

MOBILE GENETIC ELEMENTS IN CELLULAR DIFFERENTIATION, GENOME STABILITY, AND CANCER

EDITED BY: Tammy A. Morrish and Jose Luis García Pérez

PUBLISHED IN: *Frontiers in Chemistry* and *Frontiers in Molecular Biosciences*
and *Frontiers in Cell and Developmental Biology*





frontiers

Frontiers Copyright Statement

© Copyright 2007-2018 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-389-4

DOI 10.3389/978-2-88945-389-4

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

MOBILE GENETIC ELEMENTS IN CELLULAR DIFFERENTIATION, GENOME STABILITY, AND CANCER

Topic Editors:

Tammy A. Morrish, Independent Researcher, Ann Arbor, MI, United States

Jose Luis García Pérez, MRC Institute of Genetics and Molecular Medicine (IGMM), University of Edinburgh, United Kingdom; Center for Genomics and Oncological Research (GENYO), Pfizer, University of Granada, Andalusian Regional Government, Spain



Image: hermes/Shutterstock.com

The human genome, as with the genome of most organisms, is comprised of various types of mobile genetic element derived repeats. Mobile genetic elements that mobilize by an RNA intermediate, include both autonomous and non-autonomous retrotransposons, and mobilize by a “copy and paste” mechanism that relies of the presence of a functional reverse transcriptase activity. The extent to which these different types of elements are actively mobilizing varies among organisms, as revealed with the advent of Next Generation DNA sequencing (NGS).

To understand the normal and aberrant mechanisms that impact the mobility of these elements requires a more extensive understanding of how these elements interact with molecular pathways of the cell, including DNA repair, recombination and chromatin. In addition, epigenetic based-mechanisms can also influence the mobility of these elements, likely by transcriptional activation or repression in certain cell types. Studies regarding how mobile genetic elements interface and evolve with these pathways will rely on genomic studies from various model organisms. In addition, the mechanistic details of how these elements are regulated will continue to be elucidated with the use of genetic, biochemical, molecular, cellular, and bioinformatic approaches. Remarkably, the current understanding regarding the biology of these elements in the human genome, suggests these elements may impact developmental biology, including cellular differentiation, neuronal development, and immune function. Thus, aberrant changes in these molecular pathways may also impact disease, including neuronal degeneration, autoimmunity, and cancer.

Citation: Morrish, T. A., Pérez, J. L. G., eds. (2018). Mobile Genetic Elements in Cellular Differentiation, Genome Stability, and Cancer Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-389-4

Table of Contents

1. Editorial

- 05 Editorial: Mobile Genetic Elements in Cellular Differentiation, Genome Stability, and Cancer**

Tammy A. Morrish and Jose L. Garcia-Pérez

2. Retrotransposons in Somatic and Cancer Cells

- 08 Patterns of Transposable Element Expression and Insertion in Cancer**

Evan A. Clayton, Lu Wang, Lavanya Rishishwar, Jianrong Wang, John F. McDonald and I. King Jordan

- 19 The Reverse Transcriptase Encoded by LINE-1 Retrotransposons in the Genesis, Progression, and Therapy of Cancer**

Ilaria Sciamanna, Chiara De Luca and Corrado Spadafora

- 29 Crossing the LINE Toward Genomic Instability: LINE-1 Retrotransposition in Cancer**

Jacqueline R. Kemp and Michelle S. Longworth

- 38 Links between Human LINE-1 Retrotransposons and Hepatitis Virus-Related Hepatocellular Carcinoma**

Tomoyuki Honda

- 46 Guardian of the Human Genome: Host Defense Mechanisms against LINE-1 Retrotransposition**

Yasuo Ariumi

- 58 Templated Sequence Insertion Polymorphisms in the Human Genome**

Masahiro Onozawa and Peter D. Aplan

- 65 Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells**

Javier G. Pizarro and Gaël Cristofari

3. Retrotransposons in Evolution

- 74 An LTR Retrotransposon-Derived Gene Displays Lineage-Specific Structural and Putative Species-Specific Functional Variations in Eutherians**

Masahito Irie, Akihiko Koga, Tomoko Kaneko-Ishino and Fumitoshi Ishino

- 93 LTRs of Endogenous Retroviruses as a Source of Tbx6 Binding Sites**

Yukuto Yasuhiko, Yoko Hirabayashi and Ryuichi Ono

- 106 Endogenous Retroviruses: With Us and against Us**

Thomas J. Meyer, Jimi L. Rosenkrantz, Lucia Carbone and Shawn L. Chavez

- 114 Friends-Enemies: Endogenous Retroviruses Are Major Transcriptional Regulators of Human DNA**

Anton A. Buzdin, Vladimir Prassolov and Andrew V. Garazha



Editorial: Mobile Genetic Elements in Cellular Differentiation, Genome Stability, and Cancer

Tammy A. Morrish^{1*} and Jose L. Garcia-Pérez^{2,3,4}

¹ Independent Researcher, Ann Arbor, MI, United States, ² MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine (IGMM), University of Edinburgh, Edinburgh, United Kingdom, ³ Junta de Andalucía de Genómica e Investigación Oncológica (GENYO), Granada, Spain, ⁴ Centre for Genomics and Oncological Research, University of Granada, Granada, Spain

Keywords: mobile DNA, reverse transcriptase, genome stability, cellular differentiation, model organisms, retrotransposon, transposon, DNA repair

Editorial on the Research Topic

Mobile Genetic Elements in Cellular Differentiation, Genome Stability, and Cancer

The human genome, as well as the genome of most organisms, harbors various types and abundances of transposable element derived repeats (Lander, 2001; Waterston et al., 2002). The topic on: “Mobile Genetic Elements in Cellular Differentiation, Genome Stability, and Cancer,” includes a collection of original research articles and reviews, which address the impact of reverse transcriptases, including the ones coded by transposable elements, on both basic biological mechanisms and disease. In 1970, the discovery of reverse transcriptases or RNA-dependent DNA polymerases, was reported by two different laboratories (Baltimore, 1970; Temin and Mizutani, 1970). Since then numerous studies regarding retroviral reverse transcriptases have significantly contributed to the characterization and biology of many different retrovirus and retroelements. These studies continue to be of interest for the prevention and treatment of various retroviral induced human diseases and for the basic understanding of the origin of retroviruses. In addition the knowledge of reverse transcription has been harnessed for basic use in molecular biology and other applications, including recent widely used methods such as RNAseq. As retroviruses are considered exogenously derived reverse transcriptases, the subsequent discovery in 1987 of telomerase, also considered an endogenous RNA-dependent DNA polymerase, has significantly contributed to the understanding of one of the predominant mechanisms of telomere maintenance that contributes to most, but not all organisms with linear chromosomes (Greider and Blackburn, 1985; Biessmann et al., 1990). Yet, sequences encoding for endogenous RNA-dependent DNA polymerases are not limited to telomerase. The isolation and subsequent genetic, biochemical, and molecular characterization of human full-length non-Long Terminal Repeat (LTR) retrotransposons, termed Long Interspersed Elements (LINE-1) demonstrated that elements formally encode a reverse transcriptase activity (Dombroski et al., 1991; Mathias et al., 1991; Feng et al., 1996; Moran et al., 1996). Non-LTR retrotransposons are not limited to the human genome, and are present as full-length and/or truncated, rearranged, inactive remnants in many other genomes. In addition, the reverse transcriptase activities encoded by non-LTR retrotransposons share sequence identity with many other reverse transcriptases (Nakamura et al., 1997; Malik et al., 1999). Furthermore, non-LTR retrotransposons rely on the encoded reverse transcriptase for integration, typically by target-primed reverse transcription (TPRT), which was initially biochemically defined using the non-LTR retrotransposon R2Bm, from *Bombyx mori* (Luan et al., 1993). A review by Onozawa and Aplan included in this topic, describes two different types of LINE-1 reverse transcriptase-mediated template sequence insertion polymorphisms (TSIPs), or integration structures that are polymorphic in the human genome (Onozawa and Aplan). The

OPEN ACCESS

Edited and reviewed by:

Cecilia Giulivi,
University of California, Davis,
United States

*Correspondence:

Tammy A. Morrish
morrish@gmail.com

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 23 October 2017

Accepted: 20 November 2017

Published: 04 December 2017

Citation:

Morrish TA and Garcia-Pérez JL
(2017) Editorial: Mobile Genetic
Elements in Cellular Differentiation,
Genome Stability, and Cancer.
Front. Chem. 5:108.
doi: 10.3389/fchem.2017.00108

characteristics of class2 structures allude to the occurrence of additional integration mechanisms by the LINE-1 reverse transcriptase that may occur in germ cells or during embryogenesis (Onozawa and Aplan). To note, the features described in these class2 structures are consistent with previous reports of endonuclease-independent LINE-1 retrotransposition (Eickbush, 2002; Morrish et al., 2002).

Phylogenetic analysis of the reverse transcriptase domains support the idea that retroviruses and telomerase evolved from non-LTR retrotransposons, due to the gain or loss of LTR sequence and/or sequences encoding for specific domains (Xiong and Eickbush, 1988; Malik et al., 1999). These early phylogenetic studies are consistent with the provirus hypothesis proposed by Temin, that (1) retroviruses are likely derived from endogenous retrotransposons and (2) mutations that arise due to the mobility of retrotransposons could potentially activate oncogenes or inactivate tumor suppressor genes, perhaps contributing to tumorigenesis (Temin, 1971; Shimotohno et al., 1980). As LINE-1 elements are active in tumors, yet transcriptionally repressed in many somatic cell types, there was much interest to understand the extent that LINE-1 retrotransposition contributes to tumorigenesis (Solyom et al., 2012; Shukla et al., 2013; Doucet-O'Hare et al., 2015; Ewing et al., 2015; Rodic et al., 2015). Included in this topic is original research using bioinformatic approaches to examine LINE-1 expression and insertion profiles using RNA-seq data from normal and primary tumor samples collected using the Cancer Genome Atlas (TCGA) (Clayton et al.). Here the authors examined the expression and integration differences in breast invasive carcinoma, head and neck squamous carcinoma, and lung adenocarcinoma and their analysis indicates two cases of LINE-1 mediated insertions near two different tumor suppressor genes, including an Alu insertion into the *CBL* gene in breast invasive carcinoma and a LINE-1 insertion into the first exon of the *BAALC* gene in a head and neck squamous cell carcinoma. Again, these findings are consistent with the provirus hypothesis. However these tumors may also harbor mutations in “host” genes that regulate LINE-1 retrotransposition. A number of reviews were included in this topic that address recent studies on LINE-1 retrotransposition in cancer (Honda; Kemp and Longworth ; Sciamanna et al.). In addition, identifying cellular genes and pathways that regulate LINE-1 transcription and activity is an active area of research, and two reviews discuss the current understanding regarding the regulation of LINE-1 retrotransposition in somatic cells, which may become dysregulated in cancer (Ariumi; Pizarro and Cristofari). The topic also includes two original research articles on the impact of endogenous retroviruses on genome evolution. In the article by Irie et al., the authors use dN/dS analysis and molecular approaches to validate their findings regarding the contribution of the *sushi-ichi* retrotransposon during the evolution of the zinc finger protein-encoding gene

SIRH11/ZCCHC16 and the impact of this gene during eutherian brain evolution. In addition, another research article examines the evolution of the Tbx6 transcription binding sites, (ORRA1-ORRA1D), which are LTRs derived from the endogenous retroviruses, MaLRs (Yasuhiko et al.). The authors examine the impact on transcription of genes harboring these Tbx6 binding sites, using the Tbx6 knockout mouse. Their findings are coupled with biochemical and bioinformatic approaches. Finally two reviews nicely described the host cellular factors that impact the transcriptional dynamics of ERVs in the human genome (Buzdin et al.; Meyer et al.).

Overall the articles that were received for this topic: “Mobile Genetic Elements in Cellular Differentiation, Genome Stability, and Cancer” predominantly focus on the evolution of endogenous reverse transcriptases (RT), including the LINE-1 encoded RT, and the endogenous retroviruses ERVs and MaLR. These articles also summarize the findings in the field regarding these reverse transcriptases in normal biology and disease. These summaries and newly reported findings are consistent with the provirus hypothesis (Temin, 1971; Shimotohno et al., 1980; Shimotohno and Temin, 1981). Identification of additional host factors and cellular pathways that contribute to LINE-1 retrotransposition will help further elucidate the provirus hypothesis, as not all LINE-1 insertions occur in tumor suppressor or oncogenes. In addition, further studies regarding exogenous and endogenous reverse transcriptases will continue to shed light on the growing knowledge surrounding reverse transcription in the RNA world.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

Funding was provided by a Howard Temin Pathway to Independence Award, Grant Number K99/R00CA154889 from the National Cancer Institute (TM) and the deArce Koch Memorial Endowment Fund from the University of Toledo (TM). JG-P's lab is supported by CICE-FEDER-P12-CTS-2256, Plan Nacional de I+D+I 2008–2011 and 2013–2016 (FIS-FEDER-PI14/02152), PCIN-2014-115-ERA-NET NEURON II, the European Research Council (ERC-Consolidator ERC-STG-2012-233764), by an International Early Career Scientist grant from the Howard Hughes Medical Institute (IECS-55007420), by The Wellcome Trust-University of Edinburgh Institutional Strategic Support Fund (ISFF2) and by a private donation by Ms. Francisca Serrano (Trading y Bolsa para Torpes, Granada, Spain).

REFERENCES

- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226, 1209–1211. doi: 10.1038/2261209a0
- Biessmann, H., Mason, J. M., Ferry, K., d'Hulst, M., Valgeirsdottir, K., Traverse, K. L., et al. (1990). Addition of telomere-associated HeT DNA sequences “heals” broken chromosome ends in *Drosophila*. *Cell* 61, 663–673. doi: 10.1016/0092-8674(90)90478-W

- Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F., and Kazazian, H. H. Jr. (1991). Isolation of an active human transposable element. *Science* 254, 1805–1808. doi: 10.1126/science.1662412
- Doucet-O'Hare, T. T., Rodic, N., Sharma, R., Darbari, I., Abril, G., Choi, J. A., et al. (2015). LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4894–E4900. doi: 10.1073/pnas.1502474112
- Eickbush, T. H. (2002). Repair by retrotransposition. *Nat. Genet.* 31, 126–127. doi: 10.1038/ng897
- Ewing, A. D., Gacita, A., Wood, L. D., Ma, F., Xing, D., Kim, M. S., et al. (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* 25, 1536–1545. doi: 10.1101/gr.196238.115
- Feng, Q., Moran, J. V., Kazazian, H. H. Jr., and Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905–916. doi: 10.1016/S0092-8674(00)81997-2
- Greider, C. W., and Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* 43, 405–413. doi: 10.1016/0092-8674(85)90170-9
- Lander, E. S. (2001). Initial impact of the sequencing of the human genome. *Nature* 470, 187–197. doi: 10.1038/nature09792
- Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605. doi: 10.1016/0092-8674(93)90078-5
- Malik, H. S., Burke, W. D., and Eickbush, T. H. (1999). The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* 16, 793–805. doi: 10.1093/oxfordjournals.molbev.a026164
- Mathias, S. L., Scott, A. F., Kazazian, H. H. Jr., Boeke, J. D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* 254, 1808–1810. doi: 10.1126/science.1722352
- Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., and Kazazian, H. H. Jr. (1996). High frequency retrotransposition in cultured mammalian cells. *Cell* 87, 917–927. doi: 10.1016/S0092-8674(00)81998-4
- Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., et al. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31, 159–165. doi: 10.1038/ng898
- Nakamura, T. M., Morin, G. B., Chapman, K. B., Weinrich, S. L., Andrews, W. H., Lingner, J., et al. (1997). Telomerase catalytic subunit homologs from fission yeast and human. *Science* 277, 955–959. doi: 10.1126/science.277.5328.955
- Rodic, N., Steranka, J. P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., et al. (2015). Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat. Med.* 21, 1060–1064. doi: 10.1038/nm.3919
- Shimotohno, K., Mizutani, S., and Temin, H. M. (1980). Sequence of retrovirus provirus resembles that of bacterial transposable elements. *Nature* 285, 550–554. doi: 10.1038/285550a0
- Shimotohno, K., and Temin, H. M. (1981). Evolution of retroviruses from cellular movable genetic elements. *Cold Spring Harb Symp. Quant. Biol.* 45 (Pt 2), 719–730. doi: 10.1101/SQB.1981.045.01.090
- Shukla, R., Upton, K. R., Munoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111. doi: 10.1016/j.cell.2013.02.032
- Solyom, S., Ewing, A. D., Rahrmann, E. P., Doucet, T., Nelson, H. H., Burns, M. B., et al. (2012). Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* 22, 2328–2338. doi: 10.1101/gr.145235.112
- Temin, H. M. (1971). The provirus hypothesis: speculations on the significance of RNA-directed DNA synthesis for normal development and for carcinogenesis. *J. Natl. Cancer Inst.* 46, 3–7.
- Temin, H. M., and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226, 1211–1213. doi: 10.1038/2261211a0
- Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562. doi: 10.1038/nature01262
- Xiong, Y. E., and Eickbush, T. H. (1988). Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. *Cell* 55, 235–246. doi: 10.1016/0092-8674(88)90046-3

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 © 2017 Morrish and Garcia-Pérez. 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) or licensor 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.



Patterns of Transposable Element Expression and Insertion in Cancer

Evan A. Clayton^{1,2}, Lu Wang^{3,4}, Lavanya Rishishwar^{3,4,5}, Jianrong Wang⁶,
John F. McDonald^{1,2} and I. King Jordan^{3,4,5*}

¹ Integrated Cancer Research Center, School of Biology, Georgia Institute of Technology, Atlanta, GA, USA, ² Ovarian Cancer Institute, Atlanta, GA, USA, ³ School of Biology, Georgia Institute of Technology, Atlanta, GA, USA, ⁴ PanAmerican Bioinformatics Institute, Cali, Colombia, ⁵ Applied Bioinformatics Laboratory, Atlanta, GA, USA, ⁶ Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI, USA

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Formerly affiliated with University of
Toledo, USA

Reviewed by:

David Ray,
Mississippi State University, USA
David E. Symer,
Ohio State University Comp. Cancer
Ctr., USA
Tara Theresa Doucet-O'Hare,
National Institutes of Health, USA

*Correspondence:

I. King Jordan
king.jordan@biology.gatech.edu

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Molecular Biosciences

Received: 24 August 2016

Accepted: 31 October 2016

Published: 16 November 2016

Citation:

Clayton EA, Wang L, Rishishwar L,
Wang J, McDonald JF and Jordan IK
(2016) Patterns of Transposable
Element Expression and Insertion in
Cancer. *Front. Mol. Biosci.* 3:76.
doi: 10.3389/fmolb.2016.00076

Human transposable element (TE) activity in somatic tissues causes mutations that can contribute to tumorigenesis. Indeed, TE insertion mutations have been implicated in the etiology of a number of different cancer types. Nevertheless, the full extent of somatic TE activity, along with its relationship to tumorigenesis, have yet to be fully explored. Recent developments in bioinformatics software make it possible to analyze TE expression levels and TE insertional activity directly from transcriptome (RNA-seq) and whole genome (DNA-seq) next-generation sequence data. We applied these new sequence analysis techniques to matched normal and primary tumor patient samples from the Cancer Genome Atlas (TCGA) in order to analyze the patterns of TE expression and insertion for three cancer types: breast invasive carcinoma, head and neck squamous cell carcinoma, and lung adenocarcinoma. Our analysis focused on the three most abundant families of active human TEs: Alu, SVA, and L1. We found evidence for high levels of somatic TE activity for these three families in normal and cancer samples across diverse tissue types. Abundant transcripts for all three TE families were detected in both normal and cancer tissues along with an average of ~80 unique TE insertions per individual patient/tissue. We observed an increase in L1 transcript expression and L1 insertional activity in primary tumor samples for all three cancer types. Tumor-specific TE insertions are enriched for private mutations, consistent with a potentially causal role in tumorigenesis. We used genome feature analysis to investigate two specific cases of putative cancer-causing TE mutations in further detail. An Alu insertion in an upstream enhancer of the *CBL* tumor suppressor gene is associated with down-regulation of the gene in a single breast cancer patient, and an L1 insertion in the first exon of the *BAALC* gene also disrupts its expression in head and neck squamous cell carcinoma. Our results are consistent with widespread somatic activity of human TEs leading to numerous insertion mutations that can contribute to tumorigenesis in a variety of tissues.

Keywords: LINE-1, L1, Alu, SVA, retrotransposons, bioinformatics, mutation, tumorigenesis

INTRODUCTION

More than 50% of the human genome sequence is derived from transposable element (TE) insertions (Lander et al., 2001; de Koning et al., 2011). The vast majority of TE-derived sequences in the human genome correspond to relatively ancient insertions that are no longer capable of transposition (Mills et al., 2007). However, there are several families of human TEs that remain

active to this day. The most abundant families of active TEs in the human genome are the Alu and SVA short interspersed nuclear elements (SINEs) along with the L1 Long Interspersed Nuclear Element (LINE) family (Kazazian et al., 1988; Batzer and Deininger, 1991; Batzer et al., 1991; Brouha et al., 2003; Ostertag et al., 2003; Wang et al., 2005). Alu and SVA SINEs are non-autonomous TEs that are mobilized via the transpositional machinery encoded by the autonomous L1 family of LINEs. Recent evidence indicates that a handful of HERV-K endogenous retroviral elements also remain active in the human genome (Wildschutte et al., 2016).

Active TE families are of great interest since they have the ability to generate *de novo* mutations, many of which have been linked to human disease (Hancks and Kazazian, 2012; Solyom and Kazazian, 2012). For instance, TE insertions have been shown to contribute to the etiology of a variety of different cancer types (Belancio et al., 2010a; Carreira et al., 2014). Numerous recent studies have used a combination of next-generation sequence analysis, followed by validation with PCR and/or Sanger sequencing, to elucidate connections between TE activity and cancer (Solyom et al., 2012; Shukla et al., 2013; Tubio et al., 2014; Doucet-O'Hare et al., 2015; Ewing et al., 2015). L1 insertions in particular have been implicated as potential cancer causing mutations in those and other studies (Morse et al., 1988; Miki et al., 1992; Iskow et al., 2010; Lee et al., 2012; Scott et al., 2016). L1 activity is thought to promote tumor development by causing genomic instability, via impaired chromosomal pairing during mitosis, and/or by disrupting coding or regulatory sequences (Kemp and Longworth, 2015).

Many of the studies that have related TEs to cancer have considered TE expression, at the transcript or protein level, and TE insertional activity separately. A number of different cancer types are positive for L1 transcript expression (Belancio et al., 2010b), and L1 proteins have been shown to be ubiquitously expressed in both normal and tumor samples from the same individuals (Bratthauer and Fanning, 1992, 1993; Bratthauer et al., 1994; Asch et al., 1996; Doucet-O'Hare et al., 2015, 2016). There is also evidence suggesting that L1 protein expression can be limited to tumor tissues and thereby serve as a useful cancer biomarker; nearly half of all human cancers are exclusively immunoreactive for L1-ORF1 encoded proteins (Rodic et al., 2014). The expression of L1 proteins in tumors has been shown to affect the expression of a number of cancer-related genes, including the down-regulation of tumor suppressors (Rangasamy et al., 2015). With respect to TE insertional activity, studies on matched normal and tumor tissues have found that novel L1 insertions occur at high frequencies in lung cancer genomes (Iskow et al., 2010). Such insertions frequently occur in oncogenes and tumor suppressors, underscoring their putative role in tumorigenesis (Lee et al., 2012).

A principal challenge when interpreting cancer genomes is distinguishing between so-called passenger and driver mutations. While passenger mutations are present in cancer genomes, they are not considered to contribute to cancer progression; instead, they are simply somatic mutations that arise during carcinogenesis and are carried along during clonal expansion. Driver mutations, on the other hand, are causal mutations that

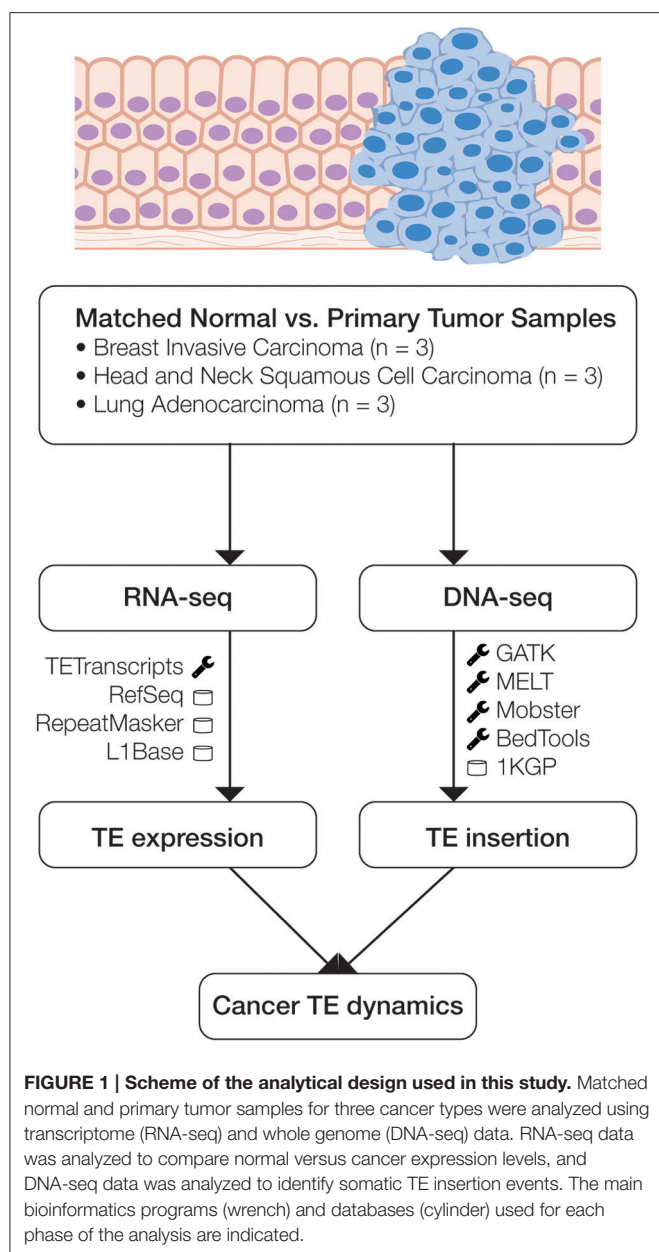
are directly implicated in carcinogenesis and the promotion of cancer growth (Stratton et al., 2009; Marx, 2014; Pon and Marra, 2015). To date, only a few studies have directly implicated TE insertions as cancer driver mutations. One such study analyzed 19 hepatocellular carcinoma genomes utilizing the RC-Seq methodology (Baillie et al., 2011) and discovered two separate L1 insertions that initiate tumorigenesis via distinct oncogenic pathways (Shukla et al., 2013). This study found L1 insertions in two different tumor suppressor genes: Mutated in Colorectal Cancers (*MCC*) and Suppression of Tumorigenicity (*ST18*). Most recently, a role for L1 insertional activity was conclusively demonstrated for colorectal cancer caused by an insertion in the *APC* tumor suppressor gene (Scott et al., 2016). This paper describes a somatic L1 insertion into one copy of the *APC* gene that, when coupled with a point mutation in the other copy of the gene, initiates tumorigenesis through the two hit colorectal cancer pathway.

Owing to parallel developments in genomics and bioinformatics, it is now possible to jointly analyze the patterns of TE transcript expression and TE insertional activity in human cancers. The Cancer Genome Atlas (TCGA) provides access to both transcriptome sequence data (RNA-seq) and whole genome sequence data (DNA-seq) for a number of matched normal and primary tumor sample pairs from individual patients (Weinstein et al., 2013). In addition, recently developed bioinformatics algorithms allow for the detection of TE transcripts directly from RNA-seq data (Jin et al., 2015) as well as for the characterization of novel TE insertions from DNA-seq data (Thung et al., 2014; Sudmant et al., 2015). We took advantage of these developments in order to evaluate the patterns of both TE expression and insertional activity in three cancer types: breast invasive carcinoma, head, and neck squamous cell carcinoma, and lung adenocarcinoma (**Figure 1** and Supplementary Figure 1). We observed a simultaneous increase of L1 transcript expression and L1 insertional activity for primary tumor samples for all three cancers, and we evaluate individual cases of TE insertions that are implicated as potential cancer causing mutations.

MATERIALS AND METHODS

Genome and Transcriptome Sequence Data

Whole genome sequence data (DNA-seq), transcriptome sequence data (RNA-seq) and patient metadata for matched normal and primary tumor tissue samples from nine cancer patients were acquired from The TCGA (Weinstein et al., 2013) via the Cancer Genomics Hub (CGHub) using the download client GeneTorrent (Maltbie et al., 2013). The nine participants included three breast invasive carcinoma patients, three head and neck squamous cell carcinoma patients and three lung adenocarcinoma patients (**Table 1**). DNA-seq and RNA-seq data were accessed as BAM files of paired-end Illumina sequence data aligned against the human genome reference sequence (build hg19). BAM files containing sequence alignments were validated for quality using FASTQC (Andrews, 2011), and autosomes were



extracted from the BAM files for downstream analysis using SAMtools (Li et al., 2009).

Gene and Transposable Element (TE) Expression Levels

Gene and TE expression levels were measured using RNA-seq data for the nine matched normal and primary tumor tissue samples. Gene expression levels were quantified as read counts mapped to NCBI RefSeq gene annotations (Pruitt et al., 2012). TE expression levels—for Alu, L1 and SVA elements—were quantified using reads mapped to RepeatMasker annotations, which were subsequently analyzed with the TETranscripts package (Jin et al., 2015). The TETranscripts program uses an expectation maximization (EM) algorithm to choose optimal

unique TE locations for multi-mapped reads, thereby allowing for accurate expression level measurements for active TE families. The TETranscripts method was recently shown to yield more reliable measures of TE transcription levels compared to previously published methods, such as HTSeq-count, Cufflinks, and RepEnrich (Trapnell et al., 2010; Criscione et al., 2014; Anders et al., 2015). The L1Base database was used to identify the genomic locations of 145 full length, intact elements from the most recently active L1 subfamily (Penzkofer et al., 2005). The set of full-length intact L1 sequences from the L1Base was generated by performing a BLAST search using the human genomic DNA sequences against the L1 template sequence (Penzkofer et al., 2005). L1Base was used to facilitate measures of active L1 element expression by limiting our analysis to RNA-seq reads that map to full-length, intact L1 sequences which retain the potential to be transpositionally active. This was done in an effort to ensure that the reads we analyzed were taken from potentially active L1 elements as opposed to older fixed elements, which could represent read-through transcripts initiated from nearby genomic promoters. The expression levels of these potentially active L1 elements were analyzed separately using the TETranscripts method.

Differential expression levels between normal and cancer tissue pairs, for genes and TEs, were evaluated by comparing distributions of \log_{10} transformed RNA-seq expression levels characterized as described above. The statistical significance levels of the observed differential expression between normal and cancer pairs were evaluated by comparing these distributions using the non-parametric Kolmogorov-Smirnov test. Statistical comparisons were done separately for each tissue (cancer) type: breast invasive carcinoma, head and neck squamous cell carcinoma and lung adenocarcinoma.

Transposable Element Insertion Detection

The genomic locations of novel TE insertions from matched normal and primary tumor tissue samples were predicted based on discordant read-pair mapping of DNA-seq data (Ewing, 2015) (Table 2). A scheme of our TE insertion detection analysis pipeline is shown in Supplementary Figure 2. DNA-seq BAM files were realigned according to GATK's standard indel realignment method (Van der Auwera et al., 2013) to facilitate TE insertion detection. The programs MELT (Sudmant et al., 2015) and Mobster (Thung et al., 2014) were used together for TE insertion detection. These two programs were selected owing to their previously demonstrated superior performance for human TE insertion detection (Rishishwar et al., 2016). Only TE insertion sites that were found by both methods (i.e., the intersection of the predictions) were used for subsequent analysis. TE insertion predictions made by the individual programs were considered to represent the same insertion if they were found within ± 100 bp of each other. An additional filtering step was applied based on the number of mapped sequence reads (coverage) that support each TE insertion prediction. Only predictions with a minimum coverage of 5 reads and a maximum coverage of 4X the average sequencing depth of the sample were used for subsequent analysis. These upper and lower cut-off thresholds were empirically chosen based on the observed distributions

TABLE 1 | TCGA whole genome (DNA-seq) and transcriptome (RNA-seq) data sources for the patients analyzed in this study.

ID	TCGA barcode	Cancer type	Sex	Age	Sample type ^a	Seq depth	Read len.
Breast 1	TCGA-BH-A0B3-11B-21D-A128-09	Breast invasive carcinoma	F	53	NT-W	42.4	100
	TCGA-BH-A0B3-11B-21R-A089-07				NT-R	5.5	50
	TCGA-BH-A0B3-01A-11D-A128-09				TP-W	40.2	100
	TCGA-BH-A0B3-01B-21R-A089-07				TP-R	5.4	50
Breast 2	TCGA-BH-A0BW-11A-12D-A314-09		F	71	NT-W	54.1	100
	TCGA-BH-A0BW-11A-12R-A115-07				NT-R	7	50
	TCGA-BH-A0BW-01A-11D-A10Y-09				TP-W	46.1	100
	TCGA-BH-A0BW-01A-12R-A115-07				TP-R	7.3	50
Breast 3	TCGA-BH-A0DT-11A-12D-A12B-09		F	41	NT-W	63.3	100
	TCGA-BH-A0DT-11A-12R-A12D-07				NT-R	7.7	50
	TCGA-BH-A0DT-01A-21D-A12B-09				TP-W	79.9	100
	TCGA-BH-A0DT-01A-21R-A12D-07				TP-R	6.6	50
Head 1	TCGA-CV-7255-11A-01D-2276-10	Head and neck squamous cell carcinoma	F	32	NT-W	6.9	101
	TCGA-CV-7255-11A-01R-2016-07				NT-R	7.5	48
	TCGA-CV-7255-01A-11D-2276-10				TP-W	5.8	101
	TCGA-CV-7255-01A-11R-2016-07				TP-R	7.1	48
Head 2	TCGA-CV-7416-11A-01D-2334-08		F	29	NT-W	7.7	101
	TCGA-CV-7416-11A-01R-2081-07				NT-R	5.9	48
	TCGA-CV-7416-01A-11D-2334-08				TP-W	28.6	101
	TCGA-CV-7416-01A-11R-2081-07				TP-R	6	48
Head 3	TCGA-CV-6959-11A-01D-1911-02		M	48	NT-W	38.3	51
	TCGA-CV-6959-11A-01R-1915-07				NT-R	8.5	48
	TCGA-CV-6959-01A-11D-1911-02				TP-W	31.4	51
	TCGA-CV-6959-01A-11R-1915-07				TP-R	6.6	48
Lung 1	TCGA-44-6776-11A-01D-1853-02	Lung adenocarcinoma	F	60	NT-W	38.9	51
	TCGA-44-6776-11A-01R-1858-07				NT-R	5.4	48
	TCGA-44-6776-01A-11D-1853-02				TP-W	6.9	51
	TCGA-44-6776-01A-11R-1858-07				TP-R	7.4	48
Lung 2	TCGA-50-5932-11A-01D-1753-08		M	75	NT-W	34.6	101
	TCGA-50-5932-11A-01R-1755-07				NT-R	4.2	48
	TCGA-50-5932-01A-11D-1753-08				TP-W	44.5	101
	TCGA-50-5932-01A-11R-1755-07				TP-R	7.4	48
Lung 3	TCGA-55-6984-11A-01D-1945-08		F	NA	NT-W	36.2	101
	TCGA-55-6984-11A-01R-1949-07				NT-R	4.9	48
	TCGA-55-6984-01A-11D-1945-08				TP-W	41	101
	TCGA-55-6984-01A-11R-1949-07				TP-R	5.2	48

^aNT-D, Normal tissue DNA-seq; NT-R, Normal tissue RNA-seq; TP-D, Tumor primary DNA-seq; TP-R, Tumor primary RNA-seq.

of the numbers of discordant mapped read pairs used to call individual TE insertions. Read count distributions were computed individually for each program (MELT, Mobster) used and for each sample (Supplementary Figure 3). The resulting distributions were typically bimodal with a lower peak (i.e., with lower read count support) that we considered to be enriched for potential false positive TE insertion calls. The lower cut-off threshold of 5 reads was chosen to minimize such false positives,

and the upper cut-off threshold was chosen to remove calls made in genomic regions that show anomalously high numbers of mapped reads, which tend to be enriched for ambiguously mapped reads.

The number of observed versus expected counts of unique L1 insertions were compared for matched normal and primary tumor tissue samples. The observed counts were taken from the TE detection pipeline, and the expected counts were computed

TABLE 2 | Numbers of MELT and Mobster predicted TE insertions in matched normal (N) and primary tumor (T) samples across 9 individuals.

Participant ID	TE insertions in matched normal tissue				TE insertions in tumor primary tissue			
	Alu	SVA	L1	Total	Alu	SVA	L1	Total
Breast 1	913	28	127	1069	853	33	110	997
Breast 2	1004	21	121	1147	1160	54	143	1358
Breast 3	1012	63	139	1215	952	60	136	149
Head 1	984	72	140	1197	741	66	107	915
Head 2	945	25	131	1102	832	26	138	997
Head 3	860	36	108	1005	819	41	112	973
Lung 1	716	29	92	838	780	36	113	930
Lung 2	806	25	103	935	701	20	94	816
Lung 3	856	21	110	988	746	14	100	861

as the ratio of unique insertions seen in matched normal vs. primary tissue for all TEs multiplied by the total number of observed L1 insertions. The significance of the difference between the observed versus expected counts of unique L1 insertions was evaluated using the Fisher's exact test. Counts of TE insertions for matched normal and primary tumor tissue samples were characterized based on their frequencies from the 1000 Genomes Project (1KGP) (Sudmant et al., 2015) and grouped into three distinct frequency bins. The distributions of TE insertion counts across the three frequency bins were compared for matched normal and cancer samples for the different tissue types analyzed here, and the significance of the differences between these distributions were evaluated using the Kolmogorov-Smirnov test.

TE Insertion Genome Feature Analysis

The genomic locations of novel TE insertions were considered with respect to several genomic features using the BEDTools program (Quinlan, 2014): RefSeq genes (Pruitt et al., 2012), COSMIC tumor suppressor genes (Forbes et al., 2015), and enhancer elements defined by chromatin states (Roadmap Epigenomics et al., 2015). The population allele frequencies of the predicted TE insertions were computed from the Phase 3 release of the 1KGP (Sudmant et al., 2015) as previously described (Rishishwar et al., 2015).

RESULTS AND DISCUSSION

TE Expression Levels in Matched Normal vs. Primary Tumor Tissue Samples

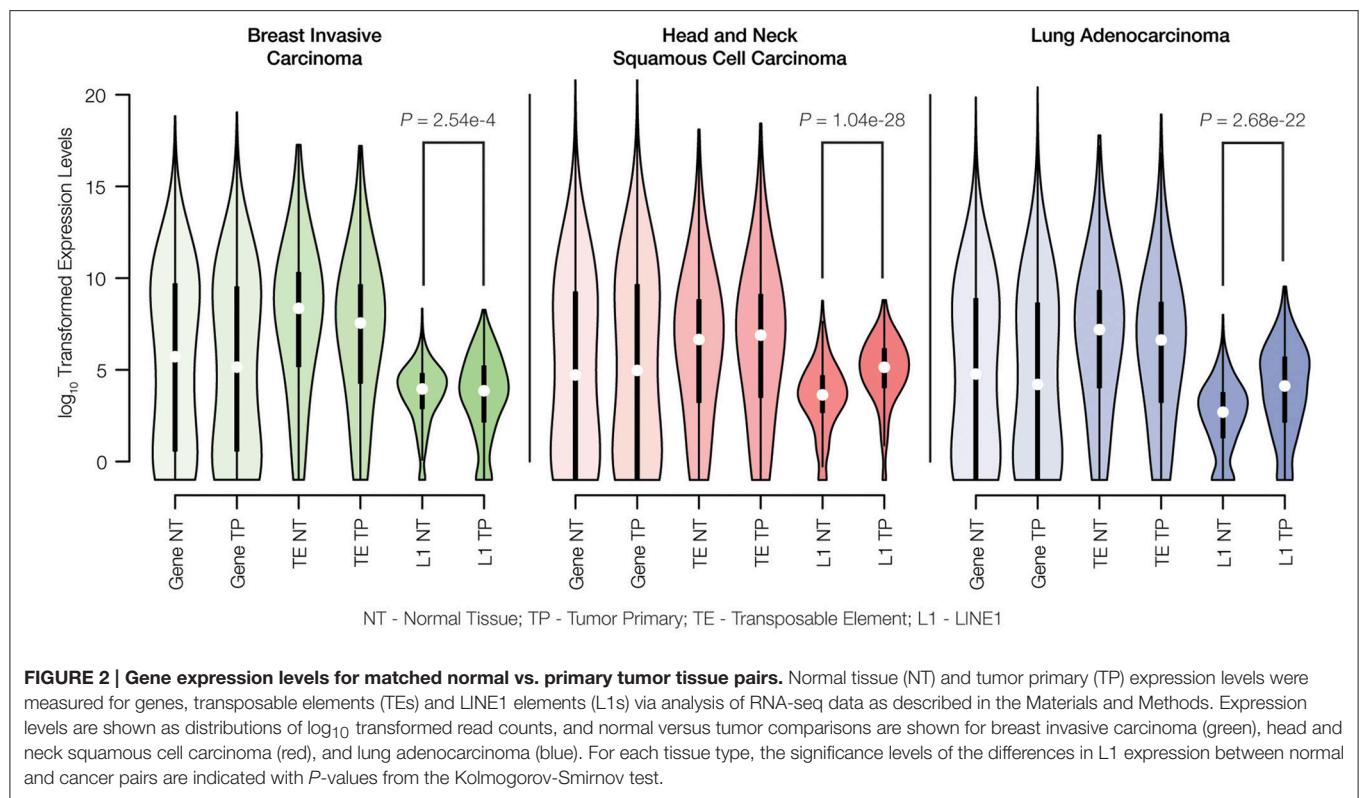
RNA-seq data were used to evaluate the differences in TE expression levels between matched normal and primary tumor tissue samples as described in the Materials and Methods. The observed differences in gene expression levels between normal and tumor tissue were compared to differences in TE expression levels for breast invasive carcinoma, head, and neck squamous cell carcinoma and lung adenocarcinoma. There are no significant differences observed for the distributions of gene expression levels between matched normal and primary tumor tissue pairs for any of the three cancer types analyzed here (Figure 2). Similarly, when all three families of potentially active

TEs (Alu, L1, and SVA) are considered together, there is no significant difference seen for the overall levels of expression between matched normal and tumor tissue. However, when full-length, potentially active L1 sequences are considered alone, we observe statistically significant increases in L1 expression levels for all three cancer types.

The methods that we used to characterize TE expression levels include several analytical controls aimed to ensure that only genuine TE-initiated transcripts, from members of potentially active families, are measured. Nevertheless, the lack of a difference between normal and tumor expression levels observed when all three active TE families were considered together could reflect technical difficulties with identifying *bona fide* TE transcripts that are initiated from element promoters as opposed to TE sequences that are passively expressed as part of longer genic transcripts. This is particularly true for Alu elements, many of which are found in the introns of human genes and transcribed as read-through transcripts initiated from RNA Pol II gene promoters (Deininger, 2011). Our confidence in the ability to measure L1-initiated transcripts is higher owing to the focus on previously identified full-length, intact elements that are located in intergenic regions. In any case, the up-regulation of L1s in cancer that we observed has potential implications for increased TE insertional activity for all three families, since L1 encoded proteins are responsible for the *cis* retrotransposition of L1s as well as the *trans* activation of Alu and SVA elements (Batzer and Deininger, 2002; Hancks and Kazazian, 2010). We analyzed the same pairs of matched normal and primary tumor tissues to evaluate whether the observed increase in L1 expression corresponds to increased transpositional activity of human TEs.

Novel TE Insertions in Matched Normal and Primary Tumor Tissue Samples

It is now possible to characterize the genomic locations and copy numbers of individual TE insertions from whole genome DNA-seq data owing to recent developments in computational genomics software (Ewing, 2015; Rishishwar et al., 2016). This technological advance is exemplified by the recent Phase 3 release of the 1KGP, which includes a complete genome-wide census of polymorphic TE insertion sites for 2504 individuals across



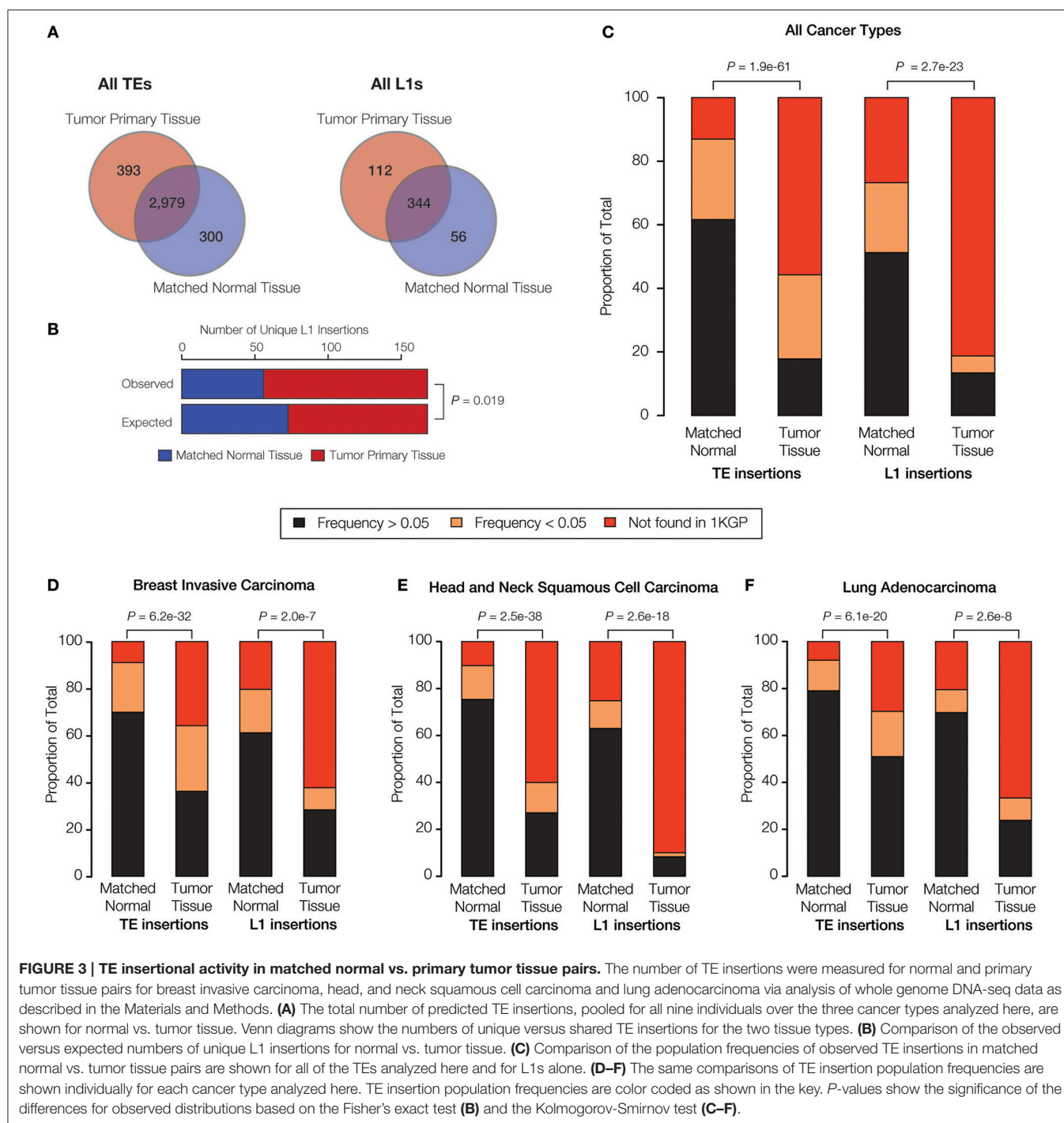
26 human populations (Sudmant et al., 2015). We analyzed whole genome DNA-seq data using computational methods for TE insertion detection (see Materials and Methods) in order to compare TE insertional activity between matched normal versus primary tumor tissue samples.

When all three families of active human TEs are considered together, we observed a total of 3672 TE insertions across the nine individuals analyzed for normal and cancer tissue pairs, 693 of which are unique insertions found in only one individual and one tissue type. In other words, we observe an average of ~77 unique somatic TE insertions per person, i.e., “private” TE insertions. This estimate is similar to the value of ~90 unique (presumably germline) TE insertions that we previously observed for individuals from the 1KGP (Rishishwar et al., 2015). A large majority of the observed TE insertions—81% for all TEs and 62% for L1s alone—are shared between the normal and tumor tissue types of an individual, suggesting that they represent germline insertions (Figure 3A). There are 1.3x more unique TE insertions seen for tumor compared to normal tissue, and this effect is more pronounced for L1s alone, which are 2x more abundant in tumor tissue samples. Accordingly, there is a statistically significant excess of observed versus expected L1 insertions in tumor versus normal tissue ($P = 0.019$) (Figure 3B). These results are consistent with a potential role for L1 transpositional activity in tumorigenesis for the cancer types analyzed here, as has been previously suggested for several different cancers (Morse et al., 1988; Iskow et al., 2010; Lee et al., 2012; Scott et al., 2016).

Given the relatively high level of L1 insertional activity in the tumor tissue samples analyzed here, we tested whether tumor-specific L1 insertions are found at lower frequencies among the (presumably) healthy donors from the 1KGP compared to L1 insertions found in matched normal tissue. The idea was to evaluate whether the tumor-specific L1 insertions represent mutations that are private, and thereby more likely to be deleterious or disease-causing. To do this, individual TE insertions were classified as high frequency (>0.05), low frequency (<0.05) or private (absent) according to their previously characterized population (allele) frequencies from the 1KGP (Rishishwar et al., 2015; Sudmant et al., 2015).

When all three cancer types are considered together, there is a statistically significant excess of private and low frequency TE insertions observed for tumor compared to normal tissue ($P = 1.9e-61$) (Figure 3C). This effect is even more pronounced when L1 insertions are considered alone ($P = 2.7e-23$). The same pattern of an increased frequency of private L1 insertions in tumor tissue is observed ($P < 2.0e-7$) when all three cancer types are analyzed for sets of patients (Figures 3D–F) and when samples for individual patients are analyzed separately (Supplementary Figure 4). The strongest effect is seen for head and neck squamous cell carcinoma. The pattern of a significant excess of private L1 insertions in tumor compared to normal tissue, observed for all three cancer types studied here, provides further evidence in support of a possible role for L1 activity in tumorigenesis.

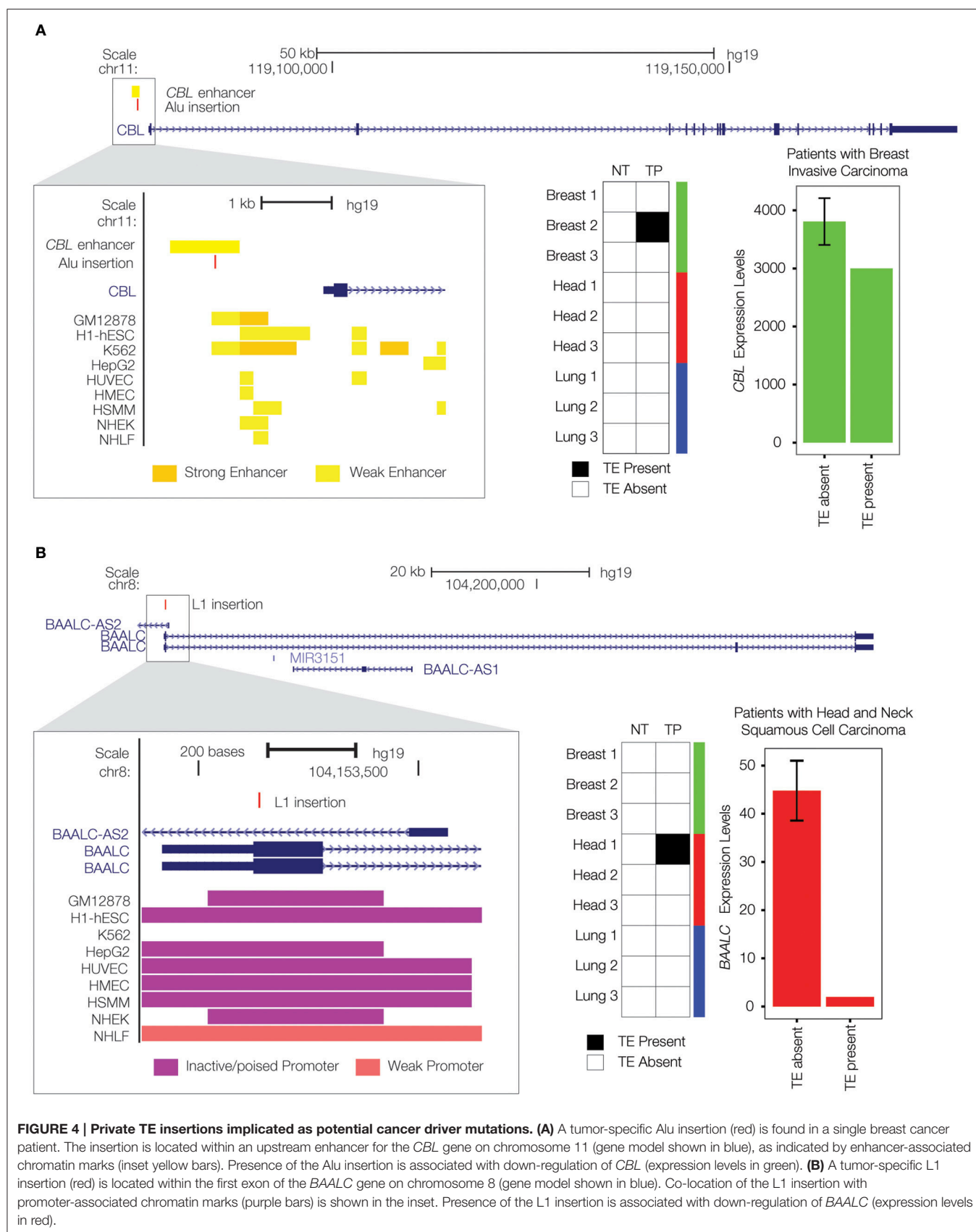
It should be noted TE insertions found in low copy numbers may not be detectable using next-generation sequence analysis,



whereas such insertions may be uncovered using more sensitive PCR-based approaches. False negatives of this kind will be more prevalent at low levels of sequence coverage. We have tried to control for this by using relatively high sequence coverage (~35X) studies here, but the conservative lower read count cut-off of 5 reads per TE insertion call that we used may still lead to missing TE insertion calls. Sequence based predictions can also yield false-positive TE insertion calls. In an effort to deal with this issue, we have only used high-confidence calls produced by

two independent programs—MELT and Mobster—that we have recently shown to be most reliable for the detection of human TE insertions (Rishishwar et al., 2016).

One other potential problem with the sequence based analysis relates to the base pair resolution with which TE insertions can be called via computational analysis of next-generation sequence data. Currently, the most accurate programs for calling TE insertions from next-generation sequence data do not yet allow for the insertions to be precisely located to genomic regions



at single base pair resolution. To account for this fact, TE insertions called within a window of ± 100 bp are considered to be co-located (Supplementary Figure 2). It is possible that this approximation can lead to multiple TE insertion events being collapsed into a single event. Subsequent experimental confirmation of individual TE insertion calls of interest (e.g., potentially tumorigenic TE insertions) should help to provide certainty with respect to both their validity and their precise genomic locations.

Potentially Tumorigenic TE Insertions

Having established a potential role for transpositional activity in tumorigenesis using the genome-wide approaches described above, we wanted to search for specific examples where individual TE insertions could be implicated as possible cancer driver mutations. To do so, we performed an integrated analysis of TE insertion, gene expression and chromatin data (see Materials and Methods) in an effort to identify the cancer-specific TE insertions that are most likely to play a causal role in tumorigenesis. We considered TE insertions that are co-located with either exons or regulatory elements of previously characterized tumor suppressor genes to have the highest likelihood of being functionally relevant. We observed a total of 141 intragenic (35.9%) insertions and 246 intronic insertions (62.6%) out of the 393 total cancer-specific insertions in our dataset. None of these intergenic or intronic cancer-specific TE insertions were found to disrupt any known functional (regulatory) sequence element. Thus, consistent with previous studies, the vast majority of TE insertions that we observed are not likely to affect gene function or expression in cancer. We did find 4 exonic TE insertions, along with 2 insertions located in regulatory elements, for known tumor suppressor genes (1.5% of the total). Here, we focus on two of these potential cases of cancer driver TE insertions, which could prove to be of interest to the TE and/or cancer research communities.

There is a private, breast cancer tumor-specific Alu insertion that is located within an upstream enhancer element that helps to regulate the expression of the *Cbl* Proto-Oncogene (*CBL*) gene (Figure 4A). *CBL* is classified as a tumor suppressor gene by the COSMIC database (Forbes et al., 2015). It has been found to be mutated or translocated in a number of cancers including acute myeloid leukemia (Abbas et al., 2008; Naramura et al., 2011; Aranaz et al., 2013); mutations in *CBL* are also the cause of Noonan syndrome-like disorder (Martinelli et al., 2010). The *CBL* encoded protein functions as a negative regulator of signal transduction pathways (Schmidt and Dikic, 2005), activation of which have been associated with cancer (Sever and Brugge, 2015). The tumor-specific Alu enhancer insertion that we characterized is associated with down-regulation of *CBL* expression, consistent with a potential role in tumorigenesis via the activation of signal transduction pathways associated with cell proliferation (Sever and Brugge, 2015).

We also found a private L1 insertion that was unique to a head and neck squamous cell carcinoma tissue sample, located within the first exon of the Brain and Acute Leukemia, Cytoplasmic (*BAALC*) gene (Figure 4B). As its name implies, the *BAALC* gene is expressed in the brain and related neural tissues, and it was first identified by association with acute myeloid

leukemia where it was shown to be overexpressed (Damiani et al., 2013; Zhou et al., 2015). TE insertions within exons are extremely rare and would presumably have a dramatic effect on gene function. Indeed, this particular insertion is associated with nearly complete inactivation of the *BAALC* gene. This is consistent with previous results showing that the presence of fixed L1 insertions genome-wide is strongly associated with the down-regulation of human gene expression (Han et al., 2004). A recent study has demonstrated that *BAALC* can inhibit extracellular signal-regulated kinase (ERK) mediated monocytic differentiation of AML cells (Morita et al., 2015). Thus, down-regulation of *BAALC* would presumably result in a loss of control over cellular differentiation, consistent with a possible role in tumorigenesis. A recent study discovered a role for the change in methylation status of a cancer-specific L1 insertion in tumorigenesis (Scott et al., 2016); this could be an additional mechanism by which the *BAALC* L1 insertion observed here exerts a regulatory effect.

CONCLUSION

The results of our analysis show a surprisingly high level of somatic TE activity in the human genome. Abundant transcripts from members of all three active human TE families analyzed here—Alu, SVA and L1—can be identified for both normal and cancer tissue samples. In addition, after filtering for high confidence TE insertion calls, we identified an average of close to 80 unique insertions for each tissue among the individual patients in our study. Thus, active human TE families retain the ability to transpose in somatic tissue thereby generating substantial levels of cellular heterogeneity among diverse tissues.

We also observe a correlated increase in both transcript expression levels and transpositional activity for L1 elements in cancer tissue samples when compared to matched normal tissue. Increased cancer expression of L1 elements is particularly relevant for TE insertional activity, since the L1 transpositional machinery is responsible for transposing non-autonomous Alu and SVA elements in *trans* along with L1 elements in *cis*. Our results are consistent with previous studies showing expression of L1 transcripts in lung cancer (Belancio et al., 2010b) and expression of L1 ORF1p in breast cancer (Harris et al., 2010), and tumor-specific L1 insertions have also previously been found in breast (Morse et al., 1988), head and neck (Helman et al., 2014), and lung tumors (Helman et al., 2014). We confirmed the presence of numerous tumor-specific L1 insertions in these three cancer types and identify two potentially tumorigenic TE insertions, an Alu insertion in the enhancer region of the tumor suppressor gene *CBL* and an L1 insertion in the first exon of the *BAALC* gene. These results underscore the potential for somatic TE activity to generate cellular heterogeneity and to contribute to the etiology of cancer across a wide range of human tissues.

ETHICS STATEMENT

Ethical approval was not required for this study on restricted access, de-identified data in accordance with the guidelines of the Cancer Genome Atlas (TCGA). Access to the data was approved by the data access committee of the TCGA.

AUTHOR CONTRIBUTIONS

EC, LW, and LR performed all of the analyses described in the study. JW contributed to the genome feature analysis. IJ and JM conceived of designed and supervised the study. All authors contributed to the drafting and revision of the manuscript.

FUNDING

EC and LW were supported by the Georgia Tech Bioinformatics Graduate Program. LR and IJ were supported by the IHRC-Georgia Tech Applied Bioinformatics Laboratory (ABiL).

