 mice and that the Asian lineage ZIKV from South America was isolated in higher viral burden than those from Oceania in ocular tissues (93).

Brain, Spleen, and Testicular Infections

Regarding tissue tropism by the Asian and African strains in animal models, the reports are conflicting. Dowall et al. demonstrated that viral RNA and ZIKV-induced histopathology with both MP1751 (African lineage) and PRVABC59 (Asian lineage) were highest in brain, spleen, and testis of A129, although the histological changes were more prominent in animals infected with the African lineage. The histopathology was minimal in heart, liver, kidney, and lung, although the Asian lineage caused no measurable clinical features (94). The Natal RGN strain from

northeastern Brazil was isolated from the brain of a fetus with microcephaly and contained half of all mutations in the NS1 gene, suggesting that tissue-specific evolution of ZIKV has contributed to the emergence of Congenital Zika syndrome (2).

Lineage-Dependent Differences in Animal and Cell Culture Models

There are limited studies characterizing ZIKV strains and most studies utilize one strain exclusively. Phenotypic differences among African and Asian isolates have been reported in both *in vitro* and *in vivo* models, demonstrating the importance of considering ZIKV isolate, passage history, cell type, or mouse model when interpreting results. Asian strains of ZIKV have been analyzed for differential infectivity in many human and non-human cell lines. These include cells from ovary, kidney, liver, brain, lung, and keratinocytes. The latter of these is important to understand since the skin is the first barrier encountered during mosquito bites, as well as the first defense against ZIKV entry. Analyses from these *in vitro* studies demonstrate that infection after 48 h produced differences between cell lines in the amount of intracellular NS1, as well as amount of virus release, and the extent of infection did not directly correlate to IFN response (95).

Substantial differences are also found between African and Asian strains *in vitro* among mammalian and insect cell lines (27). A low-passage African isolate from mosquito reached higher titers than two low-passage Asian strains at all observed time points (0–36 dpi) in cell lines from four diverse vertebrate hosts and five insect cell lines (27). Similarly, Vielle et al. reported strain-specific infection profiles in Vero cells, Aedes cells, and human monocytoic DCs (MoDCs). The authors used five African and Asian lineage strains isolated from various hosts; MR766 (U-1947) from monkey, MP1751 (U-1962) from a pool of *A. africanus* mosquito, PF13/25013-18: FP-2013 (French Polynesia) from human serum, PR-2015 (Puerto Rico) from human serum, and G-2016 (Guadeloupe) from human semen. The low passage U-1962 and U-1947 are very distant phylogenetically. The African lineage U-1962 and the Asian lineage PR-2015 showed highest rates of infection in Vero cells compared to the other strains. Similar findings for the U-1962 were observed in the mosquito cell line. In contrast, *in vitro* studies of human MoDCs showed similar susceptibility to infection, activation/maturation, expression of type I and III IFNs or cell death between lineages. The authors reported that NS5 of U-1962 showed polymorphism compared to the other strains of the study, but none of the residues were putative STAT2 binding residues, suggesting that the levels of expression of mutations were independent of mutations in the NS5 sequence of the U-1962 strain (96).

Additionally, infectivity studies comparing the two lineages revealed that African strains of ZIKV can infect human neural progenitor cells and produce both higher titers of progeny virus, and also induce higher levels of cellular apoptosis (34, 97, 98). A study by Hamel et al. demonstrated similar findings using human astrocyte cell cultures, where they found African ZIKV strain HD78788 can reach higher infectious titers 24 h postinfection of human astrocytes and also induces less innate antiviral gene transcription than Asian strain H/PF/2013 (99). This trend was

further confirmed in human dendritic cells, which are one of the primary cell types naturally infected by ZIKV (96).

While mice may not naturally become infected by ZIKV, they can be used as models for pathogenesis, and similar reports of differential lineage-specific infection characteristics have been published regarding *in vivo* mouse experiments. Zhang et al. conducted a study juxtaposing an older Asian lineage strain from Cambodia, with a more contemporary American strain from Venezuela to investigate potential differences in neuro-virulence between the two (100). They found that compared to the Cambodia strain, neonatal mice infected with Venezuelan strain of ZIKV demonstrated more neuronal cell death, obliteration of oligodendrocyte development, and an increase in the amount of CD68 and Iba1 positive microglia/macrophages in brain tissues. On the other hand, Qian et al. reported that African and Asian lineages showed similar levels of brain-development disruption in an organoid development model, thus prompting further questions regarding further analyses within the Asian lineages (101). This change between Asian and American, or alternatively between pre-epidemic and epidemic strains of ZIKV, may be attributed to new mutations between the two (102). Indeed, sequencing data comparing over 20 strains of ZIKV reveals at least 15 amino acid changes between epidemic and pre-epidemic strains, as well as the generation of a 9 amino acid bulge, rather than an external loop structure at the 3-prime UTR region of the NS5 sequence.

Mouse background strain, transgenic line, and pharmacological manipulation in wild-type strains all produce variable results using the same dose and strain of ZIKV from either lineage. The animal model chosen is one potential complication of comparing ZIKV isolates, but so is *in vitro* and/or *in vivo* passage history of isolates. There is a tradeoff between passage number and viral fitness in either vertebrate or insect hosts. A study by Haddow et al. suggests that high passage number of traditional African strains, such as MR766, in both Vero cells and suckling mouse brains has resulted in a distinct loss of glycosylation sites, which may thus affect pathology in other organs (32). The pathology of a low-passage African isolate from mosquito and two low-passage Asian strains were compared *in vivo* using two mouse models, the A129 mouse (deficient in type-I interferon receptor, *Ifnar1*^{-/-}) and the IFN-I antibody blockade mouse. The A129 mouse is commonly used to study ZIKV because the virus has been shown to infect cells by targeting human STAT2 to suppress IFN signaling, and it has been proposed that since it does not bind murine STAT2, it cannot infect mice unless the type I IFN receptor is knocked out or blocked (1). Of additional note, the majority of investigations that utilize ZIKV grow their viral stocks in Vero cells, which do not produce IFN type I in response to a viral infection, and thus allows them to be permissive to ZIKV infections (103, 104).

According to Smith et al., the African isolate caused more severe clinical pathology and lethality in both mouse models, suggesting enhanced virulence of the African strain compared to both Asian strains. Significant phenotypic differences were also observed between the two Asian strains (CPC-0740 and SV0127-14) used in the study; SV0127-14 produced 10- to 100-fold lower titers in all cell types compared to CPC-0740, and it produced

only mild clinical symptoms and 10% mortality in *Ifnar1*^{-/-} mice versus 90% mortality with the more virulent CPC-0740 (27). The IFN-I antibody mouse model was far less susceptible than the *Ifnar1*^{-/-} model, producing zero mortality and no clinical symptoms with CPC-0740. The same result was found using a more recent Asian isolate from Puerto Rico, PRVABC59 in the IFN-I antibody blockade mouse (27).

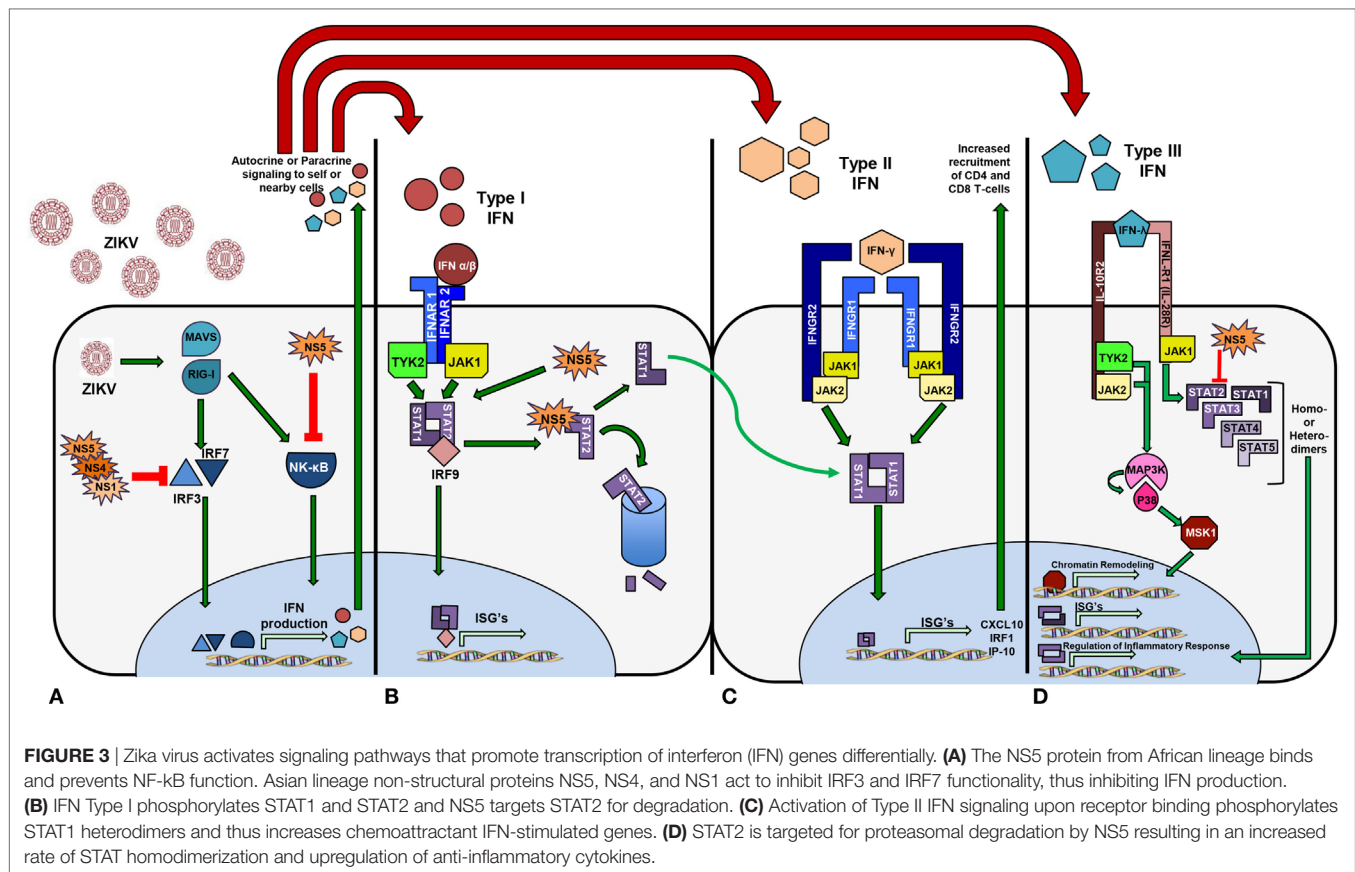
Similar to previous studies comparing African and Asian strains, Dowall et al. reported that A129 mice tolerated infections with an Asian strain well, while an African strain was lethal, with morbidity and mortality worsening in a dose-dependent manner. Interestingly, although the Asian strain produced no clinical symptoms, viral RNA levels were detected in various tissues, including brain, spleen, lungs, and kidneys and viral burden was detected in secretions, albeit the magnitude and time course of the Asian and African strain differed, with detection levels produced earlier using African infections. Moreover, seroreactivity revealed detectable antibody responses in the Asian infected A129 mice despite no clinical signs of illness (1).