REFERENCES

- Abbas, S., Rotmans, G., Löwenberg, B., and Valk, P. J. (2008). Exon 8 splice site mutations in the gene encoding the E3-ligase CBL are associated with core binding factor acute myeloid leukemias. *Haematologica* 93, 1595–1597. doi: 10.3324/haematol.13187
- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Andrews, S. (2011). *FastQC A Quality Control Tool for High Throughput Sequence Data*. Cambridge: Babraham Institute.
- Aranaz, P., Miguélez, I., Hurtado, C., Erquiaga, I., Larrayoz, M. J., Calasanz, M. J., et al. (2013). CBL RING finger deletions are common in core-binding factor acute myeloid leukemias. *Leuk. Lymphoma* 54, 428–431. doi: 10.3109/10428194.2012.709629
- Asch, H. L., Eliacin, E., Fanning, T. G., Connolly, J. L., Bratthauer, G., and Asch, B. (1996). Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol. Res.* 8, 239–247.
- Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., et al. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479, 534–537. doi: 10.1038/nature10531
- Batzler, M. A., and Deininger, P. L. (1991). A human-specific subfamily of Alu sequences. *Genomics* 9, 481–487. doi: 10.1016/0888-7543(91)90414-A
- Batzler, M. A., and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3, 370–379. doi: 10.1038/nrg798
- Batzler, M. A., Gudi, V. A., Mena, J. C., Foltz, D. W., Herrera, R. J., and Deininger, P. L. (1991). Amplification dynamics of human-specific (HS) Alu family members. *Nucleic Acids Res.* 19, 3619–3623. doi: 10.1093/nar/19.13.3619
- Belancio, V. P., Roy-Engel, A. M., and Deininger, P. L. (2010a). All y'all need to know 'bout retroelements in cancer. *Semin. Cancer Biol.* 20, 200–210. doi: 10.1016/j.semcancer.2010.06.001
- Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R., and Deininger, P. (2010b). Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* 38, 3909–3922. doi: 10.1093/nar/gkq132
- Bratthauer, G. L., Cardiff, R. D., and Fanning, T. G. (1994). Expression of LINE-1 retrotransposons in human breast cancer. *Cancer* 73, 2333–2336.
- Bratthauer, G. L., and Fanning, T. G. (1992). Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* 7, 507–510.
- Bratthauer, G. L., and Fanning, T. G. (1993). LINE-1 retrotransposon expression in pediatric germ cell tumors. *Cancer* 71, 2383–2386.
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285. doi: 10.1073/pnas.0831042100
- Carreira, P. E., Richardson, S. R., and Faulkner, G. J. (2014). L1 retrotransposons, cancer stem cells and oncogenesis. *FEBS J.* 281, 63–73. doi: 10.1111/febs.12601
- Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M., and Neretti, N. (2014). Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* 15:583. doi: 10.1186/1471-2164-15-583

ACKNOWLEDGMENTS

The results published here are in whole or part based upon data generated by The Cancer Genome Atlas managed by the NCI and NHGRI. Information about TCGA can be found at <http://cancergenome.nih.gov>. The authors thank Emily Norris for feedback on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmolb.2016.00076/full#supplementary-material>

- Damiani, D., Tiribelli, M., Franzoni, A., Michelutti, A., Fabbro, D., Cavallin, M., et al. (2013). BAALC overexpression retains its negative prognostic role across all cytogenetic risk groups in acute myeloid leukemia patients. *Am. J. Hematol.* 88, 848–852. doi: 10.1002/ajh.23516
- Deininger, P. (2011). Alu elements: know the SINEs. *Genome Biol.* 12:236. doi: 10.1186/gb-2011-12-12-236
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384
- Doucet-O'Hare, T. T., Rodic, N., Sharma, R., Darbari, I., Abril, G., Choi, J. A., et al. (2015). LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4894–E4900. doi: 10.1073/pnas.1502474112
- Doucet-O'Hare, T. T., Sharma, R., Rodic, N., Anders, R. A., Burns, K. H., and Kazazian, H. H. Jr. (2016). Somatic Acquired LINE-1 Insertions in Normal Esophagus Undergo Clonal Expansion in Esophageal Squamous Cell Carcinoma. *Hum. Mutat.* 37, 942–954. doi: 10.1002/humu.23027
- Ewing, A. D. (2015). Transposable element detection from whole genome sequence data. *Mob. DNA* 6, 24. doi: 10.1186/s13100-015-0055-3
- Ewing, A. D., Gacita, A., Wood, L. D., Ma, F., Xing, D., Kim, M. S., et al. (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* 25, 1536–1545. doi: 10.1101/gr.196238.115
- Forbes, S. A., Beare, D., Gunasekaran, P., Leung, K., Bindal, N., Boutselakis, H., et al. (2015). COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* 43(Database issue), D805–D811. doi: 10.1093/nar/gku1075
- Han, J. S., Szak, S. T., and Boeke, J. D. (2004). Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* 429, 268–274. doi: 10.1038/nature02536
- Hancks, D. C., and Kazazian, H. H. Jr. (2010). SVA retrotransposons: evolution and genetic instability. *Semin. Cancer Biol.* 20, 234–245. doi: 10.1016/j.semcancer.2010.04.001
- Hancks, D. C., and Kazazian, H. H. Jr. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.gde.2012.02.006
- Harris, C. R., Normart, R., Yang, Q., Stevenson, E., Haffty, B. G., Ganesan, S., et al. (2010). Association of nuclear localization of a long interspersed nuclear element-1 protein in breast tumors with poor prognostic outcomes. *Genes Cancer* 1, 115–124. doi: 10.1177/1947601909360812
- Helman, E., Lawrence, M. S., Stewart, C., Sougnez, C., Getz, G., and Meyerson, M. (2014). Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Gen. Res.* 24, 1053–1063. doi: 10.1101/gr.163659.113
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020
- Jin, Y., Tam, O. H., Paniagua, E., and Hammell, M. (2015). TETranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* 31, 3593–3599. doi: 10.1093/bioinformatics/btv422

- Kazazian, H. H. Jr., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332, 164–166. doi: 10.1038/332164a0
- Kemp, J. R., and Longworth, M. S. (2015). Crossing the LINE Toward Genomic Instability: LINE-1 Retrotransposition in Cancer. *Front. Chem.* 3:68. doi: 10.3389/fchem.2015.00068
- 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, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J. III, et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971. doi: 10.1126/science.1222077
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Maltbie, D., Ganeshalingam, L., and Allen, P. (2013). *System and Method for Secure, High-Speed Transfer of Very Large Files*. Google Patents.
- Martinelli, S., De Luca, A., Stellacci, E., Rossi, C., Checquolo, S., Lepri, F., et al. (2010). Heterozygous germline mutations in the CBL tumor-suppressor gene cause a Noonan syndrome-like phenotype. *Am. J. Hum. Genet.* 87, 250–257. doi: 10.1016/j.ajhg.2010.06.015
- Marx, V. (2014). Cancer genomes: discerning drivers from passengers. *Nat. Methods* 11, 375–379. doi: 10.1038/nmeth.2891
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., et al. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–645.
- Mills, R. E., Bennett, E. A., Iskow, R. C., and Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends Genet.* 23, 183–191. doi: 10.1016/j.tig.2007.02.006
- Morita, K., Masamoto, Y., Kataoka, K., Koya, J., Kagoya, Y., Yashiroda, H., et al. (2015). BAALC potentiates oncogenic ERK pathway through interactions with MEK1 and KLF4. *Leukemia* 29, 2248–2256. doi: 10.1038/leu.2015.137
- Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M., and Astrin, S. M. (1988). Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 333, 87–90. doi: 10.1038/333087a0
- Naramura, M., Nadeau, S., Mohapatra, B., Ahmad, G., Mukhopadhyay, C., Sattler, M., et al. (2011). Mutant Cbl proteins as oncogenic drivers in myeloproliferative disorders. *Oncotarget* 2, 245–250. doi: 10.18632/oncotarget.233
- Ostertag, E. M., Goodier, J. L., Zhang, Y., and Kazazian, H. H. Jr. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* 73, 1444–1451. doi: 10.1086/380207
- Penzkofer, T., Dandekar, T., and Zemojtel, T. (2005). L1Base: from functional annotation to prediction of active LINE-1 elements. *Nucleic Acids Res.* 33(Database issue), D498–D500. doi: 10.1093/nar/gki044
- Pon, J. R., and Marra, M. A. (2015). Driver and passenger mutations in cancer. *Annu. Rev. Pathol.* 10, 25–50. doi: 10.1146/annurev-pathol-012414-040312
- Pruitt, K. D., Tatusova, T., Brown, G. R., and Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 40(Database issue), D130–D135. doi: 10.1093/nar/gkr1079
- Quinlan, A. R. (2014). BEDTools: the Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics* 47, 11.12.1–11.12.34. doi: 10.1002/0471250953.bi1112s47
- Rangasamy, D., Lenka, N., Ohms, S., Dahlstrom, J. E., Blackburn, A. C., and Board, P. G. (2015). Activation of LINE-1 Retrotransposon Increases the Risk of Epithelial-Mesenchymal Transition and Metastasis in Epithelial Cancer. *Curr. Mol. Med.* 15, 588–597. doi: 10.2174/1566524015666150831130827
- Rishishwar, L., Marino-Ramirez, L., and Jordan, I. K. (2016). Benchmarking computational tools for polymorphic transposable element detection. *Brief. Bioinform.* doi: 10.1093/bib/bbw072. [Epub ahead of print].
- Rishishwar, L., Tellez Villa, C. E., and Jordan, I. K. (2015). Transposable element polymorphisms recapitulate human evolution. *Mob. DNA* 6, 21. doi: 10.1186/s13100-015-0052-6
- Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330. doi: 10.1038/nature14248
- Rodic, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., et al. (2014). Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* 184, 1280–1286. doi: 10.1016/j.ajpath.2014.01.007
- Schmidt, M. H., and Dikic, I. (2005). The Cbl interactome and its functions. *Nat. Rev. Mol. Cell Biol.* 6, 907–918. doi: 10.1038/nrm1762
- Scott, E. C., Gardner, E. J., Masood, A., Chuang, N. T., Vertino, P. M., and Devine, S. E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res.* 26, 745–755. doi: 10.1101/gr.201814.115
- Sever, R., and Brugge, J. S. (2015). Signal transduction in cancer. *Cold Spring Harb. Perspect. Med.* 5:a006098. doi: 10.1101/cshperspect.a006098
- Shukla, R., Upton, K. R., Muñoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111. doi: 10.1016/j.cell.2013.02.032
- Solyom, S., Ewing, A. D., Rahrmann, E. P., Doucet, T., Nelson, H. H., Burns, M. B., et al. (2012). Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* 22, 2328–2338. doi: 10.1101/gr.145235.112
- Solyom, S., and Kazazian, H. H. (2012). Mobile elements in the human genome: implications for disease. *Genome Med.* 4:12. doi: 10.1186/gm311
- Stratton, M. R., Campbell, P. J., and Futreal, P. A. (2009). The cancer genome. *Nature* 458, 719–724. doi: 10.1038/nature07943
- Sudmant, P. H., Rausch, T., Gardner, E. J., Handsaker, R. E., Abyzov, A., Huddleston, J., et al. (2015). An integrated map of structural variation in 2,504 human genomes. *Nature* 526, 75–81. doi: 10.1038/nature15394
- Thung, D. T., de Ligt, J., Vissers, L. E., Steehouwer, M., Kroon, M., de Vries, P., et al. (2014). Mobster: accurate detection of mobile element insertions in next generation sequencing data. *Genome Biol.* 15:488. doi: 10.1186/s13059-014-0488-x
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621
- Tubio, J. M., Li, Y., Ju, Y. S., Martincorena, I., Cooke, S. L., Tojo, M., et al. (2014). Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345:1251343. doi: 10.1126/science.1251343
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11 10 11–33. doi: 10.1002/0471250953.bi1110s43
- Wang, H., Xing, J., Grover, D., Hedges, D. J., Han, K., Walker, J. A., et al. (2005). SVA elements: a hominid-specific retroposon family. *J. Mol. Biol.* 354, 994–1007. doi: 10.1016/j.jmb.2005.09.085
- Weinstein, J. N., Collisson, E. A., Mills, G. B., Shaw, K. M., Ozenberger, B. A., Ellrott, K., et al. (2013). The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120. doi: 10.1038/ng.2764
- Wildschutte, J. H., Williams, Z. H., Montesin, 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
- Zhou, J. D., Yang, L., Zhang, Y. Y., Yang, J., Wen, X. M., Guo, H., et al. (2015). Overexpression of BAALC: clinical significance in Chinese de novo acute myeloid leukemia. *Med. Oncol.* 32:386. doi: 10.1007/s12032-014-0386-9

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 © 2016 Clayton, Wang, Rishishwar, Wang, McDonald and Jordan. 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) or licensor 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 Reverse Transcriptase Encoded by LINE-1 Retrotransposons in the Genesis, Progression, and Therapy of Cancer

Ilaria Sciamanna¹, Chiara De Luca¹ and Corrado Spadafora^{2*}

¹ Istituto Superiore di Sanità, Rome, Italy, ² Institute of Translational Pharmacology, National Research Council of Italy, Rome, Italy

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Independent Investigator, USA

Reviewed by:

Amy Hulme,
Missouri State University, USA
Wenfeng An,
South Dakota State University, USA

*Correspondence:

Corrado Spadafora
cspadafora@tin.it

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 17 August 2015

Accepted: 26 January 2016

Published: 11 February 2016

Citation:

Sciamanna I, De Luca C and
Spadafora C (2016) The Reverse
Transcriptase Encoded by LINE-1
Retrotransposons in the Genesis,
Progression, and Therapy of Cancer.
Front. Chem. 4:6.
doi: 10.3389/fchem.2016.00006

In higher eukaryotic genomes, Long Interspersed Nuclear Element 1 (LINE-1) retrotransposons represent a large family of repeated genomic elements. They transpose using a reverse transcriptase (RT), which they encode as part of the ORF2p product. RT inhibition in cancer cells, either via RNA interference-dependent silencing of active LINE-1 elements, or using RT inhibitory drugs, reduces cancer cell proliferation, promotes their differentiation and antagonizes tumor progression in animal models. Indeed, the non-nucleoside RT inhibitor efavirenz has recently been tested in a phase II clinical trial with metastatic prostate cancer patients. An in-depth analysis of ORF2p in a mouse model of breast cancer showed ORF2p to be precociously expressed in precancerous lesions and highly abundant in advanced cancer stages, while being barely detectable in normal breast tissue, providing a rationale for the finding that RT-expressing tumors are therapeutically sensitive to RT inhibitors. We summarize mechanistic and gene profiling studies indicating that abundant LINE-1-derived RT can “sequester” RNA substrates for reverse transcription in tumor cells, entailing the formation of RNA:DNA hybrid molecules and impairing the overall production of regulatory miRNAs, with a global impact on the cell transcriptome. Based on these data, LINE-1-ORF2 encoded RT has a tumor-promoting potential that is exerted at an epigenetic level. We propose a model whereby LINE1-RT drives a previously unrecognized global regulatory process, the deregulation of which drives cell transformation and tumorigenesis with possible implications for cancer cell heterogeneity.

Keywords: LINE-1 retrotransposons, reverse transcriptase, tumorigenesis, differentiation therapy, cancer heterogeneity, epigenetics

INTRODUCTION: THE RETROTRANSPOSITION MACHINERY IN THE GENESIS OF GENOMIC AND EPIGENOMIC LANDSCAPES

The complete sequencing of the human genome has disclosed the unexpected finding that coding genes account for a mere 1.2% of the total genome, while the remaining portion is constituted by non-coding DNA (International Human Genome Sequencing Consortium, 2001). Branded by a historically “bad reputation”, non-coding sequences have been defined as “junk” (Ohno, 1972) or “selfish” (Orgel and Crick, 1980) DNA, a view further strengthened by the evidence that nearly 50% of the human genome is constituted by apparently functionless transposable “genetic parasites” thought to increasingly litter all chromosomes during evolution.

Two main families of transposable elements characterize eukaryotic genomes: DNA transposons, which mobilize through a “cut and paste” mechanism (Muñoz-López and García-Pérez, 2010), and retrotransposons, which mobilize instead through “copy and paste,” a process that requires the reverse transcription of RNA intermediates into cDNA copies as a preliminary step in retrotransposition (Levin and Moran, 2011), promoting the broad expansion of retroelements in eukaryotic genomes.

A key player in this mechanism is the enzyme reverse transcriptase (RT) encoded by LINE-1 retrotransposons themselves. The latter are a source of the RT activity required to promote retrotransposition in human cells (Brouha et al., 2003). LINE-1 elements actually harbor two open reading frames, ORF1 and ORF2, which respectively encode ORF1p, an RNA-binding protein, and ORF2p, with reverse transcriptase (RT) and endonuclease (EN) activities (reviewed in Babushok and Kazazian, 2007). The LINE-1-derived retrotransposition machinery, constituted by ORF1 and ORF2 proteins, has *cis*-preference for its own LINE-1 RNA (Esnault et al., 2000; Wei et al., 2001; Kulpa and Moran, 2006). LINE-1-derived RT is also used for retrotranscription /retrotransposition of other RNAs, including Alu elements (Dewannieux et al., 2003), SVA (SINE-R/VNTR/ALU) elements (Ostertag et al., 2003) and mRNAs that give rise to a large population of processed pseudogenes nearly as numerous as the original coding genes (Pink et al., 2011). It is now well established that RT-originating sequences contribute to shape genomes and constitute a large proportion of evolutionarily conserved chromosomal DNA, accounting altogether for nearly 50% of the human genome (International Human Genome Sequencing Consortium, 2001). Such extensive preservation suggests a functional importance of retrotransposons. Not surprisingly, retrotransposons are increasingly being implicated in fundamental genomic functions, in both normal and pathological contexts (Rebollo et al., 2012). Indeed, as highly dynamic components of genomes, they contribute a relentless source of genetic and epigenetic variations and novelty (Feschotte, 2008; Bourque, 2009; Beck et al., 2011) and, on the long run, a major driving force in genome evolution (Oliver and Greene, 2011). A detailed description of all functional implications of retrotransposition in genome biology and evolution would be out of scope in this article, but extensive information is discussed in excellent reviews (Feschotte, 2008; Goodier and Kazazian, 2008; Bourque, 2009; Beck et al., 2011; Oliver and Greene, 2011; Rebollo et al., 2012).

The advent of high-throughput technologies in recent years has provided an accurate localization of new genomic insertions, shifting the focus from a gene-centric to a genome-wide view. This has revolutionized the traditional paradigms of genome organization by disclosing novel and unexpectedly complex genomic landscapes. Studies now show that genomes are crowded with sequences of reverse-transcribed origin, many of which are correlated with the insurgence of a variety of pathologies (for a review Hancks and Kazazian, 2012), in particular cancer (Belancio et al., 2010).

The ENCODE Project Consortium (The ENCODE Project Consortium, 2012) showed that approximately 80% of the human genome is pervasively transcribed; actually, a relevant proportion

of small and long non-coding transcripts are functional components of genome-wide regulatory networks (Djebali et al., 2012). The groundbreaking finding of an astounding landscape of small RNAs—classified as microRNAs (miRNAs), endogenous small interfering RNAs (endo-siRNAs or siRNAs) (Piatek and Werner, 2014) and Piwi-interacting RNAs (piRNAs) (Kim et al., 2009), depending on their origin and the proteins they interact with - has unveiled an RNA-mediated regulatory network that controls the genome architecture and transcriptomic profile (Aalto and Pasquinelli, 2012; Li, 2014), influencing a multitude of biological processes. Growing data show a dual relationship between small RNAs and retroelements: on the one hand, small RNAs act as “guardians of the genome” in transposon-defense pathways aimed at repressing retroelement mobility (Yang and Kazazian, 2006; reviewed in Malone and Hannon, 2009); on the other hand, retroelements are intimately involved in their biogenesis, because a growing number of small RNAs in all three classes have a recognized retrotransposon-derived origin (Borchert et al., 2011; Watanabe et al., 2011).

Long non-coding RNA (lncRNAs) are components of the mammalian transcriptome and constitute a heterogeneous class of thousands of polymerase II-transcribed RNA species, polyadenylated, spliced, mostly localized in the nucleus (reviewed by Zhang et al., 2013). A large proportion of lncRNAs, with either oncogenic or tumor suppressor roles, are constituted by antisense RNAs; the latter, together with sense transcripts, are being identified in genome-wide regulatory networks that epigenetically fine-tune genome expression, with implications in tumorigenesis, differentiation and development (reviewed in Pelechano and Steinmetz, 2013; Fatica and Bozzoni, 2014). lncRNAs also have tight connections with transposable elements of both the DNA-based and the retroelements families, which occur within nearly 80% of mature lncRNA transcripts and account for about 30–40% of total human lncRNA sequences (Kelley and Rinn, 2012; Kapusta et al., 2013). Also of RT-derived origin are a large proportion of genomic sequences highly preserved throughout evolution and classified as conserved, highly-conserved and ultra-conserved elements (UCRs), according to the level of conservation throughout species (Bejerano et al., 2004; Woolfe et al., 2005, for a recent review see Nelson and Wardle, 2013).

From an ample survey encompassing the genomes of 29 mammalian species (Lowe and Haussler, 2012), a vision of genomes emerges as complex integrated functional systems, in which a considerable proportion of non-exonic sequences were exapted from mobile element insertions (Nishihara et al., 2006; Lowe and Haussler, 2012) to assemble large-scale regulatory circuits. Deregulation of these circuits is implicated in a variety of diseases, including cancer (Esteller, 2011).

LINE-1-ENCODED RT AS A NEW UNDERESTIMATED PLAYER IN CANCER

While retrotransposable elements are extensively studied and characterized, somewhat surprisingly the retrotransposon-encoded RT activity has long failed to attract an equivalent

attention. The RT encoded by infective retroviruses has actually been intensively studied since the time of its discovery in 1970 by Baltimore (1970) and Temin and Mizutani (1970), due to its clinical implications (Herschhorn and Hizi, 2010). In contrast, the endogenous cellular RT has long been overlooked, despite many clues implicating it in as relevant processes as embryogenesis and tumorigenesis. Decades after the discoveries of Baltimore and Temin, a body of evidence has shown that endogenous RT expression is developmentally modulated: low levels of RT, if any, are expressed in differentiated non pathological tissues; increased expression is instead typical of cells characterized by low differentiation and high proliferation, e.g., early embryos (for a review see Sciamanna et al., 2011) and transformed cells (for a review see Sinibaldi-Vallebona et al., 2011). Overall, that is consistent with the notion that LINE-1 increased expression (Chen et al., 2012a; Rodić et al., 2014) and retroelement mobilization are implicated in tumorigenesis (Hancks and Kazazian, 2012; Kaer and Speek, 2013).

In contrast to differentiated quiescent cells, tissues and cells with low differentiation and high proliferation states are sites of high RT expression and provide permissive contexts for retrotransposition. Following up on that line, several studies have pursued RT inhibition in cancer cells, either using non-nucleoside RT inhibitors (nevirapine and efavirenz; Mangiacasale et al., 2003; Landriscina et al., 2005; Sciamanna et al., 2005, 2013), or RNA interference (RNAi)-mediated downregulation of RT-encoding LINE-1 elements (Sciamanna et al., 2005; Oricchio et al., 2007). In the latter case, the RNAi assays were carried out using double-stranded siRNA oligonucleotide targeted against the ORF-1 encoding domain of human full-length, highly expressed LINE-1s (Brouha et al., 2003). Both the drug-mediated and the RNAi-mediated approaches to reduce LINE-1-derived RT were found to reduce proliferation, promote differentiation and reprogram the global transcription profiles of coding and non-coding sequences in several cancer cell lines (human melanoma, glioblastoma, osteosarcoma and prostate, colon and small cell lung carcinomas). This provided early evidence for the implication of the LINE-1-encoded RT in tumorigenesis. The inhibitory effects of efavirenz on LINE-1 reverse transcription and retrotransposition were further tested in *in vitro* assays (Dai et al., 2011), and its antiproliferative and differentiating potential have been recently confirmed in breast (Patnala et al., 2013) and pancreatic (Hecht et al., 2015) cancer cell lines. Moreover, efavirenz treatment of mice xenografted with human tumorigenic cells caused the arrest, or a significant slow down, of progression of several tumor types *in vivo* (Sciamanna et al., 2005). Importantly, RNAi-mediated LINE-1 downregulation drastically reduced the tumorigenic potential of human cancer cells in nude mice (Oricchio et al., 2007). These effects are reversible and, upon discontinuation of RT inhibitory treatments, tumor cells return to their original de-differentiated phenotype and unrestrained proliferation capacity (Sciamanna et al., 2005); these observations provided initial hints to an epigenetic role of RT.

The high levels of RT activity found in tumor cells and tissues, reported by our (Mangiacasale et al., 2003; Gualtieri et al., 2013) and other laboratories (Patnala et al., 2013), correlate well with the enhanced rate of retrotransposition observed in many

human tumors, a phenomenon that dramatically contributes to shape cancer genomes (Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012; Shukla et al., 2013; Ewing et al., 2015). In a MMTV-PyVT transgenic mouse strain (Guy et al., 1992), whose females spontaneously develop breast carcinoma, a burst in the copy number of both LINE-1 and SINE B1 elements was depicted very early at tumor onset; their copy number further increases along with tumor progression (Gualtieri et al., 2013). These data converge to indicate that tumors constitute a highly permissive environment for retrotranscription, yet do not answer the question of whether overexpression and amplification of LINE-1 elements act as oncological “drivers” or as mere “passengers” (Rodic and Burns, 2013). The findings that pharmacological inhibition of RT is sufficient to reduce cancer cell proliferation, promote differentiation and antagonize tumor progression in animal models, similar to the effects obtained by RNAi-specific downregulation of LINE-1 expression, strongly support a causative role of LINE-1-encoded RT in tumorigenesis. In an applied clinical perspective, therefore, RT can be regarded as a target and RT inhibitors as potential therapeutic agents in a novel cancer differentiation therapy. Efavirenz has recently been tested in a phase II trial with metastatic prostate cancer patients, suggesting that relatively high dosage (over 600 mg per day) may be beneficial as a novel anticancer treatment (Houédé et al., 2014).

The role of RT encoded by LINE-1 in tumorigenesis is distinct from that of RT activities produced from the other two potential sources, i.e., endogenous retroviruses (HERVs) and telomerase-associated RT (TERT). First, RNAi-mediated downregulation of HERV-K expression showed negligible effects on the rate of proliferation and differentiation of cancer cells, in contrast with the dramatic effects observed after LINE-1-specific RNAi (Oricchio et al., 2007). Second, inhibitors of LINE-1 derived RT elicit rapid changes in treated cells in our experiments (Mangiacasale et al., 2003; Sciamanna et al., 2005), differently from drugs targeting telomerase, which reduce cancer cell proliferation after a long treatment (about 120 days; Damm et al., 2001); these data therefore rule out the possibility that TERT contributes to the rapid response of cells to RT inhibitors. It should be noted, however, that LINE-1 RT is critical for telomere maintenance, given that LINE-1 knockdown in cancer cells correlates with: (i) reduced length of telomeres, (ii) decreased telomerase activity, and (iii) decreased telomerase mRNA level (Aschacher et al., 2016). Together these results reveal that LINE-1 RT has a functional impact on TERT. Thus, while TERT is not involved in the changes elicited by inhibitors targeting the retrotransposon-derived genuine RT, the level of activity of LINE-1 elements may impact on TERT. These findings again strengthen the view that LINE-1 RT is a major player in tumorigenesis.

LINE-1 ORF2-ENCODED RT ACTIVITY IN CANCER PROGRESSION

The ORF2-encoded RT has been recently assessed for its suitability as a tumor marker (Gualtieri et al., 2013) in females of the cancer-prone MMTV-PyVT described above (Guy et al.,

1992). In these females, breast cancer tissues withdrawn at different times after birth are representative of progressive cancer stages. ORF2p cannot be detected in normal breast tissue by immunohistochemistry (IHC), but increased expression is triggered very early in tumorigenesis, preceding the appearance of typical histological alterations and accepted cancer markers (e.g., Ki67 and epidermal growth factor receptor ERB2); further upregulation takes place during tumor growth. These findings correlate well with the notion that hypomethylated LINE-1 sequences, from which ORF2p is produced, are typical of cancer genomes and precancerous lesions compared to their normal tissues counterpart (Miousse and Koturbash, 2015).

The abundant expression of LINE-1 products in preneoplastic mammary tissues suggests an exploitable tool as a potential diagnostic biomarker for early cancer detection: the identification of cancer-prone foci marked by increased RT before the appearance of recognizable histological alterations, can expand the window of opportunities for therapeutical intervention, which can possibly be most effective if associated with the development of RT inhibitory treatment. Interestingly, the abundance and subcellular localization of LINE-1 products are also proposed to have prognostic value in human metastatic breast cancer (Chen et al., 2012a).

Compelling objectives of “the war on cancer” currently include the definition of novel early markers identifying cancer-prone lesions before their spreading, as well as the development of novel therapeutic approaches in possible replacement of conventional cytotoxic chemotherapy. In a recent critical reappraisal, Hanahan (2014) has pointed out that the war on cancer, if not lost, is certainly not won and has suggested that therapeutic strategies should avoid fragmenting along multiple, highly diversified narrow paths targeting many substrates, each of which is highly selective for a specific cancer. Rather, the therapeutic “bullets” ought to hit fewer targets shared by a large spectrum of cancers (Hanahan, 2014). LINE-1 ORF2-encoded RT would fulfill these criteria, representing, at the same time, an early diagnostic cancer marker, a worth pursuing therapeutic target and the driving component of a newly emerging cancer-promoting mechanism.

THE MOLECULAR BASES OF THE RT-DEPENDENT CANCER-PROMOTING MECHANISM

As briefly recalled above, retrotransposition events have had a fundamental role not only in shaping the genomic landscape, but also in directing regulatory networks aimed to fine-tune a variety of genomic functions. Data obtained in the last few years growingly indicate that the retrotransposon machinery, besides being a well-known source of genomic variations caused by new insertions (Böehne et al., 2008; Bourque, 2009), also exerts a global epigenetic regulatory role on the cellular transcriptome. LINE-1 ORF2-encoded RT is a new player in this mechanism.

Prompted by the finding that tumor cell lines are endowed with abundant LINE-1-encoded RT, Sciamanna et al. (2013)

began to address the mechanism through which RT might act by comparing the global transcription profile of melanoma cells before and after RT inhibition by microarray analysis. The results showed that RT inhibition modulates the expression of a broad range of coding genes, but also long and small non-coding sequences, including UCRs and miRNAs. miRNAs actually emerged as crucial components of the RT-dependent mechanism; indeed, a subpopulation, known to be involved in cell differentiation, cell growth, tumorigenesis and metastatic progression proved highly responsive to RT inhibition. Many miRNA-encoding genes are significantly associated with genomic regions enriched in closely spaced Alu repeats, further strengthening the link between miRNAs and retrotransposons. The physical association of pre-miRNA genomic loci with high density retroelements actually suggests that the latter can exert a regulatory “position effect” on miRNA expression (Slotkin and Martienssen, 2007). Experimental evidence supporting an orchestrating role of the RT enzyme emerged from cesium chloride density centrifugation analysis of nucleic acids extracted from melanoma and prostate carcinoma cell lines, harboring either “native” or efavirenz-inhibited RT: by buoyant density analysis, LINE-1- and Alu-containing molecules with the density of DNA:RNA hybrids were selectively identified in tumor cells, which disappeared upon treatment with efavirenz and were absent in non-transformed human fibroblasts (Sciamanna et al., 2013). Thus, the DNA:RNA hybrids are an especially abundant, if not exclusive, component of cancer cells, generated by reverse transcription of RNA templates, largely—albeit not exclusively—provided by LINE-1 and Alu transcripts. These data suggest that a cancer-promoting RT-dependent mechanism is active in tumor cells and can be blocked by inhibiting the LINE-1 RT. Based on these data, Sciamanna et al. proposed a model (Sciamanna et al., 2014) whereby the highly expressed LINE-1 RT in cancer cells can intercept RNAs and convert them in RNA:DNA hybrids via reverse transcription. Central to the model is the RT-dependent production of RNA:DNA hybrids, associated with altered functional miRNA profiles, observed under conditions of high LINE-1-derived RT in cancer cells and modulatable by RT inhibitors. A wealth of data show that miRNA expression is indeed downregulated in cancer cells, with profound implications for cell fates (Lu et al., 2005; Gaur et al., 2007; Jansson and Lund, 2012). A variety of small RNAs, including 7SL RNA (Ullu and Weiner, 1984), tRNAs (Kaçar et al., 1992), small nuclear RNAs (Doucet et al., 2015), and YRNAs (Perreault et al., 2005), are known to act as templates for reverse transcription in intermediate steps of the genesis of pseudogenes. It is not unreasonable to hypothesize that miRNA precursors may also be retrotranscribed. The observation that the production of hybrid RNA-DNA molecules is associated with aberrant miRNA profiles in cancer cells actually suggests that RT can “subtract” RNA precursors, thus preventing or impairing the formation of double stranded (ds) RNA dicer substrates for the biogenesis of mature miRNAs: this would ultimately contribute to establish favorable conditions for the onset of cancer phenotypes.

RT inhibition results in restored miRNA biogenesis, likely re-establishing their regulatory networks, consistent with its

empirically established capacity to revert the cancer phenotype (Sciamanna et al., 2013).

In agreement with this idea, a subset of LINE-1-specific siRNAs, targeting LINE-1 expression and capable to induce methylation of their promoters, are found to be down-modulated in breast cancer compared to normal cells (Chen et al., 2012b). Conversely, LINE-1 inhibition by siRNAs up-modulate the expression of miRNAs involved in tumor suppression (Ohms et al., 2014). Taken as a whole, these findings indicate an orchestrating role of LINE-1-encoded RT in setting a cancer-permissive cellular state.

Although, the LINE-1 enzymatic machinery preferentially reverse transcribes its own RNA (Esnault et al., 2000; Wei et al., 2001; Kulpa and Moran, 2006), the presence of intronless pseudogenes scattered throughout mammalian genomes points out that mRNAs transcribed from protein-coding genes are also substrates for reverse transcription by the endogenous RT (Pink et al., 2011). This suggests that the RT-depending mechanism, in addition to targeting miRNAs, can also target several more RNA classes, coding and non-coding, small- and long-RNAs, though with a possible preferential bias for those associated with, or derived from, retroelement sequences. Consistent with this view, Sciamanna et al. (2013) found that about one third of the efavirenz-downmodulated miRNAs in melanoma cells are clustered on chromosome 19 (C19MC) in a locus characterized by a high density of primate-specific Alu repeats, which were shown to have co-evolved with miRNAs coding genes (Lehnert et al., 2009). An independent study also reported that LINE-1 silencing caused a deregulated profile of miRNA expression in breast cancer cells (Ohms and Rangasamy, 2014).

In summary, LINE-1 expression and small RNA networks emerge as the balanced components of a RT-depending regulatory mechanism placed at the intersection between normally differentiated and transformed non-differentiated cellular states: when one component raises the other one decays.

It is worth stressing that the partial inactivation of miRNA function is not an exclusive feature of cancer, but is a physiological phenomenon, shared with early preimplantation embryos, a context where again miRNA pathways become transiently suppressed (Suh et al., 2010). Moreover, miRNA inactivation is concomitant with a burst of LINE-1 activity in both tumorigenesis and embryogenesis. In the next paragraph we discuss this striking analogy and suggest that physiological and pathological processes have in common the same RT-dependent mechanism.

THE RT-BASED MECHANISM AS GLOBAL REGULATOR OF DIFFERENTIATION IN TUMORIGENESIS AND EMBRYOGENESIS

In prior developmental studies, the presence of LINE-1-encoded RT activity and protein was assessed in gametes and early embryos to address the potential role of this enzyme in embryogenesis. Unexpectedly, Giordano et al. (2000) found an RT activity in mature murine spermatozoa, providing the first hint that RT might somehow be involved in early

embryogenesis. The sperm endogenous RT, far from being a nonfunctional remnant encoded by “genomic parasites,” has a full enzymatic activity able to reverse transcribe exogenous RNA molecules, taken up and internalized by spermatozoa, in cDNA copies that could then be delivered to embryos at fertilization (Giordano et al., 2000; reviewed in Spadafora, 2008). Pittoggi et al. (2003) further found that RT is also present in early embryos and is strictly required for preimplantation development: indeed, exposing zygotes to RT inhibitor, or antisense-mediated downregulation of LINE-1 (Beraldi et al., 2006), both caused a drastic arrest of development at the two- and four-cell embryo stages with globally altered gene expression profiles. Interestingly, fertilization activates a reverse transcription wave in zygotes within a few hours, which then propagates throughout the first cell division; that is concomitant with the production of new LINE-1 copies that mostly remain as non integrated extrachromosomal structures (Vitullo et al., 2012). Indirect evidence for an embryonic RT activity also emerge from reports that somatic LINE-1 retrotranspositions occur in human stem cells (Garcia-Perez et al., 2007; Coufal et al., 2009) and in very early stages of development in humans (van den Hurk et al., 2007) and rodents using transgenic murine and rat models (Kano et al., 2009). These findings indicate that the endogenous RT is active in early stages of embryogenesis, where it appears to have implications for epigenetic regulation of gene expression and to be necessary for the unfolding of the developmental program.

Cancer and embryo developmental studies convergingly point to the conclusion that an RT-based mechanism is physiologically activated in early embryogenesis and repressed in differentiated tissues; its unscheduled re-activation in somatic cells has cancer-promoting effects, yielding increased cell proliferation and loss of differentiation, in analogy with embryonic growth. It is a well-established notion that tumors and embryos share a variety of cellular, biochemical and molecular features and that genes typically expressed in embryogenesis, yet silenced in normal differentiated tissues, are re-expressed in tumors (Ma et al., 2010). These circumstances support the conclusion that tumorigenesis often recapitulates developmental patterns (Kaiser et al., 2007). In this conceptual framework, Spadafora (Spadafora, 2015) proposed that the RT-dependent mechanism is a source of the functional analogies shared by the physiological and pathological processes connecting embryogenesis and tumorigenesis.

THE GENESIS OF CANCER HETEROGENEITY

The retrotransposon machinery is highly sensitive to stressing stimuli (Hagan and Rudin, 2002; Terasaki et al., 2013). In response, LINE-1 expression can be activated at differential levels in different cells, depending on the nature and the intensity of endogenous or exogenous stressors. We propose that the differential activation of RT, including by stress, can generate the heterogeneously differentiated cell populations that typically characterize human cancers (Meachem and Morrison, 2013). It is currently unclear whether the cellular heterogeneity observed

in cancer reflects the existence of cell populations undergoing a progressive transformation “trajectory,” initiating as a primary cancer state and sequentially evolving into metastatic cells, or whether a broad array of cellular variations simultaneously arise in a single stress-responding event. Based on the data discussed above, it is tempting to speculate that the latter is the case; cells with varying degrees of malignancy—some of which may confer metastatic capacity—may concomitantly originate in a single genome-wide burst of stress-activated LINE-1-RT expression. In this hypothetical model, schematized in **Figure 1**, burst(s) of LINE-1 expression, triggered by exogenous and/or endogenous stimuli in normal cells (in green), would generate an array of cell populations endowed with various levels of LINE-1-dependent RT activity (indicated by different shades of colored cells), coinciding with the emergence of preneoplastic lesions. We propose that the different levels of LINE-1 activation correspond to different degrees of cell de-differentiation; in the process, embryonic regulatory patterns can be reactivated and induce somatic cells to revert back to embryo-like states (Kaiser et al., 2007). LINE-1 activation at low levels would exert modest de-differentiation effects, while higher levels would determine a more extensive reactivation of embryonic patterns, with the ensuing production of more aggressive “embryo-like” transformed cells. The cell populations concomitantly originating from the activation of RT expression would then differentially propagate throughout cancer progression, thus contributing to cancer heterogeneity. The model represented in **Figure 1** was inspired by the recently proposed “Big Bang” hypothesis for the genesis of human cancer, in which a single ancestral event is thought to originate the heterogeneity of cancer cell populations (Sottoriva et al., 2015), which would then progress and expand in parallel (Klein, 2009). The simultaneous genesis of cells with heterogeneous invasive potential would also offer a possible explanation for the genesis and spreading of metastatic tumors of unknown primary origin: these are a relatively rare class of metastatic tumors detected in patients in which the primary tumor cannot be identified, and account for 3–5% of all cancer diagnoses (Natoli et al., 2011; Stella et al., 2012).

In more general terms, the model predicts a relatively minor role for DNA mutations in cancer progression, as cell transformation is rather viewed as originating from an RT-mediated reactivation of “embryonic” regulatory circuits mostly acting at the epigenetic level in differentiated cells (Spadafora, 2015). Although needing further experimental testing, the model builds on emerging evidence indicating the global reach of RT onto several RNA classes (Sciamanna et al., 2013; Ohms et al., 2014; Ohms and Rangasamy, 2014) and is compatible with its reversible character by modulating RT levels (Sciamanna et al., 2005).

Retrotransposable elements also clearly impinge on genome function by generating extensive variations via insertional mutagenesis. Although large numbers of mutations are identified by high-throughput sequencing data in cancer contexts, their role(s) in tumorigenesis is often undefined (Kandoth et al., 2013): predisposing gene mutations in fact play a documented causative role only in 5–10% of human cancers (Nagy et al., 2004).

Recent excellent works have reported that the genomes of different tumor types harbor hundreds of *de novo* somatic insertions, selectively found in cancer genomes (Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012; Shukla et al., 2013; Doucet-O’Hare et al., 2015; Ewing et al., 2015; Rodić et al., 2015). Despite of these reports, however, the general implication of L1 retrotransposition events as either “driver” mutations (i.e., with a causative role in tumorigenesis), or as “passengers” (i.e., manifesting a consequence of the loss of genome regulation associated with cell transformation), remains an open question (Rodić and Burns, 2013). Insertions were documented and were attributed causative trigger in specific cases; among others, LINE-1 insertion were found within the *c-myc* gene (Morse et al., 1988), or in the tumor-suppressing gene *apc* (Miki et al., 1992) in breast and colon carcinoma, respectively; in those instances, LINE-1 insertions should have an activating (*c-myc*) or inhibitory (*apc*) role, respectively. In a different context, Alu insertions also result in neurofibromatosis type 1 (Wallace et al., 1991). It is worth recalling, however, that together LINE-1, Alu and SVA insertions account only for a marginal contribution (<0.5%) to the genesis of cancer (Callinan and Batzer, 2006). This leaves

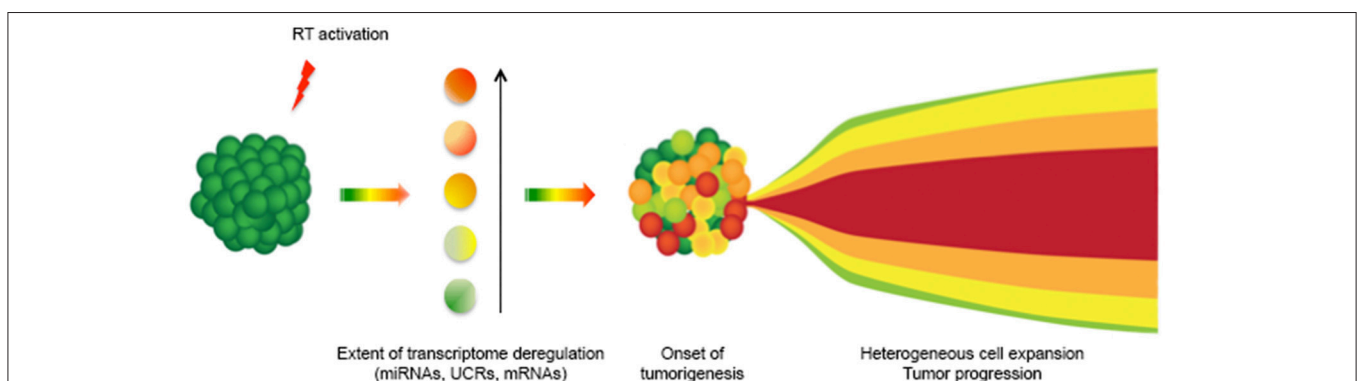


FIGURE 1 | A model for RT-dependent induction of cancer cell heterogeneity. The deregulated expression of LINE-1 elements in somatic cells (green) causes a burst of RT activity (red flash), which deregulates the transcriptome of individual cells at various levels (represented by different color shades): this originates heterogeneous cancer cell populations. In the model, cancer cell heterogeneity would therefore set in following the early burst of differentially expressed RT activity in different cells. Cancer would then progress with the expansion of various cell populations (on the right).

ample room for non-inserational mechanisms of tumorigenesis that implicate retrotransposons. In addition, the concept that new insertions might cause tumorigenesis would be hard to reconcile with the full reversibility of the “therapeutic” effect associated with LINE-1 RT inhibition in various cancer cells, observed in our and other laboratories (Sciamanna et al., 2005; Oricchio et al., 2007; Patnala et al., 2013). In our model, therefore, insertional mutagenesis, though not being totally ruled out, plays a minor role. We believe that most retrotranspositional insertions observed in many tumors reflect a failure to repress the activity of retroelements (a frequent failure in cancer), rather than being a cause of tumorigenesis.

In an extreme view, mutations may often represent a tolerated consequence of the tumor-associated global deregulation rather than the cause. The evidence summarized so far suggest that deregulated RT activity, likely acting in combination with other key epigenetic processes such as global DNA hypomethylation and chromatin remodeling, contributes to shape pro-tumorigenic expression profiles, and thus favors the phenotypic plasticity and diversity of cancer cells.

CANCER AS A REVERSIBLE “DEVELOPMENTAL” DISORDER AND DIFFERENTIATION THERAPY

As discussed above, the non-coding RNA profiles modulated by RT can globally regulate cell differentiation. Evidence is emerging that unscheduled reactivation of RT, as occurring in cancer cells, or its developmentally regulated repression, as in normal cells, are sufficient to promote cell de-differentiation or, on the contrary, stabilize the differentiated state, respectively. Tumorigenesis can be viewed as the erroneous resumption of genome-wide networks active in embryogenesis and silenced in adult life, and the differentiation process can be regarded as a sequence of transient and reversible cellular states in which RT activity is variably activated. According to this view, cancer would also be a reversible phenomenon and, as such, potentially modulatable by RT-inhibitory differentiation-inducing agents. The idea that the “normal” differentiation program can be restored to cancer cells, with the loss or attenuation of tumorigenic phenotype, has inspired much research and clinical work in the last decades. Perhaps the best known example is the development of retinoic acid-based differentiation therapy, successfully applied to treat acute promyelocytic leukemia (APML). Retinoic acid is

a powerful morphogen and differentiating agent and has been the object of intense studies in the last decades, the outcome of which cannot be exhaustively discussed here (reviewed by Tang and Gudas, 2011). Other attempts to apply the same principle to solid tumors, however, have had more limited results so far (reviewed by Leszczyniecka et al., 2001; Cruz and Matushansky, 2012). The data obtained from *in vitro* assays, preclinical tests on animal models and a recent human trial, converge in viewing the LINE-1-encoded RT as an effective target for a non-cytotoxic, differentiation-inducing cancer therapy; RT inhibition appears to be the common condition sufficient to reverse tumorigenicity and restore differentiation to a wide variety of cancer cells.

CONCLUSIONS

Growing data undermine the concept of terminal differentiation as a stably acquired condition, revealing that: (i) differentiation states should rather be viewed as transient conditions, and (ii) even in the presence of genomic alterations, epigenetics often wins over genetics (Lotem and Sachs, 2002). Epigenetic changes can effectively bypass the genetic alterations associated with, or caused by, tumorigenesis and reprogram gene expression profiles, reverting, or mitigating, the malignant phenotypes of cells. LINE-1-encoded RT is emerging as a key epigenetic regulator at the intersection between normal and pathological development. As such, the level of RT activity has the potential to shift the biological balance of cells in one or the other direction. In our view, these findings and emerging concepts, besides their clinical implications, fulfill the early prediction by Temin that endogenous RT activity plays roles both in normal development, as in embryogenesis, and in pathologies as such as cancer (Temin, 1971).

FUNDING

Work in our laboratory was supported by the Italian Ministry of Health grant n. 15ONC/8 “Endogenous Reverse Transcriptase as a tumor marker and causative agent of tumor onset and progression,” to CS.

ACKNOWLEDGMENTS

We acknowledge the skillful assistance of Cosimo Curianò with drawing preparation.

REFERENCES

- Aalto, A. P., and Pasquinelli, A. E. (2012). Small non-coding RNAs mount a silent revolution in gene expression. *Curr. Opin. Cell Biol.* 24, 333–340. doi: 10.1016/j.ceb.2012.03.006
- Aschacher, T., Wolf, B., Enzmann, F., Kienzl, P., Messner, B., Sampl, S., et al. (2016). LINE-1 induces hTERT and ensures telomere maintenance in tumour cell lines. *Oncogene* 35, 94–104. doi: 10.1038/onc.2015.65
- Babushok, D. V., and Kazazian, H. H. Jr. (2007). Progress in understanding the biology of the human mutagen LINE-1. *Hum. Mutat.* 28, 527–539. doi: 10.1002/humu.20486
- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumor viruses. *Nature* 226, 1209–1211. doi: 10.1038/2261209a0
- Beck, C. R., Garcia-Perez, J. L., Badge, R. M., and Moran, J. V. (2011). LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* 12, 187–215. doi: 10.1146/annurev-genom-082509-141802
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., et al. (2004). Ultraconserved elements in the human genome. *Science* 304, 1321–1325. doi: 10.1126/science.1098119
- Belancio, V. P., Roy-Engel, A. M., and Deininger, P. L. (2010). All y’all need to know’bout retroelements in cancer. *Semin. Cancer Biol.* 20, 200–210. doi: 10.1016/j.semcancer.2010.06.001

- Beraldi, R., Pittoggi, C., Sciamanna, I., Mattei, E., and Spadafora, C. (2006). Expression of LINE-1 retroposons is essential for murine preimplantation development. *Mol. Reprod. Dev.* 73, 279–287. doi: 10.1002/mrd.20423
- Böehne, A., Brunet, F., Galiana-Arnoux, D., Schultheis, C., and Volff, J. N. (2008). Transposable elements as drivers of genomic and biological diversity in vertebrates. *Chromosome Res.* 16, 203–215. doi: 10.1007/s10577-007-1202-6
- Borchert, G. M., Holton, N. W., Williams, J. D., Hernan, W. L., Bishop, I. P., Dembosky, J. A., et al. (2011). Comprehensive analysis of microRNA genomic loci identifies pervasive repetitive-element origins. *Mob. Genet. Elements* 1, 8–17. doi: 10.4161/mge.1.1.15766
- Bourque, G. (2009). Transposable elements in gene regulation and in the evolution of vertebrate genomes. *Curr. Opin. Genet. Dev.* 19, 607–612. doi: 10.1016/j.gde.2009.10.013
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285. doi: 10.1073/pnas.0831042100
- Callinan, P. A., and Batzer, M. A. (2006). Retrotransposable elements and human disease. *Genome Dyn.* 3, 104–115. doi: 10.1159/000092503
- Chen, L., Dahlstrom, J. E., Chandra, A., Board, P., and Rangasamy, D. (2012a). Prognostic value of LINE-1 retrotransposon expression and its subcellular localization in breast cancer. *Breast Cancer Res. Treat.* 136, 129–142. doi: 10.1007/s10549-012-2246-7
- Chen, L., Dahlstrom, J. E., Lee, S. H., and Rangasamy, D. (2012b). Naturally occurring endo-siRNA silences LINE-1 retrotransposons in human cells through DNA methylation. *Epigenetics* 7, 758–771. doi: 10.4161/epi.20706
- Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., et al. (2009). L1 retrotransposition in human neural progenitor cells. *Nature* 460, 1127–1131. doi: 10.1038/nature08248
- Cruz, F. D., and Matushansky, I. (2012). Solid tumor differentiation therapy - is it possible? *Oncotarget* 3, 559–567. doi: 10.18632/oncotarget.512
- Dai, L., Huang, Q., and Boeke, J. D. (2011). Effect of reverse transcriptase inhibitors on LINE-1 and Ty1 reverse transcriptase activities and on LINE-1 retrotransposition. *BMC Biochem.* 2:18. doi: 10.1186/1471-2091-12-18
- Damm, K., Hemmann, U., Garin-Chesa, P., Haul, N., Kauffmann, I., Priepke, H., et al. (2001). A highly selective telomerase inhibitor limiting human cancer cell proliferation. *EMBO J.* 20, 6958–6968. doi: 10.1093/emboj/20.24.6958
- Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35, 41–48. doi: 10.1038/ng1223
- Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., et al. (2012). Landscape of transcription in human cells. *Nature* 489, 101–108. doi: 10.1038/nature11233
- Doucet, A. J., Droc, G., Siol, O., Audoux, J., and Gilbert, N. (2015). U6 snRNA Pseudogenes: markers of retrotransposition dynamics in mammals. *Mol. Biol. Evol.* 32, 1815–1832. doi: 10.1093/molbev/msv062
- Doucet-O'Hare, T. T., Rodic, N., Sharma, R., Darbari, I., Abril, G., Choi, J. A., et al. (2015). LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4894–E900. doi: 10.1073/pnas.1502474112
- Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24, 363–367. doi: 10.1038/74184
- Esteller, M. (2011). Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12, 861–874. doi: 10.1038/nrg3074
- Ewing, A. D., Gacita, A., Wood, L. D., Ma, F., Xing, D., Kim, M. S., et al. (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* 25, 1536–1545. doi: 10.1101/gr.196238.115
- Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev.* 15, 7–21. doi: 10.1038/nrg3606
- Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. doi: 10.1038/nrg2337
- García-Perez, J. L., Marchetto, M. C. N., Muotri, A. R., Coufal, N. C., Gage, F. H., O'Shea, K. S., et al. (2007). LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* 16, 1569–1577. doi: 10.1093/hmg/ddm105
- Gaur, A., Jewell, D. A., Liang, Y., Ridzon, D., Moore, J. H., Chen, C., et al. (2007). Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer Res.* 67, 2456–2468. doi: 10.1158/0008-5472.CAN-06-2698
- Giordano, R., Magnano, A. R., Zaccagnini, G., Pittoggi, C., Moscufo, N., Lorenzini, R., et al. (2000). Reverse transcriptase activity in mature spermatozoa of mouse. *J. Cell Biol.* 148, 1107–1113. doi: 10.1083/jcb.148.6.1107
- Goodier, J. L., and Kazazian, H. H. Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* 135, 23–35. doi: 10.1016/j.cell.2008.09.022
- Gualtieri, A., Andreola, F., Sciamanna, I., Sinibaldi-Vallebona, P., Serafino, A., and Spadafora, C. (2013). Increased expression and copy number amplification of LINE-1 and SINE B1 retrotransposable elements in murine mammary carcinoma progression. *Oncotarget* 4, 1882–1893. doi: 10.18632/oncotarget.1188
- Guy, C. T., Cardiff, R. D., and Muller, W. J. (1992). Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* 12, 954–961. doi: 10.1128/MCB.12.3.954
- Hagan, C. R., and Rudin, C. M. (2002). Mobile genetic element activation and genotoxic cancer therapy potential clinical implications. *Am. J. Pharmacogenomics* 2, 25–35. doi: 10.2165/00129785-200202010-00003
- Hanahan, D. (2014). Rethinking the war on cancer. *Lancet* 383, 558–563. doi: 10.1016/S0140-6736(13)62226-6
- Hancks, D. C., and Kazazian, H. H. Jr. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.gde.2012.02.006
- Hecht, M., Erber, S., Harrer, T., Klinker, H., Roth, T., Parsch, H., et al. (2015). Efavirenz has the highest anti-proliferative effect of non-nucleoside reverse transcriptase inhibitors against pancreatic cancer cells. *PLoS ONE* 10:e0130277. doi: 10.1371/journal.pone.0130277
- Herschhorn, A., and Hizi, A. (2010). Retroviral reverse transcriptases. *Cell. Mol. Life Sci.* 67, 2717–2747. doi: 10.1007/s00018-010-0346-2
- Houédé, N., Pulido, M., Mourey, L., Joly, F., Ferrero, J. M., Bellera, C., et al. (2014). A phase II trial evaluating the efficacy and safety of efavirenz in metastatic castration-resistant prostate cancer. *Oncologist* 19, 1227–1228. doi: 10.1634/theoncologist.2014-0345
- International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020
- Jansson, M. D., and Lund, A. H. (2012). MicroRNA and cancer. *Mol. Oncol.* 6, 590–610. doi: 10.1016/j.molonc.2012.09.006
- Kaçar, Y., Thomann, H. U., and Gross, H. J. (1992). The first human genes for tRNA(ArgICG), tRNA(GlyUCC), and tRNA(ThrIGU) and more tRNA(Val) pseudogenes: expression and pre-tRNA maturation in HeLa cell-free extracts. *DNA Cell Biol.* 11, 781–790. doi: 10.1089/dna.1992.11.781
- Kaer, K., and Speek, M. (2013). Retroelements in human disease. *Gene* 518, 231–241. doi: 10.1016/j.gene.2013.01.008
- Kaiser, S., Park, Y. K., Franklin, J. L., Halberg, R. B., Yu, M., Jessen, W. J., et al. (2007). Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. *Genome Biol.* 8:R131. doi: 10.1186/gb-2007-8-7-r131
- Kandoth, C., McLellan, M. D., Vandin, F., Ye, K., Niu, B., Lu, C., et al. (2013). Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339. doi: 10.1038/nature12634
- Kano, H., Godoy, I., Courtney, C., Vetter, M. R., Gerton, G. L., Ostertag, E. M., et al. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 23, 1303–1312. doi: 10.1101/gad.183909
- Kapusta, A., Kronenberg, Z., Lynch, V. J., Zhuo, X., Ramsay, L. A., Bourque, G., et al. (2013). Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* 9:e1003470. doi: 10.1371/journal.pgen.1003470
- Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* 13:R107. doi: 10.1186/gb-2012-13-11-r107
- Kim, V. N., Han, J., and Siomi, M. C. (2009). Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* 10, 126–139. doi: 10.1038/nrm2632
- Klein, C. A. (2009). Parallel progression of primary tumours and metastases. *Nat. Rev. Cancer* 9, 302–312. doi: 10.1038/nrc2627

- Kulpa, D. A., and Moran, J. V. (2006). Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* 13, 655–660. doi: 10.1038/nsmb1107
- Landriscina, M., Fabiano, A., Altamura, S., Bagala, C., Piscazzi, A., Cassano, A., et al. (2005). Reverse transcriptase inhibitors downregulate cell proliferation *in vitro* and *in vivo* and restore TSH signaling and iodine uptake in human thyroid anaplastic carcinoma. *J. Clin. Endocrinol. Metab.* 90, 5663–5671. doi: 10.1210/jc.2005-0367
- Lee, E., Iskrow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J. III., et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971. doi: 10.1126/science.1222077
- Lehnert, S., Van Loo, P., Thilakarathne, P. J., Marynen, P., Verbeke, G., and Schuit, F. (2009). Evidence for co-evolution between human microRNAs and Alu-repeats. *PLoS ONE* 4:e4456. doi: 10.1371/journal.pone.0004456
- Leszczyniecka, M., Roberts, T., Dent, P., Grant, S., and Fisher, P. B. (2001). Differentiation therapy of human cancer: basic science and clinical applications. *Pharmacol. Ther.* 90, 105–196. doi: 10.1016/S0163-7258(01)00132-2
- Levin, H. L., and Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627. doi: 10.1038/nrg3030
- Li, L. C. (2014). Chromatin remodeling by the small RNA machinery in mammalian cells. *Epigenetics* 9, 1–8. doi: 10.4161/epi.26830
- Lotem, J., and Sachs, L. (2002). Epigenetics wins over genetics: induction of differentiation in tumor cells. *Semin. Cancer Biol.* 12, 339–346. doi: 10.1016/S1044-579X(02)00054-8
- Lowe, C. B., and Haussler, D. (2012). 29 Mammalian genomes reveal novel exaptations of mobile elements for likely regulatory functions in the human genome. *PLoS ONE* 7:e43128. doi: 10.1371/journal.pone.0043128
- Lu, J., Getz, G., Miska, E. A., Alvarez-Saavedra, E., Lamb, J., and Peck, D., et al. (2005). MicroRNA expression profiles classify human cancers. *Nature* 435, 834–838. doi: 10.1038/nature03702
- Ma, Y., Zhang P., Wang, F., Yang, J., Yang, Z., and Qin, H. (2010). The relationship between early embryo development and tumorigenesis. *J. Cell. Mol. Med.* 14, 2697–2701. doi: 10.1111/j.1582-4934.2010.01191.x
- Malone, C. D., and Hannon, G. J. (2009). Small RNAs as guardians of the genome. *Cell* 136, 656–668. doi: 10.1016/j.cell.2009.01.045
- Mangiacasale, R., Pittoggi, C., Sciamanna, I., Careddu, A., Mattei, E., Lorenzini, R., et al. (2003). Exposure of normal and transformed cells to nevirapine, a reverse transcriptase inhibitor, reduces cell growth and promotes differentiation. *Oncogene* 22, 2750–2761. doi: 10.1038/sj.onc.1206354
- Meachem, C. E., and Morrison, S. J. (2013). Tumour heterogeneity and cancer cell plasticity. *Nature* 501, 328–337. doi: 10.1038/nature12624
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., et al. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–645.
- Mioudou, I. R., and Koturbash, I. (2015). The Fine LINE: methylation drawing the cancer landscape. *Biomed Res. Int.* 2015:131547. doi: 10.1155/2015/131547
- Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M., and Astrin, S. M. (1988). Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 333, 87–90. doi: 10.1038/333087a0
- Muñoz-López, M., and García-Pérez, J. L. (2010). DNA Transposons: Nature and application in Genomics. *Curr. Biol.* 11, 115–128. doi: 10.2174/138920210790886871
- Nagy, R., Sweet, K., and Eng, C. (2004). Highly penetrant hereditary cancer syndromes. *Oncogene* 23, 6445–6470. doi: 10.1038/sj.onc.1207714
- Natoli, C., Ramazzotti, V., Nappi, O., Giacomini, P., Palmeri, S., Salvatore, M., et al. (2011). Unknown primary tumors. *Biochim. Biophys. Acta* 1816, 13–24. doi: 10.1016/j.bbcan.2011.02.002
- Nelson, A. C., and Wardle, F. C. (2013). Conserved non-coding elements and cis regulation: actions speak louder than words. *Development* 140, 1385–1395. doi: 10.1242/dev.084459
- Nishihara, H., Smit, A. F. A., and Okada, N. (2006). Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.* 16, 864–874. doi: 10.1101/gr.5255506
- Ohms, S., Lee, S. H., and Rangasamy, D. (2014). LINE-1 retrotransposons and let-7 miRNA: partners in the pathogenesis of cancer? *Front Genet.* 5:338. doi: 10.3389/fgene.2014.00338
- Ohms, S., and Rangasamy, D. (2014). Silencing of LINE-1 retrotransposons contributes to variation in small noncoding RNA expression in human cancer cells. *Oncotarget* 5, 4103–4117. doi: 10.18632/oncotarget.1822
- Ohno, S. (1972). So much “junk” DNA in our genome. *Brookhaven Symp. Biol.* 23, 366–370.
- Oliver, K. R., and Greene, W. K. (2011). Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. *Mob. DNA* 2:8. doi: 10.1186/1759-8753-2-8
- Orgel, L. E., and Crick, F. H. (1980). Selfish DNA: the ultimate parasite. *Nature* 284, 604–607. doi: 10.1038/284604a0
- Oricchio, E., Beraldi, R., Sciamanna, I., Tolstonog, G. V., Schumann, G. G., and Spadafora, C. (2007). Distinct roles for LINE-1 and Herv-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene* 26, 4226–4233. doi: 10.1038/sj.onc.1210214
- Ostertag, E. M., Goodier, J. L., Zhang, Y., and Kazazian, H. H. Jr. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* 73, 1444–1451. doi: 10.1086/380207
- Patnala, R., Lee, S. H., Dahlstrom, J. E., Ohms, S., Chen, L., Dheen, S. T., et al. (2013). Inhibition of LINE-1 retrotransposon-encoded reverse transcriptase modulates the expression of cell differentiation genes in breast cancer cells. *Breast. Cancer Res. Treat.* 143, 239–253. doi: 10.1007/s10549-013-2812-7
- Pelechano, V., and Steinmetz, L. M. (2013). Gene regulation by antisense transcription. *Nat. Rev.* 14, 880–893. doi: 10.1038/nrg3594
- Perreault, J., Noël, J. F., Brière, F., Cousineau, B., Lucier, J. F., et al. (2005). Retropseudogenes derived from the human Ro/SS-A autoantigen-associated hY RNAs. *Nucleic Acids Res.* 33, 2032–2041. doi: 10.1093/nar/gki504
- Piatek, M. J., and Werner, A. (2014). Endogenous siRNAs: regulators of internal affairs. *Biochem. Soc. Trans.* 42, 1174–1179. doi: 10.1042/BST20140068
- Pink, R. C., Wicks, K., Caley, D. P., Punch, E. K., Jacobs, L., and Carter, D. R. F. (2011). Pseudogenes: Pseudo-functional or key regulators in health and disease? *RNA* 17, 792–798. doi: 10.1261/rna.2658311
- Pittoggi, C., Sciamanna, I., Mattei, E., Beraldi, R., Lobascio, A. M., Mai, A., et al. (2003). Role of endogenous reverse transcriptase in murine early embryo development. *Mol. Reprod. Dev.* 66, 225–236. doi: 10.1002/mrd.10349
- Rebollo, R., Romanish, M. T., and Mager, D. L. (2012). Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* 46, 21–42. doi: 10.1146/annurev-genet-110711-155621
- Rodić, N., and Burns, K. H. (2013). Long interspersed element-1 (LINE-1): passenger or driver in human neoplasms? *PLoS Genet.* 9:e1003402. doi: 10.1371/journal.pgen.1003402
- Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., et al. (2014). Long Interspersed Element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* 184, 1280–1286. doi: 10.1016/j.ajpath.2014.01.007
- Rodić, N., Steranka, J. P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., et al. (2015). Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat. Med.* 21, 1060–1064. doi: 10.1038/nm.3919
- Sciamanna, I., Gualtieri, A., Cossetti, C., Osimo, E. F., Ferracin, M., Macchia, G., et al. (2013). A tumor-promoting mechanism mediated by retrotransposon-encoded reverse transcriptase is active in human transformed cell lines. *Oncotarget* 4, 2271–2287. doi: 10.18632/oncotarget.1403
- Sciamanna, I., Gualtieri, A., Piazza, P. V., and Spadafora, C. (2014). Regulatory roles of LINE-1-encoded reverse transcriptase in cancer onset and progression. *Oncotarget* 5, 8039–8051. doi: 10.18632/oncotarget.2504
- Sciamanna, I., Landriscina, M., Pittoggi, C., Quirino, M., Marelli, C., Beraldi, R., et al. (2005). Inhibition of endogenous reverse transcriptase antagonizes human tumor growth. *Oncogene* 24, 3923–3931. doi: 10.1038/sj.onc.1208562
- Sciamanna, I., Vitullo, P., Curatolo, A., and Spadafora, C. (2011). A reverse transcriptase-dependent mechanism is essential for murine preimplantation development. *Genes* 2, 360–373. doi: 10.3390/genes2020360
- Shukla, R., Upton, K. R., Muñoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111. doi: 10.1016/j.cell.2013.02.032
- Sinibaldi-Vallebona, P., Matteucci, C., and Spadafora, C. (2011). Retrotransposon-encoded reverse transcriptase in the genesis, progression and cellular plasticity of human cancer. *Cancers* 3, 1141–1157. doi: 10.3390/cancers3011141
- Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285. doi: 10.1038/nrg2072

- Solyom, S., Ewing, A. D., Rahrmann, E. P., Doucet, T., Nelson, H. H., Burns, M. B., et al. (2012). Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* 22, 2328–2338. doi: 10.1101/gr.145235.112
- Sottoriva, A., Kang, H., Ma, Z., Graham, T. A., Salomon, M. P., Zhao, J., et al. (2015). A Big Bang model of human colorectal tumor growth. *Nat. Genet.* 47, 209–216. doi: 10.1038/ng.3214
- Spadafora, C. (2008). Sperm-mediated 'reverse' gene transfer: a role of reverse transcriptase in the generation of new genetic information. *Hum. Reprod.* 23, 735–740. doi: 10.1093/humrep/dem425
- Spadafora, C. (2015). A LINE-1–encoded reverse transcriptase–dependent regulatory mechanism is active in embryogenesis and tumorigenesis. *Ann. N.Y. Acad. Sci.* 1341, 164–171. doi: 10.1111/nyas.12637
- Stella, G. M., Senetta, R., Cassenti, A., Ronco, M., and Cassoni, P. (2012). Cancers of unknown primary origin: current perspectives and future therapeutic strategies. *J. Translat. Med.* 10:12. doi: 10.1186/1479-5876-10-12
- Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., et al. (2010). MicroRNA function is globally suppressed in mouse oocytes and early embryos. *Curr. Biol.* 20, 271–277. doi: 10.1016/j.cub.2009.12.044
- Tang, X. H., and Gudas, L. J. (2011). Retinoids, retinoic acid receptors, and cancer. *Annu. Rev. Pathol.* 6, 345–364. doi: 10.1146/annurev-pathol-011110-130303
- Temin, H. M. (1971). Guest editorial. The provirus hypothesis: speculations on the significance of RNA-directed DNA synthesis for normal development and for carcinogenesis. *J. Natl. Cancer Inst.* 46, III–VI.
- Temin, H. M., and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* 226, 1211–1213. doi: 10.1038/2261211a0
- Terasaki, N., Goodier, J. L., Cheung, L. E., Wang, Y. J., Kajikawa, M., Kazazian, H. H. Jr., et al. (2013). *In vitro* screening for compounds that enhance human L1 mobilization. *PLoS ONE* 8:e74629. doi: 10.1371/journal.pone.0074629
- The ENCODE and Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- Ullu, E., and Weiner, A. M. (1984). Human genes and pseudogenes for the 7SL RNA component of signal recognition particle. *EMBO J.* 3, 3303–3310
- van den Hurk, J. A. J. M., Meij, I. C., Seleme, M. C., Kano, H., Nikopoulos, K., Hoefslot, L. H., et al. (2007). L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* 16, 1587–1592. doi: 10.1093/hmg/ddm108
- Vitullo, P., Sciamanna, I., Baiocchi, M., Sinibaldi-Vallebona, P., and Spadafora, C. (2012). LINE-1 retrotransposon copies are amplified during murine early embryo development. *Mol. Reprod. Dev.* 79, 118–127. doi: 10.1002/mrd.22003
- Wallace, M. R., Andersen, L. B., Saulino, A. M., Gregory, P. E., Glover, T. W., and Collins, F. S. (1991). A *de novo* Alu insertion results in neurofibromatosis type 1. *Nature* 353, 864–866. doi: 10.1038/353864a0
- Watanabe, T., Tomizawa, S., Mitsuya, K., Totoki, Y., Yamamoto, Y., et al. (2011). Role for piRNAs and noncoding RNA in *de novo* DNA methylation of the imprinted mouse Rasgrf1 locus. *Science* 332, 848–852. doi: 10.1126/science.1203919
- Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., et al. (2001). Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 21, 1429–1439. doi: 10.1128/MCB.21.4.1429-1439.2001
- Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., et al. (2005). Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 3:e7. doi: 10.1371/journal.pbio.0030007
- Yang, N., and Kazazian, H. H. Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* 13, 763–771. doi: 10.1038/nsmb1141
- Zhang, H., Chen, Z., Wang, X., Huang, Z., He, Z., and Chen, Y. (2013). Long non-coding RNA: a new player in cancer. *J. Hemat. Oncol.* 6:37. doi:10.1186/1756-8722-6-37

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 © 2016 Sciamanna, De Luca and Spadafora. 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) or licensor 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.



Crossing the LINE Toward Genomic Instability: LINE-1 Retrotransposition in Cancer

Jacqueline R. Kemp and Michelle S. Longworth *

Department of Cellular and Molecular Medicine, Lerner Research Institute of Cleveland Clinic, Cleveland, OH, USA

OPEN ACCESS

Edited by:

Tammy A. Morrish,
The University of Toledo, USA

Reviewed by:

Revati Wani,
Pfizer Inc., USA
Sheila Lutz,
Wadsworth Center, New York State
Department of Health, USA

*Correspondence:

Michelle S. Longworth
longwom@ccf.org

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 13 August 2015

Accepted: 27 November 2015

Published: 16 December 2015

Citation:

Kemp JR and Longworth MS (2015)
Crossing the LINE Toward Genomic
Instability: LINE-1 Retrotransposition
in Cancer. *Front. Chem.* 3:68.
doi: 10.3389/fchem.2015.00068

Retrotransposons are repetitive DNA sequences that are positioned throughout the human genome. Retrotransposons are capable of copying themselves and mobilizing new copies to novel genomic locations in a process called retrotransposition. While most retrotransposon sequences in the human genome are incomplete and incapable of mobilization, the LINE-1 retrotransposon, which comprises ~17% of the human genome, remains active. The disruption of cellular mechanisms that suppress retrotransposon activity is linked to the generation of aneuploidy, a potential driver of tumor development. When retrotransposons insert into a novel genomic region, they have the potential to disrupt the coding sequence of endogenous genes and alter gene expression, which can lead to deleterious consequences for the organism. Additionally, increased LINE-1 copy numbers provide more chances for recombination events to occur between retrotransposons, which can lead to chromosomal breaks and rearrangements. LINE-1 activity is increased in various cancer cell lines and in patient tissues resected from primary tumors. LINE-1 activity also correlates with increased cancer metastasis. This review aims to give a brief overview of the connections between LINE-1 retrotransposition and the loss of genome stability. We will also discuss the mechanisms that repress retrotransposition in human cells and their links to cancer.

Keywords: retrotransposons, LINE-1, genomic instability, retrotransposition, cancer

THE LINE-1 RETROTRANSPOSON IS AN ACTIVE MOBILE ELEMENT

Retrotransposons, a class of transposable elements (TE), are highly repetitive DNA sequences positioned throughout the human genome. These structural elements make use of an RNA-mediated transposition process, allowing them to move from one location in the genome to another, while the original copy remains in its original locus. The RNA-based retrotransposons are classified into the autonomous long terminal repeat (LTR) and the non-LTR containing retrotransposons. LTR containing retrotransposons, as their name implies, possess LTRs ranging from 100 bp to over 5 kb in size and are endogenous retroviruses. Long interspersed nuclear elements (LINEs), comprising 20% of the human genome are a type of non-LTR retrotransposon. Non-autonomous retrotransposons are a third class of retrotransposons, of which the short interspersed nuclear elements (SINEs) comprise ~13% of the human genome (Lander et al., 2001).

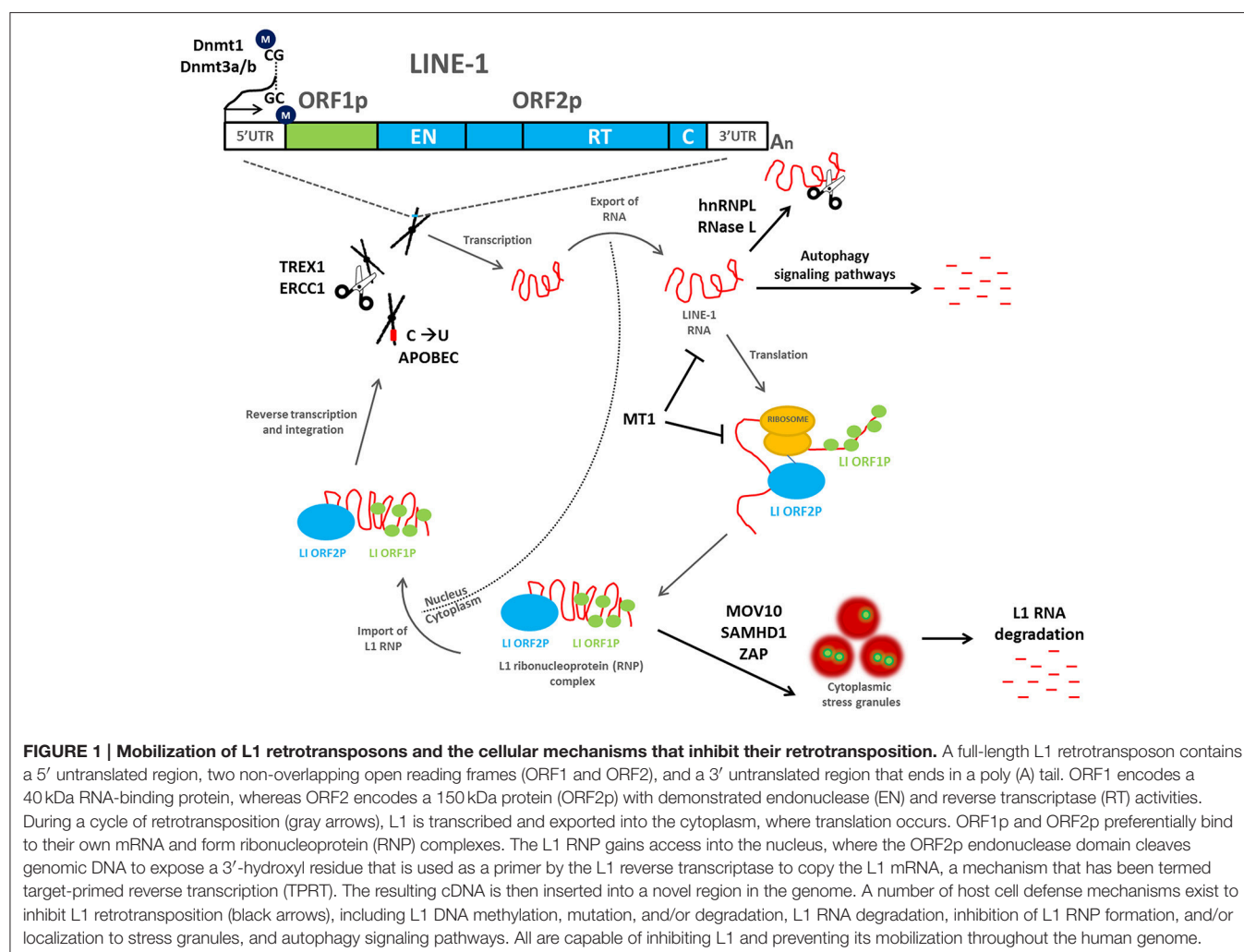
The human genome contains millions of copies of retrotransposons; however, only a single non-LTR retrotransposon family, the LINE-1 (L1) family, remains the primary source of

retrotransposition. The activity of the L1 retrotransposon has persisted over time within the human genome and its de-repression is associated with genomic instability and tumor development (Gasior et al., 2006; Lee et al., 2012). Over 100,000 L1 sequences exist in the human genome; however, most are rendered inactive by point mutations, rearrangements, or truncations (Brouha et al., 2003). It was originally estimated that the average human diploid genome contains ~80–100 active L1s that are capable of undergoing retrotransposition (Sassaman et al., 1997). Of those which are active, six were classified as “hot” L1s responsible for the bulk of L1 retrotransposition within the human genome (Brouha et al., 2003). More recently, however, three independent studies demonstrated that the occurrence of new L1 insertions is more prevalent than previously thought. Additionally, a number of the newly inserted “hot” L1s were found to be extremely polymorphic and specific to a few individuals, suggesting that L1 retrotransposition may contribute to the propensity for one individual to develop disease over another (Beck et al., 2010; Huang et al., 2010; Iskow et al., 2010).

A full-length L1 retrotransposon is ~6 kB in size and contains a 5′ untranslated region, two non-overlapping open reading

frames (ORF1 and ORF2), and a 3′ untranslated region that ends in a poly (A) tail (Swergold, 1990; Becker et al., 1993). ORF1 encodes a 40 kDa RNA-binding protein (Mathias et al., 1991), whereas ORF2 encodes a 150 kDa protein (ORF2p) with demonstrated endonuclease and reverse transcriptase activities (Mathias et al., 1991; Feng et al., 1996; Piskareva et al., 2003). Interestingly, ORF2p also contains a conserved cysteine-rich domain recently shown to have a high non-specific affinity to RNA, which may contribute to the process of reverse transcription (Piskareva et al., 2013). Various mutants of ORF1p and ORF2p, have been created and used to demonstrate that the two proteins are necessary for retrotransposition in a cell culture based assay (Moran et al., 1996; Wei et al., 2001; Kulpa and Moran, 2005; Doucet et al., 2010).

The mobility of a L1 retrotransposon is completely dependent on transcription and translation of its encoded proteins and therefore includes both nuclear and cytoplasmic events essential for retrotransposon duplication (**Figure 1**). ORF1p and ORF2p preferentially bind to their own mRNA and form ribonucleoprotein (RNP) complexes (Leibold et al., 1990; Alisch et al., 2006; Dmitriev et al., 2007; Doucet et al., 2010). ORF1p



has been demonstrated to have nucleic acid chaperone activity that is essential for the retrotransposition process (Martin et al., 2005, 2008). The L1 RNP gains access into the nucleus, where the ORF2p endonuclease domain cleaves genomic DNA to expose a 3'-hydroxyl residue that is used as a primer by the L1 reverse transcriptase to copy the L1 mRNA, a mechanism that has been termed target-primed reverse transcription (TPRT). The resulting cDNA is then inserted into a novel region in the genome (Cost et al., 2002). A nuclear localization signal has been identified in ORF2p (Goodier et al., 2004); however, it is unclear whether the L1 RNP is capable of crossing an intact nuclear membrane or whether it gains access following nuclear envelope breakdown (Kubo et al., 2006).

POTENTIAL L1-MEDIATED MECHANISMS OF TUMOR DEVELOPMENT

Many reports have demonstrated that retrotransposons can significantly impact the structure of the human genome. Retrotransposons have adverse effects on genome stability since multiple copies of the same sequence can hinder precise chromosomal pairing during mitosis and meiosis, resulting in DNA double-stranded breaks, more homologous recombination, chromosome duplication, and increased potential for inefficient repair of recombination events (Belgnaoui et al., 2006; Farkash et al., 2006; Gasior et al., 2006). A recent study identified LINE-LINE-mediated non-allelic homologous recombination as an important mechanism of structural rearrangement, contributing to genomic variability and instability (Startek et al., 2015).

L1 retrotransposition events in the human genome have been deemed responsible for ~97 disease-producing insertions (reviewed in Hancks and Kazazian, 2012). Specifically, direct insertional mutagenesis caused by L1 retrotransposition can result in disruption of coding sequence, disruption of splicing, and/or deregulation of gene expression. Symer and colleagues identified L1 element inversions, extra nucleotide insertions, exon deletions, a chromosomal inversion, and flanking sequence comobilization in the retrotransposon target site in human tissue culture cells (Symer et al., 2002). Studies have also shown that L1 acts as more than just an insertional mutagen, but also that its retrotransposition activity can result in large genomic deletions (Gilbert et al., 2002).

L1 retrotransposons exhibit a *cis*-preference, in which the L1 proteins preferentially use their own L1 RNA as the transcript for reverse transcription and integration (Wei et al., 2001; Kulpa and Moran, 2006). However, L1 proteins can also work in *trans* to promote mobilization of other RNAs, thus increasing their potential for causing genomic instability. Non-autonomous elements including SINEs (Dewannieux et al., 2003) and SVAs (Raiz et al., 2012), as well as small nuclear RNAs (e.g., U6 snRNA; Buzdin et al., 2002; Gilbert et al., 2005; Garcia-Perez et al., 2007), small nucleolar RNAs (e.g., U3 snoRNA; Weber, 2006), and messenger RNAs (Esnault et al., 2000; Wei et al., 2001) are all capable of being *trans*-mobilized via L1. In all of these cases,

retrotransposition of mRNAs results in processed pseudogenes that bear L1 structural hallmarks. These *trans*-mobilization events utilize the ORF1p and/or ORF2p to insert into the human genome and do not involve sequence specificity. Once these pseudogenes are inserted back into the genome, they usually lack introns and promoters, but contain a poly (A) 3' end and target-site duplications of varying length (Vanin, 1985; Weiner et al., 1986; Esnault et al., 2000). Interestingly, siRNAs have been shown to be expressed from pseudogenes in mouse oocytes, suggesting a potential way in which they might influence gene regulation (Tam et al., 2008). Therefore, generation of processed pseudogenes is a direct product of endogenous retrotransposition activity in the human genome that can contribute to genomic diversity and instability.

Integration of L1 in or near oncogenes or tumor suppressor genes can contribute to tumor development (Morse et al., 1988; Miki et al., 1992; Iskow et al., 2010) and progression of life-threatening cancers, including lung, colon, and breast cancer in humans (Lee et al., 2012; Criscione et al., 2014). For example, disruption of the APC gene by a somatic insertion of L1 was shown to be present in colon cancer and associated with development of colorectal tumors (Miki et al., 1992). The APC gene encodes a tumor suppressor involved in maintaining chromosomal stability during mitosis (Fodde et al., 2001b). In *Apc* deficient mouse cells, structural rearrangements, resulting from chromosomal breakage and recombination are apparent (Fodde et al., 2001a). Further, cells are defective in chromosome segregation when they carry a truncated form of *Apc* (Kaplan et al., 2001). Other tumor suppressor genes found to be disrupted by tumor-specific L1 insertions include Mutated in Colorectal Cancers (MCC) and Suppression of Tumorigenicity 18 (ST18; Shukla et al., 2013). Furthermore, since L1 machinery acts to *trans*-mobilize other RNAs, those insertions can also impact expression of genes. *Alu*, a type of SINE present in higher copy numbers than L1, can be *trans*-mobilized, leading to cancer-associated gene insertions. Sites of *Alu* insertions include the APC locus and this was associated with Desmoids tumors (Halling et al., 1999), the tumor suppressor NF-1 (neurofibromatosis type I; Wallace et al., 1991), and the BRCA1 and BRCA2 breast/ovarian cancer related genes (Miki et al., 1996; Teugels et al., 2005). SVA elements can also be mobilized by the L1 retrotransposition machinery, leading to disease (Ostertag et al., 2003). In one study, mobilization of SVA resulted in deletion of the *HLA-A* gene in three Japanese families; a number of individuals in these families were afflicted with leukemia (Takasu et al., 2007).

Telomerase reactivation, as a means to maintain telomeres, occurs in the early stages of carcinogenesis to promote cancer cell immortalization (Counter et al., 1992; Kim et al., 1994). Transcriptional regulation of hTERT, the catalytic subunit of telomerase, is a major mechanism for telomerase activation in the cancer setting. In a recent study, L1 was shown to contribute to tumor pathogenicity by inducing hTERT and helping to maintain telomeres in telomerase-positive tumor cells. Depletion of L1 resulted in reduced telomere length, suggesting that L1 is a reasonable target in the treatment of telomerase-positive cancer (Aschacher et al., 2015).

L1 EXPRESSION IN CANCERS

Given that L1 retrotransposition can lead to genomic instability and genetic heterogeneity is a common feature in tumor initiating cells, it is not surprising that expression of the L1-encoded ORF1p is reported to be a hallmark of many human cancers, with almost half (47%) of the human neoplasms examined being immunoreactive for L1 (Rodic et al., 2014). L1 positive neoplasms included invasive breast carcinomas (97% L1 positive), high-grade ovarian carcinomas (91.5% L1 positive), and pancreatic ductal adenocarcinomas (PDACs; 89% L1 positive). Carcinomas originating in the endometrium, biliary tract, esophagus, bladder, head and neck, lung, and colon were also frequently L1 immunoreactive (22.6–76.7% L1 positive; Rodic et al., 2014). In a separate study, increased ORF1p expression and novel L1 insertions in PDAC were observed in matched primary and metastatic tissues. However, the overall results showed discordant rates of retrotransposition, suggesting that while increased L1 retrotransposition may not be a direct cause of metastatic PDAC, it may contribute to gene dysregulation leading to metastasis (Rodic et al., 2015). Furthermore, activation of L1 increases the risk of epithelial-mesenchymal transition and metastasis in epithelial cancer (reviewed in Rangasamy et al., 2015) and promotes proliferation and invasion of LoVo colorectal cancer cells (Li et al., 2014) and MDA-MB-231 breast cancer cells (Yang et al., 2013). ORF1p and ORF2p levels are upregulated in breast cancers compared to normal tissues. Cytoplasmic levels of ORF1p and ORF2p are elevated in DCIS breast cancers compared to highly invasive cancers. Conversely, nuclear levels of ORF1p and ORF2p were found to be higher in invasive breast cancers and correlated with increased lymph node metastasis and poor patient survival (Harris et al., 2010; Chen et al., 2012). Furthermore, inhibition of the L1-encoded reverse transcriptase in breast cancer cells was demonstrated to reduce the rate of proliferation and promote cellular differentiation (Patnala et al., 2014). Finally, L1 activity and expression was elevated in rat chloroleukemia cells, suggesting that mobilization of this retrotransposon may contribute to the genomic instability observed in this model of blood cancer (Kirilyuk et al., 2008).

Hypomethylation of L1 DNA has been observed in various cancers and is associated with an increase in transcriptional activation and expression of L1 (Alves et al., 1996; Asch et al., 1996; Kitkumthorn et al., 2012; Murata et al., 2013; Criscione et al., 2014; Park et al., 2014). L1 hypomethylation can occur early in tumorigenesis and is associated with bladder (Patchsung et al., 2012; Salas et al., 2014), gastric (Shigaki et al., 2013; Baba et al., 2014a), colon (Ogino et al., 2008; Antelo et al., 2012; Murata et al., 2013), lung (Saito et al., 2010), and breast cancers (Park et al., 2014). L1 hypomethylation is associated with poor prognosis of lung adenocarcinoma (Ikeda et al., 2013), hepatocellular carcinoma via activation of c-Met (Zhu et al., 2014), esophageal squamous cell carcinoma (ESCC; Iwagami et al., 2013), and with inferior survival in colorectal carcinomas with high microsatellite instability (Inamura et al., 2014). Additionally, L1 hypomethylation in ESCC was shown to be significantly associated with lymph node metastasis, frequency of p53 mutation, and chromosomal instability (Kawano et al., 2014). In a separate study, L1 hypomethylation in ESCC patient

samples was associated with an increase in CDK6 expression (Baba et al., 2014b). This may contribute to the aggressiveness of tumors since CDK6 is known to promote tumor progression by stimulating proliferation and angiogenesis (Kollmann et al., 2013). Finally, hypomethylation of L1 in colorectal cancer can lead to activation of oncogenes important in metastasis, including MET, RAB1P, and CHRM3 (Hur et al., 2014). It was observed that specific L1 sequences residing within the intronic regions of these proto-oncogenes were hypomethylated and reduced methylation of specific L1 elements within the MET gene correlated with an induction of MET expression (Hur et al., 2014). However, since methylation levels of repetitive L1 elements often tightly correlate with global DNA methylation levels, it is difficult to conclude that L1 hypomethylation directly results in the increased genomic instability found in tumors.

MECHANISMS THAT INHIBIT L1 RETROTRANSPOSITION ARE OFTEN DEREGULATED IN CANCER

As the uncontrolled movement of retrotransposons throughout the genome can have deleterious consequences for genome stability and health in general, a number of defense mechanisms exist in human cells to repress their movement. These mechanisms exist at the DNA, RNA, and protein levels to inhibit L1 and retrotransposition (Figure 1).

DNA methylation status is a major determinant of gene expression changes within the human genome and is involved in various biological processes including cancer (Liu et al., 2003). As discussed above, hypomethylation of L1 DNA is associated with an increase in L1 expression. Conversely, methylation of L1 within the CpG rich 5'-UTR represses its ability to be activated and transcribed, thereby minimizing the exposure of genomic DNA to L1-associated damage (Hata and Sakaki, 1997; Weisenberger et al., 2005; Barchitta et al., 2014). DNA methylation, therefore, is a key mechanism for L1 silencing. It has been shown in mouse embryonic stem cells that methylation of the L1 promoter is maintained by DNA methyltransferases, including Dmmt1 and Dmmt3a and/or -3b (Woodcock et al., 1998; Liang et al., 2002).

Other epigenetic mechanisms have been reported to be involved in regulating L1 expression. One study showed that reporter genes introduced into human embryonic carcinoma-derived cell lines by engineered L1 retrotransposons were rapidly silenced during or shortly after their integration (Garcia-Perez et al., 2010). Treatment of the cells with histone deacetylase inhibitors reversed the silencing and ChIP experiments demonstrated that a change in the chromatin status at the L1 integration site correlated with reactivation of the reporter gene (Garcia-Perez et al., 2010). Other studies involving chromatin structure averaged global histone modifications and found that histone H3 lysine nine methylation is enriched at human retrotransposons, suggesting that histone methylation may play a role in repressing recombination of these retrotransposons (Kondo and Issa, 2003; Martens et al., 2005; Goodier and Kazazian, 2008). Low levels of the silencing histone modification H3K27me3 at L1 loci in conjunction with

L1 hypomethylation has been shown to support an active role in rectal cancer prognosis and a poor clinical outcome (Benard et al., 2013). Conversely, high levels of the activating histone modification H3K9Ac at L1 loci were associated with poor patient survival. This indicates that L1 methylation and histone modifications work closely together in determining gene expression and tumor progression (Benard et al., 2013).

Global chromatin organization is also involved in repression of *Drosophila melanogaster* retrotransposons. Studies in the fly have identified a role for the chromatin organizing complex, Condensin II in repressing retrotransposition in somatic cells and tissues. The Condensin II subunit, dCAP-D3, promotes silencing of retrotransposon-containing loci by maintaining boundaries of repressive histone modifications to repress retrotransposon transcription and ultimately inhibit retrotransposition (Schuster et al., 2013). Furthermore, decreased dCAP-D3 expression impacts chromatin structure, resulting in DNA double strand breaks within the retrotransposon sequence, an increase in homologous pairing, and an increase in global retrotransposon copy number. While global chromatin regulators have yet to be implemented in L1 repression, CAP-D3, and Condensin II are conserved and further studies are necessary to determine whether they also inhibit retrotransposition in human cells.

Epigenetic modification, however, is not the only mechanism employed by cells to inhibit retrotransposition. Exciting new evidence from multiple labs suggests that a host of cellular proteins employ distinct mechanisms to accomplish the inhibition.

One mechanism includes targeting the L1 RNA intermediate to prohibit insertion of L1 into the human genome. The ribonucleoprotein hnRNPL, which plays multiple roles in RNA metabolism, has been shown to directly interact with L1 RNA to negatively regulate retrotransposition. hnRNPL does so by decreasing the steady-state levels of the L1 RNA (Peddigari et al., 2013). Downregulation of L1 mRNA and subsequently, reduced expression of the ORF1p and ORF2p by RNase L was also shown to restrict L1 mobilization; whereas, siRNA-mediated knockdown of endogenous RNase L lead to a significant increase in L1 retrotransposition events in a human ovarian cancer cell line (Zhang et al., 2014). Similarly, the melatonin receptor 1 (MT1) inhibits retrotransposition through downregulation of L1 mRNA and ORF1p. Researchers showed that antagonists directed against MT1 abolished this effect in a dose-dependent manner (deHaro et al., 2014). Furthermore, melatonin-rich blood suppressed endogenous L1 RNA during *in situ* perfusion of tissue-isolated xenografts of human pancreatic cancer (deHaro et al., 2014).

Innate immune defenses can also inhibit retrotransposition of L1. Guo and colleagues demonstrated that autophagy degrades the L1 RNA intermediate, preventing new insertions into the genome and promoting genome stability. Degradation of retrotransposon RNA was facilitated by receptors involved in activating autophagy signaling pathways, NDP52 and p62. Interestingly, this study also showed that mice lacking *Atg6/Beclin1*, a gene critical for the formation of autophagosomes, accumulate retrotransposon RNA and new genomic insertions of L1 (Guo et al., 2014).

L1 RNP formation and safe delivery of the RNP to genomic DNA is essential for TPRT to occur; therefore, targeting the RNP for degradation is a useful mechanism to inhibit this process. The RNA helicase, MOV10 directly associates with the L1 RNP (Goodier et al., 2012) and similar to SAMHD1 (Zhao et al., 2013) inhibits L1 retrotransposition by promoting stress granule formation (Arjan-Odedra et al., 2012; Li et al., 2013); stress granules are ribonucleoprotein cytosolic foci that appear under cellular stress and often act to promote mRNA degradation (Kedersha et al., 2005). Further, the L1 ORF1p was shown in a separate study to localize in stress granules with components of RISC, suggesting a mechanism for controlling retrotransposition and the associated genomic damage (Goodier et al., 2007). More recently, the zinc-finger antiviral protein, ZAP was shown to inhibit L1 retrotransposition by binding to the L1 RNP and inhibiting accumulation of L1 RNA (Goodier et al., 2015; Moldovan and Moran, 2015). ZAP colocalizes with the RNP in cytoplasmic stress granules and interacts with a number of novel proteins, including MOV10 (Goodier et al., 2015; Moldovan and Moran, 2015).

Another mechanism to inhibit retrotransposition involves targeting the single-strand DNA that arises during the process of L1 integration, to repress its mobilization. These cellular proteins can directly promote degradation of L1, thereby inhibiting retrotransposition. For example, the APOBEC3 (A3) family of cytidine deaminases functions to inhibit L1 retrotransposition by deaminating the transiently exposed cDNA, creating C-to-U conversions (Richardson et al., 2014). This may then target the mutated retrotransposon DNA for degradation through endonuclease activity. Additionally, the endonucleases TREX1 (Stetson et al., 2008) and ERCC1/XPF (Gasior et al., 2008) can physically cleave the reverse-transcribed cDNA of L1, thereby inhibiting retrotransposition.

CONCLUSIONS

Undeniably, L1 retrotransposons are an interesting and important component of the human genome. The activity of L1 retrotransposons can generate a wide array of genomic mutations and rearrangements, with potentially serious consequences for the stability of the genome. L1s are frequently hypomethylated and expressed in human cancers and their increased activity correlates with tumor progression and metastasis. Additionally, L1-insertion-mediated interference with normal RNA processing and expression also contributes to cancer development. Further studies on L1 retrotransposition, their effects on local and global genome organization, and the identification of novel mechanisms which repress retrotransposition to prevent tumor development will broaden our understanding of the impact of retrotransposons on genetic diversity and human health.

FUNDING

This work was supported by a National Institutes of Health research grant R01GM102400 to ML. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

REFERENCES

- Alisch, R. S., Garcia-Perez, J. L., Muotri, A. R., Gage, F. H., and Moran, J. V. (2006). Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev.* 20, 210–224. doi: 10.1101/gad.1380406
- Alves, G., Tatso, A., and Fanning, T. (1996). Differential methylation of human LINE-1 retrotransposons in malignant cells. *Gene* 176, 39–44. doi: 10.1016/0378-1119(96)00205-3
- Antelo, M., Balaguer, F., Shia, J., Shen, Y., Hur, K., Moreira, L., et al. (2012). A high degree of LINE-1 hypomethylation is a unique feature of early-onset colorectal cancer. *PLoS ONE* 7:e45357. doi: 10.1371/journal.pone.0045357
- Arjan-Odedra, S., Swanson, C. M., Sherer, N. M., Wolinsky, S. M., and Malim, M. H. (2012). Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. *Retrovirology* 9:53. doi: 10.1186/1742-4690-9-53
- Asch, H. L., Eliacin, E., Fanning, T. G., Connolly, J. L., Bratthauer, G., and Asch, B. B. (1996). Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol. Res.* 8, 239–247.
- Aschacher, T., Wolf, B., Enzmann, F., Kienzl, P., Messner, B., Sampl, S., et al. (2015). LINE-1 induces hTERT and ensures telomere maintenance in tumour cell lines. *Oncogene*. doi: 10.1038/onc.2015.65. [Epub ahead of print].
- Baba, Y., Murata, A., Watanabe, M., and Baba, H. (2014a). Clinical implications of the LINE-1 methylation levels in patients with gastrointestinal cancer. *Surg. Today* 44, 1807–1816. doi: 10.1007/s00595-013-0763-6
- Baba, Y., Watanabe, M., Murata, A., Shigaki, H., Miyake, K., Ishimoto, T., et al. (2014b). LINE-1 hypomethylation, DNA copy number alterations, and CDK6 amplification in esophageal squamous cell carcinoma. *Clin. Cancer Res.* 20, 1114–1124. doi: 10.1158/1078-0432.CCR-13-1645
- Barchitta, M., Quattrocchi, A., Maugeri, A., Vinciguerra, M., and Agodi, A. (2014). LINE-1 hypomethylation in blood and tissue samples as an epigenetic marker for cancer risk: a systematic review and meta-analysis. *PLoS ONE* 9:e109478. doi: 10.1371/journal.pone.0109478
- Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., et al. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159–1170. doi: 10.1016/j.cell.2010.05.021
- Becker, K. G., Swergold, G. D., Ozato, K., and Thayer, R. E. (1993). Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. *Hum. Mol. Genet.* 2, 1697–1702. doi: 10.1093/hmg/2.10.1697
- Belgnaoui, S. M., Gosden, R. G., Semmes, O. J., and Haoudi, A. (2006). Human LINE-1 retrotransposon induces DNA damage and apoptosis in cancer cells. *Cancer Cell Int.* 6:13. doi: 10.1186/1475-2867-6-13
- Benard, A., van de Velde, C. J., Lessard, L., Putter, H., Takeshima, L., Kuppen, P. J., et al. (2013). Epigenetic status of LINE-1 predicts clinical outcome in early-stage rectal cancer. *Br. J. Cancer* 109, 3073–3083. doi: 10.1038/bjc.2013.654
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285. doi: 10.1073/pnas.0831042100
- Buzdin, A., Ustyugova, S., Gogvadze, E., Vinogradova, T., Lebedev, Y., and Sverdlov, E. (2002). A new family of chimeric retrotranscripts formed by a full copy of U6 small nuclear RNA fused to the 3' terminus of I1. *Genomics* 80, 402–406. doi: 10.1006/geno.2002.6843
- Chen, L., Dahlstrom, J. E., Chandra, A., Board, P., and Rangasamy, D. (2012). Prognostic value of LINE-1 retrotransposon expression and its subcellular localization in breast cancer. *Breast Cancer Res. Treat.* 136, 129–142. doi: 10.1007/s10549-012-2246-7
- Cost, G. J., Feng, Q., Jacquier, A., and Boeke, J. D. (2002). Human L1 element target-primed reverse transcription *in vitro*. *EMBO J.* 21, 5899–5910. doi: 10.1093/emboj/cdf592
- Counter, C. M., Avilion, A. A., LeFeuvre, C. E., Stewart, N. G., Greider, C. W., Harley, C. B., et al. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *EMBO J.* 11, 1921–1929.
- Criscione, S. W., Zhang, Y., Thompson, W., Sedivy, J. M., and Neretti, N. (2014). Transcriptional landscape of repetitive elements in normal and cancer human cells. *BMC Genomics* 15:583. doi: 10.1186/1471-2164-15-583
- deHaro, D., Kines, K. J., Sokolowski, M., Dauchy, R. T., Strevia, V. A., Hill, S. M., et al. (2014). Regulation of L1 expression and retrotransposition by melatonin and its receptor: implications for cancer risk associated with light exposure at night. *Nucleic Acids Res.* 42, 7694–7707. doi: 10.1093/nar/gku503
- Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35, 41–48. doi: 10.1038/ng1223
- Dmitriev, S. E., Andreev, D. E., Terenin, I. M., Olovnikov, I. A., Prassolov, V. S., Merrick, W. C., et al. (2007). Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol. Cell. Biol.* 27, 4685–4697. doi: 10.1128/MCB.02138-06
- Doucet, A. J., Hulme, A. E., Sahinovic, E., Kulpa, D. A., Moldovan, J. B., Kopera, H. C., et al. (2010). Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genetics* 6:e1001150. doi: 10.1371/journal.pgen.1001150
- Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24, 363–367. doi: 10.1038/74184
- Farkash, E. A., Kao, G. D., Horman, S. R., and Prak, E. T. (2006). Gamma radiation increases endonuclease-dependent L1 retrotransposition in a cultured cell assay. *Nucleic Acids Res.* 34, 1196–1204. doi: 10.1093/nar/gkj522
- Feng, Q., Moran, J. V., Kazazian, H. H. Jr., and Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905–916. doi: 10.1016/S0092-8674(00)81997-2
- Fodde, R., Kuipers, J., Rosenberg, C., Smits, R., Kielman, M., Gaspar, C., et al. (2001a). Mutations in the APC tumour suppressor gene cause chromosomal instability. *Nat. Cell Biol.* 3, 433–438. doi: 10.1038/35070129
- Fodde, R., Smits, R., and Clevers, H. (2001b). APC, signal transduction and genetic instability in colorectal cancer. *Nat. Rev. Cancer* 1, 55–67. doi: 10.1038/35094067
- Garcia-Perez, J. L., Doucet, A. J., Bucheton, A., Moran, J. V., and Gilbert, N. (2007). Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res.* 17, 602–611. doi: 10.1101/gr.5870107
- Garcia-Perez, J. L., Morell, M., Scheys, J. O., Kulpa, D. A., Morell, S., Carter, C. C., et al. (2010). Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* 466, 769–773. doi: 10.1038/nature09209
- Gasior, S. L., Roy-Engel, A. M., and Deininger, P. L. (2008). ERCC1/XPF limits L1 retrotransposition. *DNA Repair* 7, 983–989. doi: 10.1016/j.dnarep.2008.02.006
- Gasior, S. L., Wakeman, T. P., Xu, B., and Deininger, P. L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* 357, 1383–1393. doi: 10.1016/j.jmb.2006.01.089
- Gilbert, N., Lutz-Prigge, S., and Moran, J. V. (2002). Genomic deletions created upon LINE-1 retrotransposition. *Cell* 110, 315–325. doi: 10.1016/S0092-8674(02)00828-0
- Gilbert, N., Lutz, S., Morrish, T. A., and Moran, J. V. (2005). Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* 25, 7780–7795. doi: 10.1128/MCB.25.17.7780-7795.2005
- Goodier, J. L., Cheung, L. E., and Kazazian, H. H. Jr. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet.* 8:e1002941. doi: 10.1371/journal.pgen.1002941
- Goodier, J. L., and Kazazian, H. H. Jr. (2008). Retrotransposons revisited: the restraint and rehabilitation of parasites. *Cell* 135, 23–35. doi: 10.1016/j.cell.2008.09.022
- Goodier, J. L., Ostertag, E. M., Engleka, K. A., Seleme, M. C., and Kazazian, H. H. Jr. (2004). A potential role for the nucleolus in L1 retrotransposition. *Hum. Mol. Genet.* 13, 1041–1048. doi: 10.1093/hmg/ddh118
- Goodier, J. L., Pereira, G. C., Cheung, L. E., Rose, R. J., and Kazazian, H. H. Jr. (2015). The Broad-spectrum antiviral protein ZAP restricts human retrotransposition. *PLoS Genet.* 11:e1005252. doi: 10.1371/journal.pgen.1005252
- Goodier, J. L., Zhang, L., Vetter, M. R., and Kazazian, H. H. Jr. (2007). LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins,

- including components of RNA interference RNA-induced silencing complex. *Mol. Cell. Biol.* 27, 6469–6483. doi: 10.1128/MCB.00332-07
- Guo, H., Chitiprolu, M., Gagnon, D., Meng, L., Perez-Iratxeta, C., Lagace, D., et al. (2014). Autophagy supports genomic stability by degrading retrotransposon RNA. *Nat. Commun.* 5, 5276. doi: 10.1038/ncomms6276
- Halling, K. C., Lazzaro, C. R., Honchel, R., Bufile, J. A., Powell, S. M., Arndt, C. A., et al. (1999). Hereditary desmoid disease in a family with a germline Alu I repeat mutation of the APC gene. *Hum. Hered.* 49, 97–102. doi: 10.1159/000022852
- Hancks, D. C., and Kazazian, H. H. Jr. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.gde.2012.02.006
- Harris, C. R., Normart, R., Yang, Q., Stevenson, E., Haffty, B. G., Ganesan, S., et al. (2010). Association of nuclear localization of a long interspersed nuclear element-1 protein in breast tumors with poor prognostic outcomes. *Genes Cancer* 1, 115–124. doi: 10.1177/1947601909360812
- Hata, K., and Sakaki, Y. (1997). Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* 189, 227–234. doi: 10.1016/S0378-1119(96)00856-6
- Huang, C. R., Schneider, A. M., Lu, Y., Niranjana, T., Shen, P., Robinson, M. A., et al. (2010). Mobile interspersed repeats are major structural variants in the human genome. *Cell* 141, 1171–1182. doi: 10.1016/j.cell.2010.05.026
- Hur, K., Cejas, P., Feliu, J., Moreno-Rubio, J., Burgos, E., Boland, C. R., et al. (2014). Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. *Gut* 63, 635–646. doi: 10.1136/gutjnl-2012-304219
- Ikeda, K., Shiraishi, K., Eguchi, A., Shibata, H., Yoshimoto, K., Mori, T., et al. (2013). Long interspersed nucleotide element 1 hypomethylation is associated with poor prognosis of lung adenocarcinoma. *Ann. Thorac. Surg.* 96, 1790–1794. doi: 10.1016/j.athoracsur.2013.06.035
- Inamura, K., Yamauchi, M., Nishihara, R., Lochhead, P., Qian, Z. R., Kuchiba, A., et al. (2014). Tumor LINE-1 methylation level and microsatellite instability in relation to colorectal cancer prognosis. *J. Natl. Cancer Inst.* 106:dju195. doi: 10.1093/jnci/dju195
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020
- Iwagami, S., Baba, Y., Watanabe, M., Shigaki, H., Miyake, K., Ishimoto, T., et al. (2013). LINE-1 hypomethylation is associated with a poor prognosis among patients with curatively resected esophageal squamous cell carcinoma. *Ann. Surg.* 257, 449–455. doi: 10.1097/SLA.0b013e31826d8602
- Kaplan, K. B., Burds, A. A., Swedlow, J. R., Bekir, S. S., Sorger, P. K., and Näthke, I. S. (2001). A role for the Adenomatous polyposis coli protein in chromosome segregation. *Nat. Cell Biol.* 3, 429–432. doi: 10.1038/35070123
- Kawano, H., Saeki, H., Kitao, H., Tsuda, Y., Otsu, H., Ando, K., et al. (2014). Chromosomal instability associated with global DNA hypomethylation is associated with the initiation and progression of esophageal squamous cell carcinoma. *Ann. Surg. Oncol.* 21(Suppl. 4), S696–S702. doi: 10.1245/s10434-014-3818-z
- Kedersha, N., Stoecklin, G., Ayodele, M., Yacono, P., Lykke-Andersen, J., Fritzler, M. J., et al. (2005). Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol.* 169, 871–884. doi: 10.1083/jcb.200502088
- Kim, N. W., Piatyszek, M. A., Prowse, K. R., Harley, C. B., West, M. D., Ho, P. L., et al. (1994). Specific association of human telomerase activity with immortal cells and cancer. *Science* 266, 2011–2015. doi: 10.1126/science.7605428
- Kirilyuk, A., Tolstonog, G. V., Damert, A., Held, U., Hahn, S., Löwer, R., et al. (2008). Functional endogenous LINE-1 retrotransposons are expressed and mobilized in rat chloroleukemia cells. *Nucleic Acids Res.* 36, 648–665. doi: 10.1093/nar/gkm1045
- Kitkumthorn, N., Tuangsintanakul, T., Rattanatanyong, P., Tiwawech, D., and Mutirangura, A. (2012). LINE-1 methylation in the peripheral blood mononuclear cells of cancer patients. *Clin. Chim. Acta* 413, 869–874. doi: 10.1016/j.cca.2012.01.024
- Kollmann, K., Heller, G., Schnecklenleithner, C., Warsch, W., Scheicher, R., Ott, R. G., et al. (2013). A kinase-independent function of CDK6 links the cell cycle to tumor angiogenesis. *Cancer Cell* 24, 167–181. doi: 10.1016/j.ccr.2013.07.012
- Kondo, Y., and Issa, J. P. (2003). Enrichment for histone H3 lysine 9 methylation at Alu repeats in human cells. *J. Biol. Chem.* 278, 27658–27662. doi: 10.1074/jbc.M304072200
- Kubo, S., Seleme, M. C., Soifer, H. S., Perez, J. L., Moran, J. V., Kazazian, H. H. Jr., et al. (2006). L1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8036–8041. doi: 10.1073/pnas.0601954103
- Kulpa, D. A., and Moran, J. V. (2005). Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.* 14, 3237–3248. doi: 10.1093/hmg/ddi354
- Kulpa, D. A., and Moran, J. V. (2006). Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* 13, 655–660. doi: 10.1038/nsmb1107
- 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, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J. III, et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971. doi: 10.1126/science.1222077
- Leibold, D. M., Swergold, G. D., Singer, M. F., Thayer, R. E., Dombroski, B. A., and Fanning, T. G. (1990). Translation of LINE-1 DNA elements *in vitro* and in human cells. *Proc. Natl. Acad. Sci. U.S.A.* 87, 6990–6994. doi: 10.1073/pnas.87.18.6990
- Li, M. Y., Zhu, M., Feng, F., Cai, F. Y., Fan, K. C., Jiang, H., et al. (2014). Long interspersed nucleotide acid element-1 ORF-1 protein promotes proliferation and invasion of human colorectal cancer LoVo cells through enhancing ETS-1 activity. *Genet. Mol. Res.* 13, 6981–6994. doi: 10.4238/2014.April.14.13
- Li, X., Zhang, J., Jia, R., Cheng, V., Xu, X., Qiao, W., et al. (2013). The MOV10 helicase inhibits LINE-1 mobility. *J. Biol. Chem.* 288, 21148–21160. doi: 10.1074/jbc.M113.465856
- Liang, G., Chan, M. F., Tomigahara, Y., Tsai, Y. C., Gonzales, F. A., Li, E., et al. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol. Cell. Biol.* 22, 480–491. doi: 10.1128/MCB.22.2.480-491.2002
- Liu, L., Wylie, R. C., Andrews, L. G., and Tollefsbol, T. O. (2003). Aging, cancer and nutrition: the DNA methylation connection. *Mech. Ageing Dev.* 124, 989–998. doi: 10.1016/j.mad.2003.08.001
- Martens, J. H., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., et al. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* 24, 800–812. doi: 10.1038/sj.emboj.7600545
- Martin, S. L., Bushman, D., Wang, F., Li, P. W., Walker, A., Cumiskey, J., et al. (2008). A single amino acid substitution in ORF1 dramatically decreases L1 retrotransposition and provides insight into nucleic acid chaperone activity. *Nucleic Acids Res.* 36, 5845–5854. doi: 10.1093/nar/gkn554
- Martin, S. L., Cruceanu, M., Branciforte, D., Wai-Lun Li, P., Kwok, S. C., Hodges, R. S., et al. (2005). LINE-1 retrotransposition requires the nucleic acid chaperone activity of the ORF1 protein. *J. Mol. Biol.* 348, 549–561. doi: 10.1016/j.jmb.2005.03.003
- Mathias, S. L., Scott, A. F., Kazazian, H. H. Jr., Boeke, J. D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* 254, 1808–1810. doi: 10.1126/science.1722352
- Miki, Y., Katagiri, T., Kasumi, F., Yoshimoto, T., and Nakamura, Y. (1996). Mutation analysis in the BRCA2 gene in primary breast cancers. *Nat. Genet.* 13, 245–247. doi: 10.1038/ng0696-245
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., et al. (1992). Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–645.

- Moldovan, J. B., and Moran, J. V. (2015). The Zinc-finger antiviral protein ZAP inhibits LINE and Alu retrotransposition. *PLoS Genet.* 11:e1005121. doi: 10.1371/journal.pgen.1005121
- Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., and Kazazian, H. H. Jr. (1996). High frequency retrotransposition in cultured mammalian cells. *Cell* 87, 917–927. doi: 10.1016/S0092-8674(00)81998-4
- Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M., and Astrin, S. M. (1988). Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 333, 87–90. doi: 10.1038/333087a0
- Murata, A., Baba, Y., Watanabe, M., Shigaki, H., Miyake, K., Ishimoto, T., et al. (2013). Methylation levels of LINE-1 in primary lesion and matched metastatic lesions of colorectal cancer. *Br. J. Cancer* 109, 408–415. doi: 10.1038/bjc.2013.289
- Ogino, S., Noshio, K., Kirkner, G. J., Kawasaki, T., Chan, A. T., Schernhammer, E. S., et al. (2008). A cohort study of tumoral LINE-1 hypomethylation and prognosis in colon cancer. *J. Natl. Cancer Inst.* 100, 1734–1738. doi: 10.1093/jnci/djn359
- Ostertag, E. M., Goodier, J. L., Zhang, Y., and Kazazian, H. H. Jr. (2003). SVA elements are nonautonomous retrotransposons that cause disease in humans. *Am. J. Hum. Genet.* 73, 1444–1451. doi: 10.1086/380207
- Park, S. Y., Seo, A. N., Jung, H. Y., Gwak, J. M., Jung, N., Cho, N. Y., et al. (2014). Alu and LINE-1 hypomethylation is associated with HER2 enriched subtype of breast cancer. *PLoS ONE* 9:e100429. doi: 10.1371/journal.pone.0100429
- Patchsung, M., Boonla, C., Amnatrakul, P., Dissayabutra, T., Mutirangura, A., and Tosukhowong, P. (2012). Long interspersed nuclear element-1 hypomethylation and oxidative stress: correlation and bladder cancer diagnostic potential. *PLoS ONE* 7:e37009. doi: 10.1371/journal.pone.0037009
- Patnala, R., Lee, S. H., Dahlstrom, J. E., Ohms, S., Chen, L., Dheen, S. T., et al. (2014). Inhibition of LINE-1 retrotransposon-encoded reverse transcriptase modulates the expression of cell differentiation genes in breast cancer cells. *Breast Cancer Res. Treat.* 143, 239–253. doi: 10.1007/s10549-013-2812-7
- Peddigari, S., Li, P. W., Rabe, J. L., and Martin, S. L. (2013). hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic Acids Res.* 41 575–585. doi: 10.1093/nar/gks1075
- Piskareva, O., Denmukhametova, S., and Schmatchenko, V. (2003). Functional reverse transcriptase encoded by the human LINE-1 from baculovirus-infected insect cells. *Protein Expr. Purif.* 28, 125–130. doi: 10.1016/S1046-5928(02)00655-1
- Piskareva, O., Ernst, C., Higgins, N., and Schmatchenko, V. (2013). The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *FEBS Open Bio* 3, 433–437. doi: 10.1016/j.fob.2013.09.005
- Raiz, J., Damert, A., Chira, S., Held, U., Klawitter, S., Hamdorf, M., et al. (2012). The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res.* 40, 1666–1683. doi: 10.1093/nar/gkr863
- Rangasamy, D., Lenka, N., Ohms, S., Dahlstrom, J. E., Blackburn, A. C., and Board, P. G. (2015). Activation of LINE-1 Retrotransposon Increases the Risk of Epithelial-Mesenchymal Transition and Metastasis in Epithelial Cancer. *Curr. Mol. Med.* 15, 588–597. doi: 10.2174/1566524015666150831130827
- Richardson, S. R., Narvaiza, I., Planegger, R. A., Weitzman, M. D., and Moran, J. V. (2014). APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* 3:e02008. doi: 10.7554/elifelife.02008
- Rodic, N., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., Hruban, R. H., et al. (2014). Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* 184, 1280–1286. doi: 10.1016/j.ajpath.2014.01.007
- Rodic, N., Steranka, J. P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., et al. (2015). Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat. Med.* 21, 1060–1064. doi: 10.1038/nm.3919
- Saito, K., Kawakami, K., Matsumoto, I., Oda, M., Watanabe, G., and Minamoto, T. (2010). Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. *Clin. Cancer Res.* 16, 2418–2426. doi: 10.1158/1078-0432.CCR-09-2819
- Salas, L. A., Villanueva, C. M., Tajuddin, S. M., Amaral, A. F., Fernandez, A. F., Moore, L. E., et al. (2014). LINE-1 methylation in granulocyte DNA and trihalomethane exposure is associated with bladder cancer risk. *Epigenetics* 9, 1532–1539. doi: 10.4161/15592294.2014.983377
- Sassaman, D. M., Dombroski, B. A., Moran, J. V., Kimberland, M. L., Naas, T. P., DeBerardinis, R. J., et al. (1997). Many human L1 elements are capable of retrotransposition. *Nat. Genet.* 16, 37–43. doi: 10.1038/ng0597-37
- Schuster, A. T., Sarvepalli, K., Murphy, E. A., and Longworth, M. S. (2013). Condensin II subunit dCAP-D3 restricts retrotransposon mobilization in Drosophila somatic cells. *PLoS Genet.* 9:e1003879. doi: 10.1371/journal.pgen.1003879
- Shigaki, H., Baba, Y., Watanabe, M., Murata, A., Iwagami, S., Miyake, K., et al. (2013). LINE-1 hypomethylation in gastric cancer, detected by bisulfite pyrosequencing, is associated with poor prognosis. *Gastric Cancer* 16, 480–487. doi: 10.1007/s10120-012-0209-7
- Shukla, R., Upton, K. R., Muñoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111. doi: 10.1016/j.cell.2013.02.032
- Startek, M., Szafranski, P., Gambin, T., Campbell, I. M., Hixson, P., Shaw, C. A., et al. (2015). Genome-wide analyses of LINE-LINE-mediated nonallelic homologous recombination. *Nucleic Acids Res.* 43, 2188–2198. doi: 10.1093/nar/gku1394
- Stetson, D. B., Ko, J. S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* 134, 587–598. doi: 10.1016/j.cell.2008.06.032
- Swergold, G. D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 10, 6718–6729. doi: 10.1128/MCB.10.12.6718
- Symer, D. E., Connelly, C., Szak, S. T., Caputo, E. M., Cost, G. J., Parmigiani, G., et al. (2002). Human L1 retrotransposition is associated with genetic instability in vivo. *Cell* 110, 327–338. doi: 10.1016/S0092-8674(02)00839-5
- Takasu, M., Hayashi, R., Maruya, E., Ota, M., Imura, K., Kougou, K., et al. (2007). Deletion of entire HLA-A gene accompanied by an insertion of a retrotransposon. *Tissue Antigens* 70, 144–150. doi: 10.1111/j.1399-0039.2007.00870.x
- Tam, O. H., Aravin, A. A., Stein, P., Girard, A., Murchison, E. P., Cheloufi, S., et al. (2008). Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* 453, 534–538. doi: 10.1038/nature06904
- Teugels, E., De Brakeleer, S., Goelen, G., Lissens, W., Sermijn, E., and De Greve, J. (2005). De novo Alu element insertions targeted to a sequence common to the BRCA1 and BRCA2 genes. *Hum. Mutat.* 26, 284. doi: 10.1002/humu.9366
- Vanin, E. F. (1985). Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* 19, 253–272. doi: 10.1146/annurev.gen.19.120185.001345
- Wallace, M. R., Andersen, L. B., Saulino, A. M., Gregory, P. E., Glover, T. W., and Collins, F. S. (1991). A de novo Alu insertion results in neurofibromatosis type 1. *Nature* 353, 864–866. doi: 10.1038/353864a0
- Weber, M. J. (2006). Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet.* 2:e205. doi: 10.1371/journal.pgen.002005
- Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., et al. (2001). Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 21, 1429–1439. doi: 10.1128/MCB.21.4.1429-1439.2001
- Weiner, A. M., Deininger, P. L., and Efstratiadis, A. (1986). Nonviral retrotransposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55, 631–661. doi: 10.1146/annurev.bi.55.070186.003215
- Weisenberger, D. J., Campan, M., Long, T. I., Kim, M., Woods, C., Fiala, E., et al. (2005). Analysis of repetitive element DNA methylation by MethyLight. *Nucleic Acids Res.* 33, 6823–6836. doi: 10.1093/nar/gki987
- Woodcock, D. M., Linsenmeyer, M. E., and Warren, W. D. (1998). DNA methylation in mouse A-repeats in DNA methyltransferase-knockout ES cells and in normal cells determined by bisulfite genomic sequencing. *Gene* 206, 63–67. doi: 10.1016/S0378-1119(97)00566-0
- Yang, Q., Feng, F., Zhang, F., Wang, C., Lu, Y., Gao, X., et al. (2013). LINE-1 ORF-1p functions as a novel HGF/ETS-1 signaling pathway co-activator and promotes the growth of MDA-MB-231 cell. *Cell. Signal.* 25, 2652–2660. doi: 10.1016/j.cellsig.2013.08.029
- Zhang, A., Dong, B., Doucet, A. J., Moldovan, J. B., Moran, J. V., and Silverman, R. H. (2014). RNase L restricts the mobility of engineered

- retrotransposons in cultured human cells. *Nucleic Acids Res.* 42, 3803–3820. doi: 10.1093/nar/gkt1308
- Zhao, K., Du, J., Han, X., Goodier, J. L., Li, P., Zhou, X., et al. (2013). Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutieres syndrome-related SAMHD1. *Cell Rep.* 4, 1108–1115. doi: 10.1016/j.celrep.2013.08.019
- Zhu, C., Utsunomiya, T., Ikemoto, T., Yamada, S., Morine, Y., Imura, S., et al. (2014). Hypomethylation of long interspersed nuclear element-1 (LINE-1) is associated with poor prognosis via activation of c-MET in hepatocellular carcinoma. *Ann. Surg. Oncol.* 21(Suppl. 4), S729–S735. doi: 10.1245/s10434-014-3874-4

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 © 2015 Kemp and Longworth. 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) or licensor 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.



Links between Human LINE-1 Retrotransposons and Hepatitis Virus-Related Hepatocellular Carcinoma

Tomoyuki Honda^{1,2*}

¹ Department of Viral Oncology, Institute for Virus Research, Kyoto University, Kyoto, Japan, ² Division of Virology, Department of Microbiology and Immunology, Osaka University Graduate School of Medicine, Suita, Japan

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Formerly affiliated with University of
Toledo, USA

Reviewed by:

Kaushendra Tripathi,
Mitchell Cancer Institute, USA
Gerald Günther Schumann,
Paul-Ehrlich-Institut, Germany

*Correspondence:

Tomoyuki Honda
thonda@virus.med.osaka-u.ac.jp

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 18 December 2015

Accepted: 22 April 2016

Published: 11 May 2016

Citation:

Honda T (2016) Links between
Human LINE-1 Retrotransposons and
Hepatitis Virus-Related Hepatocellular
Carcinoma. *Front. Chem.* 4:21.
doi: 10.3389/fchem.2016.00021

Hepatocellular carcinoma (HCC) accounts for approximately 80% of liver cancers, the third most frequent cause of cancer mortality. The most prevalent risk factors for HCC are infections by hepatitis B or hepatitis C virus. Findings suggest that hepatitis virus-related HCC might be a cancer in which LINE-1 retrotransposon, often termed L1, activity plays a potential role. Firstly, hepatitis viruses can suppress host defense factors that also control L1 mobilization. Secondly, many recent studies also have indicated that hypomethylation of L1 affects the prognosis of HCC patients. Thirdly, endogenous L1 retrotransposition was demonstrated to activate oncogenic pathways in HCC. Fourthly, several L1 chimeric transcripts with host or viral genes are found in hepatitis virus-related HCC. Such lines of evidence suggest a linkage between L1 retrotransposons and hepatitis virus-related HCC. Here, I briefly summarize current understandings of the association between hepatitis virus-related HCC and L1. Then, I discuss potential mechanisms of how hepatitis viruses drive the development of HCC via L1 retrotransposons. An increased understanding of the contribution of L1 to hepatitis virus-related HCC may provide unique insights related to the development of novel therapeutics for this disease.

Keywords: L1, retrotransposon, hepatitis C virus (HCV), hepatitis B virus (HBV), hepatocellular carcinoma

INTRODUCTION

Liver cancer, 80% of which is hepatocellular carcinoma (HCC), accounts for 9% of all cancer deaths worldwide (Jemal et al., 2011; Tateishi and Omata, 2012). The major causative agents of HCC are hepatitis viruses, such as hepatitis B virus (HBV) or hepatitis C virus (HCV) (Jemal et al., 2011; Tateishi and Omata, 2012). HBV belongs to the *Hepadnaviridae* family, which has a relaxed circular DNA (rcDNA) as a viral genome (Beck and Nassal, 2007; Nguyen et al., 2008). HCV belongs to the *Flaviviridae* family, which has a nonsegmented, positive-stranded RNA as a viral genome (Hijikata et al., 1991; Grakoui et al., 1993; Aly et al., 2012). Both viruses cause chronic infections, with approximately 350 and 170 million people worldwide affected by chronic HBV and HCV infections, respectively (Parkin, 2006; Aly et al., 2012). It is now clear that chronic HBV and HCV infections play critical roles in the development of HCC (Jemal et al., 2011; Forner et al., 2012; Tateishi and Omata, 2012). However, the precise mechanisms of hepatocarcinogenesis in chronic hepatitis virus infections are still unclear.

Long interspersed nuclear element-1 (LINE-1 or L1) retrotransposons are genetic elements that constitute approximately 17% of the human genome (Lander et al., 2001). Because most L1s are 5' truncated, most of them are defective, while 80–100 copies are still retrotransposition-competent and utilize a “copy-and-paste” mechanism to retrotranspose to new genomic loci (Brouha et al., 2003; Beck et al., 2010). Aberrantly expressed or dysregulated L1s are considered a major source of endogenous mutagenesis in humans (Levin and Moran, 2011; Burns and Boeke, 2012). L1 retrotransposition occurs in germ cells, pluripotent stem cells, at early stages of human embryonic development (van den Hurk et al., 2007; Beck et al., 2011; Levin and Moran, 2011; Klawitter et al., 2016) and in somatic cells, such as neuronal progenitor cells or cancer cells (Muotri et al., 2005; Iskow et al., 2010). Many epidemiological studies suggest a linkage between L1 and cancers (Shukla et al., 2013; Rodić et al., 2014; Harada et al., 2015). However, in most cases, it is unclear whether L1s are activated in normal cells before clonal expansion or in cancer cells at the later stage of carcinogenesis (Goodier, 2014).

Among cancers, hepatitis virus-related HCC is considered to be a cancer in which L1 might be involved (Shukla et al., 2013). Firstly, by far the majority of L1 *de novo* insertions detected in cancer tissues has been found in cancers of epithelial origin (Goodier, 2014). Secondly, HBV and HCV have a potential to suppress host defense mechanisms that can also control L1 retrotransposition (Gale and Foy, 2005; Chang et al., 2012; Yu et al., 2015). Thirdly, endogenous L1 retrotransposition was demonstrated to activate oncogenic pathways in HCC. Fourthly, several L1 chimeric transcripts with host or viral genes are found in hepatitis virus-related HCC (Lau et al., 2014). Here, I will summarize potential linkages between hepatitis virus-related HCC and L1s. Firstly, I will review how HBV could affect L1 retrotransposon activity. I will then introduce current understandings of the relationship between HCV and L1. Finally, I will discuss possible L1-mediated mechanisms that may induce HCC. Understandings of possible links between virus-related HCC and L1 may open a new avenue for the development of novel therapeutics for this disease.

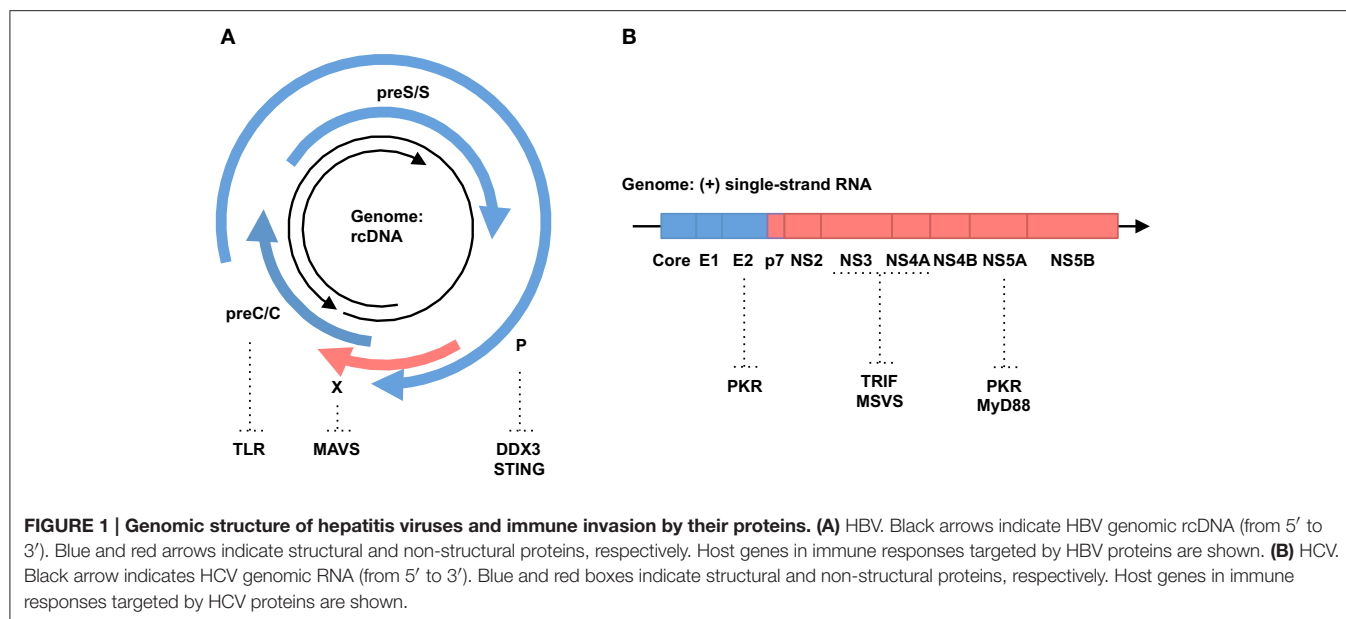
A POTENTIAL LINK BETWEEN HBV AND L1 IN HCC

The 3.2-kb HBV genome encodes four, partly overlapping open reading frames (ORFs): preC/C (core and Hepatitis B e-Antigen [HBeAg]), P (viral polymerase), preS/S (Hepatitis B surface Antigen [HBsAg]) and X (non-structural protein [HBx]) genes (Figure 1A). In the nucleus, the genome is converted into covalently closed circular DNA (cccDNA). From this cccDNA, all viral RNAs, including pregenomic RNA (pgRNA) as a replication intermediate and viral mRNAs, are transcribed. Viral proteins such as core and polymerase proteins and pgRNAs are assembled into the nucleocapsid within the cytoplasm. In the nucleocapsid, pgRNA is reverse transcribed into rcDNA. All these HBV-related nucleic acids have the potential to trigger innate immune responses in infected cells (Ait-Goughoulte et al., 2010). If

these immune responses cannot clear HBV, the virus establishes a chronic infection, which is known to increase the risk of developing liver cirrhosis and HCC (Gonzalez and Keeffe, 2011).

Type I interferons (IFNs) play a major role in anti-viral immunity (Katze et al., 2002). Association of IFNs with IFN receptors activates JAK1 and signal transducer and activator of transcription 1/2 (STAT1/2). Then, these proteins interact with interferon regulatory factor 9 (IRF9) and form a potent transcription factor, upregulating the expression of several hundreds of IFN-stimulated genes (ISGs). These ISGs suppress viral replication and spread through various mechanisms described elsewhere (Katze et al., 2002). IFN is used to control HBV replication, indicating that IFN is a restriction factor (Dienstag, 2008). For example, tetherin, an IFN-inducible transmembrane protein, inhibits HBV virion secretion (Yan et al., 2015). Zinc finger antiviral protein (ZAP) is upregulated in IFN-treated cells and restricts HBV replication through downregulation of pgRNA (Mao et al., 2013). On the other hand, HBV has a variety of strategies to counteract IFN signaling (Figure 1A). HBsAg, HBeAg and HBV virions inhibit Toll-like receptor (TLR)-mediated antiviral responses (Visvanathan et al., 2007; Wu et al., 2009; Vincent et al., 2011; Woltman et al., 2011). HBV polymerase suppresses IRF3 activation by interacting with the host RNA helicase, DDX3 (Wang and Ryu, 2010; Yu et al., 2010). HBV polymerase also disrupts ubiquitination of the stimulator of interferon genes (STING) and blocks innate immune responses against cytoplasmic DNA (Liu et al., 2015). Expression of HBx protein inhibits virus-induced expression of the IFN gene by promoting the decay of mitochondrial antiviral signaling protein (MAVS) (Wei et al., 2010; Kumar et al., 2011). Furthermore, HBV abrogates IFN signal transduction by impairing either STAT1 nuclear import or phosphorylation (Christen et al., 2006, 2007; Lütgehetmann et al., 2011). All the listed mechanisms that suppress IFN signaling could also activate L1 retrotransposon, because IFN has been shown to inhibit the expression and retrotransposition of L1 (Yu et al., 2015). The mechanisms underlying the inhibitory effect of IFN on L1 remain unclear. However, MOV10 is an attractive candidate to mediate this inhibitory effect, because MOV10 is an IFN-inducible gene and suppresses L1 retrotransposition (Schoggins et al., 2011; Goodier et al., 2012). Collectively, immune suppression by HBV may activate the expression and retrotransposition of L1 elements.

In addition, HBV may also modulate L1 expression epigenetically. L1 retrotransposition activity is usually suppressed in most somatic cells by host DNA methyltransferase-mediated DNA methylation of its promoter (Ishizu et al., 2012; Castro-Diaz et al., 2014). In cancer cells, global DNA hypomethylation occurs at various genomic loci including those containing DNA repeats and/or retrotransposons (Ehrlich, 2002; Hatzia Apostolou and Iliopoulos, 2011). Many studies have reported hypomethylation of the L1 loci in HCC and HBV infection (Shitani et al., 2012; Zhang C. et al., 2013; Gao et al., 2014; Zhu et al., 2014). In particular, L1 hypomethylation is likely to be linked to poor outcomes of HCC (Gao et al., 2014; Zhu et al., 2014). Given global hypomethylation occurs in the host genome (including the L1 loci) during HBV infection, this may upregulate L1



expression, potentially removing an obstacle to L1 transposition in liver cells. In addition, some chimeric transcripts, such as HBx-L1, are detected in HCC and associated with a poor prognosis, further supporting the link between HBV-related HCC and L1 (Lau et al., 2014).

A POTENTIAL LINK BETWEEN HCV AND L1 IN HCC

The 9.6-kb HCV genome contains a single ORF, encoding a polyprotein precursor of approximately 3000 amino acids. The polyprotein is cleaved by host and viral proteases, producing structural (core, E1 and E2) and non-structural (P7, NS2, NS3, NS4A, NS4B, NS5A and NS5B) proteins (**Figure 1B**). The replication of HCV starts with the synthesis of a full-length, negative-stranded RNA intermediate, which in turn works as a template for the *de novo* production of positive-stranded genomic RNA. Thus, HCV replicates without a known DNA intermediate stage. HCV genomic RNA is highly structured and contains double-stranded regions in various portions (Tuplin et al., 2002; Zhang S. et al., 2013). Double-stranded RNAs (dsRNAs) are also generated during the replication cycle of HCV. Such dsRNAs are potent inducers of innate immune responses, mainly through TLR3 and retinoic acid inducible gene-I (RIG-I) signal pathways (Li et al., 2012). However, the immune responses induced by HCV are not strong enough to eradicate the virus (Battaglia and Hagemeyer, 2000).