CURRENT UNDERSTANDING OF ZIKV AND THE HOST IMMUNE RESPONSE

Zika virus hid from the public eye for several decades since its discovery in the 1940s, as many cases of ZIKV infections are believed to have either been subclinical or misdiagnosed as a different flavivirus infection, such as DENV. As ZIKV spread from Africa across Asia and into the Polynesian and Micronesian islands, ZIKV shifted from a mild pathogenesis and largely subclinical symptomatology to the neuro-virulent Asian lineage with a higher incidence of congenital abnormalities. The specific pathways activated by viral infection inevitably steer the innate immune response toward differential patterns and intensities of cellular and humoral responses. ZIKV is a member of the flavivirus family, and thus shares similar cell signaling pathways with other viruses within this group, which directly antagonize the IFN response system of the host innate immune response, but through a species-specific mechanism (Figure 3). Thus, to fully comprehend the challenges that ZIKV poses to human immunity and to recognize fully efficacious vaccine candidates, a detailed understanding of how ZIKV directly evades and antagonizes host innate and adaptive responses is vital.

Type I IFN Responses

Type I IFN refers to the classic IFN α/β signaling pathway, whereby viral antigen, or a pathogen-associated molecular patterns, are recognized by pathogen recognition receptors initiating an intracellular protein cascade that culminates in protein translocation into the nucleus and subsequent transcriptional activation of DNA for IFN- α and IFN- β (105). When nearby cells have IFN α/β bind to their surface receptor, IFN alpha receptor (IFNAR), a series of phosphorylation events, involving Janus kinase 1, tyrosine kinase 2, and the many signal transducer and activator of transcription (STAT) proteins ultimately result in the transcription of IFN-stimulated genes (ISGs), which have antiviral properties through a broad variety of mechanisms.



Lineage Similarities

Many flaviviruses directly antagonize different stages of the type I IFN pathway *via* species independent and diverse mechanisms that all centrally rely on genomic non-structural protein 5 (NS5). Of the seven genomic NS proteins and sub-proteins, the NS5 protein functions as the viral polymerase enzyme, and in RNA capping (106–108). Dengue virus NS5 protein has been shown to inhibit human STAT2 function by means of an E3 ubiquitin ligase, called UBR4 that targets STAT2 for proteasomal degradation. ZIKV similarly inhibits human STAT protein, but can do so independently of UBR4 (109, 110). ZIKV inhibition of STAT2 has been demonstrated in several studies where primary human dendritic or endothelial cell infections resulting in lower expression of pro-inflammatory cytokines, such as interleukin 6 (IL-6), IFN α/β , and chemokine C-C ligand 5 (CCL5) (111, 112). Additionally, NS1 and NS4b have both demonstrated the ability to inhibit the production of IFN α/β after direct stimulation with poly I:C (synthetic double stranded RNA) by blocking the formation of the TBK1 (TANK binding kinase 1) complex, which allows for oligomerization of interferon regulator factors (IRFs) (107, 113, 114).

Primary human skin fibroblasts infected with the French Polynesia isolate H/PF/2013 mounted innate immune responses by increasing the expression of RIG-I, MDA5, TLR3 leading to upregulation of Type I IFNs and ISGs as well as CXCL10 and CCL5 (115). The same strain led to IFN- β production and

induced apoptosis of infected lung epithelial cells A549 (116). Overall, both lineages have demonstrated similar mechanisms of Type I IFN activation and upregulation, and similar pathways of STAT2 inhibition and targeting for degradation.

Importantly, ZIKV inhibits human STAT2 function *via* a mechanism similar to DENV, but does not similarly inhibit murine STAT2. These two proteins share 64% sequence homology, which may thus account for the dissimilar protein interactions. This single protein non-homology between human and mouse STAT2 has led to a domination of immune-deficient murine models as the primary model for pathogenic and vaccine studies (117, 118). These models are primarily A129 and AG129 systems, which lack the IFNAR protein, required for the production of ISGs. Despite many investigators' claims that immune competent models for ZIKV will lack symptomatic expression, replication, and similarity to human infections, several recent publications have demonstrated that ZIKV can be modeled in immune competent mice using either a C57BL/6 or BALB/c background (119). The success of immune competent systems may be directly attributed to ZIKV inhibitory effects on STAT3, 4, 5, or 6. This knowledge, however, still requires investigation and remains a knowledge gap within the field.

Lineage Differences

Bowen et al. found that dendritic cells from donors are productively infected *in vitro* by both lineages, but with different kinetics.

The African lineage had faster replication and infection magnitude, and unlike the Asian lineage, caused cell death. All strains antagonized STAT1 and STAT2. Asian strains used included PRVABC59 and P6-740, while African strains included the Ugandan prototype strain (MR766) and a low-passage Senegalese strain (DakAr41524). They found that the viral kinetics varied with the source of the monocyte isolate, but that related strains varied similarly. All strains induced IFN β gene transcription, and a decrease in STAT2 phosphorylation was seen in both, but more pronounced in the African lineage (111).

In a study by Simonin et al. who infected human primary neural cells with African and Asian lineage strains, the African strain induced upregulation of at least 19 genes including RIG-I, MDA-5, and TLR-3 and induction of type 1 and 2 IFN was higher, associated with enhanced levels of inflammatory cytokines such as IL-6 or tumor necrosis factor (TNF). The only downregulated gene was CXCL8, a mediator of inflammatory responses. The Asian strain did not show any significant upregulation of genes; instead, four genes (CXCL8, CXCL10, CASP1, CTSS) were downregulated. Even at higher MOIs, the cytokine response to Asian strain was weak. In addition, neural cell infection by both strains showed similar differences in viral infectivity and cytokine production (97). In contrast, when McGrath et al. infected human neural stem cells from two individual patients with Asian and African lineages of ZIKV and conducted a transcriptomics analysis, they found increased expression of IFN- α , IL-2, TNF- α , IFN- γ genes as well as genes involved in complement, apoptosis, and STAT5 signaling pathways in cells infected with the Asian isolate (120). Patients displaying neurological symptoms during the ZIKV epidemic in Brazil demonstrated higher blood concentrations of pro-inflammatory cytokines, such as IL-6, IL-7, and IL-8, as well as higher levels of chemotactic molecules, such as IP-10 and MCP-1 (121). These findings were recapitulated in a non-human primate model (122) and in mouse studies modeling ZIKV infections using homologous strains.

Ultimately, this inhibition of type I IFN production and signaling results in an attenuated innate response to infection and may alter T cell-specific responses (123, 124). Presently, less than a handful of reports have examined the host-induced immune responses to genetically evolved ZIKV. Tripathi et al. examined strain-specific differences using *Ifnar1*^{-/-} and *Stat2*^{-/-} C57BL/6 mice. They looked at three viruses from the Asian lineage (P6-740, FSS13025, PRVABC59) and two from the African lineage (MR766; DakAr41519). They found that the African strains conferred faster onset of disease and higher mortality and in both mouse models. Infection with the African strains was marked by more severe neurological symptoms, while neurological symptoms in Asian infections were more prolonged. While both strains induced host inflammatory responses, the African isolates elicited higher levels of several cytokines and markers of T cell infiltration (IL-6, CXCL10, TNF- α , IFN- γ , CCL3, CCL4, CCL5, CXCL9, GZMB, CCL2, CCL7, CXCL1, CXCL2, IL-1b, IL-15, CD4, CD8, CCR5, CXCR3, CCR2, CCR5) (125).

Foo et al. infected human blood monocytes with the Uganda strain MR766 or the French Polynesia strain H/PF/2013. They found that both strains productively infected CD14 monocytes, but that infection with Asian viruses led to the expansion of

non-classical monocytes, resulting in a M2-skewed immunosuppressive phenotype, marked by IL-10 production. African lineages on the other hand, induced pro-inflammatory M1-skewed responses, inducing CXCL10. They also found that blood from pregnant women was more susceptible to infection. Infection with virus from the African lineage led to higher viral burdens, and increased levels of IFN- β , STAT, OAS, IRF, and NF- κ B. The Asian lineage had higher expansion of CD14^{lo}CD16⁺ non-classical monocytes, despite having a lower viral load. In general, the African strain promoted cytokines and immunomodulatory genes involved with inflammation (CXCL10, IL-23A, CD64, CD80, IL-18, IDO, SOCS1, CCR7), while the Asian strain was associated with the activation of immunosuppressive genes (IL-10, Arg1, CD200R, CD163, CD23, CCL22, VEGFa). These results were confirmed using a second strain of ZIKV from each lineage (IbH30656 for African and PRVABC59 for Asian). Pregnancy enhanced infection of both lineages in CD14 monocytes, and they found a similar pattern with the African lineage having a higher viral load. Blood from the first and second semesters of pregnancy demonstrated considerably higher CD14^{lo}CD16⁺ non-classical monocyte levels upon infection with the Asian strain, but blood in the third trimester had similar levels to non-pregnant blood. The African strain, however, produced a slight increase in non-classical monocytes during all three trimesters. Unlike monocytes from non-pregnant women, monocytes from pregnant women were more reactive to the Asian lineage than the African one. The Asian strain additionally induced genes associated with adverse pregnancy outcomes in the first two trimesters of pregnancy (ADAMTS9 and fibronectin 1) (73).

While both Asian and African lineages have demonstrated similar agonism and antagonism in Type I IFN signaling, Asian lineage strains have additionally shown a secondary method of IFN antagonism. Asian strains isolated from South American countries appear to directly activate IRF3, IRF7, and IRF9 through NS1, NS4, and NS5 viral proteins (112). African lineages have not demonstrated this ability as of yet, which implicates the differential amino acid residues as key binding factors for innate immune mediators.

Type II IFN Responses

Unlike the type I IFN response, which inhibits ZIKV infection, Chaudhary et al. demonstrated that IFN- γ (a type II IFN) cannot only decrease ZIKV infectivity in mammalian cell culture, but can also upregulate transcription of innate inflammatory and chemoattractant cytokines, such as IRF1 and CXCL10, respectively. This suppression of type I responses, but activation of type II responses is also demonstrated by the NS2A and NS4B proteins. This phenomenon of variable activation of type I and type II IFN responses has also been documented in human clinical cases of ZIKV infections (126). A study performed by Kam et al., sought to fully characterize immune biomarkers that were associated with ZIKV infections in 95 human clinical cases from Brazil. The authors demonstrated an increase of IFN- γ among human febrile cases of ZIKV and differential cytokine expression between febrile patients and those with neurological complications. Between these two groups, patients with neurological complications showed similar levels of IFN- γ and decreased

levels of anti-inflammatory cytokines, such as IL-10 and IP-10 (121). ZIKV NS5 was shown to generate the differential type I and type II responses during infection, by specifically inhibiting IFN- β signaling, and simultaneously functioning as a prominent activator of IFN- γ signaling.

Interferon- γ plays a critical role in host antiviral responses and increased levels of IFN- γ can be associated with host natural killer (NK) cell response, as they secrete high levels of this cytokine during infection. A study by V. Costa et al. proposed that DENV infection is controlled by NK cells specifically through the production of IFN- γ , and that these NK cells are activated by DENV-infected dendritic cells (DCs) (127). While this specific mechanism of IFN- γ inhibition has not been demonstrated yet for ZIKV, it has been shown that ZIKV does infect antigen-presenting cells upon infection *via* mosquito bite. Cimini et al. found that the amount of IFN- γ secreted by CD4⁺ T-cells is reduced (128). Given that CD4⁺ and CD8⁺ T-cells have proven to be the dominant drivers in ZIKV clearance during human infection (129), this alteration in cell cytokine secretion generates several questions regarding the mechanisms that ZIKV uses to subvert the host immune response and facilitate its replication.

Ngono et al. compared CD8 T cell responses to two Zika strains from the African (MR766) and Asian (FSS13025) lineages in wild-type C57BL/6 mice treated with an IFNAR blocking antibody, and in LysMCre⁺IFNAR^{fl/fl} C57BL/6 (H-2^b) mice (lacking IFNAR in certain myeloid cells). They found that the viral load of the African lineage strain decreased 3 days postinfection, but that the Asian lineage strain did not. Both strains elicited similar levels of granzyme B⁺ CD8⁺ T cells in both mouse models. They additionally identified epitopes recognized by IFN- γ secreting CD8⁺ T cells and found that in both strains the major epitope was E protein derived. In LysMCre⁺IFNAR^{fl/fl} mice, they identified 14 peptides specific to the African lineage, three specific to the Asian, and 12 shared by both, with all proteins being targeted, except for NS1 and NS2b in the Asian lineage. The Asian and African strains both resulted in a sixfold and fivefold increase in CD44⁺CD62L⁺CD8⁺ T cells, respectively, indicating a strong CD8 response. Both strains showed similar CD8⁺T cell kinetics, with the percentage of IFN- γ ⁺ CD8⁺ T cells being highest at day 7 postinfection. It is important to note that the MR766 isolate was serially passaged in mouse brains, possibly affecting its behavior (124).

Collectively, the majority of the investigations regarding ZIKV and the Type II IFN response have been done using Asian lineage viruses. Thus, there is an immense knowledge gap concerning African lineage ZIKV strains and how they may directly affect the Type II IFN response. While it can be inferred that both African and Asian lineages both benefit from the increase in STAT1 being freely able to generate homodimers and thus promote ISGs, African lineage ZIKV strains have not specifically demonstrated this ability, and thus it remains unknown.

Type III IFN Responses

Discovered in the early 2000s, type III IFNs comprises four variants of IFN- λ (numbered 1–4). The receptor for this type of IFN is unique because, rather than being ubiquitously expressed on nucleated cells like IFNARs, it is selectively expressed on

epithelial cell surfaces. Thus, IFN- λ plays a distinct role in the protection of epithelial barriers. Additionally, while expressed on epithelia surfaces, other cell types can respond to type III IFN signaling, such as those in the central nervous system (130, 131). While the role of IFN- λ has been studied during WNV and YFV infections, there is limited information regarding IFN- λ and ZIKV infections. During DENV and YFV infection, the depletion of type III interferons results in impaired CD4⁺ and CD8⁺ T-cell activation, and thus also negatively impacts viral clearance. Additionally, in mouse models deficient for type III IFN signaling infected with YFV, type III IFN signaling results in decreased blood/brain barrier maintenance and thus allows for viral neuro-invasion (132). Of the studies for ZIKV infections investigating type III IFN responses to infection, many focus on maternal and fetal infections with emphasis on the fetal/maternal blood barrier (133–137).

The placenta is the organ that separates the fetal and maternal blood supply, primarily through the chorionic villi, where fetal and maternal blood are spatially separated by 3–4 cell layers. After blastocyst implantation in the uterine wall, trophoblast cells multiply and differentiate into variable cell types. One such cell type, the syncytiotrophoblast, forms the outer epithelial layer of the chorionic villi where the majority of fetal/maternal blood exchange occurs (138). The syncytiotrophoblast layer is the primary epithelial defense in the fetal/maternal blood barrier and the first cells ZIKV encounters during fetal infection. Indeed, the type III interferons produced by syncytiotrophoblasts allow for autocrine protection, and subsequently prevent ZIKV from infecting the fetus (139).

Provided that the cells of the fetal/maternal blood interface are resistant to ZIKV infections based on their ability to secrete IFN- λ , several studies have focused on uncovering the mechanism by which ZIKV gains entry into the amniotic space and thus can infect the fetus (140). These studies focused on a specific type of fetal macrophage cells called Hofbauer cells (HBC) that derive from the fetus; are of monocytic origin, and are commonly found through the chorionic villi (141). These cells first appear early during human pregnancy (within the first 3 weeks), and then diminish in number between the fourth and fifth month of gestation. Among mothers and fetuses infected with ZIKV, however, HBCs have been seen to linger long in to the third trimester of pregnancy at a density higher than normally observed (142). Not only do HBCs persist at increased density, they are capable of direct infection by ZIKV. It is speculated that they can move freely between the chorionic tissues where blood exchange occurs in the placenta (143–145). Thus, by traversing from the chorionic fetal/maternal blood interface to the amniotic sac and fetus, HBCs can act as a shuttle for ZIKV to bypass the fetal/maternal blood barrier and infect the fetus.

Asian lineage strains have shown the ability to upregulate Type III IFN production and mRNA translation in both cell culture and in primary human clinical cases. African lineage strains have only demonstrated this ability to a lesser extent only in cell cultures. *In vitro* studies with ZIKV AF (MR-766) and AS 9 (FSS13025) infected human choriocarcinoma JEG-3 cells showed induction of antiviral type III IFN responses and the ISG 2'-5' oligoadenylate synthetase suggesting that IFN type III responses

produced by human placental trophoblasts confer protection against ZIKV infection (143). Interestingly, while both lineages have demonstrated an increase in both translation and transcription, there is not an increase of the amount of active Type III IFN proteins produced and detected in culture medium.

ZIKV Vaccine Development

Since the outbreaks that garnered international attention for ZIKV in 2015, the race for a vaccine to combat ZIKV has yielded several candidates, currently at various stages of development. A successful ZIKV vaccine must confer strong protection in healthy and pregnant populations, while also proving safe and efficacious in regards fetal/neonatal health. Important features to consider for the best vaccine should include both enhanced magnitude and quality of neutralizing antibodies and minimal cross-reactivity to DENV to prevent antibody-dependent enhancement (ADE) (146–149). ADE is perhaps the most critical of these, as flaviviruses are antigenically and structurally similar, non-neutralizing antibodies generated by one flavivirus can result in fatal outcomes upon secondary infection with a different flavivirus.

Recombinant envelope (E) protein subunit vaccines have been tested as both whole protein and single domain subunit candidates. Many of the E protein subunit vaccines published have required multiple booster administrations to achieve high antibody titers, and rely on the use of adjuvants to amplify the humoral response induced (150, 151). The use of only domain III of the E protein has proven sufficient to neutralize ZIKV in cell culture systems and in multiple mouse models when proteins are generated using the French Polynesian strain of ZIKV, Asian lineage, as a template (152). Combining the knowledge of how T-cells mediate viral clearance of ZIKV infections with epitope predictions, Pradhan et al. demonstrate that subunit vaccines using the NS2B, NS3, and NS4A proteins have potential as neutralizing vaccines that mediate a strong CD4⁺ T-cell response through *in silico* analyses (153).