Although it is thought that non-retroviral RNA viruses are not integrated into host genomic DNA, we and others have demonstrated that they do become integrated into the host genome via host retrotransposon machineries (Geuking et al., 2009; Horie et al., 2010). Likewise, HCV cDNA is reportedly detected in patients infected with HCV (Zemer et al., 2008). Because the involvement of HIV was ruled out

in all the HCV cDNA-positive patients, it is hypothesized that host retrotransposons might be involved (Zemer et al., 2008). However, the retrotransposons responsible for this phenomenon remain unidentified and the involvement of retroviruses other than HIV is not ruled out. Because the 3'UTR of HCV is not polyadenylated, the contribution of L1, whose substrates are usually polyadenylated, to this phenomenon seems to be unlikely. However, several reports propose alternative retrotransposition mechanisms by L1, termed internal priming or twin priming, where a poly-A tail is not required to prime reverse transcription (Ostertag and Kazazian, 2001; Srikanta et al., 2009). These alternative mechanisms may explain how HCV RNA could be reverse transcribed by L1, despite lacking a poly-A tail. It is also unknown whether HCV cDNA is integrated into the host genome or exists as extrachromosomal DNA. A recent report showed that fragments homologous to HCV genes are present in the rabbit and hare genomes, which might suggest the possibility that cDNA of an HCV ancestor has been integrated into the host genome (Silva et al., 2012). These observations imply that some linkages between HCV and retrotransposon activity might exist.

Most HCV-infected patients develop a chronic infection, suggesting that HCV has developed successful strategies to evade host immune responses (Gale and Foy, 2005) (**Figure 1B**). For instance, the HCV NS3/4A protease cleaves the Toll/IL-1 receptor domain-containing adaptor inducing IFN- β (TRIF) adaptor protein and MAVS to impair TLR3 and retinoic acid-inducible gene-I (RIG-I) signaling pathways, respectively (Foy et al., 2005; Li K. et al., 2005; Li X.-D. et al., 2005). NS5A and E2 proteins suppress the signaling of the interferon-dependent induced protein kinase R (PKR), a key molecule in the innate immune system (Gale et al., 1997; Taylor et al., 1999). The interferon sensitivity-determining region (ISDR) in the NS5A protein interacts with the death domain of myeloid differentiation primary response 88 protein (MyD88), a major adaptor protein in TLR signaling, and impairs its signaling (Abe

et al., 2007). All these mechanisms that suppress IFN responses against HCV could in turn activate retrotransposons, such as L1, in infected cells, because IFN and IFN-inducible genes, such as MOV10, are shown to suppress retrotransposition of L1s as described above (Schoggins et al., 2011; Goodier et al., 2012; Yu et al., 2015).

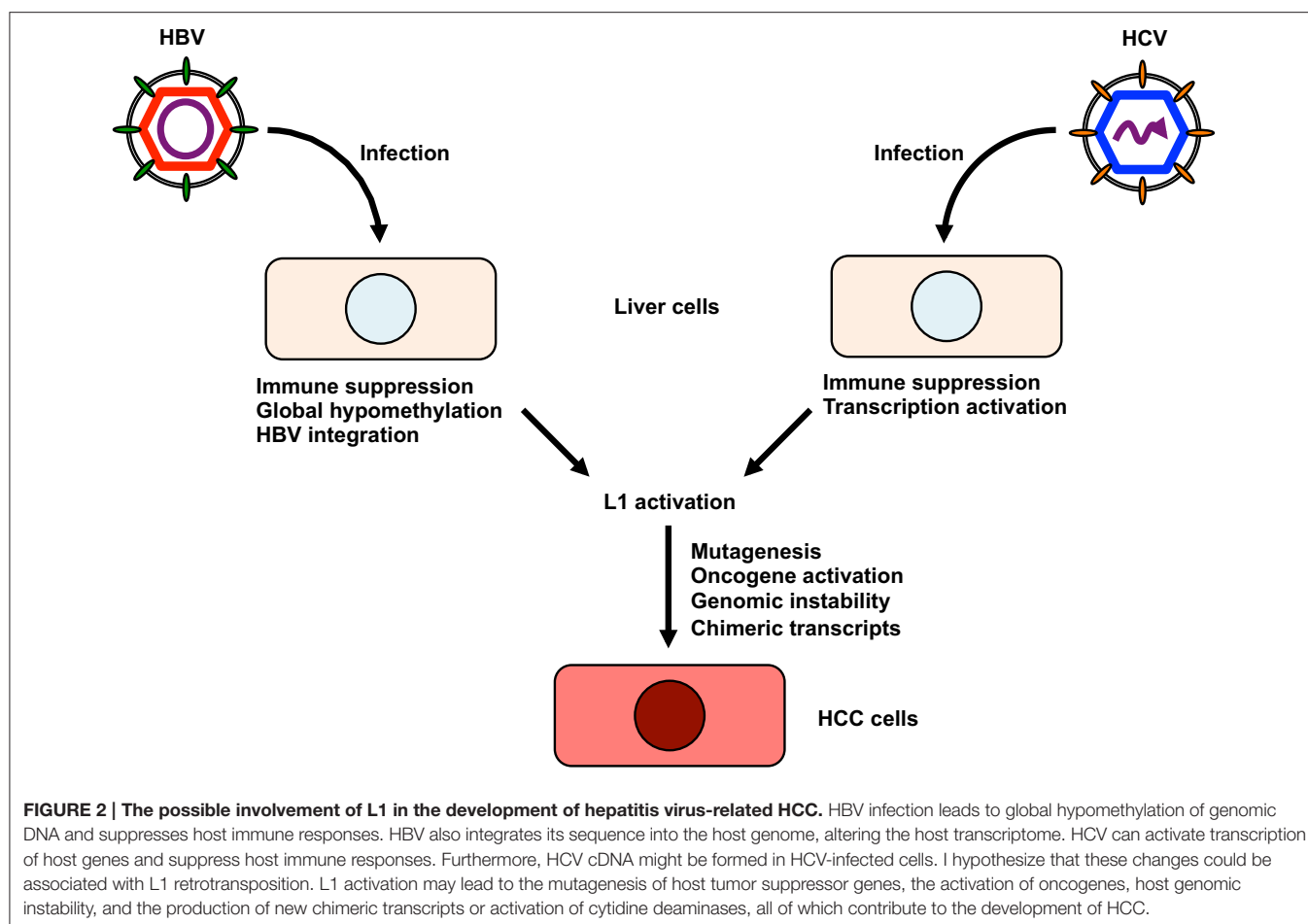
The HCV core protein has an oncogenic potential (Moriya et al., 1998; Shimotohno, 2000). One mechanism put forward for this is that the core protein modulates host gene expression pathways which may activate oncogene expression (Shrivastava et al., 1998; Marusawa et al., 1999; Shimotohno, 2000; Watashi et al., 2001; Ray et al., 2002). In addition to the core protein, NS5A protein also stimulates NF- κ B signaling (Ray et al., 1995; Gong et al., 2001; Park et al., 2002; Waris et al., 2003). Similarly, HCV proteins may stimulate the expression of L1 retrotransposons. Indeed, the infectious HCV virion reportedly activates HIV long terminal repeats (LTR) and upregulates gene transcription (Sengupta et al., 2013). However, studies investigating whether HCV proteins have the potential to stimulate L1 expression and/or retrotransposition have not been reported so far.

POSSIBLE MECHANISMS OF L1 INVOLVEMENT IN HCC DEVELOPMENT

Although a definitive role for L1 activity in contributing to HCC etiology has not been established thus far, investigating a possible link between L1 activation and the development of HCC would be of considerable interest for a number of reasons (**Figure 2**). Firstly, L1s, when aberrantly expressed or dysregulated, can be major sources of endogenous mutagenesis in humans as described above (Levin and Moran, 2011; Burns and Boeke, 2012). Any potential disruption of tumor suppressor genes by L1 retrotransposition could contribute to the development of HCC. Indeed, L1 was shown to be a crucial source of mutations that can reduce the tumor-suppressive capacity of somatic cells (Shukla et al., 2013). A subset of L1 *de novo* insertions identified in cancer tissue occurred at genes commonly mutated in cancer (Lee et al., 2012). Secondly, L1 *de novo* insertions can affect the expression of nearby genes and the genes in which they inserted (Lee et al., 2012; Shukla et al., 2013). Intragenic L1 insertions usually coincide with reduced gene expression (Lee et al., 2012). For example, L1 insertions into the tumor suppressor mutated in colorectal cancer (MCC) gene coincides with its downregulation (Shukla et al., 2013). MCC is expressed in liver and suppresses the oncogenic β -catenin/Wnt signaling pathway frequently activated in HCC (Fukuyama et al., 2008). If an L1 insertion occurs close to an oncogene, L1 could enhance oncogene expression, resulting in the development of HCC. For example, the telomerase reverse transcriptase (TERT) gene is one of the most common genes associated with L1 *de novo* insertion (Ding et al., 2012; Lau et al., 2014). Since aberrant expression of TERT is associated with tumor development, L1 insertion near the TERT locus may have a role in carcinogenesis (Cohen et al., 2007; Nault et al., 2015). L1 insertion at the transcriptional repressor suppression of tumorigenicity 18 (ST18) gene activates its expression (Shukla et al., 2013). ST18 is a candidate oncogene in liver, because the

expression of ST18 is upregulated in several liver cancer cells and in tumors in a mouse-model for inflammation-driven HCC (Shukla et al., 2013). Thirdly, L1 provides sites that could lead to genomic rearrangements (Burwinkel and Kilimann, 1998). Such genomic rearrangements contribute to genomic instability (Burwinkel and Kilimann, 1998; Ehrlich, 2002). Fourthly, L1 retrotransposition could contribute to new splice donor or acceptor sites, which could alter the host transcriptome and might enhance HCC progression (Singer et al., 2010). Lastly, L1 retrotransposition occasionally creates new chimeric transcripts, which might enhance the progression to HCC. An example of this mechanism is the L1-MET transcript, a chimeric transcript that consists of the c-MET oncogene and an intronic L1 sequences (Zhu et al., 2014). The expression of L1-MET has been shown to be correlated with that of c-MET (Zhu et al., 2014). Because L1-MET is associated with poor prognosis in cancer (Wolff et al., 2010; Hur et al., 2014), L1-MET might be associated with a poor prognosis for HCC via the activation of c-MET signaling (Zhu et al., 2014).

Taken together, I conclude two potential roles for L1 elements in the development of hepatitis virus-related HCC. The first relates to a chimeric transcript specific to HBV-related HCC, HBx-L1, which can be detected in more than 20% of HBV-related HCC and correlates with a poor outcome (Lau et al., 2014). The promoter of the HBx gene transcribes HBx-L1 from the locus that is normally silent in the genome. Knockdown of HBx-L1 reduces migratory and invasive properties of HBV-positive HCC cells. HBx-L1 overexpression confers growth advantage and promotes cell migration and invasion regardless of its chimeric protein-coding potential, suggesting that HBx-L1 is a long non-coding RNA that promotes HCC phenotypes. Furthermore, it has been shown that HBx-L1 affects β -catenin/Wnt signaling, a major pathway in the oncogenesis of HBV-related HCC, confirming its role in HCC (Whittaker et al., 2010; Lau et al., 2014). In addition to this, I hypothesize the other possible role of L1 as a potent inducer of the expression of cytidine deaminases, such as activation-induced cytidine deaminase (AID) and apolipoprotein B mRNA editing enzyme, catalytic polypeptide 3 (APOBEC3). Since transgenic mice expressing AID genes invariably induce tumors, this suggests that cytidine deaminases may have a carcinogenic potential (Okazaki et al., 2003; Takai et al., 2009). APOBEC3 is a protein family of seven proteins in human: APOBEC3A, B, C, DE, F, G, and H (Schumann et al., 2010; Vieira and Soares, 2013). Members of the APOBEC3 protein family restrict replication of not only retroviruses such as HIV, but also retrotransposons, HBV and HCV (Harris and Liddament, 2004; Vieira and Soares, 2013). Among APOBEC3 proteins, APOBEC3G seems to have a major role in HIV restriction (Chaipan et al., 2013; Vieira and Soares, 2013). APOBEC3G also has the restriction activity against LTR retrotransposons in the mouse genome (Esnault et al., 2005; Schumacher et al., 2008). All members of the human APOBEC3 protein family of cytidine deaminases restrict L1 retrotransposition with APOBEC3A, B, C and F having the strongest inhibitory effect (Muckenfuss et al., 2006; Kinomoto et al., 2007). For HBV and HCV, APOBEC3G is a major restriction factor (Vartanian et al., 2010; Peng et al., 2011; Kitamura et al., 2013). HBV and HCV somehow stimulate



the expression of cytidine deaminases (Vartanian et al., 2010). Furthermore, L1 activation reportedly increases the expression of the mouse APOBEC3 gene in mouse embryonic fibroblasts (Yu et al., 2015). Taken together, hepatitis viruses, directly and/or maybe indirectly via L1 activation, induce the expression of cytidine deaminases, which may hyperedit host genomes, resulting in the accumulation of deleterious mutations in the genome and the development of HCC (Okazaki et al., 2003; Takai et al., 2009; Vartanian et al., 2010).

CONCLUSION AND PERSPECTIVE

Presented lines of evidence suggest potential links between hepatitis virus infection and L1 retrotransposon activity. Especially, L1 hypomethylation or some L1 chimeric transcripts are associated with poor prognosis of HCC, suggesting that it can have a significant effect on HCC phenotypes and supporting the idea that HCC is a cancer in which L1 plays a role. However, knowledge of how L1 activation by chronic hepatitis virus infection enhances the development of HCC is still limiting. Further accumulation of examples of recurrent L1 insertion sites in the host genome or recurrent chimeric transcripts specific to hepatitis virus-related HCC will be promising ways

to understand L1 involvement in HCC etiology. Single cell analyses of L1 retrotransposition events and expression in tumor cells and surrounding normal cells may enhance these processes. Understanding the potential roles of L1 in HCC may open avenues to developing novel therapeutics, such as RNA interference against HCC-specific L1 chimeric transcripts.

AUTHOR CONTRIBUTIONS

TH wrote the manuscript and approved it for publication.

ACKNOWLEDGMENTS

I would like to thank Makoto Hijikata and Nicholas F. Parrish for helpful discussions and Keizo Tomonaga for his support and encouragement. Preparation of this paper was supported in part by KAKENHI Grant Number 15K08496 from Japan Society for the Promotion of Science (JSPS), and grants from the Takeda Science Foundation, Senri Life Science Foundation, Suzuken Memorial Foundation, The Shimizu Foundation for Immunology and Neuroscience Grant for 2015 and The NOVARTIS Foundation (Japan) for the Promotion of Science.

REFERENCES

- Abe, T., Kaname, Y., Hamamoto, I., Tsuda, Y., Wen, X., Taguwa, S., et al. (2007). Hepatitis C virus nonstructural protein 5A modulates the toll-like receptor-MYD88-dependent signaling pathway in macrophage cell lines. *J. Virol.* 81, 8953–8966. doi: 10.1128/JVI.00649-07
- Ait-Goughoulte, M., Lucifora, J., Zoulim, F., and Durantel, D. (2010). Innate antiviral immune responses to hepatitis B virus. *Viruses* 2, 1394–1410. doi: 10.3390/v2071394
- Aly, H. H., Shimotohno, K., Hijikata, M., and Seya, T. (2012). *In vitro* models for analysis of the hepatitis C virus life cycle. *Microbiol. Immunol.* 56, 1–9. doi: 10.1111/j.1348-0421.2011.00403.x
- Battaglia, A. M., and Hagmeyer, K. O. (2000). Combination therapy with interferon and ribavirin in the treatment of chronic hepatitis C infection. *Ann. Pharmacother.* 34, 487–494. doi: 10.1345/aph.19183
- Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., et al. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159–1170. doi: 10.1016/j.cell.2010.05.021
- Beck, C. R., Garcia-Perez, J. L., Badge, R. M., and Moran, J. V. (2011). LINE-1 elements in structural variation and disease. *Annu. Rev. Genomics Hum. Genet.* 12, 187–215. doi: 10.1146/annurev-genom-082509-141802
- Beck, J., and Nassal, M. (2007). Hepatitis B virus replication. *World J. Gastroenterol.* 13, 48–64. doi: 10.3748/wjg.v13.i1.48
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285. doi: 10.1073/pnas.0831042100
- Burns, K. H., and Boeke, J. D. (2012). Human transposon tectonics. *Cell* 149, 740–752. doi: 10.1016/j.cell.2012.04.019
- Burwinkel, B., and Kilimann, M. W. (1998). Unequal homologous recombination between LINE-1 elements as a mutational mechanism in human genetic disease. *J. Mol. Biol.* 277, 513–517. doi: 10.1006/jmbi.1998.1641
- Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., et al. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* 28, 1397–1409. doi: 10.1101/gad.241661.114
- Chaipan, C., Smith, J. L., Hu, W.-S., and Pathak, V. K. (2013). APOBEC3G restricts HIV-1 to a greater extent than APOBEC3F and APOBEC3DE in human primary CD4⁺ T cells and macrophages. *J. Virol.* 87, 444–453. doi: 10.1128/JVI.00676-12
- Chang, J., Block, T. M., and Guo, J.-T. (2012). The innate immune response to hepatitis B virus infection: implications for pathogenesis and therapy. *Antiviral Res.* 96, 405–413. doi: 10.1016/j.antiviral.2012.10.001
- Christen, V., Duong, F., Bernsmeier, C., Sun, D., Nassal, M., and Heim, M. H. (2006). Inhibition of alpha Interferon signaling by hepatitis B virus. *J. Virol.* 81, 159–165. doi: 10.1128/JVI.01292-06
- Christen, V., Treves, S., Duong, F. H. T., and Heim, M. H. (2007). Activation of endoplasmic reticulum stress response by hepatitis viruses up-regulates protein phosphatase 2A. *Hepatology* 46, 558–565. doi: 10.1002/hep.21611
- Cohen, S. B., Graham, M. E., Lovrecz, G. O., Bache, N., Robinson, P. J., and Reddel, R. R. (2007). Protein composition of catalytically active human telomerase from immortal cells. *Science* 315, 1850–1853. doi: 10.1126/science.1138596
- Dienstag, J. L. (2008). Hepatitis B virus infection. *N. Engl. J. Med.* 359, 1486–1500. doi: 10.1056/NEJMra0801644
- Ding, D., Lou, X., Hua, D., Yu, W., Li, L., Wang, J., et al. (2012). Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet.* 8:e1003065. doi: 10.1371/journal.pgen.1003065
- Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. *Oncogene* 21, 5400–5413. doi: 10.1038/sj.onc.1205651
- Esnault, C., Heidmann, O., Delebecque, F., Dewannieux, M., Ribet, D., Hance, A. J., et al. (2005). APOBEC3G cytidine deaminase inhibits retrotransposition of endogenous retroviruses. *Nature* 433, 430–433. doi: 10.1038/nature03238
- Forner, A., Llovet, J. M., and Bruix, J. (2012). Hepatocellular carcinoma. *Lancet* 379, 1245–1255. doi: 10.1016/S0140-6736(11)61347-0
- Foy, E., Li, K., Sumpter, R. Jr., Loo, Y.-M., Johnson, C. L., Wang, C., et al. (2005). Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-1 signaling. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2986–2991. doi: 10.1073/pnas.0408707102
- Fukuyama, R., Niculaita, R., Ng, K. P., Obusez, E., Sanchez, J., Kalady, M., et al. (2008). Mutated in colorectal cancer, a putative tumor suppressor for serrated colorectal cancer, selectively represses beta-catenin-dependent transcription. *Oncogene* 27, 6044–6055. doi: 10.1038/ncr.2008.204
- Gale, M. Jr., and Foy, E. M. (2005). Evasion of intracellular host defence by hepatitis C virus. *Nature* 436, 939–945. doi: 10.1038/nature04078
- Gale, M. J. Jr., Korth, M. J., Tang, N. M., Tan, S. L., Hopkins, D. A., Dever, T. E., et al. (1997). Evidence that hepatitis C virus resistance to interferon is mediated through repression of the PKR protein kinase by the nonstructural 5A protein. *Virology* 230, 217–227. doi: 10.1006/viro.1997.8493
- Gao, X. D., Qu, J. H., Chang, X. J., Lu, Y. Y., Bai, W. L., Wang, H., et al. (2014). Hypomethylation of long interspersed nuclear element-1 promoter is associated with poor outcomes for curative resected hepatocellular carcinoma. *Liver Int.* 34, 136–146. doi: 10.1111/liv.12264
- Geuking, M. B., Weber, J., Dewannieux, M., Gorelik, E., Heidmann, T., Hengartner, H., et al. (2009). Recombination of retrotransposon and exogenous RNA virus results in nonretroviral cDNA integration. *Science* 323, 393–396. doi: 10.1126/science.1167375
- Gong, C., Waris, G., Tanveer, R., and Siddiqui, A. (2001). Human hepatitis C virus NS5A protein alters intracellular calcium levels, induces oxidative stress, and activates STAT-3 and NF-kappa B. *Proc. Natl. Acad. Sci. U.S.A.* 98, 9599–9604. doi: 10.1073/pnas.171311298
- Gonzalez, S. A., and Keffe, E. B. (2011). Chronic viral hepatitis: epidemiology, molecular biology, and antiviral therapy. *Front. Biosci.* 16, 225–250. doi: 10.2741/3685
- Goodier, J. L. (2014). Retrotransposition in tumors and brains. *Mob. DNA* 5:11. doi: 10.1186/1759-8753-5-11
- Goodier, J. L., Cheung, L. E., and Kazazian, H. H. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet.* 8:e1002941. doi: 10.1371/journal.pgen.1002941
- Grakoui, A., Wychowski, C., Lin, C., Feinstone, S. M., and Rice, C. M. (1993). Expression and identification of hepatitis C virus polyprotein cleavage products. *J. Virol.* 67, 1385–1395.
- Harada, K., Baba, Y., Ishimoto, T., Chikamoto, A., Kosumi, K., Hayashi, H., et al. (2015). LINE-1 methylation level and patient prognosis in a database of 208 hepatocellular carcinomas. *Ann. Surg. Oncol.* 22, 1280–1287. doi: 10.1245/s10434-014-4134-3
- Harris, R. S., and Liddament, M. T. (2004). Retroviral restriction by APOBEC proteins. *Nat. Rev. Immunol.* 4, 868–877. doi: 10.1038/nri1489
- Hatziaepostolou, M., and Iliopoulos, D. (2011). Epigenetic aberrations during oncogenesis. *Cell. Mol. Life Sci.* 68, 1681–1702. doi: 10.1007/s00018-010-0624-z
- Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., and Shimotohno, K. (1991). Gene mapping of the putative structural region of the hepatitis C virus genome by *in vitro* processing analysis. *Proc. Natl. Acad. Sci. U.S.A.* 88, 5547–5551. doi: 10.1073/pnas.88.13.5547
- Horie, M., Honda, T., Suzuki, Y., Kobayashi, Y., Daito, T., Oshida, T., et al. (2010). Endogenous non-retroviral RNA virus elements in mammalian genomes. *Nature* 463, 84–87. doi: 10.1038/nature08695
- Hur, K., Cejas, P., Feliu, J., Moreno-Rubio, J., Burgos, E., Boland, C. R., et al. (2014). Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. *Gut* 63, 635–646. doi: 10.1136/gutjnl-2012-304219
- Ishizu, H., Siomi, H., and Siomi, M. C. (2012). Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes Dev.* 26, 2361–2373. doi: 10.1101/gad.203786.112
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020
- Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., and Forman, D. (2011). Global cancer statistics. *CA. Cancer J. Clin.* 61, 69–90. doi: 10.3322/caac.20107
- Katze, M. G., He, Y., and Gale, M. (2002). Viruses and interferon: a fight for supremacy. *Nat. Rev. Immunol.* 2, 675–687. doi: 10.1038/nri888
- Kinamoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., et al. (2007). All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res.* 35, 2955–2964. doi: 10.1093/nar/gkm181
- Kitamura, K., Wang, Z., Chowdhury, S., Simadu, M., Koura, M., and Muramatsu, M. (2013). Uracil DNA glycosylase counteracts APOBEC3G-induced

- hypermethylation of hepatitis B viral genomes: excision repair of covalently closed circular DNA. *PLoS Pathog.* 9:e1003361. doi: 10.1371/journal.ppat.1003361
- Klawitter, S., Fuchs, N. V., Upton, K. R., Muñoz-Lopez, M., Shukla, R., Wang, J., et al. (2016). Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat. Commun.* 7:10286. doi: 10.1038/ncomms10286
- Kumar, M., Jung, S. Y., Hodgson, A. J., Madden, C. R., Qin, J., and Slagle, B. L. (2011). Hepatitis B virus regulatory HBx protein binds to adaptor protein IPS-1 and inhibits the activation of beta interferon. *J. Virol.* 85, 987–995. doi: 10.1128/JVI.01825-10
- 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
- Lau, C.-C., Sun, T., Ching, A. K. K., He, M., Li, J.-W., Wong, A. M., et al. (2014). Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell* 25, 335–349. doi: 10.1016/j.ccr.2014.01.030
- Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971. doi: 10.1126/science.1222077
- Levin, H. L., and Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627. doi: 10.1038/nrg3030
- Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C. M., Ikeda, M., et al. (2005). Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the toll-like receptor 3 adaptor protein TRIF. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2992–2997. doi: 10.1073/pnas.0408824102
- Li, K., Li, N. L., Wei, D., Pfeffer, S. R., Fan, M., and Pfeffer, L. M. (2012). Activation of chemokine and inflammatory cytokine response in hepatitis C virus-infected hepatocytes depends on toll-like receptor 3 sensing of hepatitis C virus double-stranded RNA intermediates. *Hepatology* 55, 666–675. doi: 10.1002/hep.24763
- Li, X.-D., Sun, L., Seth, R. B., Pineda, G., and Chen, Z. J. (2005). Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17717–17722. doi: 10.1073/pnas.0508531102
- Liu, Y., Li, J., Chen, J., Li, Y., Wang, W., Du, X., et al. (2015). Hepatitis B virus polymerase disrupts K63-linked ubiquitination of STING to block innate cytosolic DNA-sensing pathways. *J. Virol.* 89, 2287–2300. doi: 10.1128/JVI.02760-14
- Lütgehetmann, M., Bornscheuer, T., Volz, T., Allweiss, L., Bockmann, J.-H., Pollok, J. M., et al. (2011). Hepatitis B virus limits response of human hepatocytes to interferon- α in chimeric mice. *Gastroenterology* 140, 2074–2083, 2083.e1–2. doi: 10.1053/j.gastro.2011.02.057
- Mao, R., Nie, H., Cai, D., Zhang, J., Liu, H., Yan, R., et al. (2013). Inhibition of hepatitis B virus replication by the host zinc finger antiviral protein. *PLoS Pathog.* 9:e1003494. doi: 10.1371/journal.ppat.1003494
- Marusawa, H., Hijikata, M., Chiba, T., and Shimotohno, K. (1999). Hepatitis C virus core protein inhibits fas- and tumor necrosis factor alpha-mediated apoptosis via NF-kappaB activation. *J. Virol.* 73, 4713–4720.
- Moriya, K., Fujie, H., Shintani, Y., Yotsuyanagi, H., Tsutsumi, T., Ishibashi, K., et al. (1998). The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice. *Nat. Med.* 4, 1065–1067. doi: 10.1038/2053
- Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, J., Cichutek, K., et al. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.* 281, 22161–22172. doi: 10.1074/jbc.M601716200
- Muotri, A. R., Chu, V. T., Marchetto, M. C. N., Deng, W., Moran, J. V., and Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910. doi: 10.1038/nature03663
- Nault, J.-C., Datta, S., Imbeaud, S., Franconi, A., Mallet, M., Couchy, G., et al. (2015). Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* 47, 1187–1193. doi: 10.1038/ng.3389
- Nguyen, D. H., Ludgate, L., and Hu, J. (2008). Hepatitis B virus-cell interactions and pathogenesis. *J. Cell. Physiol.* 216, 289–294. doi: 10.1002/jcp.21416
- Okazaki, I., Hiai, H., Kakazu, N., Yamada, S., Muramatsu, M., Kinoshita, K., et al. (2003). Constitutive expression of AID leads to tumorigenesis. *J. Exp. Med.* 197, 1173–1181. doi: 10.1084/jem.20030275
- Ostertag, E. M., and Kazanietz, H. H. (2001). Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* 11, 2059–2065. doi: 10.1101/gr.205701
- Park, K.-J., Choi, S.-H., Lee, S. Y., Hwang, S. B., and Lai, M. M. C. (2002). Nonstructural 5A protein of hepatitis C virus modulates tumor necrosis factor alpha-stimulated nuclear factor kappa B activation. *J. Biol. Chem.* 277, 13122–13128. doi: 10.1074/jbc.M111599200
- Parkin, D. M. (2006). The global health burden of infection-associated cancers in the year 2002. *Int. J. Cancer* 118, 3030–3044. doi: 10.1002/ijc.21731
- Peng, Z.-G., Zhao, Z.-Y., Li, Y.-P., Wang, Y.-P., Hao, L.-H., Fan, B., et al. (2011). Host apolipoprotein B messenger RNA-editing enzyme catalytic polypeptide-like 3G is an innate defensive factor and drug target against hepatitis C virus. *Hepatology* 53, 1080–1089. doi: 10.1002/hep.24160
- Ray, R. B., Lagging, L. M., Meyer, K., Steele, R., and Ray, R. (1995). Transcriptional regulation of cellular and viral promoters by the hepatitis C virus core protein. *Virus Res.* 37, 209–220. doi: 10.1016/0168-1702(95)00034-N
- Ray, R. B., Steele, R., Basu, A., Meyer, K., Majumder, M., Ghosh, A. K., et al. (2002). Distinct functional role of hepatitis C virus core protein on NF-kappaB regulation is linked to genomic variation. *Virus Res.* 87, 21–29. doi: 10.1016/S0168-1702(02)00046-1
- Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., et al. (2014). Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* 184, 1280–1286. doi: 10.1016/j.ajpath.2014.01.007
- Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P., et al. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* 472, 481–485. doi: 10.1038/nature09907
- Schumacher, A. J., Haché, G., Macduff, D. A., Brown, W. L., and Harris, R. S. (2008). The DNA deaminase activity of human APOBEC3G is required for Ty1, MusD, and human immunodeficiency virus type 1 restriction. *J. Virol.* 82, 2652–2660. doi: 10.1128/JVI.02391-07
- Schumann, G. G., Gogvadze, E. V., Osanai-Futahashi, M., Kuroki, A., Münk, C., Fujiwara, H., et al. (2010). Unique functions of repetitive transcriptomes. *Int. Rev. Cell Mol. Biol.* 285, 115–188. doi: 10.1016/B978-0-12-381047-2.00003-7
- Sengupta, S., Powell, E., Kong, L., and Blackard, J. T. (2013). Effects of HCV on basal and tat-induced HIV LTR activation. *PLoS ONE* 8:e64956. doi: 10.1371/journal.pone.0064956
- Shimotohno, K. (2000). Hepatitis C virus and its pathogenesis. *Semin. Cancer Biol.* 10, 233–240. doi: 10.1006/scbi.2000.0322
- Shitani, M., Sasaki, S., Akutsu, N., Takagi, H., Suzuki, H., Nojima, M., et al. (2012). Genome-wide analysis of DNA methylation identifies novel cancer-related genes in hepatocellular carcinoma. *Tumour Biol.* 33, 1307–1317. doi: 10.1007/s13277-012-0378-3
- Shrivastava, A., Manna, S. K., Ray, R., and Aggarwal, B. B. (1998). Ectopic expression of hepatitis C virus core protein differentially regulates nuclear transcription factors. *J. Virol.* 72, 9722–9728.
- Shukla, R., Upton, K. R., Muñoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111. doi: 10.1016/j.cell.2013.02.032
- Silva, E., Marques, S., Osório, H., Carvalheira, J., and Thompson, G. (2012). Endogenous hepatitis C virus homolog fragments in European rabbit and hare genomes replicate in cell culture. *PLoS ONE* 7:e49820. doi: 10.1371/journal.pone.0049820
- Singer, T., McConnell, M. J., Marchetto, M. C. N., Coufal, N. G., and Gage, F. H. (2010). LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci.* 33, 345–354. doi: 10.1016/j.tins.2010.04.001
- Srikanta, D., Sen, S. K., Conlin, E. M., and Batzer, M. A. (2009). Internal priming: an opportunistic pathway for L1 and Alu retrotransposition in hominins. *Gene* 448, 233–241. doi: 10.1016/j.gene.2009.05.014
- Takai, A., Toyoshima, T., Uemura, M., Kitawaki, Y., Marusawa, H., Hiai, H., et al. (2009). A novel mouse model of hepatocarcinogenesis triggered by AID causing deleterious p53 mutations. *Oncogene* 28, 469–478. doi: 10.1038/onc.2008.415
- Tateishi, R., and Omata, M. (2012). Hepatocellular carcinoma in 2011: genomics in hepatocellular carcinoma—a big step forward. *Nat. Rev. Gastroenterol. Hepatol.* 9, 69–70. doi: 10.1038/nrgastro.2011.255
- Taylor, D. R., Shi, S. T., Romano, P. R., Barber, G. N., and Lai, M. M. (1999). Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science* 285, 107–110. doi: 10.1126/science.285.5424.107
- Tuplin, A., Wood, J., Evans, D. J., Patel, A. H., and Simmonds, P. (2002). Thermodynamic and phylogenetic prediction of RNA secondary structures in the coding region of hepatitis C virus. *RNA* 8, 824–841. doi: 10.1017/S1355838202554066

- van den Hurk, J. A. J. M., Meij, I. C., Seleme, M., del, C., Kano, H., and Nikopoulos, K., Hoefsloot, L. H., et al. (2007). L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* 16, 1587–1592. doi: 10.1093/hmg/ddm108
- Vartanian, J.-P., Henry, M., Marchio, A., Suspène, R., Aynaud, M.-M., Guétard, D., et al. (2010). Massive APOBEC3 editing of hepatitis B viral DNA in cirrhosis. *PLoS Pathog.* 6:e1000928. doi: 10.1371/journal.ppat.1000928
- Vieira, V. C., and Soares, M. A. (2013). The role of cytidine deaminases on innate immune responses against human viral infections. *Biomed Res. Int.* 2013:683095. doi: 10.1155/2013/683095
- Vincent, I. E., Zannetti, C., Lucifora, J., Norder, H., Protzer, U., Hainaut, P., et al. (2011). Hepatitis B virus impairs TLR expression and function in plasmacytoid dendritic cells. *PLoS ONE* 6:e26315. doi: 10.1371/journal.pone.0026315
- Visvanathan, K., Skinner, N. A., Thompson, A. J. V., 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
- Wang, H., and Ryu, W.-S. (2010). Hepatitis B virus polymerase blocks pattern recognition receptor signaling via interaction with DDX3: implications for immune evasion. *PLoS Pathog.* 6:e1000986. doi: 10.1371/journal.ppat.1000986
- Waris, G., Livolsi, A., Imbert, V., Peyron, J.-F., and Siddiqui, A. (2003). Hepatitis C virus NS5A and subgenomic replicon activate NF-kappa B via tyrosine phosphorylation of IkappaBalpha and its degradation by calpain protease. *J. Biol. Chem.* 278, 40778–40787. doi: 10.1074/jbc.M303248200
- Watashi, K., Hijikata, M., Marusawa, H., Doi, T., and Shimotohno, K. (2001). Cytoplasmic localization is important for transcription factor nuclear factor-kappa B activation by hepatitis C virus core protein through its amino terminal region. *Virology* 286, 391–402. doi: 10.1006/viro.2001.0990
- Wei, C., Ni, C., Song, T., Liu, Y., Yang, X., Zheng, Z., et al. (2010). The hepatitis B virus X protein disrupts innate immunity by downregulating mitochondrial antiviral signaling protein. *J. Immunol.* 185, 1158–1168. doi: 10.4049/jimmunol.0903874
- Whittaker, S., Marais, R., and Zhu, A. X. (2010). The role of signaling pathways in the development and treatment of hepatocellular carcinoma. *Oncogene* 29, 4989–5005. doi: 10.1038/onc.2010.236
- Wolff, E. M., Byun, H.-M., Han, H. F., Sharma, S., Nichols, P. W., Siegmund, K. D., et al. (2010). Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS Genet.* 6:e1000917. doi: 10.1371/journal.pgen.1000917
- Woltman, A. M., Op den Brouw, M. L., Biesta, P. J., Shi, C. C., and Janssen, H. L. A. (2011). Hepatitis B virus lacks immune activating capacity, but actively inhibits plasmacytoid dendritic cell function. *PLoS ONE* 6:e15324. doi: 10.1371/journal.pone.0015324
- Wu, J., Meng, Z., Jiang, M., Pei, R., Trippler, M., Broering, R., et al. (2009). Hepatitis B virus suppresses toll-like receptor-mediated innate immune responses in murine parenchymal and nonparenchymal liver cells. *Hepatology* 49, 1132–1140. doi: 10.1002/hep.22751
- Yan, R., Zhao, X., Cai, D., Liu, Y., Block, T. M., Guo, J.-T., et al. (2015). The interferon-inducible protein tetherin inhibits hepatitis B virus virion secretion. *J. Virol.* 89, 9200–9212. doi: 10.1128/JVI.00933-15
- Yu, Q., Carbone, C. J., Katlinskaya, Y. V., Zheng, H., Zheng, K., Luo, M., et al. (2015). Type I interferon controls propagation of long interspersed element-1. *J. Biol. Chem.* 290, 10191–10199. doi: 10.1074/jbc.M114.612374
- Yu, S., Chen, J., Wu, M., Chen, H., Kato, N., and Yuan, Z. (2010). Hepatitis B virus polymerase inhibits RIG-I- and toll-like receptor 3-mediated beta interferon induction in human hepatocytes through interference with interferon regulatory factor 3 activation and dampening of the interaction between TBK1/IKKepsilon and. *J. Gen. Virol.* 91, 2080–2090. doi: 10.1099/vir.0.020552-0
- Zemer, R., Kitay Cohen, Y., Naftaly, T., and Klein, A. (2008). Presence of hepatitis C virus DNA sequences in the DNA of infected patients. *Eur. J. Clin. Invest.* 38, 845–848. doi: 10.1111/j.1365-2362.2008.02029.x
- Zhang, C., Fan, L., Fan, T., Wu, D., Gao, L., Ling, Y., et al. (2013). Decreased PADI4 mRNA association with global hypomethylation in hepatocellular carcinoma during HBV exposure. *Cell Biochem. Biophys.* 65, 187–195. doi: 10.1007/s12013-012-9417-3
- Zhang, S., Kodys, K., Li, K., and Szabo, G. (2013). Human type 2 myeloid dendritic cells produce interferon-λ and amplify interferon-α in response to hepatitis C virus infection. *Gastroenterology* 144, 414–425.e7. doi: 10.1053/j.gastro.2012.10.034
- Zhu, C., Utsunomiya, T., Ikemoto, T., Yamada, S., Morine, Y., Imura, S., et al. (2014). Hypomethylation of long interspersed nuclear element-1 (LINE-1) is associated with poor prognosis via activation of c-MET in hepatocellular carcinoma. *Ann. Surg. Oncol.* 21(Suppl. 4), S729–S735. doi: 10.1245/s10434-014-3874-4

Conflict of Interest Statement: The author declares 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 © 2016 Honda. 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) or licensor 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.



Guardian of the Human Genome: Host Defense Mechanisms against LINE-1 Retrotransposition

Yasuo Ariumi*

Ariumi Project Laboratory, Center for AIDS Research and International Research Center for Medical Sciences, Kumamoto University, Kumamoto, Japan

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Formerly affiliated with University of
Toledo, USA

Reviewed by:

Geoff Faulkner,
Mater Medical Research Institute,
Australia
Nemanja Rodic,
Yale University, USA

*Correspondence:

Yasuo Ariumi
ariumi@kumamoto-u.ac.jp

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 14 February 2016

Accepted: 14 June 2016

Published: 28 June 2016

Citation:

Ariumi Y (2016) Guardian of the
Human Genome: Host Defense
Mechanisms against LINE-1
Retrotransposition.
Front. Chem. 4:28.
doi: 10.3389/fchem.2016.00028

Long interspersed element type 1 (LINE-1, L1) is a mobile genetic element comprising about 17% of the human genome, encoding a newly identified ORF0 with unknown function, ORF1p with RNA-binding activity and ORF2p with endonuclease and reverse transcriptase activities required for L1 retrotransposition. L1 utilizes an endonuclease (EN) to insert L1 cDNA into target DNA, which induces DNA double-strand breaks (DSBs). The ataxia-telangiectasia mutated (ATM) is activated by DSBs and subsequently the ATM-signaling pathway plays a role in regulating L1 retrotransposition. In addition, the host DNA repair machinery such as non-homologous end-joining (NHEJ) repair pathway is also involved in L1 retrotransposition. On the other hand, L1 is an insertional mutagenic agent, which contributes to genetic change, genomic instability, and tumorigenesis. Indeed, high-throughput sequencing-based approaches identified numerous tumor-specific somatic L1 insertions in variety of cancers, such as colon cancer, breast cancer, and hepatocellular carcinoma (HCC). In fact, L1 retrotransposition seems to be a potential factor to reduce the tumor suppressive property in HCC. Furthermore, recent study demonstrated that a specific viral-human chimeric transcript, HBx-L1, contributes to hepatitis B virus (HBV)-associated HCC. In contrast, host cells have evolved several defense mechanisms protecting cells against retrotransposition including epigenetic regulation through DNA methylation and host defense factors, such as APOBEC3, MOV10, and SAMHD1, which restrict L1 mobility as a guardian of the human genome. In this review, I focus on somatic L1 insertions into the human genome in cancers and host defense mechanisms against deleterious L1 insertions.

Keywords: LINE-1, retrotransposition, DNA double-strand breaks (DSBs), DNA repair, tumor suppressor, HBV, epigenetic regulation, somatic insertion

INTRODUCTION

Long interspersed element type 1 (LINE-1, L1) is an active and autonomous non-long terminal repeat (LTR) retrotransposon composing about 17% of the human genome and L1 is an essential evolutionary force (DeBerardinis et al., 1998; Ostertag and Kazazian, 2001; Cordaux and Batzer, 2009; Hancks and Kazazian, 2012). However, only 100 copies out of ~500,000 copies still remain active (Brouha et al., 2003; Mills et al., 2007; Beck et al., 2010). The remaining L1s are 5' truncated and defective. Furthermore, L1 provides the *trans*-acting functions required for the retrotransposition of non-autonomous retrotransposons such as short interspersed

element (SINE), which includes Alu repeats in humans, SINE-VNTR-Alu (SVA), and processed pseudogenes (Esnault et al., 2000; Dewannieux et al., 2003; Hancks et al., 2011).

L1 encodes three open reading frames, a newly identified ORF0 with unknown function, ORF1p with RNA-binding and nucleic acid chaperon activities, and ORF2p with AP-like endonuclease (EN) and reverse transcriptase (RT) activities required for L1 retrotransposition (Mathias et al., 1991; Martin and Bushman, 2001; Ostertag and Kazazian, 2001; Hancks and Kazazian, 2012; Denli et al., 2015). ORF0 is the primate-specific ORF in the anti-sense 5' untranslated region (UTR) of L1 (Denli et al., 2015). ORF0 predominantly localizes in nuclear PML-adjacent foci and enhances L1 mobility. ORF1p and ORF2p preferentially assemble with L1 RNA and form a ribonucleoprotein (RNP) in the cytoplasmic foci (Goodier et al., 2007; Doucet et al., 2010). Although retroviruses and LTR-retrotransposons utilize a long terminal repeat (LTR) to synthesize full-length transcripts, L1 instead utilizes an internal promoter in the 5'UTR of L1 (Swergold, 1990). Several transcription factors including SOX11 (Tchenio et al., 2000), YY1 (Becker et al., 1993; Athanikar et al., 2004), RUNX3 (Yang et al., 2003), and p53 (Harris et al., 2009) positively regulate the L1 transcription. On the other hand, SOX2 (Muotri et al., 2005) and SRY (Tchenio et al., 2000) as well as several epigenetic factors negatively regulate the L1 transcription (Table 1).

L1 integrates into the genome by target-primed reverse transcription (TPRT) (Luan et al., 1993) after the L1-RNP complex enters the nucleus. During TPRT, the L1 EN creates a nicked DNA that serves as a primer for reverse transcription of L1 RNA, leading to integration of L1 cDNA into the human genome (Feng et al., 1996). A typical L1 EN cleavage site is 5'-TTT/AA-3' (Feng et al., 1996; Cost and Boeke, 1998). Thus, L1 insertion generates DNA double-strand breaks (DSBs) as well as L1 structural hallmarks such as frequent 5' truncations, 3' poly(A) tails and variable length target site duplications (TSDs) in the target DNA. L1 can alter the mammalian genome in many ways upon retrotransposition, since the insertion of L1 into the human genome may cause genomic instability, genetic disorders, and cancers through insertional mutagenesis (Kazazian et al., 1988; Morse et al., 1988; Miki et al., 1992; Narita et al., 1993; Holmes et al., 1994; Gilbert et al., 2002; Morrish et al., 2002; Symer et al., 2002; Belancio et al., 2008; Beck et al., 2011; Hancks and Kazazian, 2012; Bundo et al., 2014; Kines et al., 2014; Figure 1). So far, >100 disease-causing retrotransposon insertions have been identified in humans [26 L1, 61 Alu, 12 SVA, 4 poly(A)] (Figure 1).

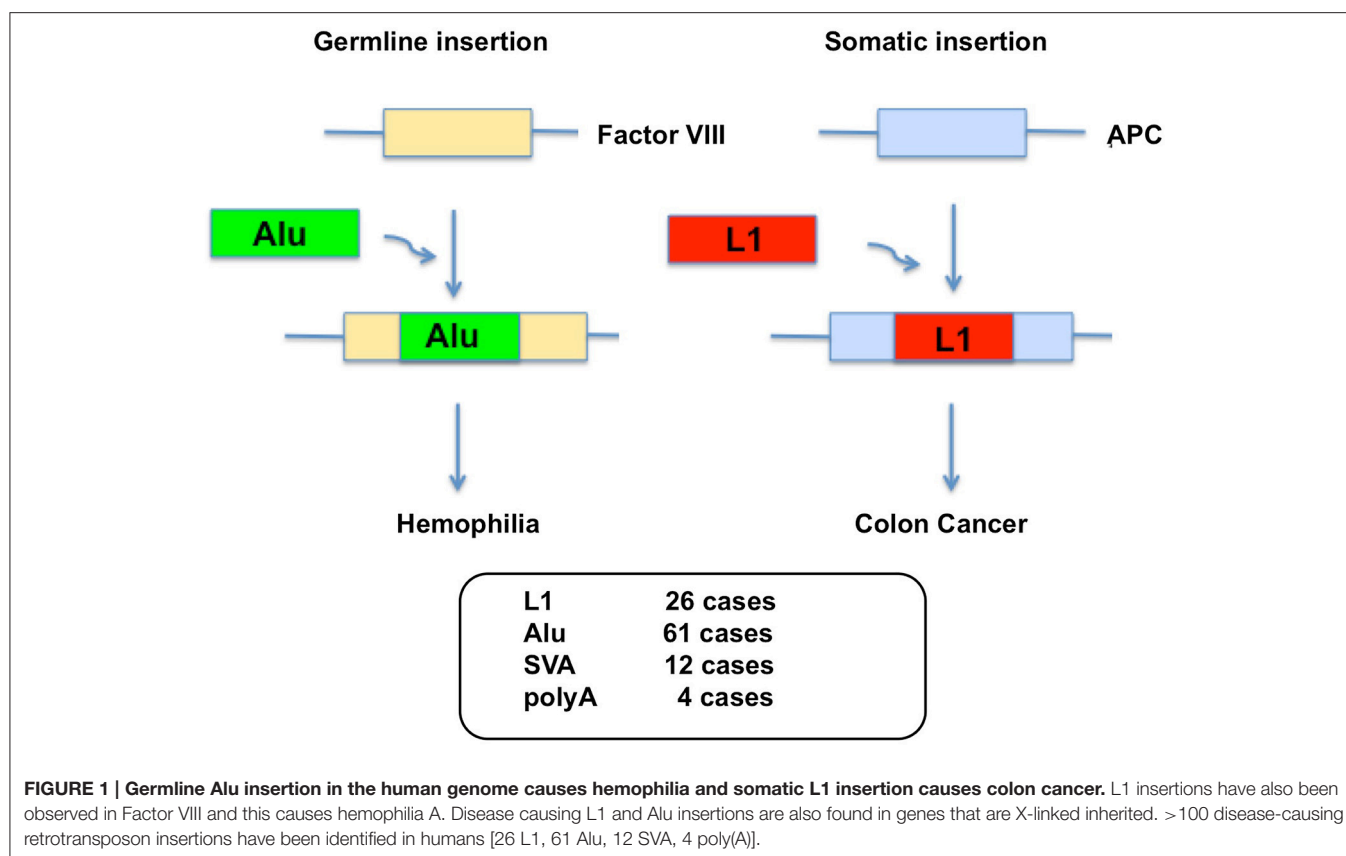
L1-MEDIATED DSBs INDUCTION AND DNA REPAIR MACHINERY

L1 is known to induce DSBs in target DNA by L1 EN activity (Gasior et al., 2006). The ataxia-telangiectasia mutated (ATM) is activated by DSBs and subsequently phosphorylates downstream substrates including p53, Chk2, BRCA1 and the MRE11-Rad50-NBS1 (MRN) complex, resulting in the activation of DNA damage checkpoint and cell cycle arrest (Harper

TABLE 1 | Host factors regulating the L1 transcription.

Host factors	Functions	References
POSITIVE FACTOR		
RNA polymerase II	RNA polymerase	Swergold, 1990; Beck et al., 2011
SOX11	Transcription factor	Tchenio et al., 2000
YY1	Transcription factor	Becker et al., 1993; Athanikar et al., 2004
RUNX3	Transcription factor	Yang et al., 2003
p53	Transcription factor	Harris et al., 2009
NEGATIVE FACTOR		
MeCP2	Methyl-CpG-binding protein	Yu et al., 2001; Muotri et al., 2010
KAP1/TRIM28	Cofactor of KRAB zinc finger protein	Rowe et al., 2010; Castro-Diaz et al., 2014
SETDB1/ESET	Histone methyltransferase	Matsui et al., 2010
DNMT1, DNMT3a, DNMT3b	DNA methyltransferase	Liang et al., 2002
ZNF93	KRAB zinc finger protein	Jacobs et al., 2014
PLZF	Transcription factor	Puszyk et al., 2013
SIRT6	Mono-ADP-ribosyl transferase	Van Meter et al., 2014
SOX2	Transcription factor	Muotri et al., 2005
SRY	Transcription factor	Tchenio et al., 2000
p53	Tumor suppressor	Wylie et al., 2016
Rb	Tumor suppressor	Montoya-Durango et al., 2009, 2016
HDAC1	Histone deacetylase	Montoya-Durango et al., 2009
HDAC2	Histone deacetylase	Montoya-Durango et al., 2009, 2016
E2F	Transcription factor	Montoya-Durango et al., 2009
NuRD	Nucleosomal and remodeling deacetylase	Montoya-Durango et al., 2016

and Elledge, 2007; Ciccia and Elledge, 2010; Shiloh, 2014). Accordingly, L1 retrotransposition was increased in ATM-deficient cells, indicating ATM signaling pathway modulates L1 retrotransposition (Coufal et al., 2011). In contrast, the E6 protein from β -human papillomavirus (β -HPV 5 and 8) reduces ATM protein levels and attenuates L1 retrotransposition, suggesting that ATM is needed for efficient L1 retrotransposition (Wallace et al., 2013). Thus, the DNA damage response may modulate L1 retrotransposition. Notably, L1 can integrate into preformed DSBs generated independently of L1 EN, resulting in retrotransposon-mediated DNA repair (Morrish et al., 2002). Furthermore, host DNA repair machinery may also impact L1 retrotransposition. Gasior et al reported that DNA repair enzyme ERCC1/XPF heterodimer limits L1 retrotransposition (Gasior et al., 2008). Importantly, deficiencies of non-homologous end-joining (NHEJ) repair pathway such as Ku70, Artemis, and DNA ligase IV (LigIV) decrease retrotransposition frequencies of human L1 in chicken DT40 cells, suggesting that the NHEJ repair



pathway is required for efficient L1 retrotransposition (Suzuki et al., 2009).

L1 RETROTRANSPOSITION IN CANCERS

Somatic L1 insertions are seldom observed in normal tissues except hippocampus (Baillie et al., 2011; Evrony et al., 2012; Upton et al., 2015). Although most L1 retrotransposition was thought to occur in the germline, somatic L1 insertions were also found to occur in variety of tumors, including breast cancer, colon cancer, hepatocellular carcinoma (HCC), and lung cancer (Miki et al., 1992; Liu et al., 1997; Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012; Shukla et al., 2013; Carreira et al., 2014; Helman et al., 2014; Ewing et al., 2015; **Table 2**). First, three L1 insertion candidates were reported in human tumors (Morse et al., 1988; Miki et al., 1992; Liu et al., 1997). However, two insertions described by Liu et al. (1997) and Morse et al. (1988) lack all of the hallmark features of a true somatic retrotransposition event, such as L1 endonuclease cleavage site, the presence of L1 poly(A) tail, target-site duplication (TSD), 5' truncation and inversion, and 3' transduction (Holmes et al., 1994; Moran et al., 1999; Goodier et al., 2000; Pickeral et al., 2000; Szak et al., 2002; **Table 2**). These insertions may be derived from recombination events, L1 EN-independent insertions (Morrish et al., 2002), or other atypical integration mechanisms of L1 retrotransposition. Indeed, an L1 insertion disrupts the adenomatous polyposis

coli (APC) gene in a colon cancer, indicating the disruption of a tumor suppressor gene caused by somatic L1 insertion (Miki et al., 1992). Accordingly, a recent study identified a novel somatic insertion in the APC gene and a hot spot for L1 insertion on Chromosome 17, suggesting that the L1 insertion initiates colorectal cancer (CRC) by mutating the APC gene through the classic two-hit CRC pathway (Scott et al., 2016). Furthermore, high-throughput sequencing-based approaches identified numerous somatic tumor-specific insertions in cancers (Miki et al., 1992; Liu et al., 1997; Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012; Shukla et al., 2013; Carreira et al., 2014; Helman et al., 2014; Tubio et al., 2014). Indeed, Lee et al. identified the L1 insertions in cadherin-12 (CDH12), roundabout, axon guidance receptor, homolog 2 (ROBO2), NRXN3, FPR2, COL11A1, NEGR1, NTM, and CTNNA2 (Lee et al., 2012). As well, Solyom et al. identified several tumor-specific insertions in colorectal tumors including odd Oz/ten-m homolog 3 (ODZ3), ROBO2, protein tyrosine phosphatase, receptor type, M (PTPRM), pericentriolar material 1 (PCM1), CDH11, and runt-related transcription factor 1 (RUNX1T1) of colorectal cancers (Solyom et al., 2012). All insertions were severely 5' truncated. Interestingly, these genes are associated with cell-adhesion functions and both groups could identify the L1 insertions in the same ROBO2 genes, suggesting the potential role of cell-adhesion genes in L1 insertion-mediated colorectal tumorigenesis. In addition, Tubio et al. analyzed the somatic L1 retrotransposition activities in 290 cancers and noticed insertions

TABLE 2 | L1 insertions in cancers.

Tumor type	L1 insertions and target genes	References
Barrett's esophagus (BE)	+ 46 somatic L1 insertions	Doucet-O'Hare et al., 2015
Breast cancer	+ Myc (*)	Morse et al., 1988
Colorectal cancer	+ APC	Miki et al., 1992; Scott et al., 2016
	+ ODZ3, ROBO2, PTPRM, PCM1, CDH11, RUNX1T1	Solyom et al., 2012
	+ 25 somatic L1 insertions ROBO2, CDH12, NRXN3, FPR2, COL11A1, NEGR1, NTM, CTNNA2	Lee et al., 2012
	+ 57 somatic L1 insertions CYLD, HDAC9	Ewing et al., 2015
Colonic adenoma	+ 29 somatic L1 insertions STX11, PANK1	Ewing et al., 2015
Desmoplastic small round cell tumor	+ t(11;22) translocation breakpoint(*) EWS-WT1	Liu et al., 1997
Endometrial carcinoma	+ PTEN	Helman et al., 2014
Esophageal adenocarcinoma	+ 75 somatic L1 insertions	Doucet-O'Hare et al., 2015
Familial retinoblastoma	+ RB1	Rodríguez-Martín et al., 2016
Gastric cancer	+ 23 somatic L1 insertions ELOVL4, CNTNAP2, RIMS2	Ewing et al., 2015
Glioblastoma	–	Iskow et al., 2010; Lee et al., 2012
Hepatocellular carcinoma (HCC)	+ MCC, ST18	Shukla et al., 2013
	HBV integration in L1 (HBx-LINE1)	Lau et al., 2014
Head and neck carcinoma	+	Helman et al., 2014
Lung cancer	+ 9 somatic L1 insertions	Iskow et al., 2010
Medulloblastoma	–	Iskow et al., 2010
Multiple Myeloma	–	Lee et al., 2012
Neurofibromatosis type 1 (NF1)	+ NF1 3 L1 insertions	Wimmer et al., 2011
Ovarian tumors	+ 13 somatic L1 insertions	Lee et al., 2012
Pancreatic cancer	+ 24 somatic L1 insertions SOX6, APAF1, GDNF	Ewing et al., 2015
Pancreatic ductal adenocarcinoma (PDAC)	+ 465 somatic L1 insertions In 20 PDAC cases	Rodić et al., 2015
Prostate tumors	+	Lee et al., 2012

*Lack of the hallmark features of a true somatic retrotransposition event (Morse et al., 1988; Liu et al., 1997).

occurring during cancer development. 53% of the patients have at least one somatic L1 retrotransposition event, of which 24% were 3' transductions, most frequently colorectal cancers (93%)

and lung cancers (75%), suggesting that 3' transduction are potentially mutagenic. Somatic L1 retrotranspositions tend to insert in intergenic or heterochromatin regions of the cancer genome (Tubio et al., 2014). Furthermore, somatic L1 insertions participate in the dynamics of many tumor genomes and lead to driver mutations. Surprisingly, L1 insertion was reported in colonic adenoma, a known cancer precursor, suggesting that widespread somatic L1 retrotransposition occurs early during development of gastrointestinal (GI) tumors, probably before dysplastic growth (Ewing et al., 2015). Similarly, a recent study demonstrated that L1 retrotransposition is active in esophageal adenocarcinoma and its precursor, Barrett's esophagus (BE), indicating that somatic L1 insertions occur early in BE and esophageal adenocarcinoma. Notably, two L1 insertions were detected in normal esophagus, indicating that some L1 insertions may occur in normal squamous epithelium cells (Doucet-O'Hare et al., 2015). In this regard, most of the new somatic insertions are truncated, and would not mobilize again. So mutations arising from insertions in the normal precursor esophageal or benign BE would be contributing to tumorigenesis. Otherwise, only a rare full-length somatic insertion has the potential to contribute to mutation during the various stages of transition to tumorigenesis. In addition, L1 insertions in pancreatic ductal adenocarcinoma (PDAC) were reported with discordant rate of retrotransposition between primary and metastatic sites, suggesting that L1 insertions in gastrointestinal neoplasms occur discontinuously. Thus, somatic L1 insertions contribute to genetic and phenotypic heterogeneity in PDAC (Rodić et al., 2015). Interestingly, somatic insertions were identified in epithelial tumors but not in blood or brain cancers (Lee et al., 2012). However, we raise awareness regarding the following limitations of this study. For example, the sample size was small and the normal tissue was not from the same patient. In addition, in this study they only examined multiple myeloma and did not look at the entire spectrum of blood based cancers. In this regard, ten-eleven-translocation (TET) 2, a DNA demethylation-related protein, is frequently mutated in myeloid and lymphoid tumors (Ko et al., 2015). The TET family that oxidizes 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC) in DNA, leads to the DNA demethylation. Since DNA methylation has a pivotal regulatory role in L1 silencing, TET2 may impact L1 mobility. Therefore, L1 insertions may be suppressed in such hematological cancers. Intriguingly, several somatic insertions occur in genes that are commonly mutated in cancers such as tumor suppressor gene. These insertions disrupt the expression of target genes, and are biased toward regions of cancer-specific DNA hypomethylation (Lee et al., 2012). Indeed, recent studies identified somatic L1 insertion in tumor suppressor genes, such as APC and PTEN (Miki et al., 1992; Helman et al., 2014). As well, the first case of familial retinoblastoma (Rb) caused by a *de novo* insertion of a full-length L1 into intron 14 of the Rb gene, resulting in the aberrant and non-canonical mRNA splicing of the Rb gene, was reported (Rodríguez-Martín et al., 2016). Furthermore, 18 retrotransposon insertions [14 Alu, 3 L1, and 1 poly(A)] were identified in neurofibromatosis type 1 (NF1) gene (Wimmer et al., 2011).

Although still debated, cell division seems to be required for efficient L1 retrotransposition (Shi et al., 2007; Xie et al., 2013). In fact, retrotransposition was strongly inhibited in the cells arrested in the G₁, S, G₂, or M phase of cell cycle. The reduction in L1 transcript abundance limits retrotransposition in non-dividing cells, suggesting that inhibition of retrotransposition in non-dividing cells protects somatic cells from accumulation of deleterious mutations caused by L1 insertions (Shi et al., 2007). In contrast, there is an opposite report that L1 retrotransposition was detected in non-dividing and primary human somatic cells using adenovirus-L1 hybrid vector, even though they detected L1 retrotransposition in G₁/S- but not in G₀-arrested cells (Kubo et al., 2006). In addition, retrotransposition was also inhibited during cellular senescence in primary human fibroblasts. So far, several biomarkers of cellular senescence have been identified such as senescence-associated β -galactosidase (SA- β -Gal), p53/p21, p16^{INK4a}, senescence-associated heterochromatin foci (SAHF), senescence-associated secretory phenotype (SASP), autophagy, telomere-induced foci/DNA damage response (DDR), and cell cycle arrest (Kuilman et al., 2010) and the reduction in L1 retrotransposition may be a biomarker of cellular senescence. Thus, cell cycle may affect L1 retrotransposition.

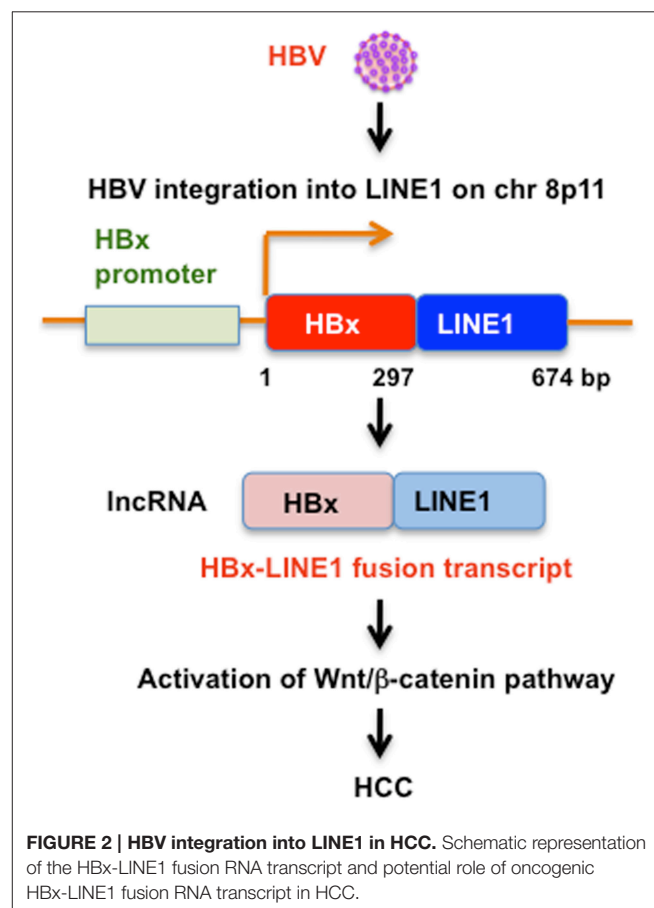
L1 protein expression is a common feature of many types of high-grade malignant tumor, yet is rarely detected in early stage of tumorigenesis (Rodić et al., 2014). L1 promoter is normally silenced by methylation in normal somatic cells (Woodcock et al., 1997; Schulz et al., 2006). In contrast, L1 promoter is hypomethylated (Baba et al., 2014), and expression of L1 is elevated in many tumors. In fact, L1 expression was detected in human breast carcinomas and testicular cancers (Bratthauer and Fanning, 1992; Bratthauer et al., 1994; Nangia-Makker et al., 1998). L1 ORF1p protein is detected in a variety of tumor cells including breast cancer, colon cancer, pancreatic ductal adenocarcinoma, and HCC but not in normal somatic cells (Bratthauer et al., 1994; Asch et al., 1996; Rodić et al., 2014). Thus, L1 ORF1p expression seems to be a hallmark of many human cancers as a highly specific tumor marker.

In addition to expression of L1, a hallmark of tumor cells is an activated telomere maintenance mechanism that allows prolonged survival of the malignant tumor cells. In more than 80% of tumors, telomeres are typically maintained by telomerase. Notably, the reduced length of telomeres was reported in the L1 knockdown cells, indicating that L1 is involved in telomere maintenance in telomerase positive tumor cells (Aschacher et al., 2016). Accordingly, L1 involves in a transcriptional regulation of hTERT and upregulation of its transcription factors c-Myc and KLF-4 (Aschacher et al., 2016). Thus, L1 may contribute to the development of cancers. However, these studies were not done in alternative lengthening of telomeres (ALT)-positive tumors or telomerase negative tumors. Consequently, it is uncertain if L1 is directly contributing to telomere maintenance or if the reduction in telomere length is contributed to the reduction in telomerase levels. Indeed, the stoichiometry of telomerase is important for maintaining telomere length (Armanios et al., 2005; Goldman et al., 2005).

Chronic infection with hepatitis B virus (HBV) is a major risk for the development of HCC. HBV integration into the

human genome was found in most HBV-related HCC and it has been implicated in the development of HCC. An initial study proposed that HBV integration occurs randomly without preferred integration site (Matsubara and Tokino, 1990). However, high-throughput sequencing-based approaches identified recurrent integration sites in HCC (Ding et al., 2012). HBV integration favored chromosome 17 and preferentially integrated into human transcript units. At least, telomerase reverse transcriptase (TERT) and fibronectin 1 (FN1) genes were identified as the recurrent HBV integration sites. Furthermore, seven integrations were found in the repeat regions including L1, LTR/ERV1, and SINE/Alu (Ding et al., 2012). Similarly, a recent transcriptome sequencing study of HBV-positive HCC cell lines discovered that HBV integrates into L1 (Lau et al., 2014). Insertion of the gene encoding hepatitis B virus x (HBx) into L1 on chromosome 8p11 produces an oncogenic HBx-LINE1 chimeric RNA transcript (Lau et al., 2014; **Figure 2**). The HBx-LINE1 RNA transcript was detected in 23.3% of HCC, suggesting that HBx-LINE1 is selected for in HCC oncogenesis. The long non-coding RNA (lncRNA)-like HBx-LINE1 transcript confers cancer-promoting properties through activation of Wnt/ β -catenin signaling pathway (Lau et al., 2014).

Furthermore, endogenous L1-mediated retrotransposition was identified in the germline and somatic cells of HCC patients (Shukla et al., 2013). The germline L1 insertion in the tumor



suppressor mutated in colorectal cancers (MCC) was detected in 21.1% of HCC, resulting in the aberrant expression of MCC. Moreover, suppression of tumorigenicity 18 (ST18) was activated by a tumor-specific somatic L1 insertion (Shukla et al., 2013). Thus, L1-mediated retrotransposition seems to be a potential etiological factor in HCC.