Another candidate for ZIKV vaccines are lipid encapsulated mRNAs. While multiple publications have demonstrated that lipid enclosed mRNA vaccines are capable of antibody-based neutralization, only the work of Richner et al. address the issue of cross-reactive antibodies by specifically deleting the domain II of the fusion loop on the ZIKV E protein (154). Removal of this fusion loop domain results in minimal cross-reactivity between ZIKV and DENV, serotype 1. DNA vaccines have also been evaluated for ZIKV candidates, and one has progressed to clinical trials. DNA vaccines provide an economic advantage with their ease of production, and Larocca et al. has demonstrated a ZIKV DNA vaccine using the E and matrix (M) proteins can induce strong T-cell responses (155). An additional study by Muthumani et al. evaluated the efficacy of a DNA vaccine in both immune compromised mice and NHP and demonstrated the successful generation of neutralizing antibodies and high antibody titers (156). Of these vaccines, constructs generated as the vaccine have synthesized conserved residue among MR766 and Brazilian strains individually in an attempt to generate the largest cross-protective response. Despite this effort, data regarding non-homologous challenges and evaluations against different

strains and lineages remains scarce. This is true for the majority of vaccine candidate studies published.

Live-attenuated vaccines and chimeric vector vaccines offer a third and fourth candidate against ZIKV, as they provide a strong cellular immune response due to their active infections, and thus provide a humoral response that more closely recapitulates a natural infection. Whole inactivated virus vaccines are the front-runner candidates for future ZIKV vaccines, based on the clinical trials occurring within the United States, 75% of which are whole inactivated particles from the PRVABC59 strain of ZIKV, which poses the largest threat to the US (157). Live-attenuated vaccines also have the benefit of generating a natural cellular response without the risk of a generating disease. Shan et al. has demonstrated that deletions within the 3'-UTR of the RNA genome can generate an attenuated ZIKV clone that fails to replicate, and that this vaccine can induce sufficient protection to prevent ZIKV from causing disease (158, 159).

Knowledge Gaps and Future Studies

Despite the major concern that ADE and ZIKV antibody cross-reactivity among flavivirus plays in the development of vaccines, few studies successfully demonstrate that the neutralizing antibodies produced by their particular vaccine candidate are not cross-reactive to other flaviviruses. Thus, while the quest for neutralization is important, the future risk of ZIKV-mediated DENV ADE cannot and should not be overlooked, so that future DENV and CHIKV outbreaks can be avoided. A third consideration is that few, if any, publications directly investigate the binding avidity of these antibodies, but instead focus more on neutralization capacity, although the data strongly suggests low avidity and affinity antibodies have a higher chance of generating ADE for other flaviviruses, presenting another potential complication (152, 160, 161).

Additionally, despite the known disparities between different lineages of ZIKV, a lack of clarity remains regarding the mechanism behind neutralizing antibodies produced by a candidate vaccine or how it fairs against strains non-homologous to those in the vaccine. These neutralization assays should ideally include multiple strains from each of the three lineages: African, Asian, and American. This information is highly significant, because representative viruses from these lineages demonstrate differential neuroinvasive and inflammatory capabilities, as well as different infection profiles.

Few studies exist that focus on the impacts of ZIKV vaccination during pregnancy and even less address the issue of ZIKV sexual transmission. This is particularly important for areas with high prevalence of DENV where ZIKV cases can often be misdiagnosed and untreated for the infection. It is, therefore, of paramount importance to design effective vaccines conferring strong mucosal immunity to these high-risk groups.

CONCLUSION

Zika virus continues to pose an international threat as both a neuroinvasive virus with potentially lethal consequences and as a direct danger to pregnant mothers. While ZIKV has demonstrated a continual geographic and phylogenetic expansion,

the actual amount of genetic mutations is surprisingly limited between lineages. African lineages have shown to be more infectious and generate better inflammatory responses in a number of *in vitro* and *in vivo* experiments. The question that rises from the collection of studies that demonstrate a more virulent role for African strains is why the human impact, including neurological complications and fetal disruption, since 2015 in the Americas is produced by strains with Asian ancestry. It is also curious why the African strain is not found in any recently reported human cases. One possibility is that surveillance of infected humans in Africa is insufficient to detect severe cases of infection with the African strain.

Moreover, the African lineages exhibit an inhibitory mechanism on IFN production and signaling, which has been well documented in other flaviviruses. The Asian and American lineages, however, have evolved a secondary mechanism to prevent IFN transcription by IRF3 and IRF7 binding and preventing their translocation into the nucleus. Hence, while the African lineage has been shown to be more infectious than the Asian, it often presents as a self-limiting febrile disease, whereas the Asian and American lineages have exhibited persistent infections, with some human cases shedding active virus for upwards of 6 months (162).

Vaccine candidate research for ZIKV continues to be limited, providing little research into potential differences in vaccination

responses between circulating lineages. Additionally, these investigations have not prioritized the critical relationship between ZIKV and other flaviviruses, such as DENV, as the antibodies proven protective against ZIKV may promote more severe infections for DENV, rather than provide cross-neutralizing benefits. While our understanding of ZIKV has increased profoundly, extensive investigations are still required for a better understanding of lineage-specific dynamics and the host immune response in terms of the evolutionary trajectory of these lineages as they continue to expand geographically. Improved understanding of these topics, which currently present knowledge gaps in the field of ZIKV research, will serve as the cornerstones for designing future vaccines and antivirals that not only efficacious for ZIKV, but also safe against other flaviviruses.

AUTHOR CONTRIBUTIONS

JB, NL, RH, and IS wrote and reviewed the manuscript, and IS approved the version to be published. All authors listed have made substantial and intellectual contribution to the work.

FUNDING

The work was supported by NIH grant 5R01AI111557.

REFERENCES

- Dowall SD, Graham VA, Rayner E, Hunter L, Atkinson B, Pearson G, et al. Lineage-dependent differences in the disease progression of Zika virus infection in type-I interferon receptor knockout (A129) mice. *PLoS Negl Trop Dis* (2017) 11:e0005704. doi:10.1371/journal.pntd.0005704
- Wang L, Valderramos SG, Wu A, Ouyang S, Li C, Brasil P, et al. From mosquitoes to humans: genetic evolution of Zika virus. *Cell Host Microbe* (2016) 19:561–5. doi:10.1016/j.chom.2016.04.006
- Dick GW, Haddow AJ. Uganda S virus; a hitherto unrecorded virus isolated from mosquitoes in Uganda. I. Isolation and pathogenicity. *Trans R Soc Trop Med Hyg* (1952) 46:600–18. doi:10.1016/0035-9203(52)90021-7
- Weaver SC, Costa F, Garcia-Blanco MA, Ko AI, Ribeiro GS, Saade G, et al. Zika virus: history, emergence, biology, and prospects for control. *Antiviral Res* (2016) 130:69–80. doi:10.1016/j.antiviral.2016.03.010
- Macnamara FN. Zika virus: a report on three cases of human infection during an epidemic of jaundice in Nigeria. *Trans R Soc Trop Med Hyg* (1954) 48:139–45. doi:10.1016/0035-9203(54)90006-1
- Duffy MR, Chen T-H, Hancock WT, Powers AM, Kool JL, Lanciotti RS, et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. *N Engl J Med* (2009) 360:2536–43. doi:10.1056/NEJMoa0805715
- Oehler E, Watrin L, Larre P, Leparc-Goffart I, Lastere S, Valour F, et al. Zika virus infection complicated by Guillain-Barre syndrome – case report, French Polynesia, December 2013. *Euro Surveill* (2014) 19. doi:10.2807/1560-7917.ES2014.19.9.20720
- Buathong R, Hermann L, Thaisomboonsuk B, Rutvisuttinunt W, Klungthong C, Chinnawirotpisan P, et al. Detection of Zika virus infection in Thailand, 2012–2014. *Am J Trop Med Hyg* (2015) 93:380–3. doi:10.4269/ajtmh.15-0022
- Faria NR, Azevedo RD, Kraemer MU, Souza R, Cunha MS, Hill SC, et al. Zika virus in the Americas: early epidemiological and genetic findings. *Science* (2016) 352(6283):345–9. doi:10.1126/science.aaf5036
- Cao-Lormeau VM, Roche C, Teissier A, Robin E, Berry AL, Mallet HP, et al. Zika virus, French polynesia, South pacific, 2013. *Emerg Infect Dis* (2014) 20:1085–6. doi:10.3201/eid2006.140138
- Tognarelli J, Ulloa S, Villagra E, Lagos J, Aguayo C, Fasce R, et al. A report on the outbreak of Zika virus on Easter Island, South Pacific, 2014. *Arch Virol* (2016) 161:665–8. doi:10.1007/s00705-015-2695-5
- WHO. *Zika Virus, Microcephaly and Guillain-Barré Syndrome*. Geneva, Switzerland: WHO (2016).
- Passos SRL, Borges Dos, Santos MA, Cerbino-Neto J, Buonora SN, Souza TML, et al. Detection of Zika virus in April 2013 patient samples, Rio de Janeiro, Brazil. *Emerg Infect Dis* (2017) 23:2120–1. doi:10.3201/eid2312.171375
- Brasil P, Calvet GA, Siqueira AM, Wakimoto M, De Sequeira PC, Nobre A, et al. Zika virus outbreak in Rio de Janeiro, Brazil: clinical characterization, epidemiological and virological aspects. *PLoS Negl Trop Dis* (2016) 10:e0004636. doi:10.1371/journal.pntd.0004636
- Brasil P, Nielsen-Saines K. More pieces to the microcephaly-Zika virus puzzle in Brazil. *Lancet Infect Dis* (2016) 16:1307–9. doi:10.1016/S1473-3099(16)30372-3
- Calvet G, Aguiar RS, Melo AS, Sampaio SA, De Filippis I, Fabri A, et al. Detection and sequencing of Zika virus from amniotic fluid of fetuses with microcephaly in Brazil: a case study. *Lancet Infect Dis* (2016) 16(6):653–60. doi:10.1016/S1473-3099(16)00095-5
- Martines RB, Bhatnagar J, Keating MK, Silva-Flannery L, Muehlenbachs A, Gary J, et al. Notes from the field: evidence of Zika virus infection in brain and placental tissues from two congenitally infected newborns and two fetal losses – Brazil, 2015. *MMWR Morb Mortal Wkly Rep* (2016) 65:159–60. doi:10.15585/mmwr.mm6506e1
- Mlakar J, Korva M, Tul N, Popovic M, Poljsak-Prijatelj M, Mraz J, et al. Zika virus associated with microcephaly. *N Engl J Med* (2016) 374(8):951–8. doi:10.1056/NEJMoa1600651
- Oliveira Melo AS, Malinger G, Ximenes R, Szejnfeld PO, Alves Sampaio S, Bispo de Filippis AM. Zika virus intrauterine infection causes fetal brain abnormality and microcephaly: tip of the iceberg? *Ultrasound Obstet Gynecol* (2016) 47:6–7. doi:10.1002/uog.15831
- Sarno M, Sacramento GA, Khouri R, Do Rosario MS, Costa F, Archanjo G, et al. Zika virus infection and stillbirths: a case of hydrops fetalis, hydranencephaly and fetal demise. *PLoS Negl Trop Dis* (2016) 10:e0004517. doi:10.1371/journal.pntd.0004517
- Zonneveld R, Roosblad J, Staveren JW, Wilschut JC, Vreden SG, Codrington J. Three atypical lethal cases associated with acute Zika virus infection in Suriname. *IDCases* (2016) 5:49–53. doi:10.1016/j.idcr.2016.07.009