GUARDIAN OF THE HUMAN GENOME: HOST DEFENSE MECHANISMS AGAINST L1 RETROTRANSPPOSITION

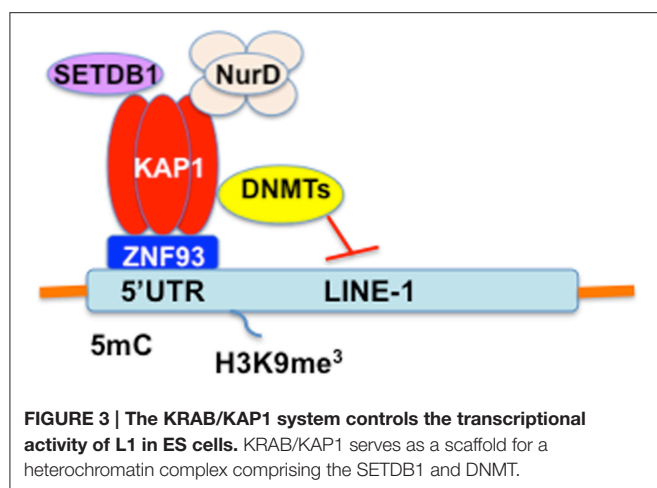
Since insertion of L1 into the human genome may cause human genetic disorders and cancer, retrotransposition must be silenced under normal conditions. To restrict deleterious retrotransposition, host cells have evolved several defense mechanisms protecting cells against retrotransposition including epigenetic regulation through DNA methylation (Burden et al., 2005; Trono, 2015), RNA silencing by RNA interference (Soifer et al., 2005; Yang and Kazazian, 2006), PIWI-interacting RNA (piRNA)-PIWI system (Aravin et al., 2007a,b; Kuramochi-Miyagawa et al., 2008; De Fazio et al., 2011; Marchetto et al., 2013), microRNA (Hamdorf et al., 2015), and host restriction factors, such as apolipoprotein B mRNA editing enzyme catalytic polypeptide-like 3 (APOBEC3), Moloney leukemia virus 10 (MOV10), and SAM domain and HD domain containing protein 1 (SAMHD1) (Table 3).

DNA methylation within the 5'UTR promoter of L1 is essential for maintaining transcriptional inactivation and for inhibiting L1 retrotransposition (Woodcock et al., 1997; Liang et al., 2002; Burden et al., 2005). L1 is highly active during early embryogenesis, while L1 is silenced early in development through epigenetic mechanisms (Table 1). Indeed, methylation of the L1 promoter is maintained by DNA methyltransferases (DNMTs), including DNMT1, DNMT3a, and DNMT3b (Liang et al., 2002). L1 retrotransposition is negatively regulated by methyl-CpG-binding protein 2 (MeCP2)-mediated DNA methylation (Yu et al., 2001; Muotri et al., 2010). In addition, nucleosomal and remodeling deacetylase (NuRD) multiprotein complex specifically enriches the L1 promoter. Rb and E2F recruit to the L1 promoter along with histone deacetylase (HDAC), including HDAC1 and HDAC2 (Montoya-Durango et al., 2009, 2016). Furthermore, KRAB-associated protein1 (KAP1, also known as TRIM28) mediates transcriptional silencing of endogenous retroelements (EREs) including L1, Alu, SVA, and human endogenous retrovirus-K (HERV-K) as well as exogenous retrovirus mouse leukemia virus (MLV) in embryonic stem (ES) cells (Wolf and Goff, 2007, 2009; Matsui et al., 2010; Rowe et al., 2010; Castro-Diaz et al., 2014; Turelli et al., 2014; Trono, 2015). Krüppel-associated box (KRAB)-containing zinc-finger protein (KRAB-ZFP/ZNF), a large family of tetrapod-restricted transcription factors, and a cofactor KAP1 serve as a scaffold for a heterochromatin complex comprising the SETDB1 (also known as ESET) histone methyltransferase, histone deacetylase, nucleosome remodeling, and DNMT activities (Trono, 2015; Figure 3). Furthermore, the protein deacylase and mono-ADP ribosyltransferase Sirtuin 6 (SIRT6) represses L1 mobility by

TABLE 3 | Host defense factors against L1.

Host factors	Functions	References
APOBEC3A	ISG, cytidine deaminase	Bogerd et al., 2006; Chen et al., 2006; Muckenfuss et al., 2006; Kinomoto et al., 2007; Niewiadomska et al., 2007
APOBEC3B	Cytidine deaminase	Bogerd et al., 2006; Muckenfuss et al., 2006; Stenglein and Harris, 2006; Kinomoto et al., 2007
APOBEC3F	ISG, cytidine deaminase anti-viral protein	Muckenfuss et al., 2006; Stenglein and Harris, 2006; Kinomoto et al., 2007; Niewiadomska et al., 2007
APOBEC3G	ISG, cytidine deaminase anti-viral protein	Kinomoto et al., 2007; Niewiadomska et al., 2007
MOV10	ISG, RNA helicase anti-HIV protein	Arjan-Odedra et al., 2012; Goodier et al., 2012, 2013; Li et al., 2013
BST-2	ISG, anti-HIV protein	Goodier et al., 2015
ISG20	ISG, anti-viral protein	Goodier et al., 2015
MAVS	ISG, innate immunity	Goodier et al., 2015
Mx2	ISG, anti-viral protein	Goodier et al., 2015
RNase L	ISG, anti-viral protein	Zhang et al., 2014
SAMHD1	ISG, anti-HIV protein	Zhao et al., 2013; Hu et al., 2015
TREX1	ISG, DNA exonuclease	Stetson et al., 2008
ZAP/PARP13	ISG, poly(ADP-ribose) polymerase	Goodier et al., 2015; Moldovan and Moran, 2015
IFN1	Induction of ISGs, anti-viral protein	Yu et al., 2015
MeCP2	Methyl-CpG-binding protein	Yu et al., 2001; Muotri et al., 2010
KAP1/TRIM28	Cofactor of KRAB zinc finger protein	Rowe et al., 2010; Castro-Diaz et al., 2014
SETDB1/ESET	Histone methyltransferase	Matsui et al., 2010
ZNF93	KRAB zinc finger protein	Jacobs et al., 2014
PLZF	Transcription factor	Puszyk et al., 2013
SIRT6	Mono-ADP-ribosyl transferase	Van Meter et al., 2014
SOX2	Transcription factor	Muotri et al., 2005
SRY	Transcription factor	Tchenio et al., 2000
p53	Tumor suppressor	Wylie et al., 2016
Rb	Tumor suppressor	Montoya-Durango et al., 2009, 2016
ATM	DNA damage sensor, kinase	Coufal et al., 2011
ERCC1/XPF	DNA repair	Gasior et al., 2008
miR-128	microRNA	Hamdorf et al., 2015
piRNA-PIWI	piRNA	De Fazio et al., 2011; Marchetto et al., 2013

ribosylating KAP1 (Van Meter et al., 2014). SIRT6 binds to the 5'UTR of L1 and ribosylates KAP1, resulting in facilitation of KAP1 interaction with the heterochromatin factor HP1 α , thereby contributing to the packaging of L1 into transcriptionally



repressive heterochromatin. Moreover, promyelocytic leukemia zinc finger (PLZF) protein, a member of the POK (POZ and Kruppel zinc finger) family of transcription factors that acts as an epigenetic regulator of stem cell maintenance in germ cells and haematopoietic stem cells, represses L1 retrotransposition in germ and progenitor cells (Puszyk et al., 2013). PLZF-mediated DNA methylation induces silencing of the L1 gene, resulting in inhibition of L1 retrotransposition. Species-specific KZNFs might recruit KAP1 to species-specific retrotransposon classes that recently invaded the host genome. In this regard, Jacobs et al. recently reported that two primate-specific ZNF91 and ZNF93 repress SVA and L1 retrotransposons, respectively (Jacobs et al., 2014). ZNF93 evolved earlier to repress the primate L1 lineage until ~12.5 million years ago when the L1PA3 subfamily escaped ZNF93-mediated restriction through the removal of the ZNF93-binding site, suggesting an evolutionary arms race between KRAB-ZNFs and retrotransposons (Jacobs et al., 2014).

Post-translational modification and subcellular localization of L1 protein seem to be important for modulation of L1 mobility. In fact, phosphorylation of ORF1p is required for L1 retrotransposition (Cook et al., 2015). L1 ORF1p contains four conserved proline-directed protein kinase (PDPKs) target sites. PDPK mutations in ORF1p could inactivate L1 mobility (Cook et al., 2015). The PDPK family includes mitogen-activated protein kinases (MAPKs) and cyclin-dependent kinases (CDKs). Although nuclear localization of L1 ORF1p and ORF2p is essential for L1 retrotransposition, L1 ORF1p predominantly localizes in punctate cytoplasmic foci in most of cases (Goodier et al., 2007; Harris et al., 2010; Chen et al., 2012). However, in several breast cancers, L1 ORF1p and ORF2p were also detected in the nucleus (Harris et al., 2010; Chen et al., 2012). Indeed, the expression of L1 is elevated in breast cancers.

Recently, APOBEC3 family of cytidine deaminases, MOV10, and SAMHD1 have been identified as restriction factors for human immunodeficiency virus type 1 (HIV-1) (Sheehy et al., 2002; Harris et al., 2003; Mangeat et al., 2003; Zhang et al., 2003; Burdick et al., 2010; Furtak et al., 2010; Wang et al., 2010; Hrecka et al., 2011; Laguette et al., 2011). APOBEC3A, APOBEC3B, and APOBEC3F but not APOBEC3G inhibit L1

retrotransposition by a DNA deaminase-independent manner, indicating a novel anti-L1 retrotransposition mechanism (Turelli et al., 2004; Bogerd et al., 2006; Chen et al., 2006; Muckenfuss et al., 2006; Stenglein and Harris, 2006; Hulme et al., 2007; Kinomoto et al., 2007; Niewiadomska et al., 2007; Schumann, 2007; Arias et al., 2012). In contrast, Kinomoto et al. and Niewiadomska et al. reported that APOBEC3G could inhibit L1 retrotransposition by a DNA deamination-independent manner (Kinomoto et al., 2007; Niewiadomska et al., 2007). Furthermore, APOBEC3G inhibits Alu retrotransposition by a DNA deaminase-independent manner (Chiu et al., 2006; Hulme et al., 2007; Bulliard et al., 2009). MOV10 RNA helicase also inhibits L1 and Alu retrotransposition (Arjan-Odedra et al., 2012; Goodier et al., 2012, 2013; Li et al., 2013). Similarly, SAMHD1 inhibits LINE-1 and Alu/SVA retrotransposition (Zhao et al., 2013). SAMHD1 inhibits L1 retrotransposition through promoting the sequestration of L1 RNP within stress granules (Hu et al., 2015). Similarly, the zinc-finger antiviral protein (ZAP) also known as PARP13, a member of poly(ADP-ribose) polymerase (PARP) family, inhibits the retrotransposition of L1, Alu, and intracisternal A particle (IAP) retrotransposons (Goodier et al., 2015; Moldovan and Moran, 2015). ZAP interacts with L1 RNA and L1 ORF1p and co-localizes with stress granules.

Type I interferons (IFN I) including IFN α and IFN β have been involved in innate immunity against viruses. In this regard, a recent study reported that L1 induces IFN1 and IFN1, in turn, inhibits L1 retrotransposition, suggesting that IFN1 controls propagation of L1 as well as maintenance of genomic integrity (Yu et al., 2015). Accordingly, several interferon-stimulated genes (ISGs), including APOBEC3, MOV10, BST-2, ISG20, MAVS, MX2, RNase L, SAMHD1, TREX1, and ZAP restrict L1 retrotransposition, indicating that ISGs are key players of the type I interferon anti-retroelement response (Turelli et al., 2004; Bogerd et al., 2006; Chen et al., 2006; Muckenfuss et al., 2006; Stenglein and Harris, 2006; Hulme et al., 2007; Kinomoto et al., 2007; Niewiadomska et al., 2007; Schumann, 2007; Stetson et al., 2008; Arias et al., 2012; Zhao et al., 2013; Zhang et al., 2014; Goodier et al., 2015; Hu et al., 2015; **Table 3**).

Small RNAs have been implicated in the regulation of L1 mobility. Piwi proteins and Piwi-interacting RNAs (piRNA) silence L1 during genome reprogramming in the embryonic male germ line (De Fazio et al., 2011; Marchetto et al., 2013). Notably, Hamdorf et al. uncovered a new mechanism in which microRNAs restrict L1 mobilization and L1-associated mutations in cancer cells, cancer-initiating cells and iPS cells (Hamdorf et al., 2015). Indeed, miR-128 represses L1 retrotransposition by binding directly to L1 RNA, suggesting a new function of microRNAs in mediating genomic stability by suppressing the mobility of endogenous retrotransposons.

Tumor suppressor p53 mutations occur in most of human cancers, however, precisely how p53 functions to mediate tumor suppression is not well understood. In this regard, p53 was reported to restrict L1 mobility and suggested that p53 restricts oncogenesis in part by restricting transposon mobility (Wylie et al., 2016). Although normal human p53 suppressed transposons, mutant p53 from cancer patients could not. In contrast, L1 activity was elevated in p53 negative

human cancers. Thus, ancestral function of p53 may be associated with transposon control as a guardian of human genome.

CONCLUSION

L1 has successfully propagated and composed 17% of the human genome, resulting in evolutionary force. Activation of the normally silent L1 is associated with a high level of cancer-associated DNA damage and genomic instability. Indeed, L1 insertions into the human genome may cause cancers through insertional mutagenesis. In fact, recent high-throughput sequencing-based approaches could identify numerous somatic tumor-specific L1 insertions in a variety of cancers (Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012; Shukla et al., 2013; Helman et al., 2014), however there is no sufficient evidence. Therefore, it should clarify the role of L1-mediated retrotransposition in human cancers. Indeed, the implication of L1 insertion events as either passenger or driver mutations with a causative role in tumorigenesis still remains to be clarified (Rodić and Burns, 2013). Intriguingly, somatic insertions were only identified in epithelial tumors (Lee et al., 2012). Accordingly, epithelial cells can be transformed to cancer stem cells (Wang et al., 2013) and metastasis is more prevalent in epithelial tumors (Gotzmann et al., 2004). Thus, epithelial cells seem to be plastic (Carreira et al., 2014). Cancer stem cells are defined as rare cells with indefinite potential for self-renewal that drive tumorigenesis (Reya et al., 2001). However, it remains to be clarified the role of L1 mobility in cancer stem cells. Recent studies focus on the relationship among L1 mobility, reprogramming, and differentiation. Indeed, reprogramming somatic cells into iPSCs activates L1 mobility (Wissing et al., 2012; Friedli et al., 2014; Klawitter et al., 2016). On the other hand, L1 mobility is enhanced in tumor cells. In this regard, the elevation of L1 protein or RNA expression levels may be useful as a diagnostic

hallmark of many human cancers and as a tumor specific marker, metastasis, and prognosis. Furthermore, recent advances in single cell analysis will be useful for comparison of the L1 mobility and the integration site of L1 at a single cell level in human cancers.

Finally, tumor suppressor proteins may be associated with transposon control to restrict deleterious retrotransposition as a guardian of the human genome. Wild-type p53 suppresses transposon mobility in normal cells, while mutant p53 in cancer cells could not, resulting in the activation of L1 mobility in cancer cells (Wylie et al., 2016). Furthermore, recent studies identified somatic L1 insertion in tumor suppressor genes, such as APC, PTEN, NF1, and Rb (Miki et al., 1992; Wimmer et al., 2011; Helman et al., 2014; Rodríguez-Martín et al., 2016). Thus, L1 insertions in the tumor suppressor genes may disrupt their functions and be associated with tumorigenesis. Altogether, host cells have evolved several defense mechanisms protecting cells against retrotransposition.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

This work was supported by a Lateral Research from the Japan Society for the Promotion of Science (JSPS), by the Research Program on Hepatitis from Japan Agency for Medical Research and Development, AMED, and by Takeda Science Foundation.

ACKNOWLEDGMENTS

I thank Ms. Kazumi Tsuruhara for secretarial assistance and her kind encouragement.

REFERENCES

- Aravin, A. A., Hannon, G. J., and Brennecke, J. (2007a). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318, 761–764. doi: 10.1126/science.1146484
- Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., and Hannon, G. J. (2007b). Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* 316, 744–747. doi: 10.1126/science.1142612
- Arias, J. F., Koyama, T., Kinomoto, M., and Tokunaga, K. (2012). Retroelements versus APOBEC3 family members: no great escape from the magnificent seven. *Front. Microbiol.* 3:275. doi: 10.3389/fmicb.2012.00275
- Arjan-Odedra, S., Swanson, C. M., Sheree, N. M., Wolinsky, S. M., and Malim, M. H. (2012). Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication if exogenous retroviruses. *Retrovirology* 9:53. doi: 10.1186/1742-4690-9-53
- Armanios, M., Chen, J. L., Chang, Y. P., Brodsky, R. A., Hawkins, A., Griffin, C. A., et al. (2005). Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenital. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15960–15964. doi: 10.1073/pnas.0508124102
- Asch, H. L., Eliacin, E., Fanning, T. G., Connolly, J. L., Brathauer, G., and Asch, B. B. (1996). Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol. Res.* 8, 239–247.
- Aschacher, T., Wolf, B., Enzmann, F., Kienzl, P., Messner, B., Sampl, S., et al. (2016). LINE-1 induces hTERT and ensures telomere maintenance in tumour cell lines. *Oncogene* 35, 94–104. doi: 10.1038/ncr.2015.65
- Athanikar, J. N., Badge, R. M., and Moran, J. V. (2004). A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* 32, 3846–3855. doi: 10.1093/nar/gkh698
- Baba, Y., Murata, A., Watanabe, M., and Baba, H. (2014). Clinical implications of the LINE-1 methylation levels in patients with gastrointestinal cancer. *Surg. Today* 44, 1807–1816. doi: 10.1007/s00595-013-0763-6
- Baillie, J. K., Barrett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., et al. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479, 534–537. doi: 10.1038/nature10531
- Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., et al. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159–1170. doi: 10.1016/j.cell.2010.05.021
- Beck, C. R., Garcia-Perez, J. L., Badge, R. M., and Moran, J. V. (2011). LINE-1 elements in structural variation and disease. *Ann. Rev. Genomics Hum. Genet.* 12, 187–215. doi: 10.1146/annurev-genom-082509-141802
- Becker, K. G., Swergold, G. D., Ozato, K., and Thayer, R. E. (1993). Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. *Hum. Mol. Genet.* 2, 1697–1702. doi: 10.1093/hmg/2.10.1697

- Belancio, V. P., Hedges, D. J., and Deininger, P. (2008). Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. *Genome Res.* 18, 343–358. doi: 10.1101/gr.5558208
- Bogerd, H. P., Wiegand, H. L., Hulme, A. E., Garcua-Perez, J. L., O'Shea, K. S., and Moran, J. V. (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
- Bratthauer, G. L., Cardiff, R. D., and Fanning, T. G. (1994). Expression of LINE-1 retrotransposons in human breast cancer. *Cancer* 73, 2333–2336.
- Bratthauer, G. L., and Fanning, T. G. (1992). Active LINE-1 retrotransposons in human testicular cancer. *Oncogene* 7, 507–510.
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285. doi: 10.1073/pnas.0831042100
- Bulliard, Y., Turelli, P., Röhrig, U. F., Zoete, V., Mangeat, B., Michielin, O., et al. (2009). Functional analysis and structural modeling of human APOBEC3G reveal the role of evolutionarily conserved elements in the inhibition of human immunodeficiency virus type 1 infection and Alu transposition. *J. Virol.* 83, 12611–12621. doi: 10.1128/JVI.01491-09
- Bundo, M., Toyoshima, M., Okada, Y., Akamatsu, W., Ueda, J., Nemoto-Miyauchi, T., et al. (2014). Increased L1 retrotransposition in the neuronal genome in schizophrenia. *Neuron* 81, 306–313. doi: 10.1016/j.neuron.2013.10.053
- Burden, A. F., Manley, N. C., Clark, A. D., Gartler, S. M., Laird, C. D., and Hansen, R. S. (2005). Hemimethylation and non-CpG methylation levels in a promoter region of human LINE-1 (L1) repeated elements. *J. Biol. Chem.* 280, 14413–14419. doi: 10.1074/jbc.M413836200
- Burdick, R., Smith, J. L., Chaipan, C., Friew, Y., Chen, J., Venkatachari, N. J., et al. (2010). P-body-associated protein Mov10 inhibits HIV-1 replication at multiple stages. *J. Virol.* 84, 10241–10253. doi: 10.1128/JVI.00585-10
- Carreira, P. E., Richardson, S. R., and Faulkner, G. J. (2014). L1 retrotransposons, cancer stem cells and oncogenesis. *FEBS J.* 281, 63–73. doi: 10.1111/febs.12601
- Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., et al. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* 28, 1397–1409. doi: 10.1101/gad.241661.114
- Chen, H., Lilley, C. E., Yu, Q., Lee, D. V., Chou, J., Narvaiza, I., et al. (2006). APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* 16, 480–485. doi: 10.1016/j.cub.2006.01.031
- Chen, L., Dahlstrom, J. E., Chandra, A., Board, P., and Rangasamy, D. (2012). Prognostic value of LINE-1 retrotransposon expression and its subcellular localization in breast cancer. *Breast Cancer Res. Treat.* 136, 129–142. doi: 10.1007/s10549-012-2246-7
- Chiu, Y. L., Witkowska, H. E., Hall, S. C., Santiago, M., Soros, V. B., Esnault, C., et al. (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.* 103, 15588–15593. doi: 10.1073/pnas.0604524103
- Ciccia, A., and Elledge, S. J. (2010). The DNA damage response: making it safe to play with knives. *Mol. Cell* 40, 179–204. doi: 10.1016/j.molcel.2010.09.019
- Cook, P. R., Jones, C. E., and Furano, A. V. (2015). Phosphorylation of ORF1p is required for L1 retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4298–4303. doi: 10.1073/pnas.1416869112
- Cordaux, R., and Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nat. Rev. Genet.* 10, 691–703. doi: 10.1038/nrg2640
- Cost, G. J., and Boeke, J. D. (1998). Targeting of human retrotransposon integration is directed by the specificity of the L1 endonuclease for regions of unusual DNA structure. *Biochemistry* 37, 18081–18093. doi: 10.1021/bi981858s
- Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Marchetto, M. C., Muotri, A. R., Mu, Y., et al. (2011). Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20382–20387. doi: 10.1073/pnas.1100273108
- DeBerardinis, R. J., Goodier, J. L., Ostertag, E. M., and Kazazian, H. H. Jr. (1998). Rapid amplification of a retrotransposon subfamily is evolving the mouse genome. *Nat. Genet.* 20, 288–290. doi: 10.1038/3104
- De Fazio, S., Bartonicek, N., Di Giacomo, M., Abreu-Goodger, C., Sankar, A., Funaya, C., et al. (2011). The endonuclease activity of Mili fuels piRNA amplification that silences LINE-1 elements. *Nature* 480, 259–263. doi: 10.1038/nature10547
- Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C., Marchetto, M. C. N., et al. (2015). Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* 163, 583–593. doi: 10.1016/j.cell.2015.09.025
- Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-1-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35, 41–48. doi: 10.1038/ng1223
- Ding, D., Lou, X., Hua, D., Yu, W., Li, L., Wang, J., et al. (2012). Recurrent targeted genes of hepatitis B virus in the liver cancer genomes identified by a next-generation sequencing-based approach. *PLoS Genet.* 8:e1003065. doi: 10.1371/journal.pgen.1003065
- Doucet, A. J., Hulme, A. E., Sahinovic, E., Kulpa, D. A., Moldovan, J. B., Kopera, H. C., et al. (2010). Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet.* 6:e1001150. doi: 10.1371/journal.pgen.1001150
- Doucet-O'Hare, T. T., Rodić, N., Sharma, R., Darbari, I., Abril, G., Choi, J. A., et al. (2015). LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4894–E4900. doi: 10.1073/pnas.1502474112
- Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24, 363–367. doi: 10.1038/74184
- Evrony, G. D., Cal, X., Lee, E., Hills, L. B., Elhosary, P. C., Lehmann, H. S., et al. (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* 151, 483–496. doi: 10.1016/j.cell.2012.09.035
- Ewing, A. D., Gacita, A., Wood, L. D., Ma, F., Xing, D., Kim, M. S., et al. (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* 25, 1536–1545. doi: 10.1101/gr.196238.115
- Feng, Q., Moran, J. V., Kazazian, H. H. Jr., and Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905–916. doi: 10.1016/S0092-8674(00)81997-2
- Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Diaz, N., Rowe, H. M., et al. (2014). Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res.* 24, 1251–1259. doi: 10.1101/gr.172809.114
- Furtak, V., Mulky, A., Rawlings, S. A., Kozhaya, L., Lee, K., Kewalramani, V. N., et al. (2010). Perturbation of the P-body component Mov10 inhibits HIV-1 infectivity. *PLoS ONE* 5:e9081. doi: 10.1371/journal.pone.0009081
- Gasior, S. L., Roy-Engel, A. M., and Deininger, P. L. (2008). ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst)* 7, 983–989. doi: 10.1016/j.dnarep.2008.02.006
- Gasior, S. L., Wakeman, T. P., Xu, B., and Deininger, P. L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* 357, 1383–1393. doi: 10.1016/j.jmb.2006.01.089
- Gilbert, N., Lutz-Prigge, S., and Moran, J. V. (2002). Genomic deletions created upon LINE-1 retrotransposition. *Cell* 110, 315–325. doi: 10.1016/S0092-8674(02)00828-0
- Goldman, F., Bouarich, R., Kulkarni, S., Freeman, S., Du, H.-Y., Harrington, L., et al. (2005). The effect of TERC haploinsufficiency on the inheritance of telomere length. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17119–17124. doi: 10.1073/pnas.0505318102
- Goodier, J. L., Cheung, L. E., and Kazazian, H. H. Jr. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet.* 8:e1002941. doi: 10.1371/journal.pgen.1002941
- Goodier, J. L., Cheung, L. E., and Kazazian, H. H. Jr. (2013). Mapping the LINE-1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res.* 41, 7401–7419. doi: 10.1093/nar/gkt512
- Goodier, J. L., Ostertag, E. M., and Kazazian, H. H. Jr. (2000). Transduction of 3'-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* 9, 653–657. doi: 10.1093/hmg/9.4.653
- Goodier, J. L., Pereira, G. C., Cheung, L. E., Rose, R. J., and Kazazian, H. H. Jr. (2015). The broad-spectrum antiviral protein ZAP restricts human retrotransposition. *PLoS Genet.* 11:e1005252. doi: 10.1371/journal.pgen.1005252
- Goodier, J. L., Zhang, L., Vetter, M. R., and Kazazian, H. H. Jr. (2007). LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol. Cell. Biol.* 27, 6469–6483. doi: 10.1128/MCB.00332-07
- Gotzmann, J., Mikula, M., Eger, A., Schulte-Hermann, R., Foisner, R., Beug, H., et al. (2004). Molecular aspects of epithelial cell plasticity: implications for

- local tumor invasion and metastasis. *Mutat. Res.* 566, 9–20. doi: 10.1016/S1383-5742(03)00033-4
- Hamdorf, M., Idica, A., Zisoulis, D. G., Gamelin, L., Martin, C., Sanders, K. J., et al. (2015). miR-128 represses L1 retrotransposition by binding directly to L1 RNA. *Nat. Struct. Mol. Biol.* 22, 824–831. doi: 10.1038/nsmb.3090
- Hancks, D. C., Goodier, J. L., Mandal, P. K., Cheung, L. E., and Kazazian, H. H. Jr. (2011). Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum. Mol. Genet.* 20, 3386–3400. doi: 10.1093/hmg/ddr245
- Hancks, D. C., and Kazazian, H. H. Jr. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.gde.2012.02.006
- Harper, J. W., and Elledge, S. J. (2007). The DNA damage response: ten years after. *Mol. Cell* 28, 739–745. doi: 10.1016/j.molcel.2007.11.015
- Harris, C. R., Dewan, A., Zupnick, A., Normart, R., Gabriel, A., Prives, C., et al. (2009). p53 responsive elements in human retrotransposons. *Oncogene* 28, 3857–3865. doi: 10.1038/onc.2009.246
- Harris, C. R., Normart, R., Yang, Q., Stevenson, E., Haffty, B. G., Ganesan, S., et al. (2010). Association of nuclear localization of a long interspersed nuclear element-1 protein in breast tumors with poor prognostic outcomes. *Genes Cancer* 1, 115–124. doi: 10.1177/1947601909360812
- Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., et al. (2003). DNA deamination mediates innate immunity to retroviral infection. *Cell* 113, 803–809. doi: 10.1016/S0092-8674(03)00423-9
- Helman, E., Lawrence, M. S., Stewart, C., Sougnez, C., Getz, G., and Meyerson, M. (2014). Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* 24, 1053–1063. doi: 10.1101/gr.163659.113
- Holmes, S. E., Dombrowski, B. A., Krebs, C. M., Boehm, C. D., and Kazazian, H. H. Jr. (1994). A new retrotransposable human L1 element from the LRE2 locus on chromosome 1q produces a chimaeric insertion. *Nat. Genet.* 7, 143–148. doi: 10.1038/ng0694-143
- Hrecka, K., Hao, C., Gierszewska, M., Swanson, S. K., Kesik-Brodacka, M., Srivastava, S., et al. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature* 474, 658–661. doi: 10.1038/nature10195
- Hu, S., Li, J., Xu, F., Mei, S., Duff, Y. L., Yin, L., et al. (2015). SAMHD1 inhibits LINE-1 retrotransposition by promoting stress granule formation. *PLoS Genet.* 11:e1005367. doi: 10.1371/journal.pgen.1005367
- Hulme, A. E., Bogerd, H. P., Cullen, B. R., and Moran, J. V. (2007). Selective inhibition of Alu retrotransposition by APOBEC3G. *Gene* 390, 199–205. doi: 10.1016/j.gene.2006.08.032
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020
- Jacobs, F. M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A. D., Katzman, S., et al. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516, 242–245. doi: 10.1038/nature13760
- Kazazian, H. H. Jr., Wong, C., Youssoufian, H., Scott, A. F., Philips, D. G., and Antonarakis, S. E. (1988). Haemophilia A resulting from *de novo* insertion of L1 sequence represents a novel mechanism for mutation in man. *Nature* 332, 164–166. doi: 10.1038/332164a0
- Kines, K. J., Sokolowski, M., deHaro, D. L., Christian, C. M., and Belancio, V. P. (2014). Potential for genomic instability associated with retrotranspositionally-incompetent L1 loci. *Nucleic Acids Res.* 42, 10488–10502. doi: 10.1093/nar/gku687
- Kinamoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., et al. (2007). ALL APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res.* 35, 2955–2964. doi: 10.1093/nar/gkm181
- Klawitter, S., Fuchs, N. V., Upton, K. R., Muñoz-Lopez, M., Shukla, R., Wang, J., et al. (2016). Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat. Commun.* 7:10286. doi: 10.1038/ncomms10286
- Ko, M., An, J., Pastor, W. A., Koralov, S. B., Rajewsky, K., and Rao, A. (2015). TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunol. Rev.* 263, 6–21. doi: 10.1111/imr.12239
- Kubo, S., Seleme, M. C., Soifer, H. S., Perez, J. L., Moran, J. V., Kazazian, H. H. Jr., et al. (2006). L1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8036–8041. doi: 10.1073/pnas.0601954103
- Kuilman, T., Michaloglou, C., Mooi, W. J., and Peeper, D. S. (2010). The essence of senescence. *Genes Dev.* 24, 2463–2479. doi: 10.1101/gad.1971610
- Kuramochi-Miyagawa, S., Watanabe, T., Gotoh, K., Totoki, Y., Toyoda, A., Ikawa, M., et al. (2008). DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. *Genes Dev.* 22, 908–917. doi: 10.1101/gad.1640708
- Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Segal, E., et al. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature* 474, 654–657. doi: 10.1038/nature10117
- Lau, C. C., Sun, T., Ching, A. K., He, M., Li, J. W., Wong, A. M., et al. (2014). Viral-human chimeric transcript predisposes risk to liver cancer development and progression. *Cancer Cell* 25, 335–349. doi: 10.1016/j.ccr.2014.01.030
- Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J. III, et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971. doi: 10.1126/science.1222077
- Li, X., Zhang, J., Jia, R., Cheng, V., Xu, X., Qiao, W., et al. (2013). The MOV10 helicase inhibits LINE-1 mobility. *J. Biol. Chem.* 288, 21146–21160. doi: 10.1074/jbc.M113.465856
- Liang, G., Chan, M. F., Tomigahara, Y., Tsai, Y. C., Gonzales, F. A., Li, E., et al. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol. Cell. Biol.* 22, 480–491. doi: 10.1128/MCB.22.2.480-491.2002
- Liu, J., Nau, M. M., Zucman-Rossi, J., Powell, J. I., Allegra, C. J., and Wright, J. J. (1997). LINE-1 element insertion at the t(11;22) translocation breakpoint of a desmoplastic small round cell tumor. *Genes Chromosomes Cancer* 18, 232–239.
- Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605. doi: 10.1016/0092-8674(93)90078-5
- Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003). Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. *Nature* 424, 99–103. doi: 10.1038/nature01709
- Marchetto, M. C. N., Narvaiza, I., Denli, A. M., Benner, C., Lazzarini, T. A., Nathanson, J. L., et al. (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* 503, 525–529. doi: 10.1038/nature12686
- Martin, S. L., and Bushman, F. D. (2001). Nucleic acid chaperon activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell Biol.* 21, 467–475. doi: 10.1128/MCB.21.2.467-475.2001
- Mathias, S. L., Scott, A. F., Kazazian, H. H. Jr., Boeke, J. D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* 254, 1808–1810. doi: 10.1126/science.1722352
- Matsubara, K., and Tokino, T. (1990). Integration of hepatitis B virus DNA and its implications for hepatocarcinogenesis. *Mol. Biol. Med.* 7, 243–260.
- Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., et al. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464, 927–931. doi: 10.1038/nature08858
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., et al. (1992). Disruption of the APC gene by a retrotransposon insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–645.
- Mills, R. E., Bennett, E. A., Iskow, R. C., and Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends Genet.* 23, 183–191. doi: 10.1016/j.tig.2007.02.006
- Moldovan, J. B., and Moran, J. V. (2015). The zinc-finger antiviral protein ZAP inhibits LINE and Alu retrotransposition. *PLoS Genet.* 11:e1005121. doi: 10.1371/journal.pgen.1005121
- Montoya-Durango, D. E., Liu, Y., Teneng, I., Kalbfleisch, T., Lacy, M. E., Steffen, M. C., et al. (2009). Epigenetic control of mammalian LINE-1 retrotransposition by retinoblastoma proteins. *Mutat. Res.* 665, 20–28. doi: 10.1016/j.mrfmmm.2009.02.011
- Montoya-Durango, D. E., Ramos, K. A., Bojang, P., Ruiz, L., Ramos, I. N., and Ramos, K. S. (2016). LINE-1 silencing by retinoblastoma proteins is effected through the nucleosomal and remodeling deacetylase multiprotein complex. *BMC Cancer* 16:38. doi: 10.1186/s12885-016-2068-9

- Moran, J. V., DeBerardinis, R. J., and Kazazian, H. H. Jr. (1999). Exon shuffling by L1 retrotransposition. *Science* 283, 1530–1534. doi: 10.1126/science.283.5407.1530
- Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., et al. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31, 159–165. doi: 10.1038/ng898
- Morse, B., Rotherg, P. G., South, V. J., Spandorfer, J. M., and Astrin, S. M. (1988). Insertional mutagenesis of the myc locus by a LINE-1 sequence in a human breast carcinoma. *Nature* 333, 87–90. doi: 10.1038/333087a0
- Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Löwer, J., Cichutek, K., et al. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.* 281, 22161–22172. doi: 10.1074/jbc.M601716200
- Muotri, A. R., Chu, V. T., Marchetto, M. C., Deng, W., Moran, J. V., and Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910. doi: 10.1038/nature03663
- Muotri, A. R., Marchetto, M. C., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., et al. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446. doi: 10.1038/nature09544
- Nangia-Makker, P., Sarvis, R., Visscher, D. W., Bailey-Penrod, J., Raz, A., and Sarkar, F. H. (1998). Galectin-3 and L1 retrotransposons in human breast carcinomas. *Breast Cancer Res. Treat.* 49, 171–183. doi: 10.1023/A:1005913810250
- Narita, N., Nishio, H., Kitoh, Y., Ishikawa, Y., Minami, R., Nakamura, H., et al. (1993). Insertion of a 5' truncated L1 element into the 3' end of exon 44 of the dystrophin gene resulted in skipping of the exon during splicing in a case of Duchenne muscular dystrophy. *J. Clin. Invest.* 91, 1862–1867. doi: 10.1172/JCI116402
- Niewiadomska, A. M., Tian, C., Tan, L., Wang, T., Sarkis, P. T. N., and Yu, X. F. (2007). Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. *J. Virol.* 81, 9577–9583. doi: 10.1128/JVI.02800-06
- Ostertag, E. M., and Kazazian, H. H. Jr. (2001). Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* 35, 501–538. doi: 10.1146/annurev.genet.35.102401.091032
- Pickeral, O. K., Makalowski, W., Boguski, M. S., and Boeke, J. D. (2000). Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* 10, 411–415. doi: 10.1101/gr.10.4.411
- Puszyk, W., Down, T., Grimwade, D., Chromienne, C., Oakley, R. J., Solomon, E., et al. (2013). The epigenetic regulator PLZF represses L1 retrotransposition in germ and progenitor cells. *EMBO J.* 32, 1941–1952. doi: 10.1038/emboj.2013.118
- Reya, T., Morrison, S. J., Clarke, M. F., and Weissman, I. L. (2001). Stem cells, cancer, and cancer stem cells. *Nature* 414, 105–111. doi: 10.1038/35102167
- Rodić, N., and Burns, K. H. (2013). Long interspersed element-1 (LINE-1): passenger or driver in human neoplasm? *PLoS Genet.* 9:e1003402. doi: 10.1371/journal.pgen.1003402
- Rodić, N., Sharma, R., Sharma, R., Zampella, J., Dai, L., Taylor, M. S., et al. (2014). Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* 184, 1280–1286. doi: 10.1016/j.ajpath.2014.01.007
- Rodić, N., Steranka, J. P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., et al. (2015). Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat. Med.* 21, 1060–1064. doi: 10.1038/nm.3919
- Rodríguez-Martín, C., Cidre, F., Fernández-Teijeiro, A., Gómez-Mariano, G., de la Vega, L., Ramos, P., et al. (2016). Familial retinoblastoma due to intronic LINE-1 insertion causes aberrant and noncanonical mRNA splicing of the RB1 gene. *J. Hum. Genet.* 61, 463–466. doi: 10.1038/jhg.2015.173
- Rowe, H. M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* 463, 237–240. doi: 10.1038/nature08674
- Schulz, W. A., Steinhoff, C., and Flori, A. R. (2006). Methylation of endogenous human retroelements in health and disease. *Curr. Top. Microbiol. Immunol.* 310, 211–250. doi: 10.1007/3-540-31181-5_11
- Schumann, G. G. (2007). APOBEC3 proteins: major players in intracellular defence against LINE-1-mediated retrotransposition. *Biochem. Soc. Trans.* 35, 637–642. doi: 10.1042/BST0350637
- Scott, E. C., Gardner, E. J., Masood, A., Chuang, N. T., Vertino, P. M., and Devine, S. E. (2016). A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res.* 26, 745–755. doi: 10.1101/gr.201814.115
- Sheehy, A. M., Gaddis, N. C., Choi, J. D., and Malim, M. H. (2002). Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature* 418, 646–650. doi: 10.1038/nature00939
- Shi, X., Seluanov, A., and Gorbunova, V. (2007). Cell divisions are required for L1 retrotransposition. *Mol. Cell. Biol.* 27, 1264–1270. doi: 10.1128/MCB.01888-06
- Shiloh, Y. (2014). ATM, expanding roles as a chief guardian of genome stability. *Exp. Cell Res.* 329, 154–161. doi: 10.1016/j.yexcr.2014.09.002
- Shukla, R., Upton, K. R., Munoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111. doi: 10.1016/j.cell.2013.02.032
- Soifer, H. S., Zaragoza, A., Peyvan, M., Behlke, M. A., and Rossi, J. J. (2005). A potential role for RNA interference in controlling the activity of the human LINE-1 retrotransposon. *Nucleic Acids Res.* 33, 846–856. doi: 10.1093/nar/gki223
- Solyom, S., Ewing, A. D., Rahrmann, E. P., Doucet, T., Nelson, H. H., Burns, M. B., et al. (2012). Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* 22, 2328–2338. doi: 10.1101/gr.145235.112
- Stenglein, M. D., and Harris, R. S. (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J. Biol. Chem.* 281, 16837–16841. doi: 10.1074/jbc.M602367200
- Stetson, D. B., Ko, J. S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* 134, 587–598. doi: 10.1016/j.cell.2008.06.032
- Suzuki, J., Yamaguchi, K., Kajikawa, M., Ichianagi, K., Adachi, N., Koyama, H., et al. (2009). Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet.* 5:e1000461. doi: 10.1371/journal.pgen.1000461
- Swergold, G. D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 10, 6718–6729. doi: 10.1128/MCB.10.12.6718
- Symer, D. E., Connelly, C., Szak, S. T., Caputo, E. M., Cost, G. J., Parmigiani, G., et al. (2002). Human L1 retrotransposition is associated with genetic instability *in vivo*. *Cell* 110, 327–338. doi: 10.1016/S0092-8674(02)00839-5
- Szak, S. T., Pickeral, O. K., Makalowski, W., Boguski, M. S., Landsman, D., and Boeke, J. D. (2002). Molecular archeology of L1 insertions in the human genome. *Genome Biol.* 3:research0052. doi: 10.1186/gb-2002-3-10-research0052
- Tchenio, T., Casella, J. F., and Heidmann, T. (2000). Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* 28, 411–415. doi: 10.1093/nar/28.2.411
- Trono, D. (2015). Transposable elements, polydactyl proteins, and the genesis of human-specific transcription networks. *Cold Spring Harb. Symp. Quant. Biol.* 80, 281–288. doi: 10.1101/sqb.2015.80.027573
- Tubio, J. C., Li, Y., Ju, Y. S., Martincorena, I., Cooke, S. L., Tojo, M., et al. (2014). Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345:1251343. doi: 10.1126/science.1251343
- Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., et al. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* 24, 1260–1270. doi: 10.1101/gr.172833.114
- Turelli, P., Vianin, S., and Trono, D. (2004). The innate antiretroviral factor APOBEC3G does not affect human LINE-1 retrotransposition in a cell culture assay. *J. Biol. Chem.* 279, 43371–43373. doi: 10.1074/jbc.C400334200
- Upton, K. R., Gerhardt, D. J., Jesuadian, J. S., Richardson, S. R., Sánchez-Luque, F. J., Bodea, G. O., et al. (2015). Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* 161, 228–239. doi: 10.1016/j.cell.2015.03.026
- Van Meter, M., Kashyap, M., Rezazadeh, S., Geneva, A. J., Morello, T. D., Seluanov, A., et al. (2014). SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat. Commun.* 5, 5011. doi: 10.1038/ncomms6011
- Wallace, N. A., Gasior, S. L., Faber, Z. J., Howie, H. L., Deininger, P. L., and Galloway, D. A. (2013). HPV 5 and 8 E6 expression reduces ATM protein levels and attenuates LINE-1 retrotransposition. *Virology* 443, 69–79. doi: 10.1016/j.virol.2013.04.022
- Wang, X., Han, Y., Dang, Y., Fu, W., Zhou, T., Ptak, R. G., et al. (2010). Moloney leukemia virus 10 (MOV10) protein inhibits retrovirus replication. *J. Biol. Chem.* 285, 14346–14355. doi: 10.1074/jbc.M110.109314

- Wang, Z. A., Mitrofanova, A., Bergren, S. K., Abate-Shen, C., Cardiff, R. D., Califano, A., et al. (2013). Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nat. Cell Biol.* 15, 274–283. doi: 10.1038/ncb2697
- Wimmer, K., Callens, T., Wernstedt, A., and Messiaen, L. (2011). The NF1 gene contains hotspots for L1 endonuclease-dependent *de novo* insertion. *PLoS Genet.* 7:e1002371. doi: 10.1371/journal.pgen.1002371
- Wissing, S., Muñoz-Lopez, M., Macia, A., Yang, Z., Montano, M., Collins, W., et al. (2012). Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum. Mol. Genet.* 21, 208–218. doi: 10.1093/hmg/ddr455
- Wolf, D., and Goff, S. P. (2007). TRIM28 mediates primer binding site-targeted silencing of murine leukemia virus in embryonic cells. *Cell* 131, 46–57. doi: 10.1016/j.cell.2007.07.026
- Wolf, D., and Goff, S. P. (2009). Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* 458, 1201–1204. doi: 10.1038/nature07844
- Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P., and Warren, W. D. (1997). Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J. Biol. Chem.* 272, 7810–7816. doi: 10.1074/jbc.272.12.7810
- Wylie, A., Jones, A. E., D'Brot, A., Lu, W. J., Kurtz, P., Moran, J. V., et al. (2016). P53 genes function to restrain mobile elements. *Genes Dev.* 30, 64–77. doi: 10.1101/gad.266098.115
- Xie, Y., Mates, L., Ivics, Z., Izsvák, Z., Martin, S. L., and An, W. (2013). Cell division promotes efficient retrotransposition in a stable L1 reporter cell line. *Mob. DNA* 4:10. doi: 10.1186/1759-8753-4-10
- Yang, N., and Kazazian, H. H. Jr. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* 9, 763–771. doi: 10.1038/nsmb1141
- Yang, N., Zhang, L., Zhang, Y., and Kazazian, H. H. Jr. (2003). An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.* 31, 4929–4940. doi: 10.1093/nar/gkg663
- Yu, F., Zingler, N., Schumann, G., and Strätling, W. H. (2001). Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* 29, 4493–4501. doi: 10.1093/nar/29.21.4493
- Yu, Q., Carbone, C. J., Katlinskaya, Y. V., Zheng, H., Zheng, K., Luo, M., et al. (2015). Type I interferon controls propagation of long interspersed element-1. *J. Biol. Chem.* 290, 10191–10199. doi: 10.1074/jbc.M114.612374
- Zhang, A., Dong, B., Doucet, A., Moldovan, J. B., Moran, J. V., and Silverman, R. H. (2014). RNase L restricts the mobility of engineered retrotransposons in cultured human cells. *Nucleic Acids Res.* 42, 3803–3820. doi: 10.1093/nar/gkt1308
- Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C., and Gao, L. (2003). The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. *Nature* 424, 94–98. doi: 10.1038/nature01707
- Zhao, K., Du, J., Han, X., Goodier, J. L., Li, P., Zhou, X., et al. (2013). Modulation of LINE-1 and Alu/SVA retrotransposition by Aicardi-Goutières Syndrome-related SAMHD1. *Cell Rep.* 4, 1108–1115. doi: 10.1016/j.celrep.2013.08.019

Conflict of Interest Statement: The author declares 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 © 2016 Ariumi. 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) or licensor 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.



Templated Sequence Insertion Polymorphisms in the Human Genome

Masahiro Onozawa^{1,2} and Peter D. Aplan^{1*}

¹ Genetics Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA, ² Department of Hematology, Hokkaido University Graduate School of Medicine, Sapporo, Japan

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Formerly affiliated with University of
Toledo, USA

Reviewed by:

Francesca Storici,
Georgia Institute of Technology, USA
Jeffrey Han,
Tulane University, USA

*Correspondence:

Peter D. Aplan
aplanp@mail.nih.gov

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 09 August 2016

Accepted: 27 October 2016

Published: 16 November 2016

Citation:

Onozawa M and Aplan PD (2016)
Templated Sequence Insertion
Polymorphisms in the Human
Genome. *Front. Chem.* 4:43.
doi: 10.3389/fchem.2016.00043

Templated Sequence Insertion Polymorphism (TSIP) is a recently described form of polymorphism recognized in the human genome, in which a sequence that is templated from a distant genomic region is inserted into the genome, seemingly at random. TSIPs can be grouped into two classes based on nucleotide sequence features at the insertion junctions; Class 1 TSIPs show features of insertions that are mediated via the LINE-1 ORF2 protein, including (1) target-site duplication (TSD), (2) polyadenylation 10–30 nucleotides downstream of a “cryptic” polyadenylation signal, and (3) preference for insertion at a 5′-TTTT/A-3′ sequence. In contrast, class 2 TSIPs show features consistent with repair of a DNA double-strand break (DSB) via insertion of a DNA “patch” that is derived from a distant genomic region. Survey of a large number of normal human volunteers demonstrates that most individuals have 25–30 TSIPs, and that these TSIPs track with specific geographic regions. Similar to other forms of human polymorphism, we suspect that these TSIPs may be important for the generation of human diversity and genetic diseases.

Keywords: templated sequence insertion polymorphisms (TSIPs), mitochondria, polymorphism, human migration, DNA repair, LINE-1 retrotransposon

INTRODUCTION

Maintenance of chromosomal integrity is required for the survival of all organisms, from simple prokaryotes to complex eukaryotes. This maintenance of chromosomal integrity is accomplished by DNA repair enzymes. There are a number of DNA repair systems that operate in eukaryotes, including DNA mismatch repair, DNA single-strand break repair, and DNA double-strand break (DSB) repair. DNA DSB repair can be further subdivided into repair by homologous recombination, “canonical” non-homologous end-joining (NHEJ), and “non-canonical” NHEJ (Chiruvella et al., 2013; Deriano and Roth, 2013).

Transfected plasmid DNA can be captured and used as a patch at the site experimentally induced DNA DSBs; the DNA patches typically show signs of NHEJ, such as micro-deletion, microhomology, and non-templated nucleotide addition (Lin and Waldman, 2001; Varga and Aplan, 2005; Cheng et al., 2010). Moreover, Yu and Gabriel detected mitochondrial DNA fragments at the site of HO endonuclease-induced DNA DSBs (Yu and Gabriel, 1999), demonstrating that a DNA DSB can be repaired by insertion of DNA sequences in yeast.

RNA can provide a template for DNA synthesis during telomere elongation (Autexier and Lue, 2006) or reverse transcription of retrotransposons (Baltimore, 1985). Several lines of evidence suggest that endogenous retrotransposons may also have a role in DNA DSB repair in human cells.

When introduced into yeast, the human LINE-1 ORF2 can mediate repair of HO endonuclease-induced DNA DSB via insertion of retrotransposon sequences (Teng et al., 1996), through a cDNA intermediate. Subsequently, synthetic RNA oligonucleotides were shown to be a template for DNA synthesis during repair of HO endonuclease induced DNA DSB in yeast, although the efficiency of repair with RNA oligonucleotides was orders of magnitude lower than DNA oligonucleotides (Storici et al., 2007). Finally, a role for LINE-1 retrotransposons in DNA DSB repair in mammalian cells has been predicted (Morrish et al., 2002, 2007). This prediction was based on clever experiments that showed new integrations of an endonuclease-incompetent LINE-1 retrotransposon could be identified in rodent cell lines; this form of LINE-1 integration was designated endonuclease-independent (ENI) retrotransposition (Morrish et al., 2002). Because these new integration sites lacked the typical features of LINE-1 endonuclease mediated insertions, such as Target-Site Duplications (TSDs) and poly(A) tracts, the authors hypothesized that LINE-1 sequences were used as a “patch” to repair a spontaneous DNA DSB (as opposed to one introduced by the LINE-1 endonuclease).

With the advent of next generation sequencing technologies, millions of germline variants in mammalian genomes within a species have been identified. Most characterized variants, known as polymorphisms, fall into three large categories; single nucleotide polymorphisms (SNPs), small (<50 bp) insertions or deletions, referred to collectively as short indels, and large deletions (>50 bp) (Genomes Project et al., 2012). More recently, a smaller number of polymorphic insertions of retro-elements (such as LINE-1 or Alu) have been identified through the use of sophisticated methods to detect mobile element insertions (Beck et al., 2010; Huang et al., 2010; Iskow et al., 2010). In addition, insertion of processed gene transcripts into the germline have been identified and referred to as polymorphic pseudogenes (Ewing et al., 2013). Recent reports have demonstrated that insertions of retroelements and pseudogenes represent only a fraction of the insertional polymorphisms in the human genome (Onozawa et al., 2014, 2015).

DNA DSBs CAN BE REPAIRED BY INSERTION OF SEQUENCES DERIVED FROM DISTANT REGIONS OF THE GENOME IN AN EXPERIMENTAL SYSTEM

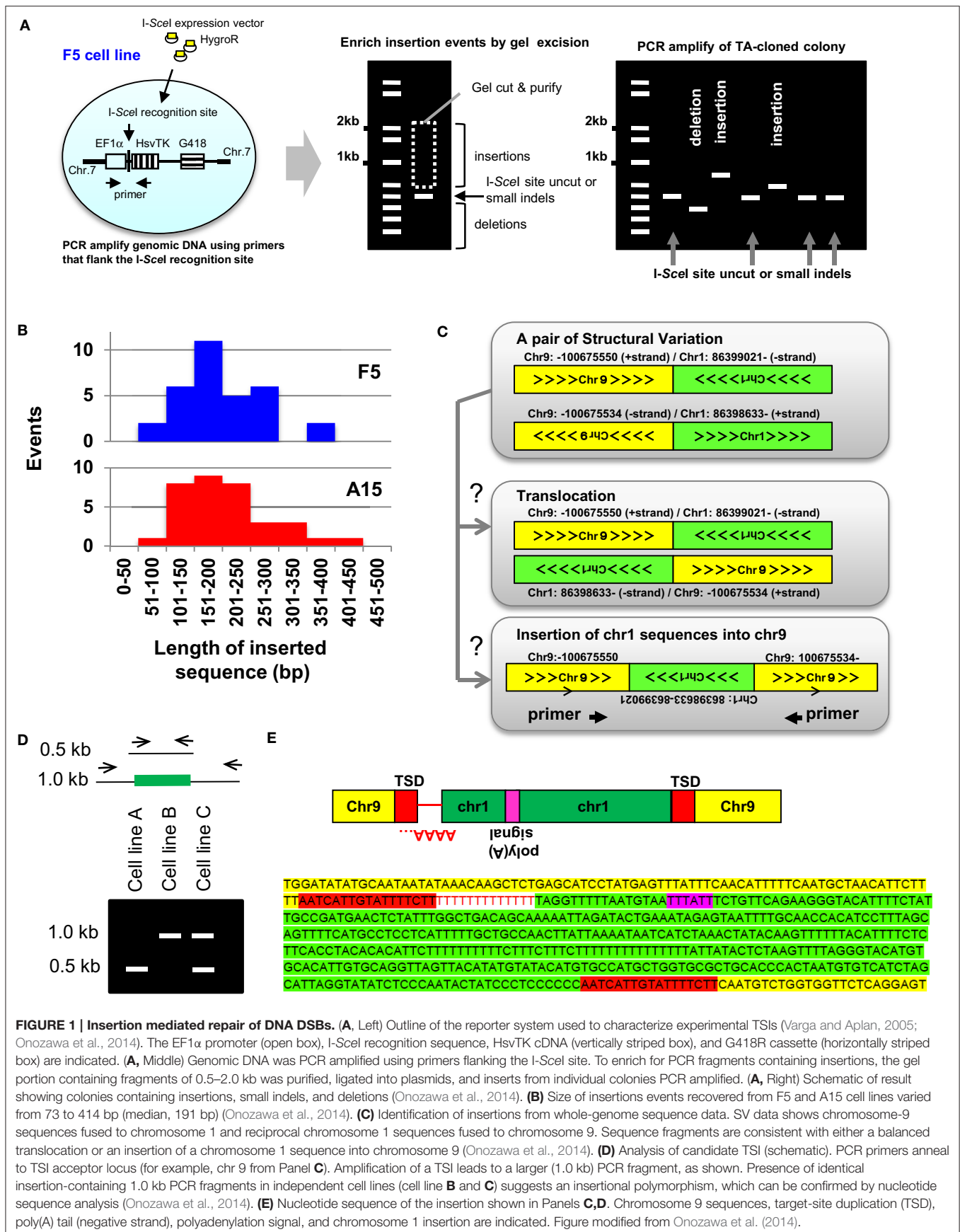
Templated Sequence Insertions (TSI) were first characterized using a cell culture based approach to study DNA DSB (Varga and Aplan, 2005). These studies modified an experimental system that had been popularized by Jasin and colleagues (Jasin and Haber, 2016), and employed a vector (designated EF1 α TK), that contained the EF1 α promoter driving expression of the herpes simplex virus thymidine kinase (HsvTK); the recognition site for the rare-cutting meganuclease I-SceI was inserted between the EF1 α promoter and HsvTK cDNA (Varga and Aplan, 2005). Expression of the HsvTK enzyme in mammalian cells confers sensitivity to the nucleoside analog ganciclovir; thus, millions of cells can quickly be screened for loss of HsvTK

expression by treatment with ganciclovir. The EF1 α TK vector was electroporated into the human leukemia cell line U937, or ovarian cancer cell line OVCAR8. Sub-clones that had integrated a single copy of the EF1 α TK vector, designated “F5” for U937 (Varga and Aplan, 2005) or “A15” for OVCAR8 (Cheng et al., 2010), were isolated (**Figure 1A**). In an attempt to induce gross chromosomal rearrangements, an I-SceI expression vector was introduced into the cell lines, followed by ganciclovir (GCV) selection of clones that had lost expression of HsvTK due to mis-repair of a DNA DSB.

Although most of the GCV-resistant clones had short deletions encompassing the HsvTK start codon, rare clones that had undergone an insertion at the DNA DSB site were identified (Varga and Aplan, 2005; Cheng et al., 2010; Onozawa et al., 2014). In these clones, the inserted sequence was not derived from nearby genomic sequences, but instead mapped to a distant region of the genome. After modifying the procedure to enrich for insertions (**Figure 1A**), 32 insertions of sequences derived from distant regions of the genome were identified in F5 subclones and 34 in A15 subclones (**Figure 1B**). The origin of these insertions mapped to 18 of 24 human chromosomes, without an obvious preference for any chromosome or chromosomal region (Onozawa et al., 2014). These insertions were designated “TSI,” or TSI, in contrast to the short, non-TSI seen at the site of NHEJ-mediated repair of a DNA DSB (Onozawa et al., 2014). The TSI junctions often showed features of NHEJ, such as microhomology and non-templated nucleotide addition. Generation of TSIs seemed to be generalizable, as they were not restricted to I-SceI induced cleavage but also found at TALEN-mediated cleavage sites (Onozawa et al., 2014).

TSIs ARE DERIVED FROM RNA

The TSIs inserted at the site of NHEJ mediated DNA DSB were not excised from the genome, since the donor sites were intact, and the donor (inserted) sequence had an additional copy compared to genomic regions flanking the donor site (Onozawa et al., 2014). The TSI sequences were enriched for transcribed sequences, suggesting that the TSI may have originated via reverse transcription of RNA. In addition, treatment of cells that expressed the I-SceI enzyme with reverse-transcriptase inhibitors suppressed the frequency of TSIs at DNA DSB site by more than three-fold. Moreover, co-transfection of murine RNA with an I-SceI expression vector into the human F5 cell line showed that reverse transcribed murine sequences were used as insertions at the I-SceI cleavage site. Finally, three insertions displayed mammalian telomere repeat (TTAGGG) $_n$ sequences, suggesting that telomerase RNA can also be used as a TSI template. Taken together, although other mechanisms, such as template switching by DNA polymerases or break-induced repair remain possible (Malkova and Haber, 2012; Morrish et al., 2013), the above observations support the hypothesis that reverse transcribed RNA can be used as a template to patch a DNA DSB (Onozawa et al., 2014). Potential sources of this reverse transcriptase activity include LINE-1 ORF2, HERVs (Hohn et al., 2013), as well as the



TERT subunit of telomerase, which has recently been suggested to possess hTR-independent RT activity (Sharma et al., 2012).

TSIs ARE NOT AN ARTEFACT OF EXPERIMENTAL, INDUCED DNA DSB, BUT CAN BE IDENTIFIED IN UNMANIPULATED HUMAN CELLS

TSIs cannot be detected from routine analysis of whole genome sequence (WGS) reads, which are typically <150 bp. However, TSIs can be identified from WGS using the principles outlined below. First, the junction of two non-homologous chromosomes is designated as a “structural variant” (SV) on short read WGS; pairs of SVs that map very closely can be ascertained by inspection of SV files (Figure 1C). For a pair of SVs to represent a Templated Sequence Insertion Polymorphism (TSIP), both fusion junctions must be located within 50 kb of one another. Second, the strand polarity must align such that an insertion is feasible. Third, each end of the SV needs to be localized to a single, unique genomic loci; SV that show multiple or imperfect alignments (<95% sequence identity) are excluded. Fourth, all highly repetitive alpha satellite sequences are discarded.

The authors screened SVs obtained from two multiple myeloma cell lines (KP6, MC1286PE1) using these criteria and identified 23 unique, verified TSIs (Onozawa et al., 2014). A typical example is shown in Figures 1C–E. Briefly, although this pair of SVs is consistent with a balanced translocation, it is also consistent with an insertion of chromosome 1 sequences into chromosome 9 (Figure 1C). Primers were generated that could amplify the putative insertion, including flanking sequences and both junctions, on a single PCR fragment (Figures 1D,E). Nucleotide sequence of the PCR product verified that the SVs were indeed produced by insertion of chromosome 1 derived sequences into chromosome 9, as opposed to a reciprocal translocation between chromosomes 1 and 9 (Figure 1E). Interestingly 8 out of 23 insertions were identical or nearly identical in the two cell lines (Onozawa et al., 2014) (a schematic example of this phenomenon is shown in Figure 1D), suggesting that these insertions represent polymorphisms in the human genome as opposed to SV acquired by the tumor cells. Consistent with the TSI definition above, these polymorphic insertions were designated templated-sequence insertion polymorphisms (TSIPs).

MOST TSIs IDENTIFIED IN NORMAL HUMAN SUBJECTS ARE POLYMORPHIC

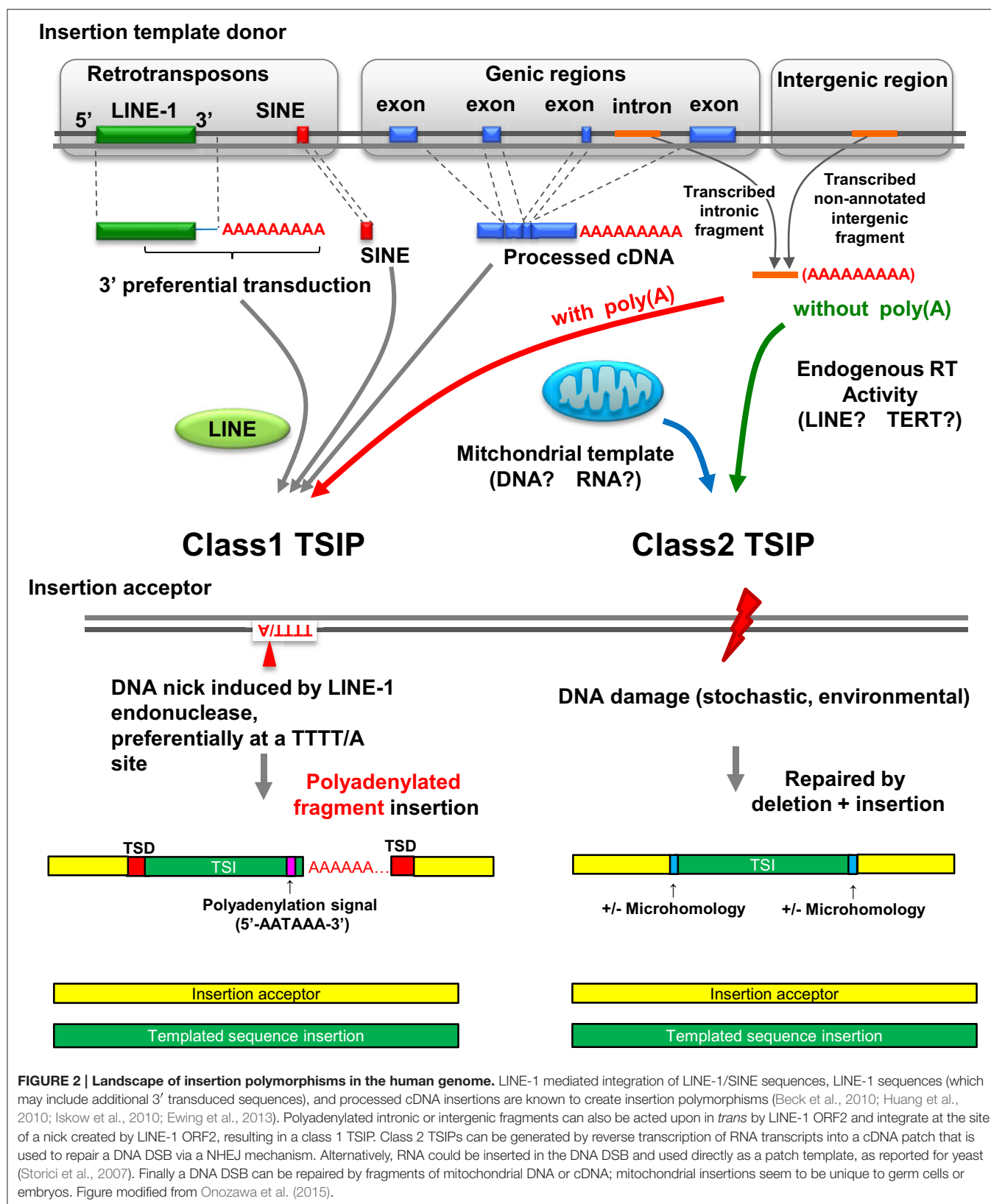
A publicly available database of WGS from 52 normal individuals of defined ethnic/geographic groups (“SV baseline genome set,” filename B37baselinejunctions.tsv, available at <http://www.completegenomics.com/sequence-data/download-data/>) contained a total of 39,595 SVs from the 52 individuals (Onozawa et al., 2015). Using the criteria set forth in section IV above, 171 candidate TSIPs were identified (Onozawa et al., 2015). Each individual had an average of 25–30 TSIPs, and TSIPs could be classified as “common” (26%; present in at least 20%

of individuals), or “rare” (74%; present in <20% of individuals). Interestingly, three TSIPs had a frequency of almost 100%, suggesting that the reference human genome (GRCh37/hg19) is based on an uncommon variant that lacks these three TSIPs. When divided into four regional “super groups” of specified geographic origin (African, Asian, European, North American), common TSIPs were present in individuals from all regions, whereas rare TSIPs tended to be restricted to individuals from a single region (Onozawa et al., 2015). There were more TSIPs per individual of African origin than other geographic regions. All of these findings are consistent with diversity identified in previous studies of mitochondrial and Y-chromosome sequences, and are consistent with patterns of human migration and the hypothesis that *Homo sapiens* originated in Africa (Cann et al., 1987; Hammer, 1995; Underhill et al., 2000).