22. Benjamin I, Fernandez G, Figueira JV, Parpacen L, Urbina MT, Medina R. Zika virus detected in amniotic fluid and umbilical cord blood in an in vitro fertilization-conceived pregnancy in Venezuela. *Fertil Steril* (2017) 107:1319–22. doi:10.1016/j.fertnstert.2017.02.112
23. Mattar S, Ojeda C, Arboleda J, Arrieta G, Bosch I, Botia I, et al. Case report: microcephaly associated with Zika virus infection, Colombia. *BMC Infect Dis* (2017) 17:423. doi:10.1186/s12879-017-2522-6
24. Teixeira MG, Costa Mda C, De Oliveira WK, Nunes ML, Rodrigues LC. The epidemic of Zika virus-related microcephaly in Brazil: detection, control, etiology, and future scenarios. *Am J Public Health* (2016) 106:601–5. doi:10.2105/AJPH.2016.303113
25. Wang L, Zhao H, Oliva SM, Zhu H. Modeling the transmission and control of Zika in Brazil. *Sci Rep* (2017) 7:7721. doi:10.1038/s41598-017-07264-y
26. Ye Q, Liu ZY, Han JF, Jiang T, Li XF, Qin CF. Genomic characterization and phylogenetic analysis of Zika virus circulating in the Americas. *Infect Genet Evol* (2016) 43:43–9. doi:10.1016/j.meegid.2016.05.004
27. Smith DR, Sprague TR, Hollidge BS, Valdez SM, Padilla SL, Bellanca SA, et al. African and Asian Zika virus isolates display phenotypic differences both in vitro and in vivo. *Am J Trop Med Hyg* (2018) 98:432–44. doi:10.4269/ajtmh.17-0685
28. Goertz GP, Abbo SR, Fros JJ, Pilman GP. Functional RNA during Zika virus infection. *Virus Res* (2017):17. doi:10.1016/j.virusres.2017.08.015
29. Akiyama BM, Laurence HM, Massey AR, Costantino DA, Xie X, Yang Y, et al. Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science* (2016) 354:1148–52. doi:10.1126/science.aah3963
30. Filomatori CV, Carballeda JM, Villordo SM, Aguirre S, Pallares HM, Maestre AM, et al. Dengue virus genomic variation associated with mosquito adaptation defines the pattern of viral non-coding RNAs and fitness in human cells. *PLoS Pathog* (2017) 13:e1006265. doi:10.1371/journal.ppat.1006265
31. Donald CL, Brennan B, Cumberworth SL, Rezeli VV, Clark JJ, Cordeiro MT, et al. Full genome sequence and sRNA interferon antagonist activity of Zika virus from Recife, Brazil. *PLoS Negl Trop Dis* (2016) 10:e0005048. doi:10.1371/journal.pntd.0005048
32. Haddow AD, Schuh AJ, Yasuda CY, Kasper MR, Heang V, Huy R, et al. Genetic characterization of Zika virus strains: geographic expansion of the Asian lineage. *PLoS Negl Trop Dis* (2012) 6:e1477. doi:10.1371/journal.pntd.0001477
33. Yokoyama S, Starmer WT. Possible roles of new mutations shared by Asian and American Zika viruses. *Mol Biol Evol* (2017) 34:525–34. doi:10.1093/molbev/msw270
34. Anfasa F, Siegers JY, Van Der Kroeg M, Mumtaz N, Stalin Raj V, De Vrij FMS, et al. Phenotypic differences between Asian and African lineage Zika viruses in human neural progenitor cells. *mSphere* (2017) 2:e292–217. doi:10.1128/mSphere.00292-17
35. Gong Z, Xu X, Han GZ. The diversification of Zika virus: are there two distinct lineages? *Genome Biol Evol* (2017) 9:2940–5. doi:10.1093/gbe/evx223
36. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM, et al. Zika virus evolution and spread in the Americas. *Nature* (2017) 546:411–5. doi:10.1038/nature22402
37. Weaver SC. Emergence of epidemic Zika virus transmission and congenital Zika syndrome: are recently evolved traits to blame? *MBio* (2017) 8:e02063–16. doi:10.1128/mBio.02063-16
38. Enfissi A, Codrington J, Roosblad J, Kazanji M, Rousset D. Zika virus genome from the Americas. *Lancet* (2016) 387:227–8. doi:10.1016/S0140-6736(16)00003-9
39. Gong Z, Gao Y, Han GZ. Zika virus: two or three lineages? *Trends Microbiol* (2016) 24:521–2. doi:10.1016/j.tim.2016.05.002
40. Faye O, Freire CC, Iamarino A, Faye O, De Oliveira JV, Diallo M, et al. Molecular evolution of Zika virus during its emergence in the 20(th) century. *PLoS Negl Trop Dis* (2014) 8:e2636. doi:10.1371/journal.pntd.0002636
41. Fritz R, Blazevic J, Taucher C, Pangerl K, Heinz FX, Stiasny K. The unique transmembrane hairpin of flavivirus fusion protein E is essential for membrane fusion. *J Virol* (2011) 85:4377–85. doi:10.1128/JVI.02458-10
42. Lin SR, Zou G, Hsieh SC, Qing M, Tsai WY, Shi PY, et al. The helical domains of the stem region of dengue virus envelope protein are involved in both virus assembly and entry. *J Virol* (2011) 85:5159–71. doi:10.1128/JVI.02099-10
43. Yuan L, Huang XY, Liu ZY, Zhang F, Zhu XL, Yu JY, et al. A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. *Science* (2017) 358:933–6. doi:10.1126/science.aam7120
44. Dai L, Song J, Lu X, Deng YQ, Musyoki AM, Cheng H, et al. Structures of the Zika virus envelope protein and its complex with a flavivirus broadly protective antibody. *Cell Host Microbe* (2016) 19:696–704. doi:10.1016/j.chom.2016.04.013
45. Kostyuchenko VA, Lim EX, Zhang S, Fibriansah G, Ng TS, Ooi JS, et al. Structure of the thermally stable Zika virus. *Nature* (2016) 533:425–8. doi:10.1038/nature17994
46. Naik NG, Wu HN. Mutation of putative N-glycosylation sites on dengue virus NS4B decreases RNA replication. *J Virol* (2015) 89:6746–60. doi:10.1128/JVI.00423-15
47. Grard C, Caron M, Mombo IM, Nkoghe D, Mbouli Ondo S, Jiolle D, et al. Zika virus in Gabon (Central Africa) – 2007: a new threat from *Aedes albopictus*? *PLoS Negl Trop Dis* (2014) 8:e2681. doi:10.1371/journal.pntd.0002681
48. Horwood P, Bande G, Dagina R, Guillaumot L, Aaskov J, Pavlin B. The threat of chikungunya in Oceania. *Western Pac Surveill Response J* (2013) 4:8–10. doi:10.5365/wpsar.2013.4.2.003
49. Musso D, Nilles EJ, Cao-Lormeau VM. Rapid spread of emerging Zika virus in the Pacific area. *Clin Microbiol Infect* (2014) 20:O595–6. doi:10.1111/1469-0691.12707
50. Lourenco-de-Oliveira R, Rua AV, Vezzani D, Willat G, Vazeille M, Mousson L, et al. *Aedes aegypti* from temperate regions of South America are highly competent to transmit dengue virus. *BMC Infect Dis* (2013) 13:610. doi:10.1186/1471-2334-13-610
51. Chouin-Carneiro T, Vega-Rua A, Vazeille M, Yebakima A, Girod R, Goindin D, et al. Differential susceptibilities of *Aedes aegypti* and *Aedes albopictus* from the Americas to Zika Virus. *PLoS Negl Trop Dis* (2016) 10:e0004543. doi:10.1371/journal.pntd.0004543
52. Dick GW, Kitchen SF, Haddow AJ. Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg* (1952) 46:509–20. doi:10.1016/0035-9203(52)90042-4
53. Hayes EB. Zika virus outside Africa. *Emerg Infect Dis* (2009) 15:1347–50. doi:10.3201/eid1509.090442
54. Berthet N, Nakoune E, Kamgang B, Selekon B, Descorps-Declere S, Gessain A, et al. Molecular characterization of three Zika flaviviruses obtained from sylvatic mosquitoes in the Central African Republic. *Vector Borne Zoonotic Dis* (2014) 14:862–5. doi:10.1089/vbz.2014.1607
55. Guedes DR, Paiva MH, Donato MM, Barbosa PP, Krokovsky L, Rocha S, et al. Zika virus replication in the mosquito *Culex quinquefasciatus* in Brazil. *Emerg Microbes Infect* (2017) 6:e69. doi:10.1038/emi.2017.59
56. Atkinson B, Hearn P, Afrough B, Lumley S, Carter D, Aarons EJ, et al. Detection of Zika virus in semen. *Emerg Infect Dis* (2016) 22(5):940. doi:10.3201/eid2205.160107
57. Musso D, Roche C, Robin E, Nhan T, Teissier A, Cao-Lormeau VM. Potential sexual transmission of Zika virus. *Emerg Infect Dis* (2015) 21(2):359–61. doi:10.3201/eid2102.141363
58. Calvez E, Guillaumot L, Millet L, Marie J, Bossin H, Rama V, et al. Genetic diversity and phylogeny of *Aedes aegypti*, the main arbovirus vector in the Pacific. *PLoS Negl Trop Dis* (2016) 10:e0004374. doi:10.1371/journal.pntd.0004374
59. Foy BD, Kobylinski KC, Chilson Foy JL, Blitvich BJ, Travassos da Rosa A, Haddow AD, et al. Probable non-vector-borne transmission of Zika virus, Colorado, USA. *Emerg Infect Dis* (2016) 5:880–2. doi:10.3201/eid1705.101939
60. Hills SL, Russell K, Hennessey M, Williams C, Oster AM, Fischer M, et al. Transmission of Zika virus through sexual contact with travelers to areas of ongoing transmission – continental United States, 2016. *MMWR Morb Mortal Wkly Rep* (2016) 65:215–6. doi:10.15585/mmwr.mm6508e2er
61. Mansuy JM, Dutertre M, Mengelle C, Fourcade C, Marchou B, Delobel P, et al. Zika virus: high infectious viral load in semen, a new sexually transmitted pathogen? *Lancet Infect Dis* (2016) 16:405. doi:10.1016/S1473-3099(16)00138-9
62. Pan American Health Organization. *PAHO Statement on Zika Virus Transmission and Prevention*. Washington DC: Pan American Health Organization (2016). Available from: http://www.paho.org/hq/index.php?option=com_content&view=article&id=11605%3A2016-paho-statement-on-zika-transmission-prevention-&catid=8424%3Acontent&lang=en (Accessed: June 18, 2018).