The investigators obtained genomic DNA from eight of the normal individuals, who had a total of 89 candidate TSIs and successfully validated 69/89 (77.5%) candidate TSIPs (Onozawa et al., 2015). Since these insertions can be polymorphic, they must be heritable, leading to the conclusion that the insertion event must have taken place in either a germ cell (sperm or egg), or early-stage embryo.

SEQUENCES USED AS TEMPLATES FOR TSIPs

Nucleotide sequence analysis of the insertion sequences revealed that partial LINE-1 elements, cDNAs (with several spliced exons), non-annotated intergenic or intronic sequences, and mitochondrial sequences were used as templates for TSIPs identified in normal individuals (Onozawa et al., 2015). Of note, although mitochondrial fragment insertions were commonly identified as TSIP donor sequences, no TSIs derived from mitochondrion were identified in experimentally induced TSIs using the F5 and A15 cell lines described above (Varga and Aplan, 2005; Cheng et al., 2010; Onozawa et al., 2014). Although speculative, it is possible that mitochondrial sequence insertions may be reproduction-specific events that take place in germ cells or early stage embryos, but do not occur, or occur only rarely, in somatic cells. Interestingly, sperm mitochondria are known to be ubiquitinated and destroyed shortly after fertilization (Sutovsky et al., 2000), leading to the hypothesis that fragmented paternal mitochondrial DNA [or reverse transcribed RNA that was encoded by mitochondrial DNA (Sharma et al., 2012)] can be used to patch a DNA DSB in a fertilized embryo, leading to a TSIP which contained mitochondrial sequence in all cells of the individual, including germ cells (Woischnik and Moraes, 2002; Onozawa et al., 2015; Zhou et al., 2016). Although no TSIPs contained telomere sequences (Onozawa et al., 2014, 2015), interstitial telomeric sequences (ITSs) have been identified in several species (Ruiz-Herrera et al., 2002), and were identified as insertions in the I-SceI experimental system (Onozawa et al., 2014), leading to the speculation that the ITSs may have resulted from telomere patches used to repair a DNA DSB (Nergadze et al., 2004, 2007; Onozawa et al., 2014).



CLASS 1 AND CLASS 2 TSIPs

TSIPs can be placed into two classes (class 1 and class 2) based on nucleotide sequences at the insertion site (**Figure 2**). Class 1 TSIPs show a duplication of recipient sequences at both insertion junctions of at least 5 bp; this feature is reminiscent of a TSD that is characteristic of insertions caused by cleavage and insertion of LINE-1 sequences. In addition, class 1 TSIs typically inserted at a preferred LINE-1 integration site (consensus sequence 5'-TTTT/A-3'), and contained a non-templated addition of 10–40 “A” residues, as well as a polyadenylation signal (5'-AATAAA-3') located 10–20 nucleotides upstream of the poly(A) track. These features strongly support an insertion mediated by LINE-1 ORF2 protein, which contains both endonuclease and reverse transcriptase activity, acting upon non-LINE-1 RNA, and inserting the sequence into a distant region of the genome (Luan et al., 1993; Piskareva et al., 2013). Moreover, since these events must have occurred in germ cells or embryos to be transmitted, it is interesting to note that LINE-1 elements are de-repressed and active in embryos (Castro-Diaz et al., 2014). We can detect no obvious physiologic function for the Class 1 TSIPs, and suspect that these are caused by a careless LINE-1 ORF2 protein causing mischief throughout the genome. Class 2 TSIPs had none of these features (i.e., no TSDs, cryptic poly(A) signal, or poly(A) tract) but instead displayed NHEJ features such as microdeletion, microhomology, and non-templated nucleotide addition at the insertion junction, similar to what one would predict if an ENI retrotransposition event (Morrish et al., 2002) used non-LINE-1 RNA as a template. Consistent with this prediction, it is well established that LINE-1 ORF2 can act in *trans* (Wei et al., 2001). All experimental DNA DSB repair events were class 2 events, and we speculate that class 2 TSIPs are caused by DNA DSB repair events that occurred in a germ cell or early stage embryo of an ancestral individual.

POTENTIAL TO CAUSE GENETIC DISEASE

There is potential for this mechanism of DNA DSB repair to cause genetic disease. Several TSIPs disrupted the coding region of a gene (Onozawa et al., 2015). Furthermore, a recent report

described a constitutional 72-bp insertion of mitochondrial sequence into the coding region of *GLI3*, leading to Pallister-Hall syndrome (Turner et al., 2003). Of note, the conception of this patient was temporally and geographically associated with high-level radioactive contamination following the Chernobyl accident (Turner et al., 2003). Although speculative, it is conceivable that a DNA DSB in the germ cell, caused by ionizing radiation, was repaired by a TSI derived from mitochondrial DNA in this individual.

CONCLUSION

TSIPs encompass several forms of insertion polymorphisms in human genomes, and are mediated via a combination of mechanisms. Class 1 TSIPs are retrotransposon-mediated events that insert polyadenylated, reverse-transcribed cDNA into seemingly random regions of the genome, whereas class 2 TSIPs are consistent with DNA DSB repair events, in which a short fragment of reverse-transcribed RNA is used as a patch to repair a DNA DSB. These TSIPs provide unique polymorphic markers, similar to SNPs and variable tandem repeats, and can be used to track population migration and evolution. Similar to retrotransposon insertions, we suspect that TSIPs, which can disrupt coding regions of the genome, may play a role in both the etiology of genetic diseases as well as mammalian evolution.

AUTHOR CONTRIBUTIONS

PA conceived and edited mini-review. MO conceived mini-review, generated figures, and wrote the first draft.

ACKNOWLEDGMENTS

This work was supported by the intramural research program of the NCI, NIH. MO was supported by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Research Activity Start-up (26890001) and Grant-in-Aid for Scientific Research (C) (16K09836).

REFERENCES

- Autexier, C., and Lue, N. F. (2006). The structure and function of telomerase reverse transcriptase. *Annu. Rev. Biochem.* 75, 493–517. doi: 10.1146/annurev.biochem.75.103004.142412
- Baltimore, D. (1985). Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* 40, 481–482. doi: 10.1016/0092-8674(85)90190-4
- Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., et al. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159–1170. doi: 10.1016/j.cell.2010.05.021
- Cann, R. L., Stoneking, M., and Wilson, A. C. (1987). Mitochondrial, D. N. A., and human evolution. *Nature* 325, 31–36. doi: 10.1038/325031a0
- Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., et al. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* 28, 1397–1409. doi: 10.1101/gad.241661.114
- Cheng, Y., Zhang, Z., Keenan, B., Roschke, A. V., Nakahara, K., and Aplan, P. D. (2010). Efficient repair of DNA double-strand breaks in malignant cells with structural instability. *Mutat. Res.* 683, 115–122. doi: 10.1016/j.mrfmmm.2009.10.016
- Chiruvella, K. K., Liang, Z., and Wilson, T. E. (2013). Repair of double-strand breaks by end joining. *Cold Spring Harb. Perspect. Biol.* 5:a012757. doi: 10.1101/cshperspect.a012757
- Deriano, L., and Roth, D. B. (2013). Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu. Rev. Genet.* 47, 433–455. doi: 10.1146/annurev-genet-110711-155540
- Ewing, A. D., Ballinger, T. J., Earl, D., Broad Institute Genome Sequencing and Analysis Program and Platform, Harris, C. C., Ding, L., et al. (2013). Retrotransposition of gene transcripts leads to structural variation in mammalian genomes. *Genome Biol.* 14:R22. doi: 10.1186/gb-2013-14-3-r22

- Genomes Project, C., Abecasis, G. R., Auton, A., Brooks, L. D., DePristo, M. A., Durbin, R. M., et al. (2012). An integrated map of genetic variation from 1,092 human genomes. *Nature* 491, 56–65. doi: 10.1038/nature11632
- Hammer, M. F. (1995). A recent common ancestry for human Y chromosomes. *Nature* 378, 376–378. doi: 10.1038/378376a0
- Hohn, O., Hanke, K., and Bannert, N. (2013). HERV-K(HML-2), the Best Preserved Family of HERVs: endogenization, expression, and implications in health and disease. *Front. Oncol.* 3:246. doi: 10.3389/fonc.2013.00246
- Huang, C. R., Schneider, A. M., Lu, Y., Niranjan, T., Shen, P., Robinson, M. A., et al. (2010). Mobile interspersed repeats are major structural variants in the human genome. *Cell* 141, 1171–1182. doi: 10.1016/j.cell.2010.05.026
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020
- Jasin, M., and Haber, J. E. (2016). The democratization of gene editing: insights from site-specific cleavage and double-strand break repair. *DNA Repair (Amst)*. 44, 6–16. doi: 10.1016/j.dnarep.2016.05.001
- Lin, Y., and Waldman, A. S. (2001). Capture of DNA sequences at double-strand breaks in mammalian chromosomes. *Genetics* 158, 1665–1674.
- Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605. doi: 10.1016/0092-8674(93)90078-5
- Malkova, A., and Haber, J. E. (2012). Mutations arising during repair of chromosome breaks. *Annu. Rev. Genet.* 46, 455–473. doi: 10.1146/annurev-genet-110711-155547
- Morrish, T. A., Bekbolysnov, D., Velliquette, D., Morgan, M., Ross, B., Wang, Y., et al. (2013). Multiple mechanisms contribute to telomere maintenance. *J. Cancer Biol. Res.* 1. Available online at: <https://www.jscimedcentral.com/CancerBiology/Articles/cancerbiology-1-1012.pdf>
- Morrish, T. A., Garcia-Perez, J. L., Stamato, T. D., Taccioli, G. E., Sekiguchi, J., and Moran, J. V. (2007). Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* 446, 208–212. doi: 10.1038/nature05560
- Morrish, T. A., Gilbert, N., Myers, J. S., Vincent, B. J., Stamato, T. D., Taccioli, G. E., et al. (2002). DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31, 159–165. doi: 10.1038/ng898
- Nergadze, S. G., Rocchi, M., Azzalin, C. M., Mondello, C., and Giulotto, E. (2004). Insertion of telomeric repeats at intrachromosomal break sites during primate evolution. *Genome Res.* 14, 1704–1710. doi: 10.1101/gr.2778904
- Nergadze, S. G., Santagostino, M. A., Salzano, A., Mondello, C., and Giulotto, E. (2007). Contribution of telomerase RNA retrotranscription to DNA double-strand break repair during mammalian genome evolution. *Genome Biol.* 8:R260. doi: 10.1186/gb-2007-8-12-r260
- Onozawa, M., Goldberg, L., and Aplan, P. D. (2015). Landscape of insertion polymorphisms in the human genome. *Genome Biol. Evol.* 7, 960–968. doi: 10.1093/gbe/evv043
- Onozawa, M., Zhang, Z., Kim, Y. J., Goldberg, L., Varga, T., Bergsagel, P. L., et al. (2014). Repair of DNA double-strand breaks by templated nucleotide sequence insertions derived from distant regions of the genome. *Proc. Natl. Acad. Sci. U.S.A.* 111, 7729–7734. doi: 10.1073/pnas.1321889111
- Piskareva, O., Ernst, C., Higgins, N., and Schmatchenko, V. (2013). The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *FEBS Open Bio* 3, 433–437. doi: 10.1016/j.fob.2013.09.005
- Ruiz-Herrera, A., García, F., Azzalin, C., Giulotto, E., Egozcue, J., Ponsà, M., et al. (2002). Distribution of intrachromosomal telomeric sequences (ITS) on *Macaca fascicularis* (Primates) chromosomes and their implication for chromosome evolution. *Hum. Genet.* 110, 578–586. doi: 10.1007/s00439-002-0730-6
- Sharma, N. K., Reyes, A., Green, P., Caron, M. J., Bonini, M. G., Gordon, D. M., et al. (2012). Human telomerase acts as a hTR-independent reverse transcriptase in mitochondria. *Nucleic Acids Res.* 40, 712–725. doi: 10.1093/nar/gkr758
- Storici, F., Bebenek, K., Kunkel, T. A., Gordenin, D. A., and Resnick, M. A. (2007). RNA-templated DNA repair. *Nature* 447, 338–341. doi: 10.1038/nature05720
- Sutovsky, P., Moreno, R. D., Ramalho-Santos, J., Dominko, T., Simerly, C., and Schatten, G. (2000). Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biol. Reprod.* 63, 582–590. doi: 10.1095/biolreprod.63.2.582
- Teng, S. C., Kim, B., and Gabriel, A. (1996). Retrotransposon reverse-transcriptase-mediated repair of chromosomal breaks. *Nature* 383, 641–644. doi: 10.1038/383641a0
- Turner, C., Killoran, C., Thomas, N. S., Rosenberg, M., Chuzhanova, N. A., Johnston, J., et al. (2003). Human genetic disease caused by *de novo* mitochondrial-nuclear DNA transfer. *Hum. Genet.* 112, 303–309. doi: 10.1007/s00439-002-0892-2
- Underhill, P. A., Shen, P., Lin, A. A., Jin, L., Passarino, G., Yang, W. H., et al. (2000). Y chromosome sequence variation and the history of human populations. *Nat. Genet.* 26, 358–361. doi: 10.1038/81685
- Varga, T., and Aplan, P. D. (2005). Chromosomal aberrations induced by double-strand DNA breaks. *DNA Repair (Amst)*. 4, 1038–1046. doi: 10.1016/j.dnarep.2005.05.004
- Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., et al. (2001). Human L1 retrotransposition: *cis* preference versus *trans* complementation. *Mol. Cell. Biol.* 21, 1429–1439. doi: 10.1128/MCB.21.4.1429-1439.2001
- Woischnik, M., and Moraes, C. T. (2002). Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* 12, 885–893. doi: 10.1101/gr.227202
- Yu, X., and Gabriel, A. (1999). Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell* 4, 873–881. doi: 10.1016/S1097-2765(00)80397-4
- Zhou, Q., Li, H., Li, H., Nakagawa, A., Lin, J. L., Lee, E. S., et al. (2016). Mitochondrial endonuclease G mediates breakdown of paternal mitochondria upon fertilization. *Science* 353, 394–399. doi: 10.1126/science.aaf4777

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 © 2016 Onozawa and Aplan. 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) or licensor 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.



Post-Transcriptional Control of LINE-1 Retrotransposition by Cellular Host Factors in Somatic Cells

Javier G. Pizarro and Gaël Cristofari*

Institute for Research on Cancer and Aging of Nice (IRCAN), Faculty of Medicine, CNRS UMR7284, INSERM U1081, University of Nice Sophia Antipolis, Nice, France

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Independent Researcher, USA

Reviewed by:

Sandy Martin,
University of Colorado School of
Medicine, USA
Lixin Dai,
Modern Meadow Inc., USA

*Correspondence:

Gaël Cristofari
gael.cristofari@unice.fr

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Cell and Developmental
Biology

Received: 14 December 2015

Accepted: 18 February 2016

Published: 07 March 2016

Citation:

Pizarro JG and Cristofari G (2016)
Post-Transcriptional Control of LINE-1
Retrotransposition by Cellular Host
Factors in Somatic Cells.
Front. Cell Dev. Biol. 4:14.
doi: 10.3389/fcell.2016.00014

Long Interspersed Element-1 (LINE-1 or L1) retrotransposons form the only autonomously active family of transposable elements in humans. They are expressed and mobile in the germline, in embryonic stem cells and in the early embryo, but are silenced in most somatic tissues. Consistently, they play an important role in individual genome variations through insertional mutagenesis and sequence transduction, which occasionally lead to novel genetic diseases. In addition, they are reactivated in nearly half of the human epithelial cancers, contributing to tumor genome dynamics. The L1 element codes for two proteins, ORF1p and ORF2p, which are essential for its mobility. ORF1p is an RNA-binding protein with nucleic acid chaperone activity and ORF2p possesses endonuclease and reverse transcriptase activities. These proteins and the L1 RNA assemble into a ribonucleoprotein particle (L1 RNP), considered as the core of the retrotransposition machinery. The L1 RNP mediates the synthesis of new L1 copies upon cleavage of the target DNA and reverse transcription of the L1 RNA at the target site. The L1 element takes benefit of cellular host factors to complete its life cycle, however several cellular pathways also limit the cellular accumulation of L1 RNPs and their deleterious activities. Here, we review the known cellular host factors and pathways that regulate positively or negatively L1 retrotransposition at post-transcriptional level, in particular by interacting with the L1 machinery or L1 replication intermediates; and how they contribute to control L1 activity in somatic cells.

Keywords: LINE-1, retrotransposon, genome evolution, repeated sequences, retrotransposition, structural variation (SV)

L1 ELEMENTS CONTRIBUTE TO THE DYNAMICS OF SOMATIC AND GERMLINE HUMAN GENOMES

The Long Interspersed Element-1 (LINE-1 or L1) retrotransposon forms 17% of our genome (Lander et al., 2001). Most L1 copies present in the reference human genome are defective but ~100 copies could be retrotransposition-competent (Brouha et al., 2003). In addition, many polymorphic L1 elements, not included in the reference genome, also have the potential to mobilize (Beck et al., 2010; Ewing, 2015; Mir et al., 2015).

L1 elements can retrotranspose in the germline, in embryonic stem cells and in the early embryo (Kazazian et al., 1988; Garcia-Perez et al., 2007; van den Hurk et al., 2007). However, L1

retrotransposons are repressed in most tested normal somatic cells except in the brain (Coufal et al., 2009; Baillie et al., 2011; Evrony et al., 2012; Richardson et al., 2014b; Upton et al., 2015). L1 mobilization impacts human genome evolution through insertional mutagenesis and sequence transduction, which occasionally results in inherited genetic diseases (Hancks and Kazazian, 2012). Somatic retrotransposition in the brain could also contribute to the etiology of some mental disorders or disabilities, such as Rett Syndrome or Ataxia Telangiectasia, characterized by increased levels of L1 mobilization (Muotri et al., 2010; Coufal et al., 2011). Moreover, somatic L1 mobilization participates to the dynamics of many tumor genomes and can lead to driver mutations (Miki et al., 1992; Iskow et al., 2010; Lee et al., 2012; Solyom et al., 2012; Shukla et al., 2013; Helman et al., 2014; Tubio et al., 2014; Doucet-O'Hare et al., 2015; Ewing et al., 2015; Rodić et al., 2015). Besides its impact as an insertional mutagen, L1 also triggers other forms of genomic alterations such as DNA double-strand breaks or chromosomal translocations, and these activities could participate to normal aging or tumorigenesis (Wallace et al., 2008; Lin et al., 2009; Belancio et al., 2010). Finally, the L1 machinery also drives the retrotransposition of Short Interspersed Elements (SINEs) and the formation of processed pseudogenes (Esnault et al., 2000; Dewannieux et al., 2003).

L1 elements and their host have co-evolved: L1s use the cellular machinery for their own replication, while the host cell has evolved multiple defense mechanisms limiting

L1 deleterious effects. Silencing L1 expression, through CpG DNA methylation and histone modifications is a major repressive mechanism, which prevents the accumulation of mutagenic events (Bourc'His and Bestor, 2004; Castro-Diaz et al., 2014; Jacobs et al., 2014). Here we review post-transcriptional cellular pathways, which regulate positively or negatively L1 retrotransposition in somatic cells, in particular by interacting with the L1 machinery or L1 replicative intermediates.

L1 REPLICATION IS MEDIATED BY A RIBONUCLEOPROTEIN PARTICLE (RNP) AND TARGET-PRIMED REVERSE TRANSCRIPTION (TPRT)

An active L1 retrotransposon comprises a 5' untranslated region (UTR), two open reading frames (ORF1 and ORF2) separated by a short inter-ORF spacer and a 3' UTR (Figure 1). An antisense ORF0 of unknown function has also been recently described in the 5' UTR (Denli et al., 2015). As a consequence of the reverse transcription and integration mechanism, L1 sequence ends with a poly(dA) stretch and is flanked by target site duplications (TSD) of variable size. The 5' UTR contains RNA polymerase II sense and antisense promoters (Swergold, 1990; Speek, 2001; Nigumann et al., 2002). The translation of the bicistronic L1 mRNA by an unconventional mechanism produces

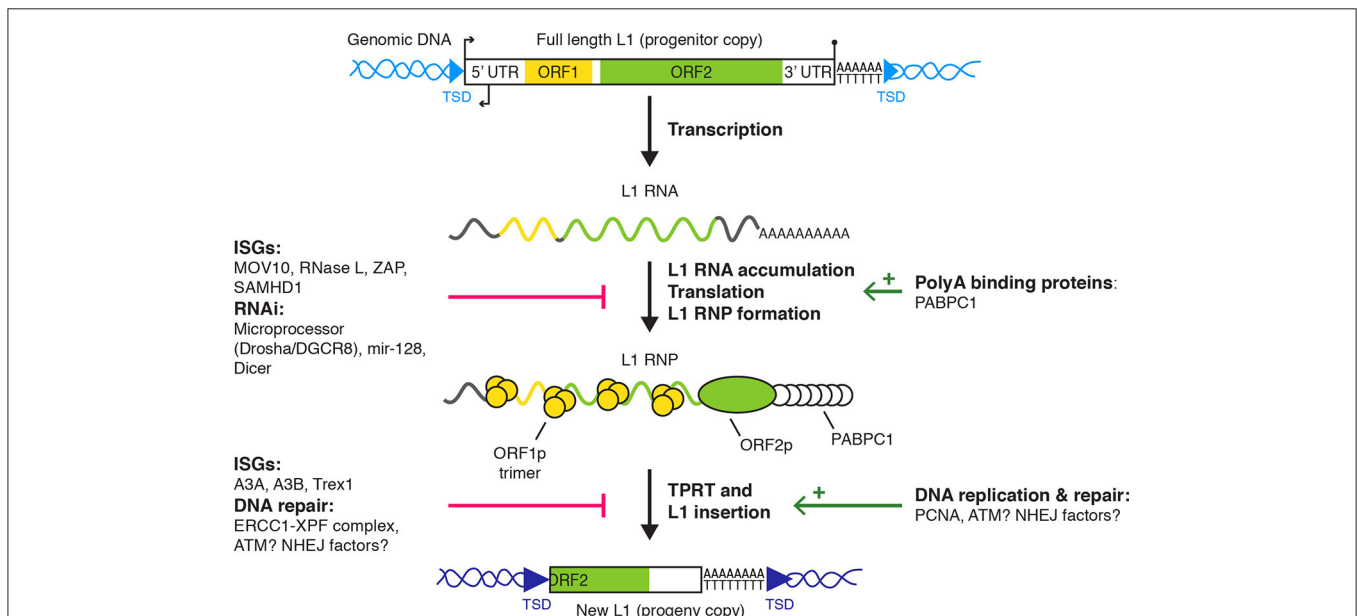


FIGURE 1 | L1 retrotransposition and cellular regulators. L1 replication starts with L1 transcription into a full length bicistronic L1 mRNA, its translation into ORF1p and ORF2p, and the assembly of an L1 RNP. For the sake of simplicity, the recently described antisense ORF0 in the 5' UTR is not depicted (Denli et al., 2015). The L1 RNP accumulates in stress granules and at least a fraction of it is imported to the nucleus (not shown) where target-primed reverse transcription (TPRT) occurs at the target DNA site. Finally, L1 insertion is resolved by an unknown mechanism (see main text for details). Only L1 regulators with a defined target/mechanism are depicted. Small broken arrows, L1 sense and antisense promoter activities; lollipop, L1 polyadenylation signal; light and dark blue arrowheads, target sites of L1 progenitor and progeny copies, respectively; red bars, negative regulation; green arrows, positive regulation. TSD, target site duplication; UTR, untranslated region; RNP, ribonucleoprotein particle; ISG, interferon-responsive genes; RNAi, RNA interference.

two proteins, named ORF1p and ORF2p (Alisch et al., 2006; Dmitriev et al., 2007). ORF1p is a 40 kDa RNA-binding protein, forming trimers and with nucleic acid chaperone activity (Martin, 1991; Holmes et al., 1992; Martin and Bushman, 2001; Martin et al., 2003; Khazina et al., 2011). ORF2p is a ~150 kDa protein with endonuclease (EN) and reverse transcriptase (RT) activities, which are critical for L1 retrotransposition (Mathias et al., 1991; Feng et al., 1996; Moran et al., 1996). ORF2p also contains a C-terminal cysteine-rich region, potentially contributing to its RNA binding capability (Piskareva et al., 2013). ORF1p and ORF2p bind the L1 mRNA to form a ribonucleoprotein particle (RNP), considered as the core of the L1 replicative complex (Hohjoh and Singer, 1996; Kolosha and Martin, 1997; Kulpa and Moran, 2005, 2006; Doucet et al., 2010; Goodier et al., 2010). This assembly occurs preferentially *in cis* (Esnault et al., 2000; Wei et al., 2001; Kulpa and Moran, 2006), through binding of ORF2p to the L1 RNA poly(A) sequence (Doucet et al., 2015). L1 RNPs accumulate in cytoplasmic foci, which colocalize with stress granules (Goodier et al., 2007, 2010; Doucet et al., 2010). The functional importance of these cytoplasmic complexes remains to be elucidated. Although cell division seems to promote retrotransposition, it is not absolutely required (Kubo et al., 2006; Shi et al., 2007; Xie et al., 2013). Thus, access of L1 RNPs to chromatin can occur independently of mitotic nuclear envelope breakdown through an unknown nuclear import mechanism.

New L1 copies are directly synthesized and inserted in the genome by a process called TPRT (Luan et al., 1993; Feng et al., 1996; Cost et al., 2002; Christensen et al., 2006). During TPRT, ORF2p binds and nicks a consensus sequence of the form 5'-TTTT/A-3' in the genomic DNA (Feng et al., 1996). This cleavage, potentially followed by additional processing steps, exposes a single-stranded T-rich DNA stretch able to partially or completely anneal to the L1 RNA poly(A) tail and to prime ORF2p-mediated reverse transcription (Kulpa and Moran, 2006; Monot et al., 2013; Viollet et al., 2014). A possible second nick, generally few nucleotides downstream of the first one, allows priming and synthesis of the second DNA strand. Finally, the L1 DNA ends are filled in and sealed, creating TSD (Luan et al., 1993; Feng et al., 1996; Cost et al., 2002). The molecular actors involved in these late stages are unknown. This process is frequently abortive, resulting in 5' truncated L1 copies.

L1 RETROTRANSPPOSITION IS REGULATED BY CELLULAR FACTORS AT MULTIPLE LEVELS

L1 activity is regulated at multiple stages of the L1 retrotransposition cycle (Figure 1). We focus here on post-transcriptional mechanisms and their molecular effectors acting in human or mammalian somatic cells and interacting with components of the L1 RNP or with L1 replication intermediates. L1 regulation in the germline, notably by Piwi-interacting RNA (piRNA), has been reviewed elsewhere (Zamudio and Bourc'his, 2010; Crichton et al., 2014) and is not detailed in the present article.

Proteomic Studies Have Revealed Cellular Partners of L1 RNPs and Potential Novel Regulators of L1 Retrotransposition Overview

Several recent studies have identified cellular partners of L1 RNPs through tagging of ORF1p, ORF2p or L1 RNA, followed by affinity chromatography and mass-spectrometry (Goodier et al., 2013; Peddigari et al., 2013; Taylor et al., 2013; Moldovan and Moran, 2015). These experimental efforts differ by the cell line, the L1 clone, the tagged component in the complex and the chromatography method used, but eventually lead to a number of common host factors (Figure 2). It should be underlined that only a fraction of the hits has been validated by co-immunoprecipitation, and only a single study used quantitative mass-spectrometry to measure the specific enrichment of the detected proteins upon elution (Taylor et al., 2013). A first step toward functional characterization generally involves retrotransposition assays in cultured cells upon depletion or overexpression of the tested factor. The outcome of these genetic assays allows a first classification into positive or negative regulators. However, many binding partners only modestly impact the levels of L1 retrotransposition in these assays, or have pleiotropic effects preventing unambiguous interpretation. With few exceptions, the majority of the tested factors are RNA-binding proteins, which copurify with ORF1p through an indirect RNA bridge, colocalize with L1 RNPs in stress granules, and inhibit L1 retrotransposition.

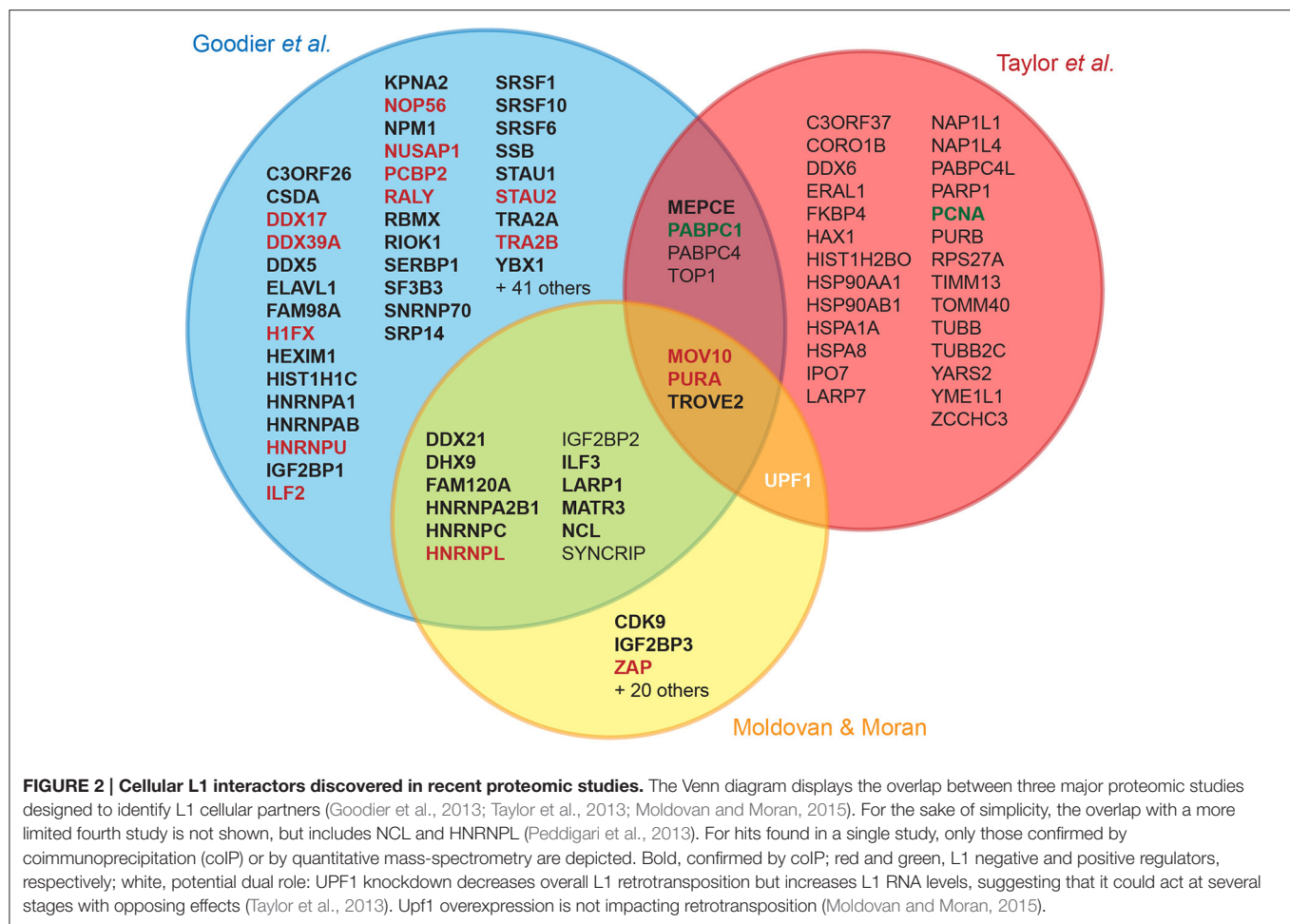
Limitations

Due to the scarcity of L1 endogenous complexes in cells, all proteomic studies rely on the overexpression of engineered L1 constructs. It is conceivable that: (i) some of the discovered partners become associated with L1 components as a result of L1 overexpression beyond physiological levels. (ii) L1 RNP stoichiometry is altered; (iii) the retrotransposition reporter cassette, which contains an intron, modifies L1 RNA cellular processing, and thus its binding partners.

Positive Regulators of L1 Retrotransposition

Poly(A) Binding Proteins Act in L1 RNP Assembly or Trafficking

Poly(A) binding proteins (PABPs) bind mRNA poly(A) tails and are involved in mRNA stability and translation initiation (Goss and Kleiman, 2013). Short hairpin RNA (shRNA)-mediated knockdown of PABPC1, reduces L1 retrotransposition with minimal effects on L1 RNA and proteins accumulation, or poly(A) tail length (Dai et al., 2012). This effect is associated with reduced L1 RNP levels and reduced nuclear accumulation of this complex, suggesting a possible—direct or indirect—role of PABPC1 in the assembly or the subcellular trafficking of the L1 RNP. Consistently, PABPC1 associates with the L1 RNP in an RNA-dependent manner, they colocalize in stress granules (Goodier et al., 2013; Taylor et al., 2013), and moderate PABPC1 overexpression stimulates retrotransposition (Dai et al., 2012).



Other PABPs have been found to associate with the L1 RNA (PABPN1, PABPC4) but addressing their specific role in L1 retrotransposition has been hampered by pleiotropic effects, or not yet tested (Dai et al., 2012; Goodier et al., 2013; Taylor et al., 2013).

PCNA is a Cofactor of TPRT

PCNA is a DNA sliding clamp acting as a processivity factor for many DNA polymerases during DNA replication or DNA damage repair (Moldovan et al., 2007). ORF2p binds PCNA through a PCNA-interacting protein (PIP) box, located between the EN and RT domains of ORF2p (Taylor et al., 2013). Mutations in ORF2p PIP box disrupt PCNA-ORF2p interaction and inhibit L1 retrotransposition. Interestingly, ORF2p mutations abrogating its EN or RT activity also disrupt PCNA-ORF2p interaction, suggesting that PCNA binding to ORF2p occurs downstream or concomitantly with TPRT.

Proline-Directed Protein Kinase(s) Regulate(s) ORF1p Function

ORF1p contains several (S/T)-P putative phosphorylation sites for proline-directed protein kinases (PDPKs), such as mitogen-activated protein kinases and cyclin-dependent kinases.

Mutations of S18, S27, T203, and T213, which are potential PDPK targets, decrease L1 retrotransposition; and these residues were found phosphorylated by mass-spectrometry in human cells (Cook et al., 2015). Interestingly, several protein kinases associate with the L1 RNP (Goodier et al., 2013; Taylor et al., 2013; Moldovan and Moran, 2015), however it remains to be demonstrated if one or several of them might directly target ORF1p. Interestingly, S18/S27 sites in ORF1p are required for binding by Pin1 prolyl isomerase (Cook et al., 2015), suggesting a scenario in which binding of Pin1 promotes ORF1p conformational change, which could affect its stability, activity or localization, or its subsequent ability to be dephosphorylated (Yeh et al., 2004; Liou et al., 2011).

Cellular Pathways Inhibiting L1 Retrotransposition at Post-Transcriptional Level

RNA Interference Pathways Prevent the Accumulation of L1 RNA

L1 RNA serves both as an mRNA to produce the L1 machinery and as a template for reverse transcription. Multiple RNA interference (RNAi) pathways act in somatic or embryonic

cells to prevent the accumulation of L1 RNA, and eventually retrotransposition.

First, the Microprocessor complex (Drosha/DGCR8), a major nuclear complex implicated in microRNA (miRNA) biosynthesis through pri-miRNA processing, is also able to bind L1 RNA *in vivo*, to reduce its abundance and to limit L1 retrotransposition. In addition, it can cleave various L1 RNA fragments derived from the L1 5' UTR region *in vitro*, indicating that L1 RNA can be a direct Microprocessor substrate (Heras et al., 2013, 2014). Moreover, the miRNA pathway could also act downstream of Microprocessor to inhibit retrotransposition. Indeed, miR-128 in complex with the Argonaute (Ago) protein binds the L1 RNA in the ORF2 region leading to L1 transcript degradation (Hamdorf et al., 2015).

Second, the combined expression of sense and antisense L1 transcripts driven by L1 5' UTR promoters reduces L1 RNA stability and L1 retrotransposition (Yang and Kazazian, 2006). This process is associated with the synthesis of rasiRNA (repeat-associated small interfering RNA) consistent with a possible processing of L1 RNA duplexes, and is modestly inhibited by Dicer knockdown, suggesting an additional layer of L1 repression mediated by siRNA mechanisms. In agreement with a role of RNAi pathways in somatic L1 regulation, L1 RNPs tend to accumulate in stress granules where they colocalize with several RNAi factors and often interact with them (Goodier et al., 2007, 2013).

Innate Immunity and Interferon Response Pathways

The cellular innate immune response is one of the first lines of defense against a broad range of viral infections. It involves cellular factors with antiviral activities, among which the interferon (IFN) response pathway plays a central role (MacMicking, 2012; Ivashkiv and Donlin, 2014). This pathway leads to the activation of IFN-stimulated genes (ISG) acting as effectors and reinforcing IFN-signaling itself. A significant proportion of ISG are viral restriction factors (MacMicking, 2012), which also appear to counteract L1 retrotransposition (Goodier et al., 2015), and are described below.

Upon overexpression, several members of the APOBEC3 (A3) cytidine deaminase family inhibit L1 retrotransposition (A3A, A3B, A3C and A3F) (Bogerd et al., 2006; Chen et al., 2006; Muckenfuss et al., 2006; Stenglein and Harris, 2006; Kinomoto et al., 2007; Niewiadomska et al., 2007). A3A is a nuclear protein predominantly expressed in peripheral blood mononuclear cells (PBMCs) and is induced by IFN- β (Chen et al., 2006; Muckenfuss et al., 2006; Stenglein et al., 2010). A3A-mediated L1 inhibition depends on A3A deaminase activity and on the subsequent processing of the deaminated DNA by uracil DNA glycosylase (UNG) and apurinic/apyrimidinic endonuclease (APE) (Richardson et al., 2014a). A3B is also a nuclear protein. It is endogenously expressed in embryonic stem cells, in induced-pluripotent stem cells and in a number of cancer cell lines. Its depletion stimulates L1 retrotransposition (Wissing et al., 2011; Marchetto et al., 2013); however, catalytically dead A3B mutants still inhibit L1 retrotransposition (Wissing et al., 2011). Thus, the mechanism by which A3B represses L1 mobilization remains unknown. Similarly, reducing the expression of A3C moderately increases retrotransposition in

cancer cell lines that express detectable levels of endogenous A3C (Muckenfuss et al., 2006). As for A3B, A3C- and A3F-mediated L1 repression is deaminase-independent (Muckenfuss et al., 2006; Stenglein and Harris, 2006; Kinomoto et al., 2007; Horn et al., 2014). A3C might interfere with L1 reverse transcription or the activity of ORF2p in the L1 RNP (Horn et al., 2014).

Several other ISG products, such as MOV10, ZAP or RNase L, limit L1 replication by limiting L1 RNA accumulation. The RNA helicase MOV10 robustly copurifies with the L1 complex, colocalizes with L1 RNPs in stress granules, reduces L1 RNA half-life, and ultimately strongly inhibits retrotransposition (Arjan-Odedra et al., 2012; Goodier et al., 2012, 2013; Li et al., 2013; Taylor et al., 2013; Moldovan and Moran, 2015). Similarly, the zinc-finger antiviral protein (ZAP) associates with the L1 RNP and accumulates with it in stress granules (Goodier et al., 2015; Moldovan and Moran, 2015). Its overexpression reduces full-length L1 RNA levels, and L1 retrotransposition levels. ZAP zinc finger domain is necessary and sufficient for its anti-L1 activity. Inversely, knocking down endogenous ZAP increases L1 retrotransposition. The ribonuclease L (RNase L) degrades L1 RNA and inhibits retrotransposition although no association or colocalization was detected with the L1 RNP (Zhang et al., 2014). Other ISGs with known viral restriction activities (e.g., BST2, ISG20, MAVS, and MX2) are also strong inhibitors of L1 retrotransposition (Goodier et al., 2015), but their mechanism of action has not yet been explored.

Finally, *SAMHD1* and *TREX1* are ISGs involved in a negative feedback loop, acting as repressors of the interferon response itself. Loss-of-function mutations in these genes lead to the Aicardi-Goutières syndrome, an autoimmune disease. Both factors inhibit L1 retrotransposition (Stetson et al., 2008; Zhao et al., 2013). Trex1 (Three-prime-repair exonuclease 1) is an abundant 3'-5' DNA exonuclease and its overexpression inhibits engineered L1 retrotransposition in cultured cells (Stetson et al., 2008). Trex1-deficient cells accumulate ssDNA fragments derived from various retroelements including L1, suggesting that Trex1 can metabolize reverse transcribed L1 cDNA (Stetson et al., 2008). *SAMHD1* (SAM Domain And HD Domain 1) impairs lentivirus replication in non-dividing cells by depleting the intracellular pool of dNTPs and thereby inhibiting reverse transcription (Lahouassa et al., 2012). In contrast, *SAMHD1* inhibits L1 retrotransposition in dividing cells, through a dNTPase-independent mechanism, which might directly affect ORF2p levels, and thus inhibit L1 reverse transcription (Zhao et al., 2013).

DNA Repair Pathways

EN-mediated cleavage of the target DNA or other TPRT intermediates could lead to DNA double-strand break (DSB) or DNA lesion signaling, and activation of subsequent DNA repair pathways. Conversely, these cellular processes could also participate in the resolution of L1 integration, through L1 second strand DNA synthesis or DNA ligation.

The role of DSB signaling and non-homologous end-joining (NHEJ) pathways remains controversial. Ataxia-telangiectasia mutated (ATM) protein, a kinase activated upon DSB, was initially proposed to be required for L1 retrotransposition and L1-induced DSBs (Gasior et al., 2006; Wallace et al., 2013).

However, independent studies using ATM-deficient mice or human cell models rather suggest that ATM is a repressor of retrotransposition (Coufal et al., 2011). Similarly, knocking out NHEJ genes (e.g., Ku70/80, DNA Ligase IV or Artemis) decreases L1 retrotransposition in chicken cells (Suzuki et al., 2009). However, loss-of-function of DNA-PKcs or DNA Ligase IV in mammalian cells does not impair L1 retrotransposition (Coufal et al., 2011), indicating that NHEJ is not absolutely required for L1 retrotransposition. An interesting possibility could be that DSB signaling and repair pathways compete with the L1 machinery or other cellular factors for the resolution of L1 insertion during—or after—cDNA synthesis, leading to 5′ truncated insertions (Zingler et al., 2005; Suzuki et al., 2009; Coufal et al., 2011).

Other DNA repair pathways can also antagonize L1 replication. The ERCC1-XPF complex, which plays a role in nucleotide excision, base excision and interstrand crosslink repair pathways is a potent inhibitor of L1 retrotransposition (Gasior et al., 2008). ERCC1-XPF is an endonuclease able to specifically cleave DNA at junctions between single-stranded and double-stranded regions, a predicted structure produced by the TPRT process. Thus, it has been hypothesized that ERCC1-XPF might cut off L1 cDNA at the target site during reverse transcription.

OPEN QUESTIONS FOR THE FUTURE

- How is unspliced L1 RNA exported to the cytosol and the L1 RNP imported back to the nucleus?

REFERENCES

- Alisch, R. S., Garcia-Perez, J. L., Muotri, A. R., Gage, F. H., and Moran, J. V. (2006). Unconventional translation of mammalian LINE-1 retrotransposons. *Genes Dev.* 20, 210–224. doi: 10.1101/gad.1380406
- Arjan-Oedra, S., Swanson, C. M., Sherer, N. M., Wolinsky, S. M., and Malim, M. H. (2012). Endogenous MOV10 inhibits the retrotransposition of endogenous retroelements but not the replication of exogenous retroviruses. *Retrovirology* 9:53. doi: 10.1186/1742-4690-9-53
- Baillie, J. K., Barnett, M. W., Upton, K. R., Gerhardt, D. J., Richmond, T. A., De Sapio, F., et al. (2011). Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479, 534–537. doi: 10.1038/nature10531
- Beck, C. R., Collier, P., Macfarlane, C., Malig, M., Kidd, J. M., Eichler, E. E., et al. (2010). LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159–1170. doi: 10.1016/j.cell.2010.05.021
- Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R., and Deininger, P. (2010). Somatic expression of LINE-1 elements in human tissues. *Nucleic Acids Res.* 38, 3909–3922. doi: 10.1093/nar/gkq132
- 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
- Bourc’His, D., and Bestor, T. H. (2004). Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96–99. doi: 10.1038/nature02886
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285. doi: 10.1073/pnas.0831042100
- Castro-Diaz, N., Ecco, G., Coluccio, A., Kapopoulou, A., Yazdanpanah, B., Friedli, M., et al. (2014). Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes Dev.* 28, 1397–1409. doi: 10.1101/gad.241661.114
- Chen, H., Lilley, C. E., Yu, Q., Lee, D. V., Chou, J., Narvaiza, I., et al. (2006). APOBEC3A is a potent inhibitor of adeno-associated virus and retrotransposons. *Curr. Biol.* 16, 480–485. doi: 10.1016/j.cub.2006.01.031
- Christensen, S. M., Ye, J., and Eickbush, T. H. (2006). RNA from the 5′ end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17602–17607. doi: 10.1073/pnas.0605476103
- Cook, P. R., Jones, C. E., and Furano, A. V. (2015). Phosphorylation of ORF1p is required for L1 retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4298–4303. doi: 10.1073/pnas.1416869112
- Cost, G. J., Feng, Q., Jacquier, A., and Boeke, J. D. (2002). Human L1 element target-primed reverse transcription *in vitro*. *EMBO J.* 21, 5899–5910. doi: 10.1093/emboj/cdf592
- Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Marchetto, M. C., Muotri, A. R., Mu, Y., et al. (2011). Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. U.S.A.* 108, 20382–20387. doi: 10.1073/pnas.1100273108
- Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., et al. (2009). L1 retrotransposition in human neural progenitor cells. *Nature* 460, 1127–1131. doi: 10.1038/nature08248
- Crichton, J. H., Dunican, D. S., MacLennan, M., Meehan, R. R., and Adams, I. R. (2014). Defending the genome from the enemy within: mechanisms of retrotransposon suppression in the mouse germline. *Cell. Mol. Life Sci.* 71, 1581–1605. doi: 10.1007/s00018-013-1468-0
- Dai, L., Taylor, M. S., O’Donnell, K. A., and Boeke, J. D. (2012). Poly(A) binding protein C1 is essential for efficient L1 retrotransposition and affects L1 RNP formation. *Mol. Cell. Biol.* 32, 4323–4336. doi: 10.1128/MCB.06785-11
- How many distinct L1 RNP forms exist in the cell and throughout the L1 replication cycle?
- Do L1 components have a life outside of the L1 RNP and retrotransposition?
- How is L1 RNP assembly regulated?
- Does L1 component accumulation in stress granules reflect a host defense mechanism or an intermediate step during retrotransposition?
- Which restriction factors are the dominant ones and do they cooperate?

AUTHOR CONTRIBUTIONS

JGP drafted the manuscript. GC revised the manuscript.

FUNDING

GC is funded by the Fondation ARC pour la recherche sur le cancer, the European Research Council (ERC-2010-StG 243312, RETROGENOMICS), the French Government (National Research Agency, ANR) through the “Investments for the Future” (LABEX SIGNALIFE, ANR-11-LABX-0028-01), and the Fondation pour la Recherche Médicale (FRM DEP20131128533).

ACKNOWLEDGMENTS

We are grateful to Aurelien J. Doucet for critical reading.

- Denli, A. M., Narvaiza, I., Kerman, B. E., Pena, M., Benner, C., Marchetto, M. C., et al. (2015). Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* 163, 583–593. doi: 10.1016/j.cell.2015.09.025
- Dewannieux, M., Esnault, C., and Heidmann, T. (2003). LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35, 41–48. doi: 10.1038/ng1223
- Dmitriev, S. E., Andreev, D. E., Terenin, I. M., Olovnikov, I. A., Prassolov, V. S., Merrick, W. C., et al. (2007). Efficient translation initiation directed by the 900-nucleotide-long and GC-rich 5' untranslated region of the human retrotransposon LINE-1 mRNA is strictly cap dependent rather than internal ribosome entry site mediated. *Mol. Cell. Biol.* 27, 4685–4697. doi: 10.1128/MCB.02138-06
- Doucet, A. J., Hulme, A. E., Sahinovic, E., Kulpa, D. A., Moldovan, J. B., Kopera, H. C., et al. (2010). Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet.* 6:e1001150. doi: 10.1371/journal.pgen.1001150
- Doucet, A. J., Wilusz, J. E., Miyoshi, T., Liu, Y., and Moran, J. V. (2015). A 3' poly(A) tract is required for LINE-1 retrotransposition. *Mol. Cell* 60, 728–741. doi: 10.1016/j.molcel.2015.10.012
- Doucet-O'Hare, T. T., Rodić, N., Sharma, R., Darbari, I., Abril, G., Choi, J. A., et al. (2015). LINE-1 expression and retrotransposition in Barrett's esophagus and esophageal carcinoma. *Proc. Natl. Acad. Sci. U.S.A.* 112, E4894–E4900. doi: 10.1073/pnas.1502474112
- Esnault, C., Maestre, J., and Heidmann, T. (2000). Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24, 363–367. doi: 10.1038/74184
- Evrny, G. D., Cai, X., Lee, E., Hills, L. B., Elhosary, P. C., Lehmann, H. S., et al. (2012). Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* 151, 483–496. doi: 10.1016/j.cell.2012.09.035
- Ewing, A. D. (2015). Transposable element detection from whole genome sequence data. *Mob. DNA* 6, 24. doi: 10.1186/s13100-015-0055-3
- Ewing, A. D., Gacita, A., Wood, L. D., Ma, F., Xing, D., Kim, M. S., et al. (2015). Widespread somatic L1 retrotransposition occurs early during gastrointestinal cancer evolution. *Genome Res.* 25, 1536–1545. doi: 10.1101/gr.196238.115
- Feng, Q., Moran, J. V., Kazazian, H. H., and Boeke, J. D. (1996). Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905–916. doi: 10.1016/S0092-8674(00)81997-2
- Garcia-Perez, J. L., Doucet, A. J., Bucheton, A., Moran, J. V., and Gilbert, N. (2007). Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res.* 17, 602–611. doi: 10.1101/gr.5870107
- Gasior, S. L., Roy-Engel, A. M., and Deininger, P. L. (2008). ERCC1/XPF limits L1 retrotransposition. *DNA Repair (Amst.)* 7, 983–989. doi: 10.1016/j.dnarep.2008.02.006
- Gasior, S. L., Wakeman, T. P., Xu, B., and Deininger, P. L. (2006). The human LINE-1 retrotransposon creates DNA double-strand breaks. *J. Mol. Biol.* 357, 1383–1393. doi: 10.1016/j.jmb.2006.01.089
- Goodier, J. L., Cheung, L. E., and Kazazian, H. H. (2012). MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet.* 8:e1002941. doi: 10.1371/journal.pgen.1002941
- Goodier, J. L., Cheung, L. E., and Kazazian, H. H. (2013). Mapping the LINE1 ORF1 protein interactome reveals associated inhibitors of human retrotransposition. *Nucleic Acids Res.* 41, 7401–7419. doi: 10.1093/nar/gkt512
- Goodier, J. L., Mandal, P. K., Zhang, L., and Kazazian, H. H. (2010). Discrete subcellular partitioning of human retrotransposon RNAs despite a common mechanism of genome insertion. *Hum. Mol. Genet.* 19, 1712–1725. doi: 10.1093/hmg/ddq048
- Goodier, J. L., Pereira, G. C., Cheung, L. E., Rose, R. J., and Kazazian, H. H. (2015). The broad-spectrum antiviral protein ZAP restricts human retrotransposition. *PLoS Genet.* 11:e1005252. doi: 10.1371/journal.pgen.1005252
- Goodier, J. L., Zhang, L., Vetter, M. R., and Kazazian, H. H. (2007). LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol. Cell. Biol.* 27, 6469–6483. doi: 10.1128/MCB.00332-07
- Goss, D. J., and Kleiman, F. E. (2013). Poly(A) binding proteins: are they all created equal? *Wiley Interdiscip. Rev. RNA* 4, 167–179. doi: 10.1002/wrna.1151
- Hamdorf, M., Ilica, A., Zisoulis, D. G., Gamelin, L., Martin, C., Sanders, K. J., et al. (2015). miR-128 represses L1 retrotransposition by binding directly to L1 RNA. *Nat. Struct. Mol. Biol.* 22, 824–831. doi: 10.1038/nsmb.3090
- Hancks, D. C., and Kazazian, H. H. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.gde.2012.02.006
- Helman, E., Lawrence, M. S., Stewart, C., Sougnez, C., Getz, G., and Meyerson, M. (2014). Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* 24, 1053–1063. doi: 10.1101/gr.163659.113
- Heras, S. R., Macias, S., Cáceres, J. F., and Garcia-Perez, J. L. (2014). Control of mammalian retrotransposons by cellular RNA processing activities. *Mob. Genet. Elements* 4:e28439. doi: 10.4161/mge.28439
- Heras, S. R., Macias, S., Plass, M., Fernandez, N., Cano, D., Eyra, E., et al. (2013). The Microprocessor controls the activity of mammalian retrotransposons. *Nat. Struct. Mol. Biol.* 20, 1173–1181. doi: 10.1038/nsmb.2658
- Hohjoh, H., and Singer, M. F. (1996). Cytoplasmic ribonucleoprotein complexes containing human LINE-1 protein and RNA. *EMBO J.* 15, 630–639.
- Holmes, S. E., Singer, M. F., and Swergold, G. D. (1992). Studies on p40, the leucine zipper motif-containing protein encoded by the first open reading frame of an active human LINE-1 transposable element. *J. Biol. Chem.* 267, 19765–19768.
- Horn, A. V., Klawitter, S., Held, U., Berger, A., Vasudevan, A. A., Bock, A., et al. (2014). Human LINE-1 restriction by APOBEC3C is deaminase independent and mediated by an ORF1p interaction that affects LINE reverse transcriptase activity. *Nucleic Acids Res.* 42, 396–416. doi: 10.1093/nar/gkt898
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261. doi: 10.1016/j.cell.2010.05.020
- Ivashkiv, L. B., and Donlin, L. T. (2014). Regulation of type I interferon responses. *Nat. Rev. Immunol.* 14, 36–49. doi: 10.1038/nri3581
- Jacobs, F. M., Greenberg, D., Nguyen, N., Haeussler, M., Ewing, A. D., Katzman, S., et al. (2014). An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516, 242–245. doi: 10.1038/nature13760
- Kazazian, H. H., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332, 164–166. doi: 10.1038/332164a0
- Khazina, E., Truffault, V., Büttner, R., Schmidt, S., Coles, M., and Weichenrieder, O. (2011). Trimeric structure and flexibility of the L1ORF1 protein in human L1 retrotransposition. *Nat. Struct. Mol. Biol.* 18, 1006–1014. doi: 10.1038/nsmb.2097
- Kinamoto, M., Kanno, T., Shimura, M., Ishizaka, Y., Kojima, A., Kurata, T., et al. (2007). All APOBEC3 family proteins differentially inhibit LINE-1 retrotransposition. *Nucleic Acids Res.* 35, 2955–2964. doi: 10.1093/nar/gkm181
- Kolosh, V. O., and Martin, S. L. (1997). *In vitro* properties of the first ORF protein from mouse LINE-1 support its role in ribonucleoprotein particle formation during retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.* 94, 10155–10160. doi: 10.1073/pnas.94.19.10155
- Kubo, S., Seleme, M. C., Soifer, H. S., Perez, J. L., Moran, J. V., Kazazian, H. H., et al. (2006). L1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8036–8041. doi: 10.1073/pnas.0601954103
- Kulpa, D. A., and Moran, J. V. (2005). Ribonucleoprotein particle formation is necessary but not sufficient for LINE-1 retrotransposition. *Hum. Mol. Genet.* 14, 3237–3248. doi: 10.1093/hmg/ddi354
- Kulpa, D. A., and Moran, J. V. (2006). Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nat. Struct. Mol. Biol.* 13, 655–660. doi: 10.1038/nsmb1107
- Lahouassa, H., Daddacha, W., Hofmann, H., Ayinde, D., Logue, E. C., Dragin, L., et al. (2012). SAMHD1 restricts the replication of human immunodeficiency virus type 1 by depleting the intracellular pool of deoxynucleoside triphosphates. *Nat. Immunol.* 13, 223–228. doi: 10.1038/ni.2236
- 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, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971. doi: 10.1126/science.1222077
- Li, X., Zhang, J., Jia, R., Cheng, V., Xu, X., Qiao, W., et al. (2013). The MOV10 helicase inhibits LINE-1 mobility. *J. Biol. Chem.* 288, 21148–21160. doi: 10.1074/jbc.M113.465856

- Lin, C., Yang, L., Tanasa, B., Hutt, K., Ju, B. G., Ohgi, K., et al. (2009). Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* 139, 1069–1083. doi: 10.1016/j.cell.2009.11.030
- Liou, Y. C., Zhou, X. Z., and Lu, K. P. (2011). Prolyl isomerase Pin1 as a molecular switch to determine the fate of phosphoproteins. *Trends Biochem. Sci.* 36, 501–514. doi: 10.1016/j.tibs.2011.07.001
- Luan, D. D., Korman, M. H., Jakubczak, J. L., and Eickbush, T. H. (1993). Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605. doi: 10.1016/0092-8674(93)90078-5
- MacMicking, J. D. (2012). Interferon-inducible effector mechanisms in cell-autonomous immunity. *Nat. Rev. Immunol.* 12, 367–382. doi: 10.1038/nri3210
- Marchetto, M. C., Narvaiza, I., Denli, A. M., Benner, C., Lazzarini, T. A., Nathanson, J. L., et al. (2013). Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* 503, 525–529. doi: 10.1038/nature12686
- Martin, S. L. (1991). Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* 11, 4804–4807. doi: 10.1128/MCB.11.9.4804
- Martin, S. L., Branciforte, D., Keller, D., and Bain, D. L. (2003). Trimeric structure for an essential protein in L1 retrotransposition. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13815–13820. doi: 10.1073/pnas.2336221100
- Martin, S. L., and Bushman, F. D. (2001). Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. *Mol. Cell. Biol.* 21, 467–475. doi: 10.1128/MCB.21.2.467-475.2001
- Mathias, S. L., Scott, A. F., Kazazian, H. H., Boeke, J. D., and Gabriel, A. (1991). Reverse transcriptase encoded by a human transposable element. *Science* 254, 1808–1810. doi: 10.1126/science.1722352
- Miki, Y., Nishisho, I., Horii, A., Miyoshi, Y., Utsunomiya, J., Kinzler, K. W., et al. (1992). Disruption of the APC gene by a retrotranspositional insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–645
- Mir, A. A., Philippe, C., and Cristofari, G. (2015). euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res.* 43, D43–D47. doi: 10.1093/nar/gku1043
- Moldovan, G. L., Pfander, B., and Jentsch, S. (2007). PCNA, the maestro of the replication fork. *Cell* 129, 665–679. doi: 10.1016/j.cell.2007.05.003
- Moldovan, J. B., and Moran, J. V. (2015). The zinc-finger antiviral protein ZAP inhibits LINE and alu retrotransposition. *PLoS Genet.* 11:e1005121. doi: 10.1371/journal.pgen.1005121
- Monot, C., Kuciak, M., Viollet, S., Mir, A. A., Gabus, C., Darlix, J. L., et al. (2013). The specificity and flexibility of L1 reverse transcription priming at imperfect T-tracts. *PLoS Genet.* 9:e1003499. doi: 10.1371/journal.pgen.1003499
- Moran, J. V., Holmes, S. E., Naas, T. P., DeBerardinis, R. J., Boeke, J. D., and Kazazian, H. H. (1996). High frequency retrotransposition in cultured mammalian cells. *Cell* 87, 917–927. doi: 10.1016/S0092-8674(00)81998-4
- Muckenfuss, H., Hamdorf, M., Held, U., Perkovic, M., Lower, J., Cichutek, K., et al. (2006). APOBEC3 proteins inhibit human LINE-1 retrotransposition. *J. Biol. Chem.* 281, 22161–22172. doi: 10.1074/jbc.M601716200
- Muotri, A. R., Marchetto, M. C., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., et al. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446. doi: 10.1038/nature09544
- Niewiadomska, A. M., Tian, C., Tan, L., Wang, T., Sarkis, P. T., and Yu, X. F. (2007). Differential inhibition of long interspersed element 1 by APOBEC3 does not correlate with high-molecular-mass-complex formation or P-body association. *J. Virol.* 81, 9577–9583. doi: 10.1128/JVI.02800-06
- Nigumann, P., Redik, K., Mätlik, K., and Speck, M. (2002). Many human genes are transcribed from the antisense promoter of L1 retrotransposon. *Genomics* 79, 628–634. doi: 10.1006/geno.2002.6758
- Peddigari, S., Li, P. W., Rabe, J. L., and Martin, S. L. (2013). hnRNPL and nucleolin bind LINE-1 RNA and function as host factors to modulate retrotransposition. *Nucleic Acids Res.* 41, 575–585. doi: 10.1093/nar/gks1075
- Piskareva, O., Ernst, C., Higgins, N., and Schmatchenko, V. (2013). The carboxy-terminal segment of the human LINE-1 ORF2 protein is involved in RNA binding. *FEBS Open Bio* 3, 433–437. doi: 10.1016/j.fob.2013.09.005
- Richardson, S. R., Morell, S., and Faulkner, G. J. (2014b). L1 Retrotransposons and Somatic Mosaicism in the Brain. *Annu. Rev. Genet.* 48, 1–27. doi: 10.1146/annurev-genet-120213-092412
- Richardson, S. R., Narvaiza, I., Planegger, R. A., Weitzman, M. D., and Moran, J. V. (2014a). APOBEC3A deaminates transiently exposed single-strand DNA during LINE-1 retrotransposition. *Elife* 3:e02008. doi: 10.7554/elife.02008
- Rodić, N., Steranka, J. P., Makohon-Moore, A., Moyer, A., Shen, P., Sharma, R., et al. (2015). Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat. Med.* 21, 1060–1064. doi: 10.1038/nm.3919
- Shi, X., Seluanov, A., and Gorbunova, V. (2007). Cell divisions are required for L1 retrotransposition. *Mol. Cell. Biol.* 27, 1264–1270. doi: 10.1128/MCB.01888-06
- Shukla, R., Upton, K. R., Muñoz-Lopez, M., Gerhardt, D. J., Fisher, M. E., Nguyen, T., et al. (2013). Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111. doi: 10.1016/j.cell.2013.02.032
- Solyom, S., Ewing, A. D., Rahrmann, E. P., Doucet, T., Nelson, H. H., Burns, M. B., et al. (2012). Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res.* 22, 2328–2338. doi: 10.1101/gr.145235.112
- Speck, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* 21, 1973–1985. doi: 10.1128/MCB.21.6.1973-1985.2001
- Stenglein, M. D., Burns, M. B., Li, M., Lengyel, J., and Harris, R. S. (2010). APOBEC3 proteins mediate the clearance of foreign DNA from human cells. *Nat. Struct. Mol. Biol.* 17, 222–229. doi: 10.1038/nsmb.1744
- Stenglein, M. D., and Harris, R. S. (2006). APOBEC3B and APOBEC3F inhibit L1 retrotransposition by a DNA deamination-independent mechanism. *J. Biol. Chem.* 281, 16837–16841. doi: 10.1074/jbc.M602367200
- Stetson, D. B., Ko, J. S., Heidmann, T., and Medzhitov, R. (2008). Trex1 prevents cell-intrinsic initiation of autoimmunity. *Cell* 134, 587–598. doi: 10.1016/j.cell.2008.06.032
- Suzuki, J., Yamaguchi, K., Kajikawa, M., Ichikawa, K., Adachi, N., Koyama, H., et al. (2009). Genetic evidence that the non-homologous end-joining repair pathway is involved in LINE retrotransposition. *PLoS Genet.* 5:e1000461. doi: 10.1371/journal.pgen.1000461
- Sweragold, G. D. (1990). Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 10, 6718–6729. doi: 10.1128/MCB.10.12.6718
- Taylor, M. S., Lacava, J., Mita, P., Molloy, K. R., Huang, C. R., Li, D., et al. (2013). Affinity proteomics reveals human host factors implicated in discrete stages of LINE-1 retrotransposition. *Cell* 155, 1034–1048. doi: 10.1016/j.cell.2013.10.021
- Tubio, J. M., Li, Y., Ju, Y. S., Martincorena, I., Cooke, S. L., Tojo, M., et al. (2014). Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345:1251343. doi: 10.1126/science.1251343
- Upton, K. R., Gerhardt, D. J., Jesuadian, J. S., Richardson, S. R., Sánchez-Luque, F. J., Bodea, G. O., et al. (2015). Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* 161, 228–239. doi: 10.1016/j.cell.2015.03.026
- van den Hurk, J. A., Meij, I. C., Seleme, M. C., Kano, H., Nikopoulos, K., Hoefsloot, L. H., et al. (2007). L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* 16, 1587–1592. doi: 10.1093/hmg/ddm108
- Viollet, S., Monot, C., and Cristofari, G. (2014). L1 retrotransposition: the snap-elcrom model and its consequences. *Mob. Genet. Elements* 4:e28907. doi: 10.4161/mge.28907
- Wallace, N. A., Belancio, V. P., and Deininger, P. L. (2008). L1 mobile element expression causes multiple types of toxicity. *Gene* 419, 75–81. doi: 10.1016/j.gene.2008.04.013
- Wallace, N. A., Gasior, S. L., Faber, Z. J., Howie, H. L., Deininger, P. L., and Galloway, D. A. (2013). HPV 5 and 8 E6 expression reduces ATM protein levels and attenuates LINE-1 retrotransposition. *Virology* 443, 69–79. doi: 10.1016/j.virol.2013.04.022
- Wei, W., Gilbert, N., Ooi, S. L., Lawler, J. F., Ostertag, E. M., Kazazian, H. H., et al. (2001). Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 21, 1429–1439. doi: 10.1128/MCB.21.4.1429-1439.2001
- Wissing, S., Montano, M., Garcia-Perez, J. L., Moran, J. V., and Greene, W. C. (2011). Endogenous APOBEC3B restricts LINE-1 retrotransposition in transformed cells and human embryonic stem cells. *J. Biol. Chem.* 286, 36427–36437. doi: 10.1074/jbc.M111.251058
- Xie, Y., Mates, L., Ivics, Z., Izsvák, Z., Martin, S. L., and An, W. (2013). Cell division promotes efficient retrotransposition in a stable L1 reporter cell line. *Mob. DNA* 4:10. doi: 10.1186/1759-8753-4-10

- Yang, N., and Kazazian, H. H. (2006). L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nat. Struct. Mol. Biol.* 13, 763–771. doi: 10.1038/nsmb1141
- Yeh, E., Cunningham, M., Arnold, H., Chasse, D., Monteith, T., Ivaldi, G., et al. (2004). A signalling pathway controlling c-Myc degradation that impacts oncogenic transformation of human cells. *Nat. Cell Biol.* 6, 308–318. doi: 10.1038/ncb1110
- Zamudio, N., and Bourc'his, D. (2010). Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? *Heredity (Edinb)*. 105, 92–104. doi: 10.1038/hdy.2010.53
- Zhang, A., Dong, B., Doucet, A. J., Moldovan, J. B., Moran, J. V., and Silverman, R. H. (2014). RNase L restricts the mobility of engineered retrotransposons in cultured human cells. *Nucleic Acids Res.* 42, 3803–3820. doi: 10.1093/nar/gkt1308
- Zhao, K., Du, J., Han, X., Goodier, J. L., Li, P., Zhou, X., et al. (2013). Modulation of LINE-1 and Alu/SVA retrotransposition by aicardi-goutières syndrome-Related SAMHD1. *Cell Rep.* 4, 1108–1115. doi: 10.1016/j.celrep.2013.08.019
- Zingler, N., Willhoeft, U., Brose, H. P., Schoder, V., Jahns, T., Hanschmann, K. M., et al. (2005). Analysis of 5' junctions of human LINE-1 and Alu retrotransposons suggests an alternative model for 5'-end attachment requiring microhomology-mediated end-joining. *Genome Res.* 15, 780–789. doi: 10.1101/gr.3421505

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 © 2016 Pizarro and Cristofari. 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) or licensor 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.



An LTR Retrotransposon-Derived Gene Displays Lineage-Specific Structural and Putative Species-Specific Functional Variations in Eutherians

Masahito Irie^{1,2}, Akihiko Koga³, Tomoko Kaneko-Ishino^{1*} and Fumitoshi Ishino^{2*}

¹ Department of Nursing, School of Health Sciences, Tokai University, Isehara, Japan, ² Department of Epigenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan, ³ Department of Cellular and Molecular Biology, Primate Research Institute, Kyoto University, Inuyama, Japan

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Formerly affiliated with University of
Toledo, USA

Reviewed by:

Richard McLaughlin,
Fred Hutchinson Cancer Research
Center, USA
Annette Damert,
Babes-Bolyai University, Romania
Tom F. Moore,
University College Cork, Ireland

*Correspondence:

Tomoko Kaneko-Ishino
tkaneko@is.icc.u-tokai;
Fumitoshi Ishino
fishino.epgn@mri.tmd.ac.jp

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 12 March 2016

Accepted: 01 June 2016

Published: 23 June 2016

Citation:

Irie M, Koga A, Kaneko-Ishino T and
Ishino F (2016) An LTR
Retrotransposon-Derived Gene
Displays Lineage-Specific Structural
and Putative Species-Specific
Functional Variations in Eutherians.
Front. Chem. 4:26.
doi: 10.3389/fchem.2016.00026

Amongst the 11 eutherian-specific genes acquired from a sushi-ichi retrotransposon is the CCHC type zinc-finger protein-encoding gene *SIRH11/ZCCHC16*. Its contribution to eutherian brain evolution is implied because of its involvement in cognitive function in mice, possibly via the noradrenergic system. Although, the possibility that *Sirh11/Zcchc16* functions as a non-coding RNA still remains, dN/dS ratios in pairwise comparisons between its orthologs have provided supportive evidence that it acts as a protein. It became a pseudogene in armadillos (Cingulata) and sloths (Pilosa), the only two extant orders of xenarthra, which prompted us to examine the lineage-specific variations of *SIRH11/ZCCHC16* in eutherians. We examined the predicted *SIRH11/ZCCHC16* open reading frame (ORF) in 95 eutherian species based on the genomic DNA information in GenBank. A large variation in the *SIRH11/ZCCHC16* ORF was detected in several lineages. These include a lack of a CCHC RNA-binding domain in its C-terminus, observed in gibbons (Hylobatidae: Primates) and megabats (Megachiroptera: Chiroptera). A lack of the N-terminal half, on the other hand, was observed in New World monkeys (Platyrrhini: Primates) and species belonging to New World and African Hystricognaths (Caviomorpha and Bathyergidae: Rodents) along with Cetacea and Ruminantia (Cetartiodactyla). Among the hominoids, interestingly, three out of four genera of gibbons have lost normal *SIRH11/ZCCHC16* function by deletion or the lack of the CCHC RNA-binding domain. Our extensive dN/dS analysis suggests that such truncated *SIRH11/ZCCHC16* ORFs are functionally diversified even within lineages. Combined, our results show that *SIRH11/ZCCHC16* may contribute to the diversification of eutherians by lineage-specific structural changes after its domestication in the common eutherian ancestor, followed by putative species-specific functional changes that enhanced fitness and occurred as a consequence of complex natural selection events.

Keywords: LTR retrotransposon, genome evolution, exaptation, primate, brain functions

INTRODUCTION

Mutation and selection are two principal factors in the Darwinian theory of evolution. The domestication of long terminal repeat (LTR) retrotransposons and retroviruses is a kind of mutation that promotes macroevolution through diversification of genomic function by creating new host genes from exogenous genetic materials (Kaneko-Ishino and Ishino, 2012, 2015; Lavalie et al., 2013; Imakawa et al., 2015). In addition to the investigation of duplication of genes (Ohno, 1970; Kimura, 1983), such acquired genes afford good examples for studying macroevolution and diversification as well as serving as lineage-specific markers in phylogenetic analysis. In the human genome, there are approximately 30 LTR retrotransposon-derived genes belonging to two main groups, the sushi-ichi retrotransposon homologs (*SIRH*, also called *MART*, or *SUSHI*) and the paraneoplastic Ma antigen (*PNMA*) family (Voltz et al., 1999; Rosenfeld et al., 2001; Schüller et al., 2005; Campillos et al., 2006; Kaneko-Ishino and Ishino, 2012, 2015; Iwasaki et al., 2013). These genes are derivatives from the original LTR retrotransposons, but each member has a unique DNA sequence. Therefore, each seems to be domesticated in such a manner to have its unique function. Among the *SIRH* genes, *PEG10/SIRH1* (*Paternally expressed 10*) is a therian-specific gene, which is conserved in eutherians and marsupials and plays an essential role in early placenta formation (Ono et al., 2001, 2006; Suzuki et al., 2007). Among all the other eutherian-specific *SIRH* genes, *PEG11/RTL1/SIRH2* (*Paternally expressed 11/Retrotransposon-like 1*) and *SIRH7/LDOC1* (*Leucine zipper, downregulated in cancer 1*) also have been shown to have essential placental functions (Charlier et al., 2001; Edwards et al., 2008; Kagami et al., 2008; Sekita et al., 2008; Naruse et al., 2014), such as maintenance of fetal capillaries and the differentiation/maturation of a variety of placental cells, respectively. All of this evidence provides strong support for the contribution of *SIRH* genes to the evolution of viviparity in mammals via their eutherian-specific functions (Kaneko-Ishino and Ishino, 2012, 2015).

SIRH11/ZCCHC16 (*Zinc-finger CCHC domain-containing 16*) is an X-linked gene that encodes a CCHC type of zinc-finger protein that exhibits high sequence identity to the LTR retrotransposon Gag protein (Irie et al., 2015). It is expressed in the brain, kidney, testis and ovary in adult mice, and its deletion causes abnormal mouse behaviors related to cognition, including attention, impulsivity and working memory, possibly via the locus coeruleus–noradrenergic (LC-NA) system (Irie et al., 2015). It is proposed that phasic activation of NA neurons in the LC is linked to cognitive shifts that facilitate dynamic reorganization of target neural networks, thus permitting rapid behavioral adaptation in response to changing environmental imperatives (Berridge and Waterhouse, 2003; Bouret and Sara, 2005). Therefore, we suggest that the acquisition of *SIRH11/ZCCHC16* has played a role in eutherian brain evolution (Irie et al., 2015; Kaneko-Ishino and Ishino, 2015).

The possibility that *Sirh11/Zcchc16* functions as a non-coding RNA has not been completely excluded. The dN/dS

ratio is a good indicator of selective pressure acting on a protein-coding gene, calculated as the ratio of the number of nonsynonymous substitutions per nonsynonymous site, in a given period of time, to the number of synonymous substitutions per synonymous site, in the same period. The values <1 mean that the gene questioned is subjected to purifying selection because the former changes tend to change the protein function while latter changes have no impacts on it. In the case of *SIRH11/ZCCHC16*, dN/dS ratios in pairwise comparisons of the orthologs between the mouse and seven representative eutherian species other than xenarthran species is approximately 0.35–0.45 (<1), which suggests that *SIRH11/ZCCHC16* has undergone purifying selection after its domestication (exaptation) in the common eutherian ancestor (Irie et al., 2015).

The evolution of mammalian species is associated with several critical geological events and their associated environmental and geographical impacts. The split of the therians from the monotremes occurred 166–186 Mya, followed by the eutherian/marsupial split 160 Mya (Luo et al., 2003; Asher et al., 2004; Madsen, 2009). The domestication of *PEG10* occurred between these two periods (Suzuki et al., 2007) and all the other *SIRH* genes, such as *PEG11/RTL1*, *SIRH7/LDOC1*, and *SIRH11/ZCCHC16*, were domesticated after the eutherian/marsupial split and before the split of the three major eutherian lineages, boreoeutheria (including euarchontoglires and laurasiatheria), afrotheria and xenarthra 120 Mya that was associated with the division of the supercontinent Pangea (Edwards et al., 2008; Nishihara et al., 2009; Naruse et al., 2014; Irie et al., 2015).

After the extinction of the dinosaurs at the Kreide (Cretaceous)-Paleogene (K-Pg) boundary 65 Mya, an adaptive radiation of mammals independently took place in Eurasia, North and South America, Africa, Australia and Antarctica as well as several isolated islands (Murphy and Eduardo, 2009). The long-term isolation of Australia and South America from other continents as well as the reunion of the two continents, such as Eurasia and Africa, and South and North America, affected the subsequent evolutionary route and history of the eutherians as well as other organisms to a great extent. For example, Xenarthrans evolved and diverged on the isolated South American continent, where carnivorous marsupials and birds had long predominated (Patterson and Pascual, 1972; Murphy and Eduardo, 2009). After the Isthmus of Panama emerged ~3 Mya, the carnivorous marsupials were replaced by an invading carnivorous laurasiatherian species from North America (Patterson and Pascual, 1972). As mentioned, the domestication of all the *SIRH* genes was completed by the time of the emergence of the common eutherian ancestor, after which extensive eutherian diversification occurred in a lineage- and species-specific manner (Kaneko-Ishino and Ishino, 2012, 2015). Therefore, it is of great interest to examine the lineage-specific variations of *SIRH11/ZCCHC16* model gene to explore the extent of its involvement in the eutherian diversification process.

RESULTS

Conservation of SIRH11/ZCCHC16 in Eutherian Species

SIRH11/ZCCHC16 encodes a protein composed of approximately 300–310 amino acids (aa), with a CCHC RNA-binding domain in its C-terminus (Irie et al., 2015). Based on whole genome sequence data of 85 eutherian species from GenBank, including two xenarthran species, the *SIRH11/ZCCHC16* ORF in each species was deduced from its own DNA sequence that displayed homology with human *SIRH11/ZCCHC16*. The predicted ORFs in the 83 species, excluding the two xenarthran species with the *pseudoSIRH11/ZCCHC16*, are illustrated in **Figure 1** (see also **Figure S1**: *SIRH11/ZCCHC16* aa sequence). Although, there might be some sequence errors in the genomic information, we used it to perform an initial investigation. Ideally, DNA sequences from multiple individuals should be analyzed in every species in this type of investigation. Instead, in this work, we focused on lineage-specific cases and confirmed the mutations using genomic DNA from the same as well as additional species in some of the primate lineages. As a result, a total of 95 eutherian species were analyzed in this study. Afrotheria is the most closely related eutherian group to the xenarthrans and their *SIRH11/ZCCHC16* ORFs are highly conserved, as previously reported (**Figure 1**, the lowest of the six rows).

Mutations Leading to the Loss of a CCHC RNA-Binding Domain in the Boreotherians

Nonsense mutations leading to loss of a CCHC RNA-binding domain were observed in five boreotherian [two euarchontoglires and three laurasiatherian species, including the white-cheeked gibbon (Primates), Chinese tree shrew (Scandentia), Amur tiger (Carnivora), and two flying fox species (Chiroptera)] (**Figure 1**, **Figure S1**). It is possible that these mutations may be due to sequence errors, however, it is worth considering other possibilities, including the cases in which the mutations are lineage-specific, such as the gibbons and two megabat species (Megachiroptera). Our analyses suggest that this type of mutation changes *SIRH11/ZCCHC16* function at least in these lineages. Interestingly, in one gibbon species, the truncated *SIRH11/ZCCHC16* is suggested to have become pseudogenized and it was lost from two other gibbon species by profound structural changes.

In the case of the white-cheeked gibbon (*Nomascus leucogenys*: *Nle*), there is a four-base pair deletion leading to a frameshift and the subsequent emergence of a nonsense codon just after it (**Figure 2A**). We confirmed all of these sequence changes in another white-cheeked gibbon at a Japanese zoo. In the dN/dS ratio in the pairwise comparison among the hominoids, *SIRH11/ZCCHC16* was found to be highly constrained among humans, chimpanzees and gorillas (0.20–0.42), while the values of the gibbon (*Nle*) compared to humans and chimpanzees are higher (0.61–0.74), and those compared to the gorilla and orangutan are close to 1 (0.84–0.90). Although in this kind of approximate method, it is not possible to compare pairwise dN/dS values in a rigorous way, these results suggest that the

truncated *Nle* ORF has been subjected to a lesser degree of purifying selection compared to other hominoid members having the full-length *SIRH11/ZCCHC16* (**Table 1A**). It is also probable that the *Nle SIRH11/ZCCHC16* has lost some function by losing its CCHC RNA-binding domain (Rajavashisth et al., 1989; Curtis et al., 1997; Chen et al., 2003; Schlatter and Fussenegger, 2003; Narayanan et al., 2006; Matsui et al., 2007).

As the gibbons are a close relative of the hominids, including humans, we further analyzed two other species in the Hylobatidae family, the white-handed gibbon, *Hylobates lar*: *Hla*, and the siamang, *Symphalangus syndactylus*: *Ssy*, and found that normal *SIRH11/ZCCHC16* was absent from these two species. We set up PCR conditions using primers designed from the gibbon (*Nle*) DNA sequence. These primers worked well even in gorilla, macaque and marmoset samples but none of the expected bands were obtained from *Hla* or *Ssy* (**Figure 2B**, top and middle columns). The results were almost the same with primers designed from the conserved sequences between gibbon and human (F3, F4, R3, R4; **Figure 2B**, top and bottom columns), and we confirmed that the quality of their genomic DNA was good enough for PCR analysis by amplifying tyrosinase (*TYR*) and lactase (*LCT*) genes as controls (**Figure S2**). Finally, we performed Southern blot analysis using a gorilla PCR fragment (F1-R1: 544 bp) as a probe and confirmed that there was no corresponding band to *SIRH11/ZCCHC16* that appeared in the white-handed gibbon and siamang (**Figure 2C**), suggesting that a large deletion or profound structural change had occurred in these two gibbon species. All these results demonstrate that gibbons in at least three out of four genera do not possess the normal full-length *SIRH11/ZCCHC16* ORF as a result of deletion/structural changes or the lack of a CCHC RNA-binding domain, supporting the notion that the gibbon *SIRH11/ZCCHC16* gene is not functional and instead has become a pseudogene.

In the case of megabats, the nonsense mutation of *SIRH11/ZCCHC16* is conserved between these two species, while it is intact in the four closest microbat species (Microchiroptera) (**Figures 2D,E**), indicating that this mutation occurred relatively recently. Except for the loss of the C-terminal CCHC domain, the megabat *SIRH11/ZCCHC16* ORF looks well conserved without further nonsense or frameshift mutations. In this case, it is difficult to estimate whether the truncated ORFs are functional or not. Although the dN/dS value between two megabats exhibits 1.5, it is difficult to distinguish whether they have been subjected to positive or neutral evolution (**Table 1B**). However, it may simply reflect the fact that these two species are too close to examine in terms of the evolutionary relationship in this way, because both the dN and dS values are approximately 10 times lower than in the four microbat species. Among the Chiroptera, the values between the megabats to the microbats are slightly higher (0.63–0.72) than those of the microbats (0.53–0.63; **Table 1B**). Although, it is possible that they might possess certain functions, the function of the truncated megabat *SIRH11/ZCCHC16* have already been changed or lost, or is on the way to either of these fates, while the microbat *SIRH11/ZCCHC16* has undergone purifying selection.

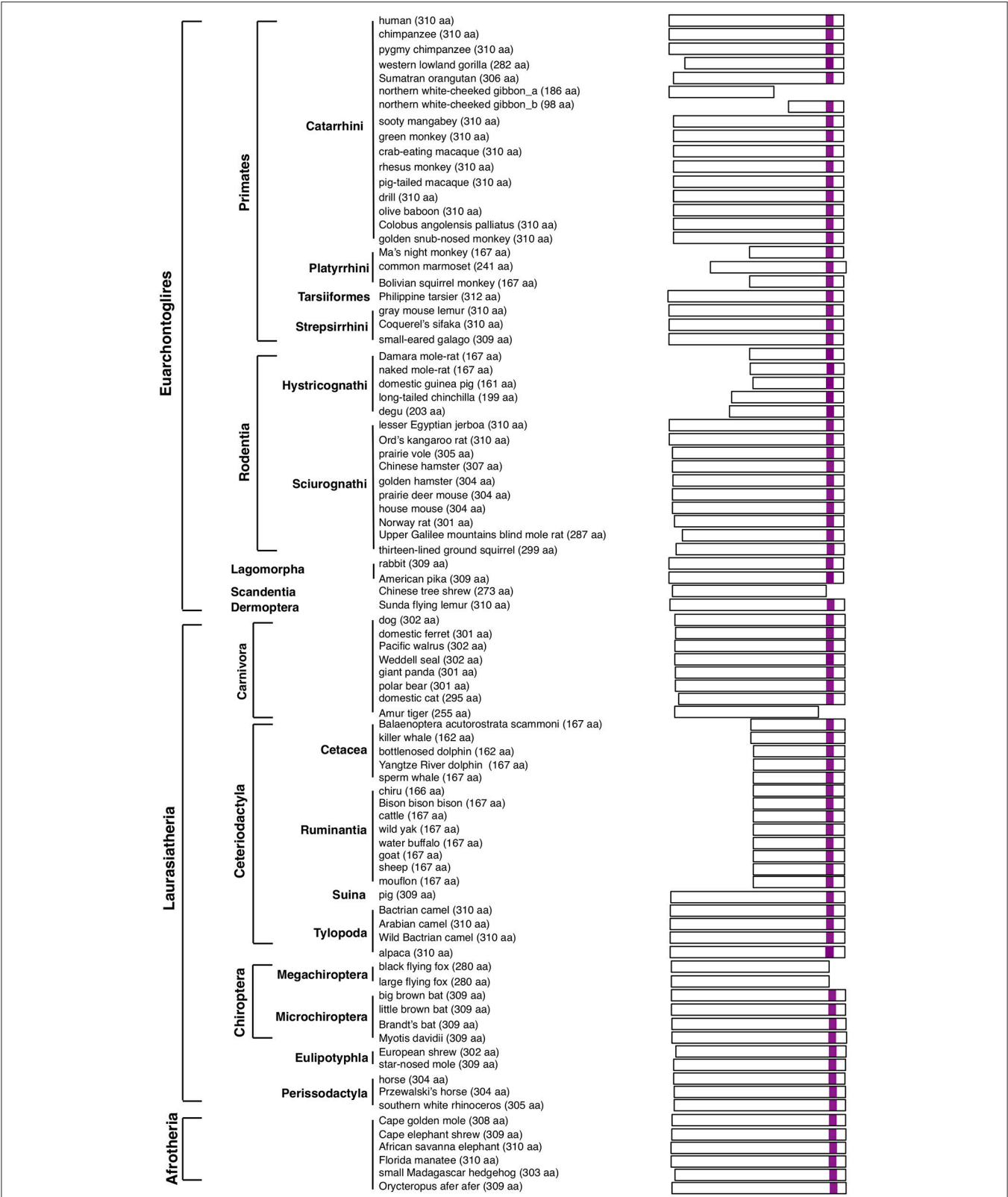


FIGURE 1 | The scheme represents expected SIRH11 ORF in 83 eutherian mammals. The purple boxes indicate CCHC RNA-binding domain.

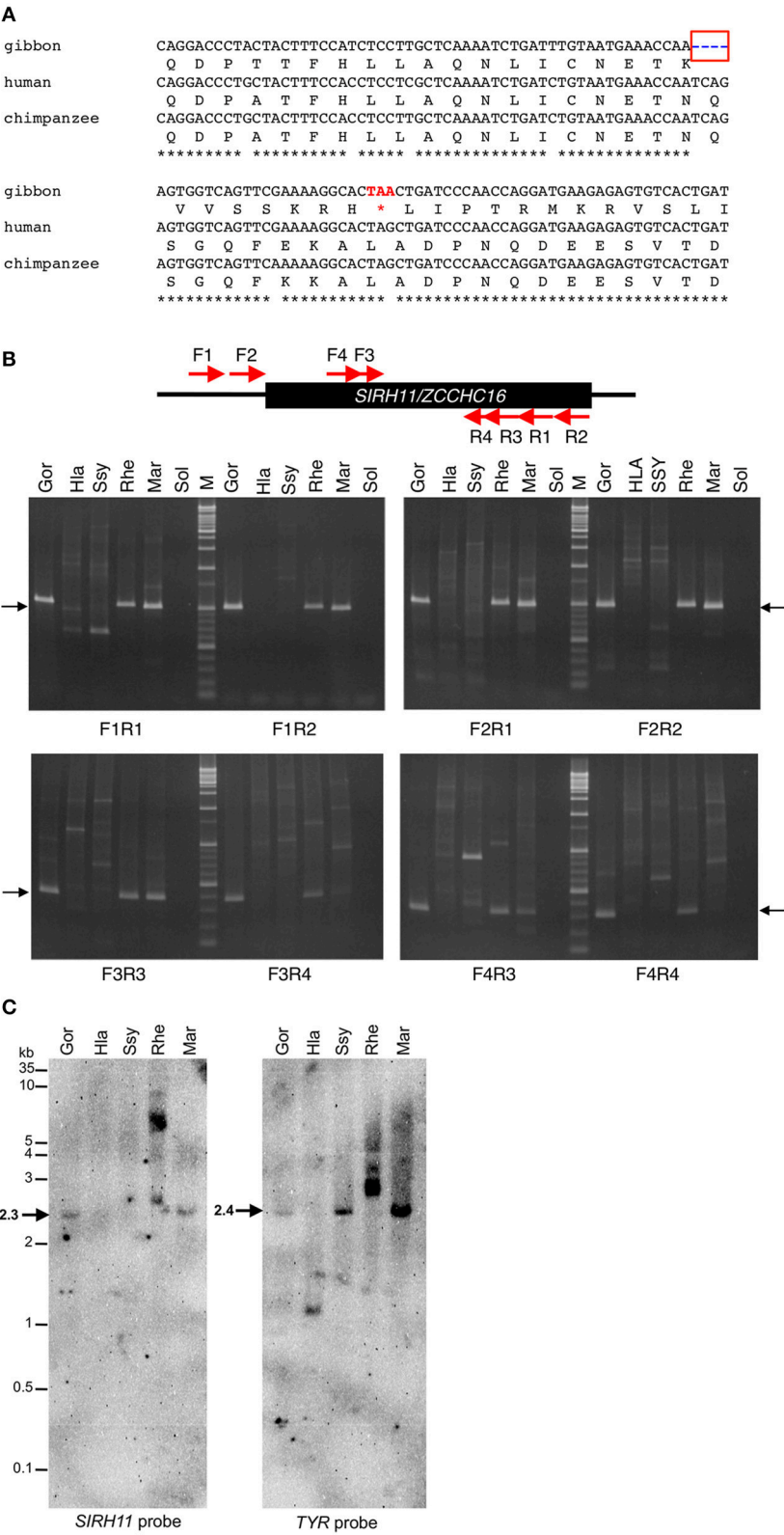


FIGURE 2 | Continued

D

black_flying_fox	MEKYTESSPTMQVEHSFFQVENPILEPQVQHLTEDNATLREQVMSALATPMMPISCSLEH
large_flying_fox	MEKYTESSPTVQVEHSFFQVENPILEPQVQHLTEDNATLREQVMSALATPMMPIFCSLEH
big_brown_bat	MEKHTESSPRLQVEHSSSLWVENLILQPQVQHLTEENTTLRDQVMPALATPMMPLLCSLKN
little_brown_bat	MEKHTESSPTLQVEHSSSLWAENLILQPQVQHLTEENTTLRDQVMPALATPMMPLLCSLKN
Brandts_bat	MEKHTESSPTLQVEHSSSLWAENLILQPQVQHLTEENTTLRDQVMPALATPMMPLLCSLKN
Myotis_davidii	MEKHTESSPTSQVEHSSSLWAENLILQPQVQHLTEENTTLKDKVMPALATPMMPLLYSLKN
	:** : .** *:*****:***:***:*** *****: **:. .
black_flying_fox	LSQFYDDTANISGFLAQVTTYLTILKIPNPADDAQVKFFFDYLSQQMESCGIISTPDQNA
large_flying_fox	LSQFYDDTANVSGFLAQVTTYLTILKIPNPADDAQVKFFFDYLSQQMESCGIISTPDQNA
big_brown_bat	LSQFRGDPANISGFLAQVTNYLTTLKNTNPIHDDQVKFFFDHLSQQMESCGMISRSEQSA
little_brown_bat	LSQFHDDPANLSGFLAQVTNYLTTLKITNPIHDAQVKLFFDYLSQQMESCGIISRSEQSA
Brandts_bat	LSQFRDDPANLSGFLAQVTNYLTTLKITNPIHDAQVKLFFDYLSQQMESCGIISRSEQSA
Myotis_davidii	LSQFPDDPANLSGFLAQVTNYLTTLKITNPIHDAQVKLFFDYLSQQMESCGMISRSEQSA
	**** * **:*****.*** ** ** . * **:***:*****:*** :*. .
black_flying_fox	LLKQYENFVLEFQQSFDESCKQEMSSLVNDKIDKGDKFSKQDATTFQLLGQNLNCNKAYQ
large_flying_fox	LLKQYENFVLEFQQSFDESCKQEMSSLVNDKIDKGDKFSKQDATTFQLLGQNLNCNKAYQ
big_brown_bat	LLKQYENFVLEFQQSFDESCKQEMNPLVNAKVDEEDNSSQDATTFQPLAQNLSCNETNE
little_brown_bat	LLKQYENFVLEFQSFDESCKQEMNPLVNAKVDEEDNSSQDATTFQPLAQNLSCNETSE
Brandts_bat	LLKQYENFVLEFQSFDESCKQEMNPLVNAKVDEEDNSSQDATTFQPLAQNLSCNETSE
Myotis_davidii	LLKQYENFVLEFQSFDESCKQAMPLMNAKVDEEDNSSQDATTFQPLAQNLSCNETNE
	*****:*** ***: * . *: * *: * *: * *:***** *.***.***: :
black_flying_fox	TDLFQKVLADSFQDEVSGTDMMDNLPDLLSVFNRTNVTTGQSCYSQR---PSSQCCLL
large_flying_fox	TDHFQKVLADSFQDEVSGTDMMDNLPDLLSVFNRTNVTTGQSCYSQR---PSSQCCLL
big_brown_bat	SDHFQEGLPDPIQDEVSGTDMMDNLPDLITQCIQLDRKHSKPELLQSEATLPVLSLIVH
little_brown_bat	SDHFQEGLPDPIQDEVSGTDMMDNLPDLITQCIQLDRKHSKPELLQSEAQLPVLSSLVH
Brandts_bat	SDHFQEGLPDPIQDEVSGTDMMDNLPDLITQCIQLDRKHSKPELLQSEAQLPVLSSLVH
Myotis_davidii	SDHFQEGLPDPIQDEMSGTDMMDNLPDLITQCIQLDRKHSKPELLQSEAQFPVLSSLVH
	:* **: * : :***:*****: : : . : : * * .. :
black_flying_fox	IPTIKSPSQVHPWKRLYSCKEASHLSLRPN--EPTRNTVCLYCS*AGHFTRDSFAKRS
large_flying_fox	IPTIKSPSQVHPWKRLYSCKEASHLSLRPN--EPTRNTVCLYCS*AGHFTRDCFAKRS
big_brown_bat	YQTLSSPSC---PSSKELIQLRGSQPSLTPAKRARQOETQLCLYCSQADHFTRDCLAKRS
little_brown_bat	YQTLSSPSC---PSSKELIQLRGSQPPLTPAKRARQOETHLCLYCSQADHFTRDCLAKRS
Brandts_bat	YQTLSSPSC---PSSKELIQLRGSQPPLTPAKRARQOETHLCLYCSQADHFTRDCLAKRS
Myotis_davidii	YQTLSSPSC---PSSKELIQLRGSQPPLTPAKRARQOETHLCLYCSQADHFTRDCLAKRS
	*:.*** * .: ..*: * * :.. :***** * *****:****
black_flying_fox	RAPARISNPAHK*
large_flying_fox	RAPARISNPAHK*
big_brown_bat	RAPARINNPAHQ*
little_brown_bat	RAPARINNPAHQ*
Brandts_bat	RAPARINNPAHQ*
Myotis_davidii	RAPARINNPAHQ*
	*****.****:*

E

black_flying_fox	TGTCTCTACTGTAGCTAAGCTGGTCACTTCACAAGA
large_flying_fox	TGTCTCTACTGTAGCTAAGCTGGTCACTTCACAAGA
big_brown_bat	TGCCTCTACTGTAGCCAAGCTGATCACTTCACAAGA
little_brown_bat	TGCCTCTACTGCAGCCAAGCTGATCACTTCACAAGA
Brandts_bat	TGCCTCTACTGCAGCCAAGCTGATCACTTCACAAGA
Myotis_davidii	TGCCTCTACTGCAGCCAAGCTGATCACTTCACAAGA

FIGURE 2 | (A) Nonsense mutation in gibbon *SIRH11/ZCCHC16*. The four bp deletion (blue in a red box) in gibbon leads to a nonsense mutation (red). Note that there is a G to A transition (DNA polymorphism) in a stop codon of gibbon (TAA) compared with human/chimpanzee and other primates (TAG). **(B)** PCR analysis of gibbon *SIRH11/ZCCHC16*. Upper panel shows the schematic representation of primer design. Lower panel shows agarose gel electrophoresis profile in each primer (Continued)

FIGURE 2 | Continued

set. The arrows represent expected band size. M, 100 bp and 1 kb ladder; Gor, gorilla; Hla, white-handed gibbon; Ssy, siamang; Rhe, rhesus macaque; Mar, common marmoset; Sol, solvent only (no DNA). **(C)** Southern blot analysis of *Hla* and *Ssy*. Left and right panels show the result of hybridization using *SIRH11* and *TYR* probes, respectively. The arrows indicate expected band size. Gor, gorilla; Hla, white-handed gibbon; Ssy, siamang; Rhe, rhesus macaque; Mar, common marmoset. **(D)** Amino acid sequence alignment of Chiroptera SIRH11/ZCCHC16. The blue asterisks in a red box indicate a common nonsense mutation in megachiroptera. The asterisks, colons and periods below the amino acids indicate identical, strongly and weakly similar residues among six species, respectively. **(E)** DNA alignments around the common TAA nonsense mutation (red).

Mutations Resulting in the Loss of the N-Terminal Half of the ORF in the Boreotherians

In three lineages in boreotheria, a deletion of the N-terminal half of SIRH11/ZCCHC16 ORF was observed [i.e., the New World monkeys (Primates: three species/three examined (3/3)), the New World and African hystricognaths (Rodentia: 3/3 and 2/2, respectively) as well as species belonging to Cetacea and Ruminantia (Cetartiodactyla: 5/5 and 7/7, respectively)] (Figure 3). In all of these species, short putative ORFs, mainly comprising 167 aa, are conserved although the causative mutations are independent of each other, reflecting their own lineage-specific events. In these cases, the pairwise dN/dS analyses suggest that there has been selective constraint in some species and perhaps a more relaxed or neutral selection in others.

Among three New World monkeys, Ma's night monkey, the Bolivian squirrel monkey and common marmoset, a common deletion of 11 aa was observed near the N-terminus, with a common nonsense mutation just after it (Figures 3A,C). The putative short ORF that starts from the next Met codon comprises 167 aa in the first two and 241 aa in the latter because of a single additional Met codon that arose in a species-specific manner (Figure S1). We further analyzed five more species, the long-haired spider monkey (*Ateles belzebuth*: Abe), common squirrel monkey (*Saimiri sciureus*: Ssc), tufted capuchin (*Cebus apella*: Cap), the Azara's owl monkey (*Aotus azarae*: Aaz) and the cotton-top tamarin (*Saguinus oedipus*: Soe) and confirmed the common deletion of 11 aa and the subsequent nonsense mutation are a lineage-specific feature (Figures 3B,C). Thus, it seems probable that these mutations emerged in the common Platyrrhini ancestor from which all the New World monkeys in South America diverged (Poux et al., 2006).

A similar situation was assumed in the case of the two closest rodent groups, the South American and African hystricognaths (Caviomorpha and Bathyergidae; Poux et al., 2006). The lost N-terminal parts contain several nonsense mutations and frameshifts, and only one nonsense mutation is conserved in all of the species (Figure 3D), indicating that this mutation emerged in a common ancestor in Africa (Poux et al., 2006). Compared with the heavily mutated N-terminal region, the C-terminal region, comprising 161–203 aa, is highly conserved.

Compared with the above two cases, DNA sequences corresponding to the N-terminal half are completely missing in the species of Cetacea and Ruminantia, indicating a large deletion occurred in the common ancestor of these two suborders, although the former has an additional frameshift event that took place in a Cetacea-specific manner (Figure 3E, Figure S3).

The remaining ORFs, each comprising 167 aa, are also highly conserved.

The pairwise dN/dS analysis within each lineage suggests that some portions of the tree display gene-wide dN/dS values that are consistent with purifying selection. For example, among five New World monkey species the truncated SIRH11/ZCCHC16 ORFs are highly constrained (dN/dS = 0.09) between the two night monkeys (the Azara's owl monkey and Ma's night monkey; Table 1C). However, those of the tufted capuchin are variable: greater than 1 (1.9 and 1.4) to the night monkeys, close to or less than 1 (0.99 and 0.80) to the common marmoset and Bolivian squirrel monkey. Those of the Bolivian squirrel monkey are consistently close to 1 (0.80–1.1), suggesting neutral evolution. Thus, it is possible that the functions of the truncated ORFs in Platyrrhini were diversified to a great extent, presumably because of species-specific adaptation after the structural change of N-terminus deletion in the common Platyrrhini ancestor.

In the case of rodents, the truncated SIRH11/ZCCHC16 ORFs in Hystricognathi are highly constrained among Rodentia except for the Damara mole rat and naked mole rat, and the Damara mole rat and long-tailed chinchilla (0.81 and 0.86, respectively; Table 1D). The cases of the Ruminantia (Table 1E) and Cetacea (Table 1F) are similar to the Platyrrhini in that the dN/dS values exhibit a large variety, some are highly constrained (between bison and cattle, as well as sheep and goats) while the others seem to be subjected to neutral evolution. All these results suggest that the function of the truncated SIRH11/ZCCHC16 ORFs diverged in a species-specific manner, possibly reflecting the functional constraints imposed by the environment after the lineage-specific structural changes.

Some species have ORFs with a small deletion of the N-terminal region, such as the western lowland gorilla, with a 28 aa deletion (Primates), and the upper Galilee mountains blind mole rat, with a 22 aa deletion (Rodentia). In the former, using three different individuals we confirmed the nonsense mutation close to the translational start site caused by a frameshift (Figure 3F).

Loss of SIRH11/ZCCHC16 in Xenathrans

We previously reported that SIRH11/ZCCHC16 in two armadillo species (*Dasypus novemcinctus* and *Tolypeutes matacus*: Cingulata) and two sloth species (*Choloepus hoffmanni* and *Choloepus didactylus*: Pilosa) were pseudogenized due to severe mutations, including multiple nonsense mutations and frameshifts (Irie et al., 2015). There are no common mutations between two armadillo species belonging to two different genera, although the genomic DNA information corresponding to the C-terminal part was lacking in GenBank for one armadillo

TABLE 1 | Pairwise dN/dS analyses on several lineages with truncated SIRH11/ZCCHC16.

dN/dS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A															
1. Human															
2. Chimpanzee	0.309														
3. pygmy_chimpanzee	0.196	0.303													
4. Western_lowland_gorilla	0.202	0.424	0.3												
5. Sumatran_orangutan	0.453	0.713	0.512	0.705											
6. Northern_white-cheeked_gibbon_a	0.703	0.741	0.613	0.902	0.836										
B															
1. Black_flying_fox															
2. Large_flying_fox	1.545														
3. Big_brown_bat	0.715	0.686													
4. Little_brown_bat	0.66	0.625	0.615												
5. Brandts_bat	0.682	0.646	0.529	0.625											
6. Myotis_davidii	0.718	0.679	0.609	0.597	0.53										
C															
1. Azaras_owl_monkey															
2. Mas_night_monkey	0.090														
3. Common_marmoset	0.749	0.657													
4. Tufted_capuchin	1.914	1.395	0.988												
5. Bolivian_squirrel_monkey	0.985	1.058	0.795	1.035											
D															
1. Damara_mole-rat															
2. Naked_mole-rat	0.813														
3. Domestic_guinea_pig	0.687	0.505													
4. Long-tailed_chinchilla	0.861	0.711	0.599												
5. Degu	0.634	0.575	0.408	0.639											
6. Lesser_Egyptian_jerboa	0.435	0.378	0.353	0.407	0.514										
7. Ords_kangaroo_rat	0.557	0.594	0.43	0.556	0.527	0.454									
8. Prairie_vole	0.485	0.492	0.549	0.63	0.53	0.439	0.582								
9. Chinese_hamster	0.514	0.556	0.509	0.54	0.602	0.385	0.462	0.546							
10. Golden_hamster	0.579	0.533	0.491	0.583	0.558	0.467	0.524	0.587	0.52						
11. Prairie_deer_mouse	0.549	0.527	0.505	0.616	0.525	0.454	0.55	0.708	0.556	0.635					
12. House_mouse	0.516	0.437	0.432	0.484	0.462	0.359	0.48	0.435	0.375	0.332	0.443				
13. Norway_rat	0.453	0.442	0.362	0.414	0.451	0.372	0.557	0.472	0.488	0.34	0.519	0.442			
14. Upper_Galilee_mountains_blind_mole_rat	0.421	0.43	0.393	0.531	0.585	0.437	0.449	0.643	0.603	0.658	0.642	0.496	0.613		
15. Thirteen-lined_ground_squirrel	0.525	0.517	0.605	0.535	0.442	0.316	0.437	0.424	0.383	0.473	0.428	0.369	0.339	0.361	
E															
1. Chiru															
2. Bison_bison_bison	0.881														
3. Cattle	1.254	0.19													
4. Wild_yak	1.313	0.323	nd												
5. Water_buffalo	1.639	0.601	0.9	0.976											
6. Goat	1.13	0.655	0.864	0.913	1.077										
7. Sheep	0.627	0.613	0.751	0.787	0.652	0.259									
8. Mouflon	0.676	0.644	0.789	0.825	0.688	0.331	nd								
F															
1. Balaenoptera_acutorostrata_scammoni															
2. killer_whale	1.415														
3. bottlenosed_dolphin	0.949	0													
4. Yangtze_River_dolphin	0.844	0.4	0.331												
5. Sperm_whale	0.574	0.722	0.575	0.466											

(A) Pairwise dN/dS analysis on Homonidae, (B) Pairwise dN/dS analysis on Chiroptera, (C) Pairwise dN/dS analysis on Platyrrhini, (D) Pairwise dN/dS analysis on Rodentia, Hystricognathi species are shown yellow. (E) Pairwise dN/dS analysis on Ruminantia, (F) Pairwise dN/dS analysis on Cetacea. The dN/dS values more than 0.80 or less than 0.21 are shown in red or blue, respectively. When the dS value is 0, it is impossible to calculate this value, therefore, indicated as nd.

A

Ma's_night_monkey	PENCTILQAENLILQPQMQYPTLDNLAKRRQVMPVLATTVMPASY*LEHLTQFHGDPASH
common marmoset	MEKCTILQAENLILQPQMQHSTLDNLAKRRQVMPVFATTVIPVSY*LEHLTQFHGDPANH
Bolivian_squirrel_monkey	MEKCTILQAENLILQPQMQHPTLDNLAKRRQVMPVLATTVMPASY*LEHLTQFHGDPANQ
	*:*****:*****:***:*.*****:
Ma's_night_monkey	SEFLAQVTTYLTALQISNPANDAQIKLFFYYLFQKLESYGIIISGPKSILLKQYENFILE
common marmoset	SELLPQMTTYLTALQISNPANDAQINLFIDYLFQKLESCGIIISGPKSTLLKQYENFILE
Bolivian_squirrel_monkey	SEFVAQVTTYLTALQICNPANDAQIKLFFDYLLQKLESCGIVSGPKDTLLKQYDSFILE
	::*:***.*****:**:***:***:*****:*****:.****
Ma's_night_monkey	FQQSLGEPTEKQEMNPLMNAKVDKGDNSSQDDPATFQLFAQNLICNETNQNGRFEKALAGP
common marmoset	FQQLLGEPTEKQEMNPLMNAKVDKGDSSQDDPATFQLLAQNMICNETNQNGQFEKALAGP
Bolivian_squirrel_monkey	FQQSLGEPTEKQEMNPLMNDKVDKGDNSSQDDPATFQLLAQNLICNETNQNDQFEKALAGP
	*** ***** *****:*****:***:***** :*****
Ma's_night_monkey	NQDEESVTDIMDNLPLDITQCIQLDKKHSVRPELLOSETQFPLLTSLIQHQALVSPTDPP
common marmoset	NQDEESVTDITDNLPLDITQCIQLDKKHSVRPELLOSETQFPLLPSLIQHQAALVSPTDPP
Bolivian_squirrel_monkey	NQDEDSVRGIMDSLPLDITQCIQLDKKHSVKPELLOSDTQFPMLTSLIQHQALVSPTDPP
	****:* * *.*****:*****:***:* *****
Ma's_night_monkey	PKKGPMQLREGQLPLTPAKRAYQOETQLCLYCNQSGHFTRDCLAKRSRALAMTNNTAHQ*
common marmoset	PKRGPMQLREGQLPLTPAKRAYQOETQLCLYCNQSGHFTRDCLAKRSRALAMTNNSSTV
Bolivian_squirrel_monkey	PKKGPMQLREGQLPLTPAKRAYQOETQLCLYCNQSGHFTRDCLAKRSRALAVTNNTAHQ*
	:****:*****:***:***:

B

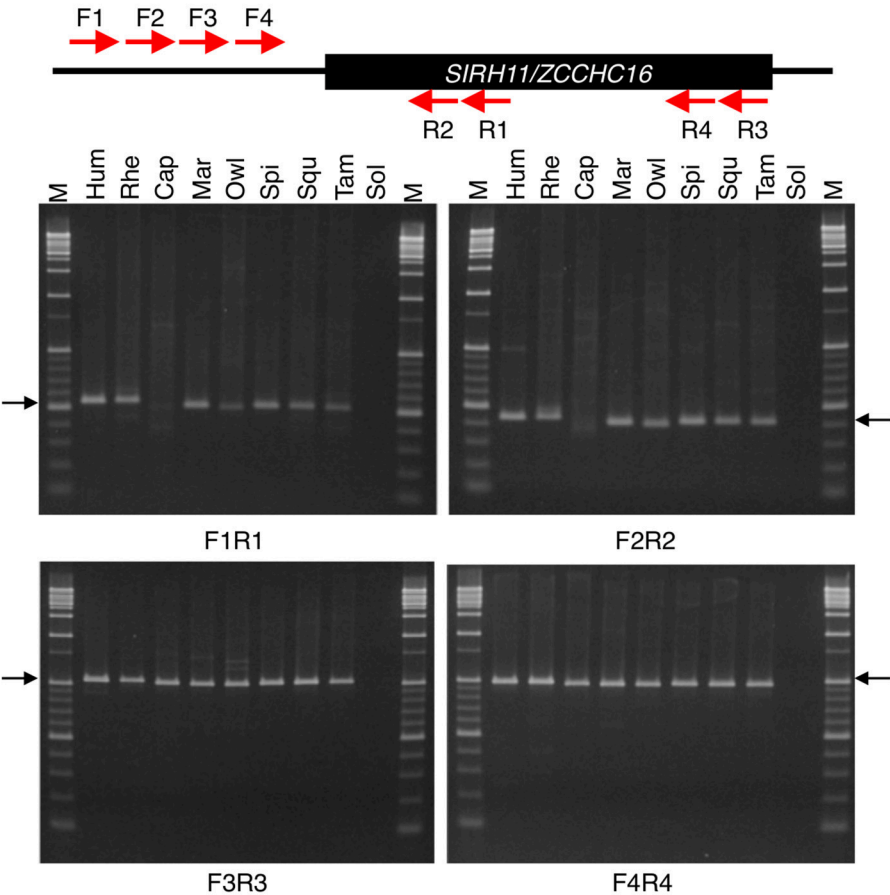


FIGURE 3 | Continued

C

human
chimpanzee
Ma's night monkey
Azara's owl monkey
common marmoset
Bolivian squirrel monkey
tufted capuchin
long-haired spider monkey

TTGGCTACTTGCATGATACACTGCCTCTCTGATCACAGAGCAATCCACTG-TCCTCAT
TTGGCTACTTGCATGATACACTGCCTCTCTGATCACAGAGCAATCCACTG-TCCTCAT
TTGGCTACTTGCATGATACACTGCCTCTCTGATCACAGAGCAATCCACTCTCGTTTCAT
TTGGCTACTTCTCATGATACCTGCTCTCTGATCACAGAGCAATCCACTGCTGTTTCAT
TTGGCTACTTCTCATGATGCTGCTCTCTGATCACAGAGCAAAACCACTGCTGTTTCAT
TTGGCTACTTCTCATGATACCTGCTCTCTGATCACAGAGCAATCCACTGCTGTTTCAT
TTGGCTACTTCTCATGATACCTGCTCTCTGATCACAGAGCAATCCACTGCTGTTTCAT
TTGGCTACTTCTCATGATACCTGCTCTCTGATCACAGAGCAATCCACTGCTGTTTCAT
TTGGCTACTTCTCATGATACCTGCTCTCTGATCACAGAGCAATCCACTGCTGTTTCAT

insertion (2 bp)

human
chimpanzee
Ma's night monkey
Azara's owl monkey
common marmoset
Bolivian squirrel monkey
tufted capuchin
long-haired spider monkey

TGGTTTCCAGTACTAGTACTCAATTCTACTCAATCCCACCTGATTCAGGAGACATAAATG
 TGGTTTCCAGTACTAGTACTCAATTCTACTCAATCCCACCTGATTCAGGAGACATAAATG
 TGGTTTCTAGTACTAGTGCCTCAATTCTACTCAATCCCACCTGATTCAGGAGACATAAATG
 TGGTTTCTAGTACTAGTGCCTCAATTCTACTCAATCCCACCTGATTCAGGAGACATAAATG
 TGGTTTCCAGTACTAGTACTCAATTCTACTCAATCCCACCTGACTCCAGGAGATATAATG
 TAGTTTCCAGTACTAGTGCCTCAATTCTACTCAATCCCACCTGATTCAGGAGACATAAATG
 TGGTTTCCAGTACTAGTACTCAATTCTACTCAATCCCACCTGATTCAGGAGACATAAATG
 TGGTTTCCAGTACTAGTACTCAATTCTACTCAATCCCACCTGATTCAGGAGACATAAATG
 TGGTTTCCAGTACTAGTACTCAATTCTACTCAATCCCACCTGATTCAGGAGACATAAATG

deletion (12 bp)

human
chimpanzee
Ma's night monkey
Azara's owl monkey
common marmoset
Bolivian squirrel monkey
tufted capuchin
Long-haired spider monkey

[illegible]

deletion (33 bp)

D

house_mouse
Damara_mole-rat
naled_mole-rat
domestic_guinea_pig
long-tailed_chinchilla
degu

```
MEKCTESLPNLNAETSFLRGGNLIQPQVOHPHTDDKPPITGQVVPAINTPVMSPGSPYSGDH
MKKCTESPTQLLIDPSPFQKENLIPQQRKYLSEGNCSSEK*XIMPALTPPVIISVPSVVKH
MKKTESPTQLPVP-YSPFQKENLIPQQRKYLSEGNCSSEK*XIMPALTPPVIISVPSVVKH
MKKTESQPCP-LR*VPTFLQRENLIQPQTOHLEENYSLRNLMPALITPVMSPVP-SFK*
MKKTESPTQLRQSPSTFLQRESDSAAANAASF*GKICSEK*XIMPALTPPVMISPLPLEH
IVKYTESLQTLXQIFLFFKEIIFWTI-----HKHIFLRNNAMPALIIAKLSGLPLEH
:   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
:   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
```

house_mouse
Damara_mole-rat
naled_mole-rat
domestic_guinea_pig
long-tailed_chinchilla
degu

TPQFHGPNPASVKGFFQVTTLYRLADLSINPADARVKHFFDYLSSQMKQCDVLSESTQNS
 FTQFHGDPDVNHFQFFVQVTTLYFPFAFKINPADARQKFFSDLSQMKQCFVGVSEYDQSI
 FTQFHGDPVNNHLLK*LAT*QLSRFPIL*--LMPRSALFSLISAEEKFWSHISI*PAY
 XLFPQGLDNLHGLFTQMIILYLTALKLSPADI*TKLSEYLFQHMNNRVSYSEYDQSI
 FSQFGDLGVNSGDFVQVLIIVLYTLD*TPSHIASBQTLQNSLYLQ*QMETGFSYSEYNO*
 FTQFLGLDVNH*GFLAQAITLYLTALKISNLIADTVKLFSYDL*QMFENFGINCEYDOCA

house_mouse
Damara_mole-rat
naled_mole-rat
domestic_guinea_pig
long-tailed_chinchilla
degu

```

LLKQYENFVLLELQSQSGFEMPQTEHTTPPMNVTVDKSN-ISPQDATNFQLHAPNLISYETNQ
PLKQENFVFLDFQPSFGAEIKQKMNPLMDKADKVRGDSTSOQDVTTTFOFFSONLSNNKTNH
STEANENFVLDQPSFGCEPIKQKMNPLMDKADKVRGDSTSOQDVTTTFOFFSONLTCNKTNH
LLKQENFVFLVAFQ*SSGEPKIQKMNPLMDKADKVRGDSTSOQDVTTTFOFFSONLSCNVNTNH
---QNECNVLDQPSGSGEPKQKMNPLMDKADKVRGDSTSHOVDTTTFOSSQSLSYETNH
PLKENENFALDFQSSSRPDEQENMLLMNAKVDKVDGFOONITTFQFSSQSLSCNETNP

```

house_mouse
Damara_mole-rat
naled_mole-rat
domestic_guinea_pig
long-tailed_chinchilla
degu

```

RDQFQNGQADPTQNIETITDMDLNPDLITQCIQLDKKHKDRPELLQSESHVPPFASFTNH
GEKFOERLSDPIODEGNTGMMRNPPDLITOCITOLDKKYNDRPELLOSETOPVPLASLIH
KEKFOERLSDPIODEKSVTGMVGNLPDLITOCITOLDKKYNDRPELLOSEVRPPLGSLPI
KEKFHRLDRCDHIODEKSVTGMENIPDLITOCITOLDKKSNKNPELLOSETOPVPLASI--
REKTHGADLPDIODEKSATAMENNPDLITOCITOLDKSNKNPELLOPENOLVPEFLNHL
KEKFOERLSDNITODOKSVTAMMANLPDLITOCITOLDKCNKEPPELLOSEAOELPFLAPFNH
:::  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *

```

house_mouse
Damara_mole-rat
naled_mole-rat
domestic_guinea_pig
long-tailed_chinchilla
degu

```

YQSF IGGV RPL PDKO RQLQGAHLPVTPAKR ARQOETQLCVVCNQA GHFTRDCLAKRSRT
HOYLSSSTGSPKKEEPL IOLRGGOPLTPAKRAHOEOTOLCLYCSOAGHFTRDCLAKRSRA
-----SSSTSPRPKEET IOLRGGOPLTPAKR ARQOETOLCLYCSOAGHFTRDCLAKRSRA
HOSSRSSTGSPKKEEPL IOLRGGOPLTPAKR ARQOESPLCLYCSOAGHFTRDCLAKRSRA
PPSFSSMCPPKKEEPIOLRGG-LPLTPAKR ARQOESOLCLYCSOAGHFTKDCLAKRSRA
*      *      *      *      *      *      *      *      *      *      *

```

```
house_mouse
Damara_mole-rat
naled_mole-rat
domestic_guinea_pig
long-tailed_chinchilla
degu
```

PARKKM*---
PERTNDTAHQ*
PERTDNTAHQ*
PARTNNPAYQ*
PAWTNNPAHQ*
PARTNNPDHQ*
 *

FIGURE 3 | Continued

FIGURE 3 | Continued

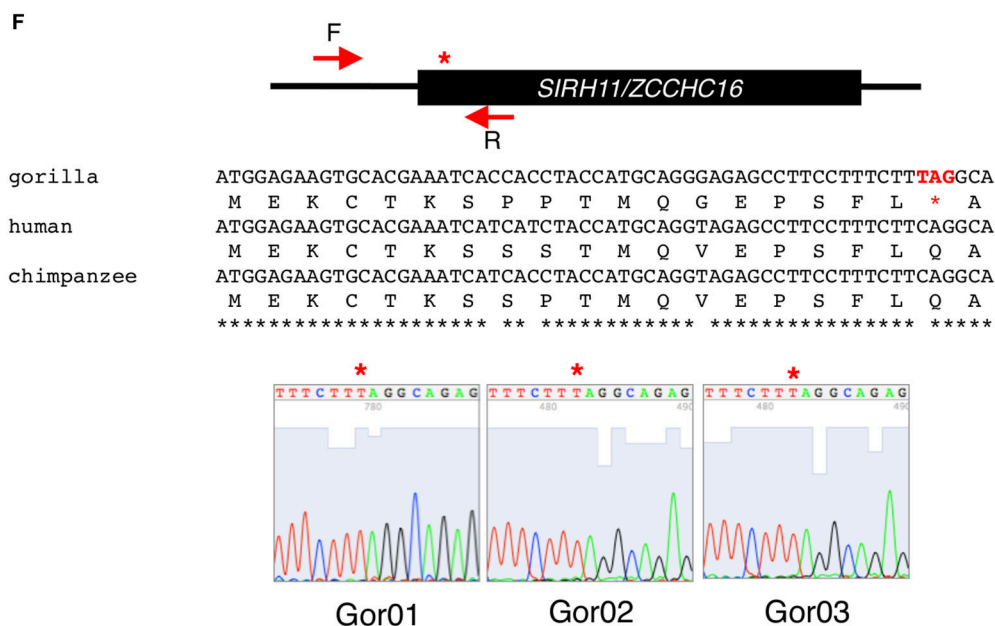


FIGURE 3 | (A) Amino acid sequence alignment of Platyrrhini *SIRH11/ZCCHC16*. The blue asterisks indicate common nonsense mutations in Platyrrhini. The underlined sequences indicate the putative short ORFs starting from a next Met codon. The asterisks, colons and periods below the amino acids indicate identical, strongly and weakly similar residues among three species, respectively. **(B)** PCR analysis of Platyrrhini *SIRH11/ZCCHC16*. Upper panel shows the schematic representation of the primer design. Lower panel shows agarose gel electrophoresis profile in each primer set. The arrows represent expected band size. M, 100 bp and 1 kb ladder; Hum, human; Rhe, rhesus macaque; Cap, tufted capuchin; Mar, common marmoset; Owl, the Azara's owl monkey; Spi, long-haired spider monkey; Squ, common squirrel monkey; Tam, cotton-top tamarin; Sol, solvent only (no DNA). **(C)** DNA sequence analysis of Platyrrhini *SIRH11/ZCCHC16*. DNA sequences of Azara's owl monkey, common marmoset, tufted capuchin, and long-haired spider monkey determined by our own experiments are also shown. Magenta boxes show lineage specific insertion or deletion. The underlined letters indicate the *SIRH11/ZCCHC16* ORF. **(D)** Amino acid sequence alignment of Hystricognathi *SIRH11/ZCCHC16*. The similar sequences among five Hystricognathi species are expressed in green. The red asterisks and Xs indicate the sites of ORF termination and frameshift, respectively. The blue asterisks in a red box indicate a common nonsense mutation in Hystricognathi. The underlined sequences indicate the putative short ORFs starting from a next Met codon. The asterisks, colons and periods below the amino acids indicate identical, strongly and weakly similar residues among six species, respectively. The house mouse sequence is used as a reference. **(E)** Amino acid sequence alignment of Cetartiodactyla *SIRH11/ZCCHC16*. The blue asterisks in red boxes indicate common nonsense mutations in Cetartiodactyla. The red asterisks and Xs indicate the sites of ORF termination and frameshift, respectively. The underlined sequences indicate the putative short ORFs starting from a next Met codon. The similar sequences among 14 Cetartiodactyla species are shown in green. The asterisks, colons and periods below the amino acids indicate identical, strongly and weakly similar residues among 14 species, respectively. The pig sequence is used as a reference. **(F)** Sequence analysis of gorilla *SIRH11/ZCCHC16*. Upper panel shows the schematic representation of the primer design and nonsense mutation site (red asterisk) in gorilla *SIRH11/ZCCHC16*. Middle panel shows the sequence comparison between gorilla, human, and chimpanzee. Lower panel represents the sequence results of three individuals.

species (*Dasypus novemcinctus*). Two sloth species belonging to the same genera exhibit a quite similar pattern of nonsense and frameshift mutations.

Then, we searched for a common nonsense mutation in another armadillo species (*Tolypeutes matacus*) along with the two sloth species again and found a promising candidate nonsense mutation in the C-terminus, 22 aa upstream of the CCHC RNA-binding domain (Figure 4A, blue asterisks in a red box, and Figure 4B). This pattern of a shared nonsense mutation is consistent with the possibility that this mutation inactivated the gene, because the CCHC domain would be critical for normal SIRH11/ZCCHC16 function, as in the case of gibbons mentioned above. In this work, we have surveyed the mutations only in the ORF region, but it should be noted that pseudogenization may have occurred through promoter or other regulatory mutations and also from a missense mutation or insertions/deletions that inactivates the gene.

The anteater is the only remaining animal group in xenarthra, belonging to Pilosa, the same order as the sloth (Delsuc and Douzery, 2009). An attempt was made to analyze *SIRH11/ZCCHC16* in the giant anteater (*Myrmecophaga tridactyla*) by genomic PCR, but the PCR did not work, even using primer sets designed with completely conserved DNA regions between armadillos (Cingulata) and sloths (Pilosa) (Figure 4C). Although, a single band larger than an expected size was seen in the two conditions (Figure 4C, F1R1, and F1R2), its DNA sequence did not have any relationship to *SIRH11/ZCCHC16* (data not shown). The quality of its genomic DNA was good enough for PCR, as shown by amplifying the tyrosinase gene (*TYR*) using primers designed from the armadillo DNA sequence (Figure S4). However, due to the limited amount and relatively low quality we were unable to perform Southern blot analysis to confirm the absence of *SIRH11/ZCCHC16* orthologs in its genome. Therefore, the final conclusion awaits the determination of

the anteater genome sequence in the future, but all of the results obtained thus far suggest that the three extant groups in xenarthra lack any functional *SIRH11/ZCCHC16* and that the pseudogenizing mutation(s) occurred in a common xenarthran ancestor.

DISCUSSION

It is of interest to determine the roles genes acquired from LTR retrotransposons play in organisms in the current form of the developmental system as well as in the course of

A

```

manatee      MEKFTESPPT-LQVEHSSLWADDLILEPQVQHLLTDDSTALRGQVMP TLATPMPVSCSLE
armadillo1   IEKYTESPST-LQLEHFSIWXENLIL*PQVRHQIEENIAPSSQVLPVLATPMPVPFSL
armadillo2   IKKYAKSPRYLCXLEHFSILAENLILQP*AQHLEETVAPSSQVLPALATSMMPVPFLLE
sloth1       MEKYTESPPS-LPVEHF*LOAENLILQPQVQHLEENTSPSG*VLPALVTPMMPIPFSL
sloth2       MEKYTESPPS-LPVEHF*LOAENLILQPQVQHLEENTSPSG*VLAALVTPMMPIPFSL
              :::::***      **:  :  ::*** *  .*  ::.  :  .  *:  .*  ***:  **

manatee      HPSQFHVDPANLSGFLAQATTYLTALKT-PNPADDVQVKFFFDYISQOVERCRILPGSDQ
armadillo1   HLTQFHDDHAHFSVG-IPGTTYLTVLKI-PSP TDDVQVKLFLITYLRS*RVVGTYLDLTX
armadillo2   HHTQFHDDPANFSVG-SQGTTYLTALKI-PNP TDDVQVKFFFDYLSQQMKSCGVLSPNQ
sloth1       HLTQFHVDPANFSVG-SQVTTYLTSLKTALNSVDNSQVKSFSFDYLSQQMKSCGVLSPD*
sloth2       HLTQFHVDPANFSVG-SQVTTYLTSLKTALNSVNNSQVKSFSFDYLSQQMKSCGVLSPD*
              *  :*** *  *.*  *****  *  .  .::  ***  *  :.  .

manatee      -STLLKQYENFVSEFQQSFGPEPTKQEMNPLVNAKLDKGND SFQYATTFQLLAENISCNK
armadillo1   ESILVKQYVNFILDFHQSFGKPPK*EMFHLVSTKIDKGDDSSLQYVTFSS---SLLKI*N
armadillo2   XSTLLKQYENFILEFQESFGKLTQEMSPVVS SKIDGDEFSLQYISTFQFHAQNLLK*
sloth1_1     -STLLKQYENFIPEYQQSFGKPTKQEMNPVNAKVDKGNDSSLPLCYHFQFLAQNLSYNE
sloth2_1     -STLLKQYENFIPEYQQSFGKPTKQEMNPVNAKVDKGNDSSLPLCYHFQFLAQNLSYNE
              *  :***  **:  ::::***:  *  *  :*.:***:::  .  .  .

manatee      TILSDQLQE-GLADPIQDEVSGTDMMENLPDLITQCIQLDRK-RSDRPELLQSEAQLPRL
armadillo1   HYEKSVLRGTGWSNL*Q---NEWYTFGQLPDLIIQCIQLNKK-HSDRPQLLHSEIQLPQL
armadillo2   KISSERYRV--VQFVMK---XRHKFGQPPDLITQCIQLDKK-HSDRPELLNAEAQLPRL
sloth1       TTM*DPLQE-GLAGPVCDEM SGK N-MDNPPDLITRCIKLDKKMXSDGPELLQSEVQLPRS
sloth2       TTM*DPLRE-GLAGPVCDEM SGK N-MDNPPDLITRCIKLDKKMXNDGPELLQSEVQLPRS
              .  :  :  :  ****  :***:::*  .*  *:*.*  ***:

manatee      ASLVHHQYLSSTGPPAKEEPTQLRGGQLPLTPAKRARQOETQFCLYCSQPGHFTKDCLA
armadillo1   TFLIYHQYFSSPTSTPPKE-----
armadillo2   TFLLYHQYFSSPTSTPPKEETILLWGGQLPLTTAK*VCQOETQLYSHCSKADQFTRDCFA
sloth1       TSLIYHQYLSSTGTLPKEKPVLLQEGQLLLTTAKXAHQOET*FCSYCSQADHFTRDCLT
sloth2       TSLIYHQYLSSTGTLPKEKPVLLQEGQVLLTTGKXAHQOET*FCSYCSQADHFTRDCLT
              :  ::::***:  *  *  **

manatee      KRSRAPARSNNPTHQ*
armadillo1   -----
armadillo2   KHS*ASAKPNN*AH*Q
sloth1       KHSRAPAKTNNPACQ-
sloth2       KHSRAPAKTNNPACQ-

```

B

```

manatee      GAAGAGCCTACACAGCTGCGGGGAGGCC
armadillo1   GAA-----
armadillo2   GAAGAAACGATATAGTTGTGGGGAGGCC
sloth1       GAAAAACCTGTATAGTTGCAGGAAGGCC
sloth2       GAAAAACCGGTATAGTTGCAGGAAGGCC

```

FIGURE 4 | Continued

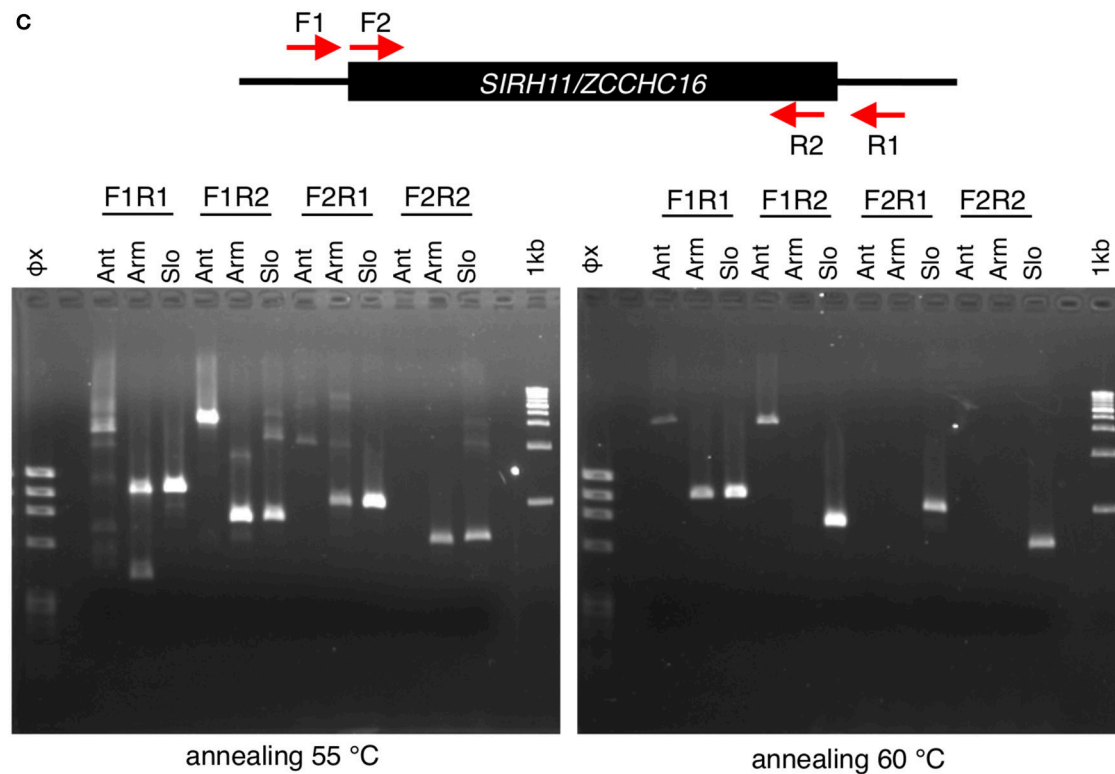


FIGURE 4 | (A) Common nonsense mutation among xenarthra species. Multiple sequence alignment was constructed using five amino acid sequences; armadillo1, *Dasyus novemcinctus*; armadillo2, *Tolypeutes matacus*; sloth1, *Choloepus hoffmanni*; sloth2, *Choloepus didactylus* and manatee, Florida manatee as a reference. The red asterisks and Xs show the sites of ORF termination and frameshift, respectively. The blue asterisks in a red box indicate a common mutation among three species. Purple characters indicate CCHC amino acids in the RNA binding domain. **(B)** DNA alignments around the common TAG nonsense mutation (red) indicated by an orange line in **(A)**. **(C)** PCR analysis of xenarthra *SIRH11/ZCCHC16* Upper panel: schematic representation of primer design to amplify xenarthra *SIRH11/ZCCHC16*. Lower panel: agarose gel electrophoresis profile in each primer set. Ant, giant anteater; Arm, southern three-banded armadillo; Slo, Linnaeus's two-toed sloth; Φx, Φx174 *HaeIII* marker; 1 kb, 1 kb ladder marker.

biological evolution. Among *SIRH* genes, *PEG10/SIRH1*, *PEG11/RTL1/SIRH2*, and *SIRH7/LDOC1* are highly conserved across eutherian species, presumably because they play essential roles in the viviparous reproduction system via placental formation, maintenance, differentiation and maturation, respectively (Ono et al., 2006; Sekita et al., 2008; Naruse et al., 2014). In this study, we found that *SIRH11/ZCCHC16* displays lineage-specific structural variations in eutherians, such as the lack of the CCHC RNA-binding domain or the N-terminal half, as well as species-specific variations in the resulting truncated ORFs. Thus, it is possible that certain *SIRH* genes, such as those concerning cognitive brain function, act as critical determinant factors in the diversification of the eutherians depending on a variety of environmental factors, such as ecological niches and the dynamics of life style as well as the evolutionary history of the species, including geological events.