63. Oster AM, Brooks JT, Stryker JE, Kachur RE, Mead P, Pesik NT, et al. Interim guidelines for prevention of sexual transmission of Zika virus – United States, 2016. *MMWR Morb Mortal Wkly Rep* (2016) 65(5):120–1. doi:10.15585/mmwr.mm6505e1
64. Oster AM, Russell K, Stryker JE, Friedman A, Kachur RE, Petersen EE, et al. Update: interim guidance for prevention of sexual transmission of Zika virus – United States, 2016. *MMWR Morb Mortal Wkly Rep* (2016) 65(12):323–5. doi:10.15585/mmwr.mm6512e3
65. Centers for Disease Control and Prevention. *CDC Encourages Following Guidance to Prevent Sexual Transmission of Zika Virus*. DeKalb County: Centers for Disease Control and Prevention (2016). Available from: <http://www.cdc.gov/media/releases/2016/s0223-zika-guidance.html> (Accessed: June 18, 2018).
66. DCHHS. *DCHHS Reports First Zika Virus Case in the Dallas County Acquired Through Sexual Transmission*. DCHHS (2016). Available from: <http://www.dallascounty.org/departments/hhs/press/documents/PR2-216DCHHSReportsFirstCaseofZikaVirusThroughSexualTransmission.pdf> (Accessed: June 18, 2018).
67. Venturi G, Zammarchi L, Fortuna C, Remoli ME, Benedetti E, Fiorentini C, et al. An autochthonous case of Zika due to possible sexual transmission, Florence, Italy, 2014. *Euro Surveill* (2016) 21:30148. doi:10.2807/1560-7917.ES.2016.21.8.30148
68. Mead PS, Duggal NK, Hook SA, Delorey M, Fischer M, Olzenak McGuire D, et al. Zika Virus shedding in semen of symptomatic infected men. *N Engl J Med* (2018) 378:1377–85. doi:10.1056/NEJMoa1711038
69. Barreto-Vieira DF, Jacome FC, Da Silva MAN, Caldas GC, De Filippis AMB, De Sequeira PC, et al. Structural investigation of C6/36 and Vero cell cultures infected with a Brazilian Zika virus. *PLoS One* (2017) 12:e0184397. doi:10.1371/journal.pone.0184397
70. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JL, Guimaraes KP, et al. The Brazilian Zika virus strain causes birth defects in experimental models. *Nature* (2016) 534:267–71. doi:10.1038/nature18296
71. Liu Y, Liu J, Du S, Shan C, Nie K, Zhang R, et al. Evolutionary enhancement of Zika virus infectivity in *Aedes aegypti* mosquitoes. *Nature* (2017) 545:482–6. doi:10.1038/nature22365
72. El Costa H, Gouilly J, Mansuy JM, Chen Q, Levy C, Cartron G, et al. ZIKA virus reveals broad tissue and cell tropism during the first trimester of pregnancy. *Sci Rep* (2016) 6:35296. doi:10.1038/srep35296
73. Foo SS, Chen W, Chan Y, Bowman JW, Chang LC, Choi Y, et al. Asian Zika virus strains target CD14(+) blood monocytes and induce M2-skewed immunosuppression during pregnancy. *Nat Microbiol* (2017) 2:1558–70. doi:10.1038/s41564-017-0016-3
74. Caires-Junior LC, Goulart E, Melo US, Araujo BHS, Alvizi L, Soares-Schanoski A, et al. Discordant congenital Zika syndrome twins show differential in vitro viral susceptibility of neural progenitor cells. *Nat Commun* (2018) 9:475. doi:10.1038/s41467-017-02790-9
75. van den Pol AN, Mao G, Yang Y, Ornaghi S, Davis JN. Zika virus targeting in the developing brain. *J Neurosci* (2017) 37:2161–75. doi:10.1523/JNEUROSCI.3124-16.2017
76. Lin MY, Wang YL, Wu WL, Wolseley V, Tsai MT, Radic V, et al. Zika virus infects intermediate progenitor cells and post-mitotic committed neurons in human fetal brain tissues. *Sci Rep* (2017) 7:14883. doi:10.1038/s41598-017-13980-2
77. Costa F, Sarno M, Khouri R, De Paula Freitas B, Siqueira I, Ribeiro GS, et al. Emergence of congenital Zika syndrome: viewpoint from the front lines. *Ann Intern Med* (2016) 164:689–91. doi:10.7326/M16-0332
78. Aliota MT, Bassit L, Bradrick SS, Cox B, Garcia-Blanco MA, Gavegnano C, et al. Zika in the Americas, year 2: what have we learned? What gaps remain? A report from the Global Virus Network. *Antiviral Res* (2017) 144:223–46. doi:10.1016/j.antiviral.2017.06.001
79. Singh MS, Marquezan MC, Omiadze R, Reddy AK, Belfort R Jr, May WN. Inner retinal vasculopathy in Zika virus disease. *Am J Ophthalmol Case Rep* (2018) 10:6–7. doi:10.1016/j.ajoc.2018.01.023
80. Tan JLL, Balne PK, Leo YS, Tong L, Ng LFP, Agrawal R. Persistence of Zika virus in conjunctival fluid of convalescence patients. *Sci Rep* (2017) 7:11194. doi:10.1038/s41598-017-09479-5
81. Fernandez MP, Parra Saad E, Ospina Martinez M, Corchuelo S, Mercado Reyes M, Herrera MJ, et al. Ocular histopathologic features of congenital Zika syndrome. *JAMA Ophthalmol* (2017) 135:1163–9. doi:10.1001/jamaophthalmol.2017.3595
82. Jampol LM, Goldstein DA. Zika virus infection and the eye. *JAMA Ophthalmol* (2016) 5:535–6. doi:10.1001/jamaophthalmol.2016.0284
83. Jampol LM, Goldstein DA. Zika virus, microcephaly, and ocular findings—reply. *JAMA Ophthalmol* (2016) 134:946. doi:10.1001/jamaophthalmol.2016.1307
84. Nayak S, Lei J, Pekosz A, Klein S, Burd I. Pathogenesis and molecular mechanisms of Zika virus. *Semin Reprod Med* (2016) 34:266–72. doi:10.1055/s-0036-1592071
85. Valentine G, Marquez L, Pammi M. Zika virus-associated microcephaly and Eye lesions in the newborn. *J Pediatric Infect Dis Soc* (2016) 5:323–8. doi:10.1093/jpids/piw037
86. Ventura CV, Maia M, Bravo-Filho V, Gois AL, Belfort R Jr. Zika virus in Brazil and macular atrophy in a child with microcephaly. *Lancet* (2016) 387:228. doi:10.1016/S0140-6736(16)00006-4
87. Ventura CV, Maia M, Dias N, Ventura LO, Belfort R Jr. Zika: neurological and ocular findings in infant without microcephaly. *Lancet* (2016) 387:2502. doi:10.1016/S0140-6736(16)30776-0
88. Ventura CV, Maia M, Ventura BV, Linden VV, Araujo EB, Ramos RC, et al. Ophthalmological findings in infants with microcephaly and presumable intra-uterus Zika virus infection. *Arq Bras Oftalmol* (2016) 79:1–3. doi:10.5935/0004-2749.20160002
89. Ventura CV, Ventura LO. Ophthalmologic manifestations associated with Zika virus infection. *Pediatrics* (2018) 141:S161–6. doi:10.1542/peds.2017-2038E
90. Ventura LO, Ventura CV, Dias NC, Vilar IG, Gois AL, Arantes TE, et al. Visual impairment evaluation in 119 children with congenital Zika syndrome. *J AAPOS* (2018) 22(3):218–22.e1. doi:10.1016/j.jaapos.2018.01.009
91. Zhao Z, Yang M, Azar SR, Soong L, Weaver SC, Sun J, et al. Viral retinopathy in experimental models of Zika infection. *Invest Ophthalmol Vis Sci* (2017) 58:4075–85. doi:10.1167/iovs.17-22016
92. Mohr EL, Block LN, Newman CM, Stewart LM, Koenig M, Semler M, et al. Ocular and uteroplacental pathology in a macaque pregnancy with congenital Zika virus infection. *PLoS One* (2018) 13:e0190617. doi:10.1371/journal.pone.0190617
93. Miner JJ, Sene A, Richner JM, Smith AM, Santeford A, Ban N, et al. Zika virus infection in mice causes panuveitis with shedding of virus in tears. *Cell Rep* (2016) 16:3208–18. doi:10.1016/j.celrep.2016.08.079
94. Dowall SD, Graham VA, Rayner E, Atkinson B, Hall G, Watson RJ, et al. A susceptible mouse model for Zika virus infection. *PLoS Negl Trop Dis* (2016) 10:e0004658. doi:10.1371/journal.pntd.0004658
95. Himmelsbach K, Hildt E. Identification of various cell culture models for the study of Zika virus. *World J Virol* (2018) 7:10–20. doi:10.5501/wjv.v7.i1.10
96. Vielle NJ, Zumkehr B, Garcia-Nicolas O, Blank F, Stojanov M, Musso D, et al. Silent infection of human dendritic cells by African and Asian strains of Zika virus. *Sci Rep* (2018) 8:5440. doi:10.1038/s41598-018-23734-3
97. Simonin Y, Loustalot F, Desmetz C, Foulongne V, Constant O, Fournier-Wirth C, et al. Zika virus strains potentially display different infectious profiles in human neural cells. *EBioMedicine* (2016) 12:161–9. doi:10.1016/j.ebiom.2016.09.020
98. Simonin Y, Van Riel D, Van De Perre P, Rockx B, Salinas S. Differential virulence between Asian and African lineages of Zika virus. *PLoS Negl Trop Dis* (2017) 11:e0005821. doi:10.1371/journal.pntd.0005821
99. Hamel R, Ferraris P, Wicht S, Diop F, Taligani L, Pompon J, et al. African and Asian Zika virus strains differentially induce early antiviral responses in primary human astrocytes. *Infect Genet Evol* (2017) 49:134–7. doi:10.1016/j.meegid.2017.01.015
100. Zhang F, Wang HJ, Wang Q, Liu ZY, Yuan L, Huang XY, et al. American strain of Zika virus causes more severe microcephaly than an old Asian strain in neonatal mice. *EBioMedicine* (2017) 25:95–105. doi:10.1016/j.ebiom.2017.10.019
101. Qian X, Nguyen HN, Song MM, Hadiono C, Ogden SC, Hammack C, et al. Brain-region-specific organoids using mini-bioreactors for modeling ZIKV exposure. *Cell* (2016) 165:1238–54. doi:10.1016/j.cell.2016.04.032
102. Zhu Z, Chan JF, Tee KM, Choi GK, Lau SK, Woo PC, et al. Comparative genomic analysis of pre-epidemic and epidemic Zika virus strains for

- virological factors potentially associated with the rapidly expanding epidemic. *Emerg Microbes Infect* (2016) 5:e22. doi:10.1038/emi.2016.48
103. Stauff CB, Gorbatshevych O, Cello J, Wimmer E, Futcher B. Comparison of African, Asian, and American Zika viruses in Swiss Webster mice: virulence, neutralizing antibodies, and serotypes. *Cold Spring Harb Lab* (2016).
 104. Willard KA, Demakovskiy L, Tesla B, Goodfellow FT, Stice SL, Murdock CC, et al. Zika virus exhibits lineage-specific phenotypes in cell culture, in *Aedes aegypti* mosquitoes, and in an embryo model. *Viruses* (2017) 9. doi:10.3390/v9120383
 105. Randall RE, Goodbourn S. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *J Gen Virol* (2008) 89:1–47. doi:10.1099/vir.0.83391-0
 106. Phoo WW, Li Y, Zhang Z, Lee MY, Loh YR, Tan YB, et al. Structure of the NS2B-NS3 protease from Zika virus after self-cleavage. *Nat Commun* (2016) 7:13410. doi:10.1038/ncomms13410
 107. Wu Y, Liu Q, Zhou J, Xie W, Chen C, Wang Z, et al. Zika virus evades interferon-mediated antiviral response through the co-operation of multiple nonstructural proteins in vitro. *Cell Discov* (2017) 3:17006. doi:10.1038/celldisc.2017.6
 108. Wang B, Thurmond S, Hai R, Song J. Structure and function of Zika virus NS5 protein: perspectives for drug design. *Cell Mol Life Sci* (2018) 75:1723–36. doi:10.1007/s00018-018-2751-x
 109. Grant A, Ponia SS, Tripathi S, Balasubramaniam V, Miorin L, Sourisseau M, et al. Zika virus targets human STAT2 to inhibit type I interferon signaling. *Cell Host Microbe* (2016) 19:882–90. doi:10.1016/j.chom.2016.05.009
 110. Liu J, Liu Y, Nie K, Du S, Qiu J, Pang X, et al. Flavivirus NS1 protein in infected host sera enhances viral acquisition by mosquitoes. *Nat Microbiol* (2016) 1:16087. doi:10.1038/nmicrobiol.2016.87
 111. Bowen JR, Quicke KM, Maddur MS, O'neal JT, McDonald CE, Fedorova NB, et al. Zika virus antagonizes type I interferon responses during infection of human dendritic cells. *PLoS Pathog* (2017) 13:e1006164. doi:10.1371/journal.ppat.1006164
 112. Mladinich MC, Schwedes J, Mackow ER. Zika virus persistently infects and is basolaterally released from primary human brain microvascular endothelial cells. *MBio* (2017) 8:e00952-17. doi:10.1128/mBio.00952-17
 113. Shu C, Sankaran B, Chaton CT, Herr AB, Mishra A, Peng J, et al. Structural insights into the functions of TBK1 in innate antimicrobial immunity. *Structure* (2013) 21:1137–48. doi:10.1016/j.str.2013.04.025
 114. Xia H, Luo H, Shan C, Muruato AE, Nunes BT, Medeiros DBA, et al. An evolutionary NS1 mutation enhances Zika virus evasion of host interferon induction. *Nat Commun* (2018) 9:414. doi:10.1038/s41467-017-02816-2
 115. Hamel R, Liegeois F, Wichit S, Pompon J, Diop F, Talignani L, et al. Zika virus: epidemiology, clinical features and host-virus interactions. *Microbes Infect* (2016) 18:441–9. doi:10.1016/j.micinf.2016.03.009
 116. Frumence E, Roche M, Krejbich-Trotot P, El-Kalamouni C, Nativel B, Rondeau P, et al. The South Pacific epidemic strain of Zika virus replicates efficiently in human epithelial A549 cells leading to IFN-beta production and apoptosis induction. *Virology* (2016) 493:217–26. doi:10.1016/j.virol.2016.03.006
 117. Rossi SL, Tesh RB, Azar SR, Muruato AE, Hanley KA, Auguste AJ, et al. Characterization of a novel murine model to study Zika virus. *Am J Trop Med Hyg* (2016) 94(6):1362–9. doi:10.4269/ajtmh.16-0111
 118. Sumathy K, Kulkarni B, Gondu RK, Ponnuru SK, Bonguram N, Eligeti R, et al. Protective efficacy of Zika vaccine in AG129 mouse model. *Sci Rep* (2017) 7:46375. doi:10.1038/srep46375
 119. Huang H, Li S, Zhang Y, Han X, Jia B, Liu H, et al. CD8(+) T cell immune response in immunocompetent mice during Zika virus infection. *J Virol* (2017) 91:e00900-17. doi:10.1128/JVI.00900-17
 120. McGrath EL, Rossi SL, Gao J, Widen SG, Grant AC, Dunn TJ, et al. Differential responses of human fetal brain neural stem cells to Zika virus infection. *Stem Cell Reports* (2017) 8:715–27. doi:10.1016/j.stemcr.2017.01.008
 121. Kam YW, Leite JA, Lum FM, Tan JLL, Lee B, Judice CC, et al. Specific biomarkers associated with neurological complications and congenital central nervous system abnormalities from Zika virus-infected patients in Brazil. *J Infect Dis* (2017) 216:172–81. doi:10.1093/infdis/jix261
 122. Silveira ELV, Rogers KA, Gumber S, Amancha P, Xiao P, Woollard SM, et al. Immune cell dynamics in rhesus macaques infected with a Brazilian strain of Zika virus. *J Immunol* (2017) 199:1003–11. doi:10.4049/jimmunol.1700256
 123. Sironi M, Forni D, Clerici M, Cagliani R. Nonstructural proteins are preferential positive selection targets in Zika virus and related flaviviruses. *PLoS Negl Trop Dis* (2016) 10:e0004978. doi:10.1371/journal.pntd.0004978
 124. Elong Ngono A, Vizcarra EA, Tang WW, Sheets N, Joo Y, Kim K, et al. Mapping and role of the CD8(+) T cell response during primary Zika virus infection in mice. *Cell Host Microbe* (2017) 21:35–46. doi:10.1016/j.chom.2016.12.010
 125. Tripathi S, Balasubramaniam VR, Brown JA, Mena I, Grant A, Bardina SV, et al. A novel Zika virus mouse model reveals strain specific differences in virus pathogenesis and host inflammatory immune responses. *PLoS Pathog* (2017) 13:e1006258. doi:10.1371/journal.ppat.1006258
 126. Chaudhary V, Yuen KS, Chan JF, Chan CP, Wang PH, Cai JP, et al. Selective activation of type II interferon signaling by Zika virus NS5 protein. *J Virol* (2017) 91:e163–117. doi:10.1128/JVI.00163-17
 127. Costa VV, Ye W, Chen Q, Teixeira MM, Preiser P, Ooi EE, et al. Dengue virus-infected dendritic cells, but not monocytes, activate natural killer cells through a contact-dependent mechanism involving adhesion molecules. *MBio* (2017) 8:e00741-17. doi:10.1128/mBio.00741-17
 128. Cimini E, Castilletti C, Sacchi A, Casetti R, Bordoni V, Romanelli A, et al. Human Zika infection induces a reduction of IFN-gamma producing CD4 T-cells and a parallel expansion of effector Vdelta2 T-cells. *Sci Rep* (2017) 7:6313. doi:10.1038/s41598-017-06536-x
 129. Parly RD, Rajah MM, Condotta SA, Taylor NG, Sagan SM, Richer MJ. Analysis of the T cell response to Zika virus and identification of a novel CD8+ T cell epitope in immunocompetent mice. *PLoS Pathog* (2017) 13:e1006184. doi:10.1371/journal.ppat.1006184
 130. Li J, Hu S, Zhou L, Ye L, Wang X, Ho J, et al. Interferon lambda inhibits herpes simplex virus type I infection of human astrocytes and neurons. *Glia* (2011) 59:58–67. doi:10.1002/glia.21076
 131. Li J, Wang Y, Wang X, Ye L, Zhou Y, Persidsky Y, et al. Immune activation of human brain microvascular endothelial cells inhibits HIV replication in macrophages. *Blood* (2013) 121:2934–42. doi:10.1182/blood-2012-08-450353
 132. Douam F, Soto Albrecht YE, Hrebikova G, Sadimin E, Davidson C, Kotenko SV, et al. Type III interferon-mediated signaling is critical for controlling live attenuated yellow fever virus infection in vivo. *MBio* (2017) 8:e819–817. doi:10.1128/mBio.00819-17
 133. Miner JJ, Cao B, Govero J, Smith AM, Fernandez E, Cabrera OH, et al. Zika virus infection during pregnancy in mice causes placental damage and fetal demise. *Cell* (2016) 165:1081–91. doi:10.1016/j.cell.2016.05.008
 134. Noronha L, Zanluca C, Azevedo ML, Luz KG, Santos CN. Zika virus damages the human placental barrier and presents marked fetal neurotropism. *Mem Inst Oswaldo Cruz* (2016) 111:287–93. doi:10.1590/0074-02760160085
 135. Yockey LJ, Varela L, Rakib T, Khoury-Hanold W, Fink SL, Stutz B, et al. Vaginal exposure to Zika virus during pregnancy leads to fetal brain infection. *Cell* (2016) 166:1247–56.e4. doi:10.1016/j.cell.2016.08.004
 136. Lin HZ, Tambyah PA, Yong EL, Biswas A, Chan SY. A review of Zika virus infections in pregnancy and implications for antenatal care in Singapore. *Singapore Med J* (2017) 58:171–8. doi:10.11622/smedj.2017026
 137. Zorrilla CD, Garcia Garcia I, Garcia Fragosol L, De La Vega A. Zika virus infection in pregnancy: maternal, fetal, and neonatal considerations. *J Infect Dis* (2017) 216:S891–6. doi:10.1093/infdis/jix448
 138. Gude NM, Roberts CT, Kalionis B, King RG. Growth and function of the normal human placenta. *Thromb Res* (2004) 114:397–407. doi:10.1016/j.thromres.2004.06.038
 139. Corry J, Arora N, Good CA, Sadovsky Y, Coyne CB. Organotypic models of type III interferon-mediated protection from Zika virus infections at the maternal-fetal interface. *Proc Natl Acad Sci U S A* (2017) 114:9433–8. doi:10.1073/pnas.1707513114
 140. Schwartz DA. Viral infection, proliferation, and hyperplasia of Hofbauer cells and absence of inflammation characterize the placental pathology of fetuses with congenital Zika virus infection. *Arch Gynecol Obstet* (2017) 295:1361–8. doi:10.1007/s00404-017-4361-5
 141. Jurado KA, Simoni MK, Tang Z, Uraki R, Hwang J, Householder S, et al. Zika virus productively infects primary human placenta-specific macrophages. *JCI Insight* (2016) 1. doi:10.1172/jci.insight.88461
 142. Weisblum Y, Oiknine-Djian E, Vorontsov OM, Haimov-Kochman R, Zakay-Rones Z, Meir K, et al. Zika virus infects early and midgestation