We showed that all the South American primates and rodents have the truncated *SIRH11/ZCCHC16* ORFs in addition to xenarthran *pseudoSIRH11/ZCCHC16*. Although this might not be of significance, it is of interest to consider the possibility that species with normal *SIRH11/ZCCHC16* function suffered a competitive disadvantage in the South American environment

in the past. South America has a unique evolutionary history in which geographical factors have played a critical role (Houle, 1999; Poux et al., 2006; Delsuc and Douzery, 2009; Murphy and Eduardo, 2009; Nishihara et al., 2009). Diversification of the three major eutherian groups, boreotheria, afrotheria and xenarthra, is supposed to largely be dependent on the division of the supercontinent Pangea, which is thought to have occurred approximately 120 Mya (Nishihara et al., 2009). Xenarthrans evolved and diverged on the isolated South American continent, where carnivorous marsupials and birds had long predominated (Patterson and Pascual, 1972; Murphy and Eduardo, 2009). After the Isthmus of Panama emerged ~3 Mya, the carnivorous marsupials were replaced by an invading carnivorous laurasiatherian species from North America (Patterson and Pascual, 1972). In the competition between the carnivorous marsupials and xenarthrans after the extinction of the dinosaurs ~65 Mya, as well as the marsupials and the carnivorous laurasiatherians ~3 Mya, the presence/absence of *SIRH11/ZCCHC16* among these groups might be a critical factor in the evolutionary outcome. For example, the extinct marsupials have no *SIRH11/ZCCHC16*, the xenarthrans have *pseudoSIRH11/ZCCHC16* and only the

laurasiatherians have a normal *SIRH11/ZCCHC16* in the South American evolutionary history.

Another issue of interest is the *SIRH11/ZCCHC16* mutations in the primates. Phylogenetic relationships, divergence times, and patterns of biogeographic descent among primate species are complex and still controversial. According to a recent molecular phylogenetic analysis using Species Supermatrix, the currently living primates last shared a common ancestor 71–63 Mya and Asia was the ancestral home of the primates. This is also true for the hominoids, suggesting that the ancestor of African apes and humans entered Africa, while the hylobatids remained in Asia (Springer et al., 2012). Among the hominoids, gibbons have lost the normal *SIRH11/ZCCHC16*: white-cheeked gibbon (*Nle*) has a truncated *SIRH11/ZCCHC16* ORF lacking the CCHC RNA-binding domain, while white-handed gibbon (*Hla*) and siamang (*Ssy*) do not have the normal *SIRH11/ZCCHC16* gene in their genomes. It is apparent that *SIRH11/ZCCHC16* function is not conserved in the latter two species, but is this the case for the truncated *SIRH11/ZCCHC16* in the former? It is typically difficult to determine whether truncated ORFs have original, similar or different functions through comparison of the amino acid sequence homology. The dN/dS analysis sometimes helps provide a useful prediction of whether they still possess some function (dN/dS < 1) or have lost their function and already become neutralized (dN/dS = 1), as previously shown in the xenarthran lineage (Irie et al., 2015). The higher dN/dS values of the gibbon (*Nle*) suggested that the truncated *Nle* ORF has been subjected to a lesser degree of purifying selection and that the *Nle* *SIRH11/ZCCHC16* has lost some function by losing its CCHC RNA-binding domain. In retroviruses, the CCHC domain forms a part of the nucleocapsid protein that functions in virus genome packaging and the early infection process (Narayanan et al., 2006). Proteins containing the CCHC zinc-finger domain are commonly known to interact with single-stranded DNAs (ssDNAs) and RNAs (Matsui et al., 2007) and play important roles in *Drosophila* as well as mammalian development via transcriptional and translational regulations (Rajavashisth et al., 1989; Curtis et al., 1997; Chen et al., 2003; Schlatter and Fussenegger, 2003). Therefore, it is probable that the CCHC zinc-finger domain is essential for the normal *SIRH11/ZCCHC16* function that confers selective advantage. In future, it will be of interest to consider the possibility that *SIRH11/ZCCHC16* contributed to brain evolution in hominoidea and also the alternative, that the loss of *SIRH11/ZCCHC16* did confer some selective advantage in the gibbons.

Lineage-specific loss of the N-terminus of *SIRH11/ZCCHC16* ORFs in all the species of New World monkeys and Hystricognathi is consistent with their evolutionary history. Our data indicates a common ancestor of the Platyrrhini in South America already had the mutation(s) leading to the N-terminal deletion. It is proposed that the common ancestor emigrated from Africa and somehow immigrated into South America ~34 Mya (Houle, 1999; Poux et al., 2006), possibly by an incidental current drift from Africa to South America that existed at that time (Houle, 1999). In the case of the two closest South American and African hystricognaths, the Caviomorpha and Bathyergidae, the results also show that the nonsense

mutation leading to the N-terminal deletion first emerged in the common ancestor in Africa. It is proposed that a common Caviomorpha ancestor, from which all the rodent species in South America diverged, emigrated from Africa by an unknown event, just as New World monkeys did (Houle, 1999). The recent discovery of (Late) Eocene primates in Santa Rosa, Peru, extends the fossil record of primates in South America back approximately 10 million years, leading to consideration of possible similarities of an intercontinental dispersal mechanism for the two mammalian groups that occurred around 36 Mya (Bond et al., 2015). However, the Eocene primates bear little resemblance to any extinct or living South American primates, but do bear striking resemblance to Eocene African anthropoids while the Santa Rosa rodents exhibit the derived status relative to the contemporaneous African rodents. Then, these authors suggested two possibilities that rodents and primates might not have had simultaneous crossing episodes or that the two groups had differing rates of diversification after their arrival in South America (Bond et al., 2015). Our results appear to support the latter idea, because the patterns of Hystricognathi and Platyrrhini *SIRH11/ZCCHC16* diversification are very different, i.e., conservative vs. highly diversified, although this might not be directly related to the morphology of the molars.

In the Chiroptera, the dN/dS analysis did not provide good evidence to indicate that *SIRH11/ZCCHC16* is subject to different types of evolutionary selection between the megabats and microbats. This may be because the numbers of species are limited, resulting in the fact that a more detailed analysis is necessary to construct a precise evolutionary view among such closely related species. It is known that between these two suborders of Chiroptera, sophisticated laryngeal echolocation system is absent in Megachiroptera (Teeling, 2005, 2009). Therefore, it will be of interest to elucidate how structural changes in *SIRH11/ZCCHC16* relate to certain neurological changes affecting differences in this behavior between these two suborders of Chiroptera. It should be noted that recent molecular data indicate that Microchiroptera is not a monophyletic group, thus, suggesting that sophisticated laryngeal echolocation in the bats either originated in the ancestor of all bats and was subsequently lost in lineages leading to the megabats or originated more than once in the microbat lineages (Teeling et al., 2000). We found that the dN/dS values exhibit a large variety in several eutherian lineages that display the N-terminus deletion. This finding suggests that the function of the truncated *SIRH11/ZCCHC16* ORFs diverged in a species-specific manner, implying that the protein contributed to diversification of eutherians by increasing evolutionary fitness although *SIRH11/ZCCHC16* itself is not an essential gene in eutherian development and growth. However, it will be necessary to carry out maximum likelihood estimates of the dN/dS values using PAML branch models or other techniques to obtain supportive evidence for this idea.

Knockout mice demonstrated that *Sirh11/Zcchc16* is involved in cognitive function, including attention, impulsivity and working memory. In mice, *Sirh11/Zcchc16* is expressed in the adult kidney, testis and ovary in addition to the brain,

but male and female KO mice exhibited normal fertility and kidney function. However, it is possible that it also plays some role in the kidney, testis, ovary and embryonic liver where *Sirh11/Zcchc16* expression was confirmed. Human *SIRH11/ZCCHC16* is expressed in similar tissues and organs, such as the adult brain, liver, kidney and testis, as shown by RT-PCR, although the levels are very low, as in the case of mice (Figure S5). Therefore, it is important to identify the roles of *SIRH11/ZCCHC16* in some other organs rather than brain in different lineages and species. It is of particular interest also to determine its function in humans because of X-linked intellectual disability and attention-deficit/hyperactivity linked phenotypes of the *Sirh11/Zcchc16* knockout mice.

MATERIALS AND METHODS

Ethics

All experiments using primate samples were performed in Kyoto University Primate Research Institute (KUPRI), in accordance with the Guidelines for Care and Use of Nonhuman Primates (Version 3; June 2010) published by KUPRI. For usage of these samples and publication of the results, we obtained permissions from the respective zoos that provided the samples.

Primate Samples

All hominoid DNA samples were extracted from liver pieces collected from animals that died of natural causes at zoos except *Nomascus leucogenys* (Nle: white-cheeked gibbon). *Nomascus* genomic DNA was isolated from its feces provided by Hirakawa Zoological Park. All New World monkey DNA samples were extracted from cultured epithelial cells originating from animals bred at KUPRI. The cultured cells were derived from a tiny piece of the ear skin of a live animal anesthetized for other purposes, such as a medical treatment or health checkup.

PCR Analysis

For gibbon *SIRH11* analysis, we prepared genomic DNA of *Hylobates lar* (Hla: white-handed gibbon) and *Symphalangus syndactylus* (Ssy: siamang). The PCR reaction was performed using PrimeSTAR GXL DNA Polymerase (TaKaRa, Japan) with the following conditions: 94°C, 2 min; 4 cycles of 98°C, 10 s, 50°C, 15 s, 68°C, 1 min; 32 cycles of 98°C, 10 s, 56°C, 15 s, 68°C, 15 s; final extension 68°C, 1 min. The following PCR primers were used: gibbon_SIRH11_F1: 5'- GGCATC TCTCCAATTCAGCTGTTAGCAACT-3', gibbon_SIRH11_R1: 5'- GGCAAGGCAATCTCTTGTGAAGTGACCACA-3', gibbon_SIRH11_F2: 5'- AGTGTCTTCTTCACAGCTAACAGC TTTGGC-3', gibbon_SIRH11_R2: 5'- CTGCAGTAGAGGCAC AAATGAGTTTCTAGC-3', gibbon_SIRH11_F3: 5'- ACATAT CTGGGCCTGACAAGAG-3', gibbon_SIRH11_R3: 5'- GGC TTGGTGTGGATCAAGG-3', gibbon_SIRH11_F4: 5'- AGC AGTCATTTGGTAAACCCAC-3', gibbon_SIRH11_R4: 5'- CAA GGAAGCCAACAATGGGAG-3'.

For Platyrrhini *SIRH11* analysis, we prepared genomic DNA of five species: Tufted capuchin (Cap), common marmoset (Mar), the Azara's owl monkey (Owl), long-haired spider monkey (Spi), common squirrel monkey (Squ) and cotton-top

tamarin (Tam). Human (Hum), rhesus macaque (Rhe) DNAs were used as controls. The PCR reaction was performed using PrimeSTAR GXL DNA Polymerase (TaKaRa, Japan) with following conditions: for F1R1 and F2R2 primer sets, 94°C, 2 min; 4 cycles of 98°C, 10 s, 50°C, 15 s, 68°C, 2 min; 30 cycles of 98°C, 10 s, 64°C, 15 s, 68°C, 40 s; final extension 68°C, 1 min; for F3R3 and F4R4 primer sets: 94°C, 2 min; 4 cycles of 98°C, 10 s, 50°C, 15 s, 68°C, 1 min; 30 cycles of 98°C, 10 s, 56°C, 15 s, 68°C, 15 s; final extension 68°C, 1 min. The following PCR primers were used: Platyrrhini_SIRH11_F1: 5'-GGCATCTCTCCA ATTCAGCTGTTAGCAACT-3', Platyrrhini_SIRH11_F2: 5'-AGTGTCTTCTTCACAGCTAACAGCTTTGGC-3', Platyrrhini_SIRH11_F3: 5'-GAGGGAGGAGAGAAAGGT ACTG-3', Platyrrhini_SIRH11_F4: 5'-TGCAGAACATTGGCC TTTTCC-3', Platyrrhini_SIRH11_R1: 5'-GGCAAGGCAATC TCTTGTGAAGTGACCACA-3', Platyrrhini_SIRH11_R2: 5'-CTGCAGTAGAGGCACAAATGAGTTTCTAGC-3', Platyrrhini_SIRH11_R3: 5'-TCTGAGCAAATTGGCAGGG TC-3', Platyrrhini_SIRH11_R4: 5'-GGTCACCATGAAACT GGGTG-3'. The PCR products were directly-sequenced.

For gorilla *SIRH11* analysis, genomic DNA was isolated from frozen liver. The PCR reaction was performed using PrimeSTAR GXL DNA Polymerase (TaKaRa, Japan) with the following conditions: 94°C, 2 min; 36 cycles of 98°C, 10 s; 55°C, 15 s; 68°C, 20 s; final extension 68°C, 60 s. The following PCR primers were used: Gorilla_SIRH11_F1: 5'-GAGGGAGGAGAG AAAGGTACTG-3' and Gorilla_SIRH11_R1: 5'-TCTGAGCAA TTGGCAGGGTC-3'.

For xenarthra *SIRH11* analysis, we prepared genomic DNA from three xenarthran species: *Tolypeutes matacus* (southern three-banded armadillo), *Choloepus didactylus* (Linnaeus's two-toed sloth), and *Myrmecophaga tridactyla* (giant anteater). Genomic DNA was isolated from frozen tissues using the DNeasy Blood & Tissues Kit (QIAGEN, Germany). The PCR reaction was performed using ExTaqHS (TaKaRa, Japan) with the following conditions: 96°C, 3 min; 30 cycles of 98°C, 10 s; 55 or 60°C, 30 s; 72°C, 60 s; final extension 72°C, 3 min. The following PCR primers were used: Xenarthra_SIRH11_F1: 5'-CTTACT GCCTGCCCATTTGGT-3', Xenarthra_SIRH11_R1: 5'-GGATT TAAAAGTTGGTGCAGG-3', Xenarthra_SIRH11_F2: 5'-GGC AGAGAATCTGATTCTA-3', Xenarthra_SIRH11_R2: 5'-GTA TTGGTGGTAGATCAGG-3'.

DNA Sequencing of *SIRH11/ZCCHC16* in Primate Species

The Gorilla_SIRH11_F1R1 PCR products described above were cloned into the pBluescript II SK (+) vector and sequenced using a forward primer: 5'-TGTAACACGACGGCCAGT-3' and a reverse primer: 5'-CAGGAAACAGCTATGACCATG-3'. Platyrrhini_SIRH11_F1R1 PCR products described above were directly-sequenced using Platyrrhini_SIRH11_F1 and R1 primers. DNA Data Bank of Japan (DDBJ) accession numbers: LC150703 for western lowland gorilla *SIRH11/ZCCHC16*, LC150704 for Tufted capuchin, LC150705 for the Azara's owl monkey and LC150706 for long hair spider monkey *SIRH11/ZCCHC16*.

Southern Blot Analysis

We prepared genomic DNA of five species: gorilla (Gor), *Hylobates lar* (Hla: white-handed gibbon), *Symphalangus syndactylus* (Ssy: siamang), rhesus macaque (Rhe), and marmoset (Mar). Twelve microgram of genomic DNA were digested by restriction enzymes, *Hind*III and *Xba*I. Southern blot analysis was performed using standard protocol (electrophoresis: submerged in 1x TAE buffer, at 1.2 V/cm, for 18 h at 4°C using 1.2% agarose gel; Treatment of DNA in gel: denaturation with 0.5 N NaOH/0.5 M NaCl for 30 min at 20–30°C, neutralization with 0.5 M Tris/0.5 M NaCl (pH7.0) for 15 min at 20–30°C; Capillary transfer to nylon membrane, Hybond-N+ (GE Healthcare): 5x SSC was supplied and absorbed by paper stack, for 4 h at 20–30°C; hybridization: 59°C for 12.5 h). The *TYR* and *SIRH11* probes were generated by genomic PCR using gorilla DNA as a template, respectively. The probe labeling, hybridization, washes, and detection were performed using AlkPhos System (GE Healthcare), per manufactures protocol.

Computational Analysis

Eighty-five eutherian mammal *SIRH11* genome sequences were downloaded from NCBI (<http://www.ncbi.nlm.nih.gov/>). Two *SIRH11* sequences we previously identified in xenarthra species, *Tolypeutes matacus* and *Choloepus didactylus*, were obtained from DDBJ accession LOC064756 and LOC064757, respectively. The *SIRH11* ORF in each species was identified by NCBI nucleotide blast search (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) using human *SIRH11* ORF sequence (GenBank Accession No. NC_000023: 112454729–112455661) as the query sequence. EMBOSS Transeq (http://www.ebi.ac.uk/Tools/st/emboss_transeq/) was used for translation nucleotide sequence to amino acids sequence. Multiple sequence alignment was constructed using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) in the default mode.

Estimation of the dN/dS Ratio

An amino acid sequence phylogenetic tree was constructed with MEGA6 (Tamura et al., 2013) using Maximum Likelihood method based on the JTT matrix based model. The codon alignment of cDNA was created with the PAL2NAL program (www.bork.embl.de/pal2nal/) (Suyama et al., 2006). The nonsynonymous/synonymous substitution rate ratio ($\omega = dN/dS$) was estimated by using CodeML (runmode: –2) in PAML (Yang, 2007).

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: TK, FI. Performed the experiments: MI, AK. Analyzed the data: MI, AK, TK, FI. Analyses of primate samples including DNA sequencing, genomic PCR and Southern blot analysis: AK. Wrote the paper: TK, FI.

FUNDING

This work was supported by funding program for Next Generation World-Leading Researchers (NEXT Program) from

the Japan Society for the Promotion of Science (JSPS) to TK, Grants-in-Aid for Scientific Research 15H04427 to AK, Grants-in-Aid for Scientific Research (S) from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan to FI, TK.

ACKNOWLEDGMENTS

We thank S. Kawada, T. Mori, and S. Kitamura at the National Museum of Nature and Science for providing us the giant anteater sample for PCR experiment. The following facilities provided us with tissue samples for PCR analyses, through the GAIN (Great Apes Information Network): Sapporo Maruyama Zoo, gorilla (Gong); Kyoto City Zoo, gorilla (Hiromi); Fukuoka City Zoological Garden, gorilla (Willy); Hirakawa Zoological Park, siamang (Lian); Fukuoka City Zoological Garden, white-handed gibbon (Taki). Fecal samples of white-cheeked gibbon (Monjiro) were provided to us by Hirakawa Zoological Park. We greatly appreciate cooperation by these facilities and their curators. We are also grateful to Yuki Enomoto of Kyoto University and Kana Mori of Hirakawa Zoological Park for technical assistance. Pacific Edit reviewed the manuscript prior to submission.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fchem.2016.00026>

Figure S1 | Alignment of amino acid sequences of SIRH11/ZCCHC16 from 83 eutherian mammals (related to Figure 1).

Figure S2 | PCR analysis of gibbon *TYR* and *LCT* (related to Figure 2B). Left panel shows the scheme of PCR primer design. Right panel shows agarose gel electrophoresis profile in each primer set. The arrows represent expected band size. To confirm the quality of genomic DNAs of Hla and Ssy, we amplified tyrosinase (*TYR*) and lactase (*LCT*) genes using the PCR primers designed on the basis of the human sequence. For *TYR* gene amplification, the PCR reaction was performed using PrimeSTAR GXL DNA Polymerase (TaKaRa, Japan) with the following conditions: 94°C, 2 min; 4 cycles of 98°C, 10 s, 50°C, 15 s, 68°C, 2 min; 30 cycles of 98°C, 10 s, 64°C, 15 s, 68°C, 40 s; 68°C, 1 min. The following PCR primers were used: human_TYR_F: 5'-TAAGAGAAGCTCTATTCTGACACTAC CTC-3' and human_TYR_R: 5'-AGCTGGTGCTTCATGGGC AAAATCAATGTC-3'. For *LCT* gene amplification, the PCR reaction was performed using PrimeSTAR GXL DNA Polymerase (TaKaRa, Japan) with the following conditions: 94°C, 2 min; 4 cycles of 98°C, 10 s, 54°C, 15 s, 55°C, 2 min; 30 cycles of 98°C, 10 s, 64°C, 15 s, 68°C, 50 s; 68°C, 1 min. The following PCR primers were used: human_LCT_F: 5'-AGTTCGAAAGAGATTGTCTACACGGGA-3' and human_LCT_R: 5'-AGCTCTGTTCATTGCCGTGGAAGGCCACGA-3'.

Figure S3 | DNA sequence alignment of Cetartiodactyla *SIRH11/ZCCHC16* (related to Figure 3E). Among 14 species, corresponding DNA sequences upstream and around the first Met of pig SIRH11/ZCCHC16 and those of the short ORFs of the other 13 Cetartiodactyla species are highly conserved, demonstrating that a large deletion occurred in this lineage.

Figure S4 | PCR analysis of xenarthran *TYR* (related to Figure 4B). Left panel shows the scheme of PCR primer design. The PCR primers, F1 and R1, were designed in consensus sequences between the armadillo (*Dasypus novemcinctus*) and sloth (*Choloepus hoffmanni*) *TYR*. Right panel shows agarose gel electrophoresis profile in each primer set. The arrows represent expected band size. To confirm the quality of anteater genomic DNA, we amplified the tyrosinase (*TYR*) gene in three xenarthra species, *Tolypeutes matacus* (southern three-banded armadillo), *Choloepus didactylus* (Linnaeus's two-toed sloth) and *Myrmecophaga tridactyla* (giant anteater). The PCR reaction was performed using

ExTaqHS (TaKaRa, Japan) with the following conditions: 96°C, 3 min; 30 cycles of 98°C, 10 s; 55°C, 30 s; 72°C, 60 s; final extension 72°C, 3 min. The following PCR primers were used: Xenarthra_TYR_F1: 5'-GTTAGTCATGTGCTTTTCAGA AG-3' and Xenarthra_TYR_R1: 5'-CCAGGTGCTTCATGAGCAAAAT-3'.

Figure S5 | Expression of human *SIRH11/ZCCHC16* in adult tissues and organs. Upper and lower panels show *SIRH11/ZCCHC16* and *ACTB* agarose gel electrophoresis profiles, respectively. HEK293T genome was amplified as a control for PCR. Human total RNA was purchased from Clontech (Human total RNA

Master Panel II, #636643). The cDNA was made from total RNA (1 µg) using Revertra Ace qPCR RT Master Mix (TOYOBO, Japan). 10 ng of cDNA was used for RT-PCR analysis. The PCR reaction was performed using ExTaqHS polymerase (TaKaRa, Japan) with following condition: 96°C, 1 min; 30 cycles (for *ACTB*) or 35 cycles (for *SIRH11/ZCCHC16*) of 98°C, 10 s; 60°C, 30 s; 72°C, 30 s; final extension 72°C, 3 min. The following primer sequences were used: hACTB_F: 5'-AAGTGTGACGTGGACATCCG-3' and hACTB_R: 5'-GATCCACAT CTGCTGGAAGG-3'; hSIRH11_F: 5'-GGTGACCCCTGCCAATTGCTC-3' and hSIRH11_R: 5'-AGGTACTCTTGTGAGCCAG-3'.

REFERENCES

- Asher, R. J., Horovitz, I., and Sánchez-Villagra, M. R. (2004). First combined cladistic analysis of marsupial mammal interrelationships. *Mol. Phylogenet. Evol.* 33, 240–250. doi: 10.1016/j.ympev.2004.05.004
- Berridge, C. W., and Waterhouse, B. D. (2003). The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes. *Brain Res. Rev.* 42, 33–84. doi: 10.1016/S0165-0173(03)00143-7
- Bond, M., Tejedor, M. F., Campbell, K. E. Jr., Chornogubsky, L., Novo, N., and Goin, F. (2015). Eocene primates of South America and the African origins of New World monkeys. *Nature* 520, 538–541. doi: 10.1038/nature14120
- Bouret, S., and Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenergic function. *Trends Neurosci.* 28, 574–582. doi: 10.1016/j.tins.2005.09.002
- Campillos, M., Doerks, T., Shah, P. K., and Bork, P. (2006). Computational characterization of multiple Gag-like human proteins. *Trends Genet.* 22, 585–589. doi: 10.1016/j.tig.2006.09.006
- Charlier, C., Segers, K., Wagenaar, D., Karim, L., Berghmans, S., Jaillon, O., et al. (2001). Human–Ovine Comparative Sequencing of a 250-kb Imprinted Domain Encompassing the Callipyge (cplg) Locus and Identification of Six Imprinted Transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. *Genome Res.* 11, 850–862. doi: 10.1101/gr.172701
- Chen, W., Liang, Y., Deng, W., Shimizu, K., Ashique, A. M., Li, E., et al. (2003). The zinc-finger protein CNBP is required for forebrain formation in the mouse. *Development* 130, 1367–1379. doi: 10.1242/dev.00349
- Curtis, D., Treiber, D., Tao, F., Zamore, P., Williamson, J., and Lehmann, R. (1997). A CCHC metal-binding domain in Nanos is essential for translational regulation. *EMBO J.* 16, 834–843. doi: 10.1093/emboj/16.4.834
- Delsuc, F., and Douzery, J. P. (2009). “Armadosillos, anteaters, and sloths (Xenarthra),” in *The Timetree of Life*, eds S. B. Hedges and S. Kumar (New York, NY: Oxford University Press), 475–478.
- Edwards, C. A., Mungall, A. J., Matthews, L., Ryder, E., Gray, D. J., Pask, A. J., et al. (2008). The Evolution of the DLK1-DIO3 Imprinted Domain in Mammals. *PLoS Biol.* 6:e135. doi: 10.1371/journal.pbio.0060135
- Houle, A. (1999). The origin of platyrrhines: an evaluation of the Antarctic scenario and the floating island model. *Am. J. Phys. Anthropol.* 109, 541–559. doi: 10.1002/(SICI)1096-8644(199908)109:4<541::AID-AJPA9>3.0.CO;2-N
- Imakawa, K., Nakagawa, S., and Miyazawa, T. (2015). Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placental during mammalian evolution. *Genes Cells* 20, 771–788. doi: 10.1111/gtc.12278
- Irie, M., Yoshikawa, M., Ono, R., Iwafune, H., Furuse, T., Yamada, I., et al. (2015). Cognitive function related to the Sirh11/Zcchc16 gene acquired from an LTR Retrotransposon in Eutherians. *PLoS Genet.* 11:e1005521. doi: 10.1371/journal.pgen.1005521
- Iwasaki, S., Suzuki, S., Pelekanos, M., Clark, H., Ono, R., Shaw, G., et al. (2013). Identification of a novel PNMA-MS1 gene in marsupials suggests the LTR retrotransposon-derived PNMA genes evolved differently in marsupials and eutherians. *DNA Res.* 20, 425–436. doi: 10.1093/dnares/dst020
- Kagami, M., Sekita, Y., Nishimura, G., Irie, M., Kato, F., Okada, M., et al. (2008). Deletions and epimutations affecting the human 14q32.2 imprinted region in individuals with paternal and maternal upd(14)-like phenotypes. *Nat. Genet.* 40, 237–242. doi: 10.1038/ng.2007.56
- 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
- Kaneko-Ishino, T., and Ishino, F. (2015). Mammalian-specific genomic functions: newly acquired traits generated by genomic imprinting and LTR retrotransposon-derived genes in mammals. *Proc. Jpn. Acad. B* 91, 511–538. doi: 10.2183/pjab.91.511
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.
- Lavialle, C., Cornelis, G., Dupressoir, A., Esnault, C., Heidmann, O., Vernochet, C., et al. (2013). Paleovirology of “syncytins”, retroviral env genes exapted for a role in placental. *Philos. Trans. R. Soc. B Biol. Sci.* 368:20120507. doi: 10.1098/rstb.2012.0507
- Luo, Z.-X., Ji, Q., Wible, J. R., and Yuan, C.-X. (2003). An early Cretaceous tribosphenic mammal and metatherian evolution. *Science* 302, 1934–1940. doi: 10.1126/science.1090718
- Madsen, O. (2009). “Mammals (Mammalia),” in *The Timetree of Life*, eds S. B. Hedges and S. Kumar (New York, NY: Oxford University Press), 459–461.
- Matsui, T., Koder, Y., Endoh, H., Miyauchi, E., Komatsu, H., Sato, K., et al. (2007). RNA recognition mechanism of the minimal active domain of the human immunodeficiency virus type-2 nucleocapsid protein. *J. Biochem.* 141, 269–277. doi: 10.1093/jb/mvm037
- Murphy, W. J., and Eduardo, E. (2009). “Placental mammals (Eutheria),” in *The Timetree of Life*, eds S. B. Hedges and S. Kumar (New York, NY: Oxford University Press), 471–474.
- Narayanan, N., Gorelick, R., and DeStefano, J. (2006). Structure/function mapping of amino acids in the N-terminal zinc-finger of the human immunodeficiency virus type 1 nucleocapsid protein: residues responsible for nucleic acid helix destabilizing activity. *Biochemistry* 45, 12617–12628. doi: 10.1021/bi060925c
- Naruse, M., Ono, R., Irie, M., Nakamura, K., Furuse, T., Hino, T., et al. (2014). Sirh7/Ldoc1 knockout mice exhibit placental P4 overproduction and delayed parturition. *Development* 141, 4763–4771. doi: 10.1242/dev.114520
- Nishihara, H., Maruyama, S., and Okada, N. (2009). Retroposon analysis and recent geological data suggest near-simultaneous divergence of the three superorders of mammals. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5235–5240. doi: 10.1073/pnas.0809297106
- Ohno, S. (1970). *Evolution by Gene Duplication*, Vol. 1-2. Berlin/Heidelberg/New York, NY: Springer-Verlag.
- Ono, R., Kobayashi, S., Wagatsuma, H., Aisaka, K., Kohda, T., Kaneko-Ishino, T., et al. (2001). A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. *Genomics* 73, 232–237. doi: 10.1006/geno.2001.6494
- Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., et al. (2006). Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* 38, 101–106. doi: 10.1038/ng1699
- Patterson, B., and Pascual, R. (1972). “The fossil mammal fauna of South America,” in *Evolution, Mammals, and Southern Continents*, eds A. Keast, F. Erk, B. Glass (Albany, NY: State University of New York Press), 247–309.
- Poux, C., Chevret, P., Huchon, D., de Jong, W. W., and Douzery, E. J. P. (2006). Arrival and diversification of caviomorph rodents and platyrrhine primates in South America. *Syst. Biol.* 55, 228–244. doi: 10.1080/10635150500481390
- Rajavashisth, T., Taylor, A., Andalibi, A., Svenson, K., and Lusis, A. (1989). Identification of a zinc-finger protein that binds to the sterol regulatory element. *Science* 245, 640–643. doi: 10.1126/science.2562787

- Rosenfeld, M. R., Eichen, J. G., Wade, D. F., Posner, J. B., and Dalmau, J. (2001). Molecular and clinical diversity in paraneoplastic immunity to Ma proteins. *Ann. Neurol.* 50, 339–348. doi: 10.1002/ana.1094
- Schlatter, S., and Fussenegger, M. (2003). Novel CNBP- and La-based translation control systems for mammalian cells. *Biotechnol. Bioeng.* 81, 1–12. doi: 10.1002/bit.10549
- Schüller, M., Jenne, D., and Voltz, R. (2005). The human PNMA family: novel neuronal proteins implicated in paraneoplastic neurological disease. *J. Neuroimmunol.* 169, 172–176. doi: 10.1016/j.jneuroim.2005.08.019
- Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., et al. (2008). Role of retrotransposon-derived imprinted gene, Rtl1, in the fetomaternal interface of mouse placenta. *Nat. Genet.* 40, 243–248. doi: 10.1038/ng.2007.51
- Springer, M. S., Meredith, R. W., Gatesy, J., Emerling, C. A., Park, J., Rabosky, D. L., et al. (2012). Macroevolutionary dynamics and historical biogeography of primate diversification inferred from a species supermatrix. *PLoS ONE* 7:e49521. doi: 10.1371/journal.pone.0049521
- Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34, W609–W612. doi: 10.1093/nar/gkl315
- Suzuki, S., Ono, R., Narita, T., Pask, A. J., Shaw, G., Wang, C., et al. (2007). Retrotransposon silencing by DNA Methylation can drive mammalian genomic imprinting. *PLoS Genet.* 3:e55. doi: 10.1371/journal.pgen.0030055
- Tamura, K., Stecher, G., Peterson, D., Filipinski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Teeling, E. C. (2005). A molecular phylogeny for bats illuminates biogeography and the fossil record. *Science* 307, 580–584. doi: 10.1126/science.1105113
- Teeling, E. C. (2009). “Bats (Chiroptera),” in *The Timetree of Life*, eds S. B. Hedges and S. Kumar (New York, NY: Oxford University Press), 499–503.
- Teeling, E. C., Scally, M., Kao, D. J., Romagnoli, M. L., Springer, M. S., and Stanhope, M. J. (2000). Molecular evidence regarding the origin of echolocation and flight in bats. *Nature* 403, 188–192. doi: 10.1038/35003188
- Voltz, R., Gultekin, S. H., Rosenfeld, M. R., Gerstner, E., Eichen, J., Posner, J. B., et al. (1999). A serologic marker of paraneoplastic limbic and brain-stem encephalitis in patients with testicular cancer. *N. Engl. J. Med.* 340, 1788–1795. doi: 10.1056/NEJM199906103402303
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

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 © 2016 Irie, Koga, Kaneko-Ishino and Ishino. 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) or licensor 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.



LTRs of Endogenous Retroviruses as a Source of Tbx6 Binding Sites

Yukuto Yasuhiko, Yoko Hirabayashi and Ryuichi Ono *

Division of Cellular and Molecular Toxicology, Biological Safety Research Centre, National Institute of Health Sciences, Tokyo, Japan

Retrotransposons are abundant in mammalian genomes and can modulate the gene expression of surrounding genes by disrupting endogenous binding sites for transcription factors (TFs) or providing novel TFs binding sites within retrotransposon sequences. Here, we show that a (C/T)CACACCT sequence motif in ORR1A, ORR1B, ORR1C, and ORR1D, Long Terminal Repeats (LTRs) of MaLR endogenous retrovirus (ERV), is the direct target of Tbx6, an evolutionary conserved family of T-box TFs. Moreover, by comparing gene expression between control mice (Tbx6 +/–) and Tbx6-deficient mice (Tbx6 –/–), we demonstrate that at least four genes, *Twist2*, *Pitx2*, *Oscp1*, and *Nfxl1*, are down-regulated with Tbx6 deficiency. These results suggest that ORR1A, ORR1B, ORR1C and ORR1D may contribute to the evolution of mammalian embryogenesis.

Keywords: endogenous retroviruses, retrotransposon, transcription factors, evolution, TBX6

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Independent Investigator,
United States

Reviewed by:

Feng Zhang,
Fudan University, China
Dustin C. Hancks,
University of Utah, United States
Yong Zhang,
Institute of Zoology (CAS), China

*Correspondence:

Ryuichi Ono
onoryu@nihs.go.jp

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 22 December 2016

Accepted: 23 May 2017

Published: 15 June 2017

Citation:

Yasuhiko Y, Hirabayashi Y and Ono R
(2017) LTRs of Endogenous
Retroviruses as a Source of Tbx6
Binding Sites. *Front. Chem.* 5:34.
doi: 10.3389/fchem.2017.00034

INTRODUCTION

About half of the mammalian genome is occupied by DNA sequences derived from transposable elements (TEs) (Lander et al., 2001; Waterston et al., 2002; Lindblad-Toh et al., 2005; de Koning et al., 2011). Retrotransposons, which mobilize via an RNA intermediate by a copy-and-paste mechanism, comprise the majority of mammalian TEs, whereas DNA transposons, which move via a cut-and-paste mechanism, comprise a small fraction and have accumulated mutations that render them immobile (Deininger et al., 2003). Most TEs are nonfunctional and are regarded as genomic parasites or junk DNA; however, a growing body of evidence suggests that retrotransposons and retrotransposon-derived genes have acquired functions essential for host survival during mammalian evolution (Yoder et al., 1997; Levin and Moran, 2011; Hancks and Kazazian, 2012).

In some cases, open reading frames from TEs are domesticated as endogenous genes during mammalian evolution. For example, *Peg10* and *Rtl1*, derived from the gag and pol proteins of the Ty3/Gypsy type retrotransposon, which is similar to *Sushi-ichi*, are highly conserved in mammals and participate in placental formation (Ono et al., 2001, 2003, 2006; Sekita et al., 2008). Similar to the gag protein of *Sushi-ichi*, the other two of the eleven *Sushi-ichi* retrotransposon homolog (*Sirh*) family genes, *Sirh7/Ldoc1* and *Sirh11/Zcchc16*, encode ORF (Open-Reading frame); they are also involved in the determination of the timing of parturition and cognitive function in the brain, respectively (Ono et al., 2011; Naruse et al., 2014; Irie et al., 2015). *Syncytins/SYNCYTINs* (mouse/human) and *FEMATRIN* (cow), derived from the envelope of endogenous retrovirus (ERV), mediate cell-cell fusion to form the syncytiotrophoblast and induce fusion with bovine endometrial cells *in vitro* (Mi et al., 2000; Dupressoir et al., 2009, 2011; Nakaya et al., 2013). *Skin aspartic protease (SAPase)*, which has a retrovirus-like aspartic protease, plays important roles in the determination of the texture of the skin by modulating the degree of hydration by processing profilaggrin (Matsui et al., 2010).

Since the discovery of TEs, it has been posited that TEs may seed regulatory elements throughout genomes and drive phenotypic differences between species via changes in transcriptional output (McClintock, 1950; Britten and Davidson, 1969; Feschotte, 2008). It has become evident that many TEs, such as long terminal repeats (LTRs) of endogenous retroviruses (ERVs), contain TF binding sites and are associated with gene expression patterns. For example, MuERV-L LTRs function as alternative promoters for protein coding genes, including *Gata4* and *Tead4*, which are important for the specification of primitive endoderm and trophoblast, respectively, in two-cell embryos (Kigami et al., 2003; Evsikov et al., 2004; Macfarlan et al., 2012). It has also been reported that MuERV-L, exclusively expressed in two-cell embryos, is captured at double-strand break (DSB) sites introduced by the CRISPR/Cas9 system in mouse zygotes (Ono et al., 2015). Some of the intracisternal A-particle (IAP) retrotransposon insertions are known to induce *de novo* metastable epi-alleles, such as *agouti viable yellow* (*Avy*), *axin fused* (*AxinFu*) and *Cdk5rap* locus (Vasicek et al., 1997; Morgan et al., 1999; Druker et al., 2004). The stochastic nature of the establishment of the epigenetic state of the 5' LTR leads to variable expressivity of the adjacent genes. Both the sense and anti-sense LINE-1 (L1) promoter can drive L1 chimeric transcripts (Criscione et al., 2016). Moreover, AS071 and AS021, two AmnSINEs present in mammals as well as birds and reptiles, are enhancers of the genes *FGF8* (fibroblast growth factor 8), 178 kb from AS071, and *SATB2*, 392 kb from AS021 (Sasaki et al., 2008). Recently, it was reported that MER41, a primate-specific endogenized gammaretrovirus, is a source of interferon γ (IFNG)-inducible binding sites (Chuong et al., 2016).

In this study, we demonstrate a potential role for ORR1A (Origin-Region Repeat 1A), ORR1B, ORR1C, and ORR1D, LTRs of the MaLR (Mammalian-Apparent Long-Terminal Repeat Retrotransposon) endogenous retrovirus-like element, in controlling gene expression via *Tbx6* binding (Smit, 1993). Because *Tbx6* functions in the regulation of early embryogenesis, including anti-neural fate regulation in the presomitic mesoderm and later somite segmentation, ORR1A, ORR1B, ORR1C, and ORR1D may have played a role in the evolution of mammalian embryogenesis (Chapman and Papaioannou, 1998; Takemoto et al., 2011).

RESULTS AND DISCUSSION

Tbx6 belongs to an evolutionarily conserved family of T-box transcription factors (TFs), known to be involved in the neural-mesodermal fate determination of axial stem cells (Chapman and Papaioannou, 1998; Takemoto et al., 2011). Previously, we revealed that *Tbx6* directly activates the expression of *Mesp2*, a segmentation and polarization factor in somitogenesis, in a Notch signal-dependent manner (Yasuhiko et al., 2006). A ligand of Notch signal, Dll1, is also a direct target of *Tbx6*, implying that *Tbx6* participates in the regulation of the Notch signaling pathway (White and Chapman, 2005). The consensus core

sequence of *Tbx6* binding sites has been reported as CACACCT or AGGTGTBRNNNN (White and Chapman, 2005). In this study, we used (C/T)CACACCT as a consensus for both reports (White and Chapman, 2005; Yasuhiko et al., 2006).

At first, the *Tbx6* binding sequence motif, (C/T) CACACCT, was identified by whole genome *in silico* screening. Furthermore, we chose the *Tbx6* binding sequence, which has at least two more *Tbx6* binding sequences within the neighboring 100 bp upstream and/or downstream regions, because we previously demonstrated that higher enhancer activity of *Tbx6* was observed when there are more than three *Tbx6* binding sequences within a narrow region. As a result, 3500 potential *Tbx6* binding sites were identified, and a characteristic feature was revealed (**Figure 1A**; Supplementary Table 1).

Approximately 70.0% of potential *Tbx6* binding sites comprise repeat sequences (**Figure 1A**). Specifically, 85.7% of the potential *Tbx6*-binding-repeat sequences were within ORR1A, ORR1B, ORR1C, and ORR1D, LTRs of the MaLR that span 679 independent ORR1s-LTRs, while SINES and LINES represent only 2% of the *Tbx6*-binding repeat sequences (Bao et al., 2015; Supplementary Table 1).

There are 166,375 Repeatmasker annotated ORR1s, including partial sequences, in the mouse genome (MM10), and 20% of them have at least one *Tbx6* binding site (**Figure 1B**). In fact, the reference sequences of ORR1s-LTRs from Repbase, which are consensus sequences of ORR1s, have one or two *Tbx6* binding sequence motifs (**Figure 1C**). These data suggest that the tandem insertion of these LTRs or degenerated LTR sequences with more than three *Tbx6* binding sequence motifs might be good targets for *Tbx6* to bind *in vivo*. Furthermore, potential *Tbx6*-binding ORR1s have more than three *Tbx6* binding motifs within themselves or share the *Tbx6* binding motifs with neighboring sequences.

Tbx6-binding ORR1s more than 300 bp in length were selected, and the consensus sequences including three *Tbx6*-binding motifs and the absolute distance from each ORR1 to the nearest mouse reference gene were determined (**Figures 1D,E**; Supplementary Table 1). The strong interaction between *Tbx6* and the consensus sequence of *Tbx6*-binding ORR1s were confirmed by electrophoretic mobility shift assay (EMSA), while the interactions disappeared by introducing mutations into the *Tbx6*-binding motif one by one (**Figure 1F**). The finding that three *Tbx6* binding motifs rather than one or two *Tbx6* binding motifs have stronger binding affinity was comparable to our previous report (**Figure 1G**; Yasuhiko et al., 2008). As *Tbx6*-binding ORR1s were relatively enriched near gene transcription start sites (**Figure 1E**), *Tbx6* may contribute to regulating the expression level of nearby genes until reaching 60 kb-windows. Then, to explore the influence of ORR1A, ORR1B, ORR1C, and ORR1D on the regulation of gene expression by *Tbx6*, we compared the expression level of 9 genes that are randomly selected within 50 kb of potential *Tbx6* binding sites on ORR1A, ORR1B, ORR1C, and/or ORR1D in *Tbx6* (+/–) (control) and *Tbx6* (–/–) (*Tbx6* KO) embryos at 8.0 day post-coitus (dpc). Because *Tbx6* KO embryos have morphological abnormalities after 9 dpc, we used 8.0 dpc embryos in this study to exclude secondary effects from morphological abnormalities.

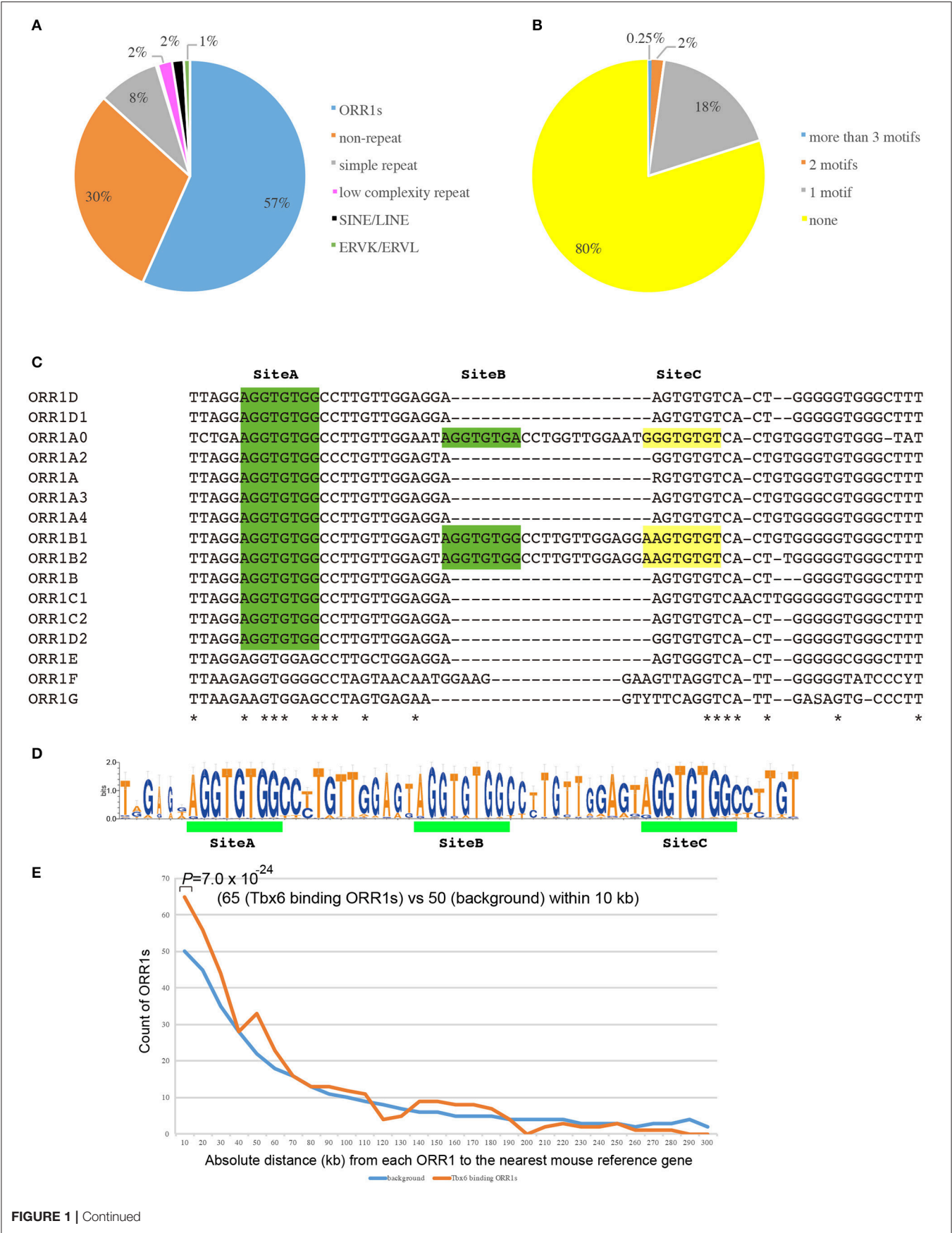


FIGURE 1 | Continued

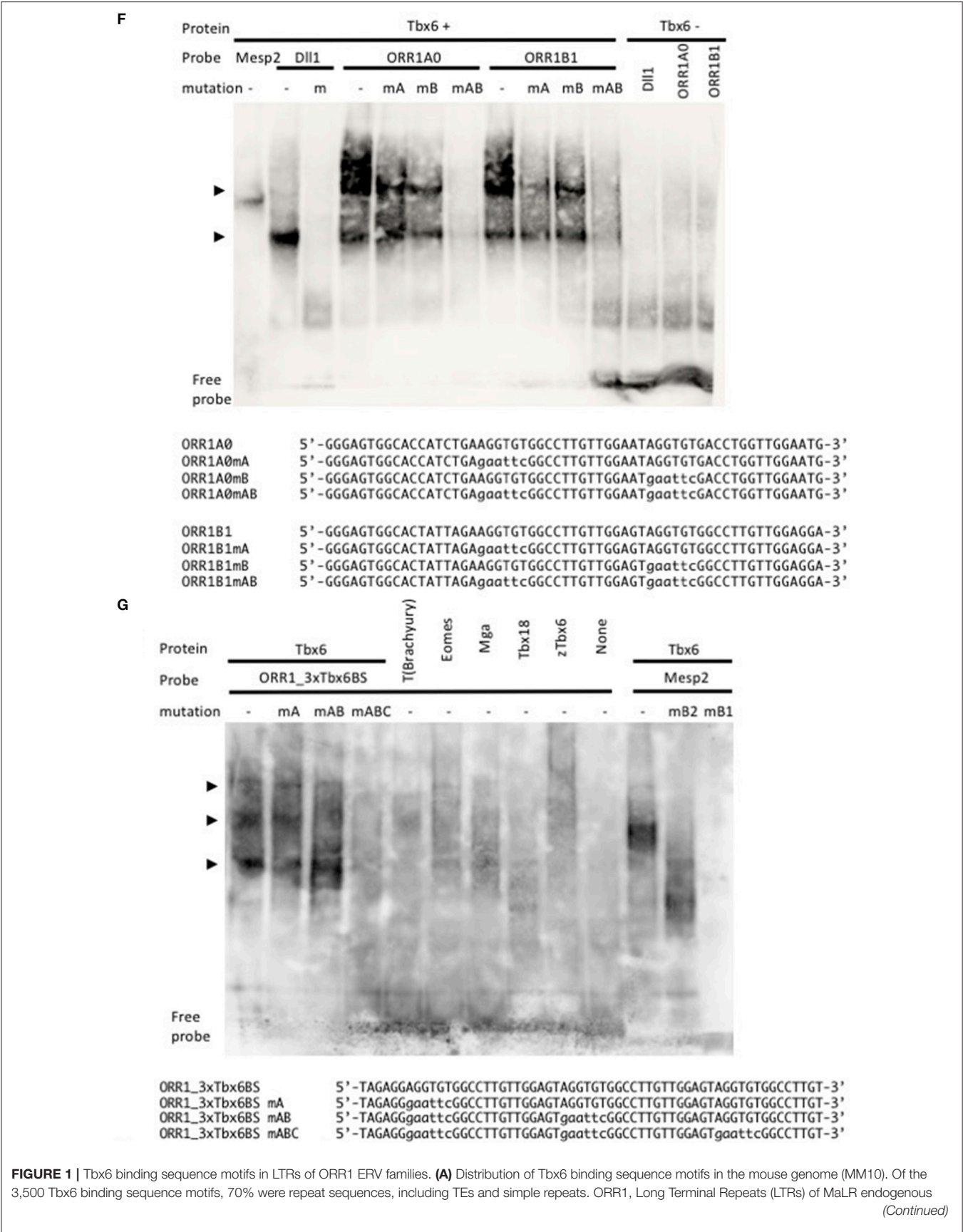


FIGURE 1 | Continued

retrovirus (ERV), occupy 57% of the total Tbx6 binding sequence motifs. **(B)** Of all the ORR1 sequences in the mouse genome, 20% of ORR1s have at least one Tbx6 binding sequence. **(C)** DNA sequence comparison between ORR1A (rodentia ancestral shared), ORR1A0 (mus musculus), ORR1A2 (muridae), ORR1A3 (muridae), ORR1A4 (muridae), ORR1B (rodentia ancestral shared), ORR1B1 (mus musculus), ORR1B2 (mus musculus), ORR1C1 (rodentia ancestral shared), ORR1C2 (rodentia ancestral shared), ORR1D1 (rodentia ancestral shared), ORR1D2 (rodentia ancestral shared), ORR1E (rodentia ancestral shared), ORR1F (muridae) and ORR1G (muridae) LTRs. Identical sequences are indicated by asterisks. The Tbx6 binding sequence motif is indicated by green boxes. Yellow boxes are indicated as a corresponding region of the Tbx6 binding sequence motif "Site C" in **Figure 1D**. **(D)** Sequence logo of the the ORR1 LTRs that had more than three Tbx6 binding sequence motifs. Three tandem "AGGTGTGs," a Tbx6 binding sequence motif, are highly conserved between ORR1 LTRs, which have more than three Tbx6 binding sequence motifs. **(E)** Frequency histogram of the absolute distance from each ORR1 to the nearest mouse reference gene. The background expectation is derived from the genome-wide ORR1s distribution. Statistical significance of the observed enrichment within the first 10 kb of the nearest mouse reference gene was assessed by a binominal test. **(F)** Site A and site B sequences independently bind to Tbx6 in an electromobility shift assay (EMSA); however, the binding affinity is much higher with the presence of both sites A and B. Sequences of oligonucleotide probes were shown below the gel image. Mutated nucleotides were depicted in lower case. **(G)** Triple Tbx6 binding sequence motif shows the highest binding affinity to Tbx6, while other T-box TFs, including T (Brachyury), Eomes, Mga, Tbx18, and zebrafish Tbx6 (zTbx6), have no affinity. Arrowheads in **(F,G)**: Positions of the bands resulted from multiple Tbx6 binding to ORR1 sequences. Sequences of oligonucleotide probes were shown below the gel image. Mutated nucleotides were depicted in lower case.

As expected, four genes, *Twist2*, *Pitx2*, *Oscp1*, and *Nfixl1*, were down-regulated, although the expression of five other genes, *Enpep*, *Prdm2*, *Corin*, *Pdpn* and *Map4k4*, was not altered significantly (**Figure 2**). It has been reported that enhancer activity could be blocked by the epigenetic repressive marks of the neighboring regions, such as histone deacetylation and trimethylation of K9 and K27 on histone H3 (H3K27me3 and H3K9me3) or an insulator, a genetic boundary element blocking the interaction between enhancers and promoters (Roth et al., 2001; Schmidl et al., 2009; Greer and Shi, 2012; Downen et al., 2014). It might be possible that five genes whose expression levels were not altered by Tbx6 deficiency could be blocked by epigenetic modifications or unknown silencers.

Our analysis revealed the rodent-specific ORR1 family of ERVs to be a source of Tbx6 binding sites. Furthermore, Tbx6-binding ORR1s are enriched near genes which might be associated with several biological process and molecular pathways (**Figure 3**). In the human genome, there are 2,927 potential TBX6 binding motifs; however, the majority of sites are not in LTRs but in simple repeat sequences or Alu (Supplementary Table 2). Although the source of Tbx6/TBX6 binding sequences is different between species, each mammalian species might shape their Tbx6/TBX6 binding sequence through mammalian evolution. Our analysis and other reports, including the primate-specific MER41 family as IFNG-inducible binding sites and AmnSINE1s as mammalian enhancers, raised the possibility that TE-derived regulatory elements influence lineage-specific mammalian evolution (Sasaki et al., 2008; Chuong et al., 2016).

MATERIALS AND METHODS

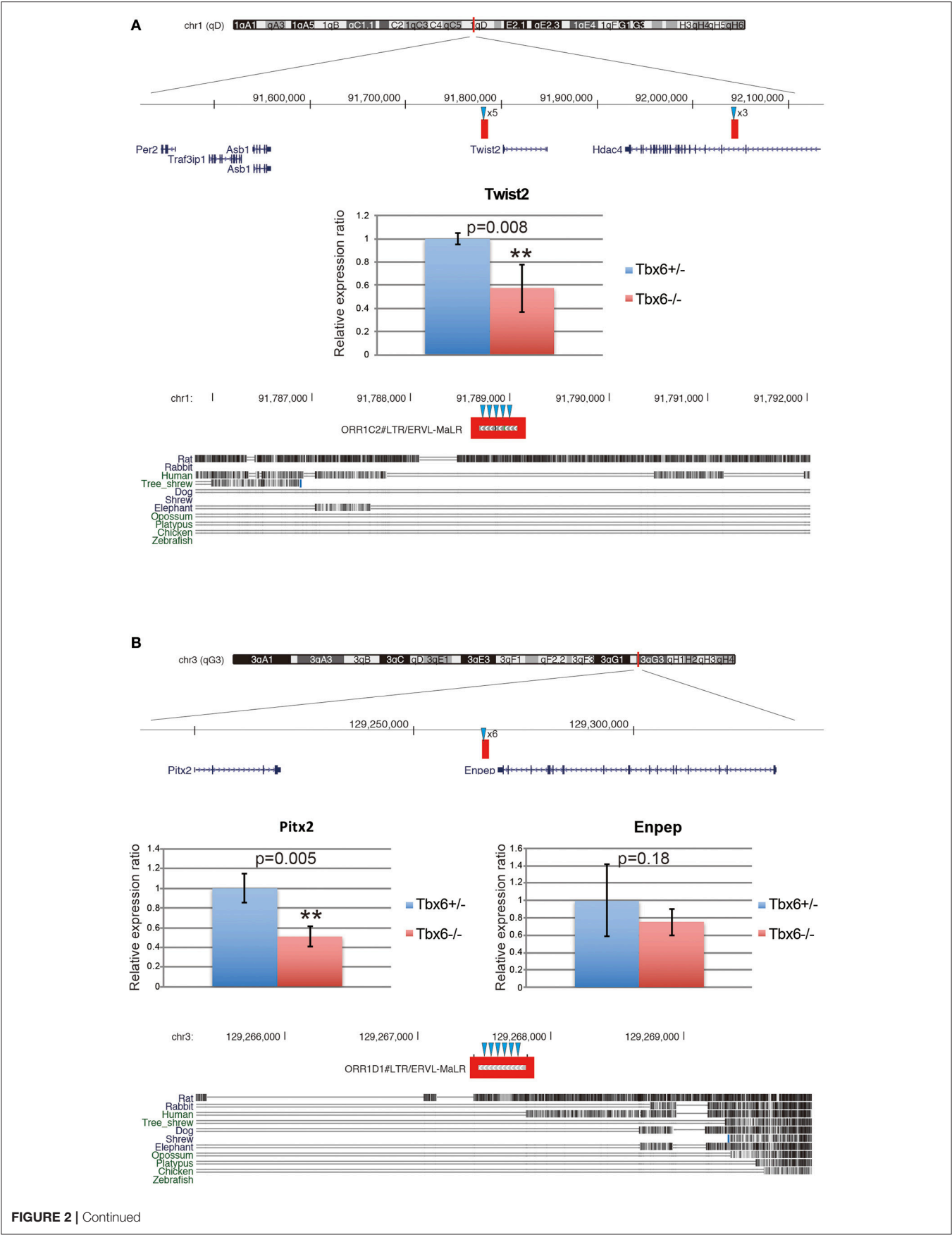
Bioinformatic Analyses

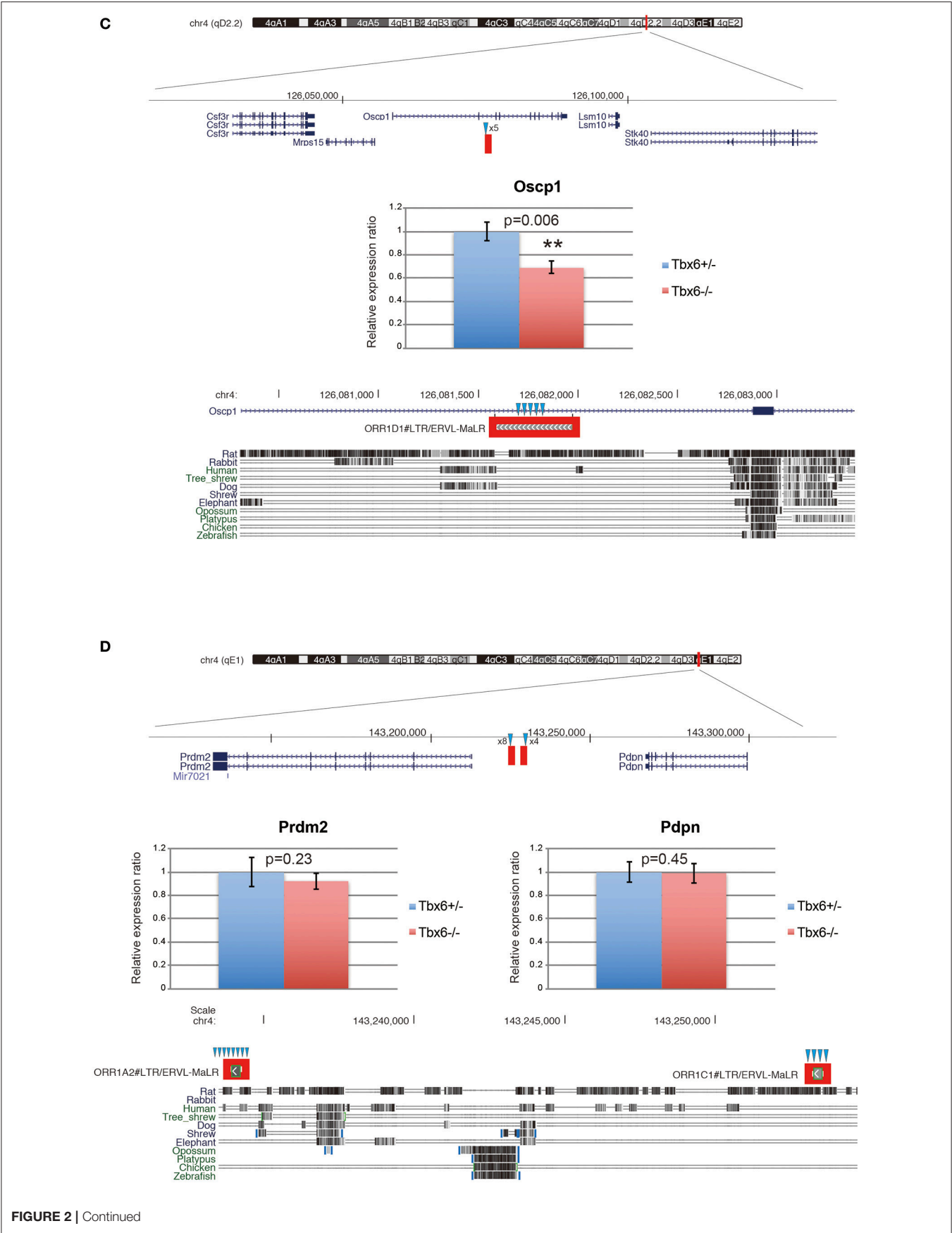
(C/T)CACACCT, Tbx6 binding sequence motifs, were identified in the mouse whole genome (MM10) and human whole genome (hg19) and filtered out when there were not two more Tbx6 binding sequences within the neighboring 100 bp upstream and/or downstream regions using gggenome (<https://gggenome.dbcls.jp>). All the TE sequences were downloaded from Repeatmasker track (mouse:MM10/human:hg19) of the UCSC genome browser (<https://genome.ucsc.edu>). The

Intersect intervals program (<https://usegalaxy.org/>) was used to identify the TEs that have potential Tbx6 binding sequences using potential Tbx6 binding sites identified as a query against Repeatmasker annotated TEs. The ClosestBed program (<https://usegalaxy.org/>) was used to find the closest mouse reference genes (MM10) and to identify the absolute distance between the potential Tbx6 binding motif and its closest reference gene. These distances were grouped by 10 kb-bin sizes. The expected background was determined by randomly sampling an equal number of the remaining 78,042 annotated ORR1s that did not have more than three Tbx6 binding motifs. Sampling was repeated 100 times, and the mean number of elements was used as the expected value for comparison to the potential Tbx6 binding ORR1s. Statistical significance was determined for the first 10-kb bin by a binominal test as previously described (Chuong et al., 2016). Gene ontology of the closest reference genes within 50 kb-windows of potential Tbx6-binding motifs were determined by the GREAT program <http://bejerano.stanford.edu/great/public/html/index.php> (**Figure 3**). The consensus sequence of the potential Tbx6-binding motifs was identified by ClustalW program (for alignment: <http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja>) and Sequence Logo program (for generation of sequence logos: <http://weblogo.berkeley.edu>). A gene ontology/signal pathway panel of Tbx6-binding ORR1s.

Electrophoretic Mobility Shift Assay (EMSA)

Full sequences of ORFs of mouse *Tbx6* (NM_011538.2), *T* (*Brachyury*; NM_009309.2), *Eomes* (NM_010136.3), *Tbx18* (NM_023814.4), and T-box-coding fragment of *Mga* (NM_013720.2) were PCR amplified and cloned in pCS2+ (Rupp et al., 1994) vector. Expression vector pCS2-zTbx6 for zebrafish Tbx6 translation was a gift from Dr Hiroyuki Takeda (Terasaki et al., 2006). Transcription factors were *in vitro* transcribed and translated using TnT(R) Quick Coupled Transcription/Translation System (Promega) following the manufacturer's protocol. Sequences of DNA probes were as follows: Mutated nucleotides are designated in lower case. Mesp2 and Dll1 were positive controls for the assay and described in Yasuhiko et al. (2006) and