- human maternal decidual tissues, inducing distinct innate tissue responses in the maternal-fetal interface. *J Virol* (2017) 91:e01905-16. doi:10.1128/JVI.01905-16
143. Bayer A, Lennemann NJ, Ouyang Y, Bramley JC, Morosky S, Marques ET Jr, et al. Type III interferons produced by human placental trophoblasts confer protection against Zika virus infection. *Cell Host Microbe* (2016) 19:705–12. doi:10.1016/j.chom.2016.03.008
 144. Quicke KM, Bowen JR, Johnson EL, McDonald CE, Ma H, O'neal JT, et al. Zika virus infects human placental macrophages. *Cell Host Microbe* (2016) 20:83–90. doi:10.1016/j.chom.2016.05.015
 145. Simoni MK, Jurado KA, Abrahams VM, Fikrig E, Guller S. Zika virus infection of Hofbauer cells. *Am J Reprod Immunol* (2017) 77(2):e12613. doi:10.1111/aji.12613
 146. Durbin AP. Vaccine development for Zika virus-timelines and strategies. *Semin Reprod Med* (2016) 34(5):299–304. doi:10.1055/s-0036-1592070
 147. Durbin A, Wilder-Smith A. An update on Zika vaccine developments. *Expert Rev Vaccines* (2017) 16:781–7. doi:10.1080/14760584.2017.1345309
 148. Lagunas-Rangel FA, Viveros-Sandoval ME, Reyes-Sandoval A. Current trends in Zika vaccine development. *J Virus Erad* (2017) 3:124–7.
 149. Durham DP, Fitzpatrick MC, Ndeffo-Mbah ML, Parpia AS, Michael NL, Galvani AP. Evaluating vaccination strategies for Zika virus in the Americas. *Ann Intern Med* (2018) 168(9):621–30. doi:10.7326/M17-0641
 150. Yang M, Dent M, Lai H, Sun H, Chen Q. Immunization of Zika virus envelope protein domain III induces specific and neutralizing immune responses against Zika virus. *Vaccine* (2017) 35:4287–94. doi:10.1016/j.vaccine.2017.04.052
 151. Liang H, Yang R, Liu Z, Li M, Liu H, Jin X. Recombinant Zika virus envelope protein elicited protective immunity against Zika virus in immunocompetent mice. *PLoS One* (2018) 13:e0194860. doi:10.1371/journal.pone.0194860
 152. To A, Medina LO, Mfuh KO, Lieberman MM, Wong TAS, Namekar M, et al. Recombinant Zika virus subunits are immunogenic and efficacious in mice. *mSphere* (2018) 3:e00576-17. doi:10.1128/mSphere.00576-17
 153. Pradhan D, Yadav M, Verma R, Khan NS, Jena L, Jain AK. Discovery of T-cell driven subunit vaccines from Zika virus genome: an immunoinformatics approach. *Interdiscip Sci* (2017) 9:468–77. doi:10.1007/s12539-017-0238-3
 154. Richner JM, Himansu S, Dowd KA, Butler SL, Salazar V, Fox JM, et al. Modified mRNA vaccines protect against Zika virus infection. *Cell* (2017) 168:1114–25.e10. doi:10.1016/j.cell.2017.02.017
 155. Larocca RA, Abbink P, Peron JP, Zanutto PM, Iampietro MJ, Badamchi-Zadeh A, et al. Vaccine protection against Zika virus from Brazil. *Nature* (2016) 536:474–8. doi:10.1038/nature18952
 156. Muthumani K, Griffin BD, Agarwal S, Kudchodkar SB, Reuschel EL, Choi H, et al. In vivo protection against ZIKV infection and pathogenesis through passive antibody transfer and active immunisation with a prMEnv DNA vaccine. *NPJ Vaccines* (2016) 1:16021. doi:10.1038/npjvaccines.2016.21
 157. Abbink P, Larocca RA, De La Barrera RA, Bricault CA, Moseley ET, Boyd M, et al. Protective efficacy of multiple vaccine platforms against Zika virus challenge in rhesus monkeys. *Science* (2016) 353:1129–32. doi:10.1126/science.aah6157
 158. Shan C, Muruato AE, Jagger BW, Richner J, Nunes BT, Medeiros DBA, et al. A single-dose live-attenuated vaccine prevents Zika virus pregnancy transmission and testis damage. *Nat Commun* (2017) 8:676. doi:10.1038/s41467-017-00737-8
 159. Shan C, Muruato AE, Nunes BT, Luo H, Xie X, Medeiros DBA, et al. A live-attenuated Zika virus vaccine candidate induces sterilizing immunity in mouse models. *Nat Med* (2017) 23:763–7. doi:10.1038/nm.4322
 160. Brault AC, Domi A, McDonald EM, Talmi-Frank D, Mccurley N, Basu R, et al. A Zika vaccine targeting NS1 protein protects immunocompetent adult mice in a lethal challenge model. *Sci Rep* (2017) 7:14769. doi:10.1038/s41598-017-15039-8
 161. Li XF, Dong HL, Wang HJ, Huang XY, Qiu YF, Ji X, et al. Development of a chimeric Zika vaccine using a licensed live-attenuated flavivirus vaccine as backbone. *Nat Commun* (2018) 9:673. doi:10.1038/s41467-018-02975-w
 162. Paz-Bailey G, Rosenberg ES, Doyle K, Munoz-Jordan J, Santiago GA, Klein L, et al. Persistence of Zika virus in body fluids – preliminary report. *N Engl J Med* (2017). doi:10.1056/NEJMoa1613108

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Beaver, Lelutiu, Habib and Skountzou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership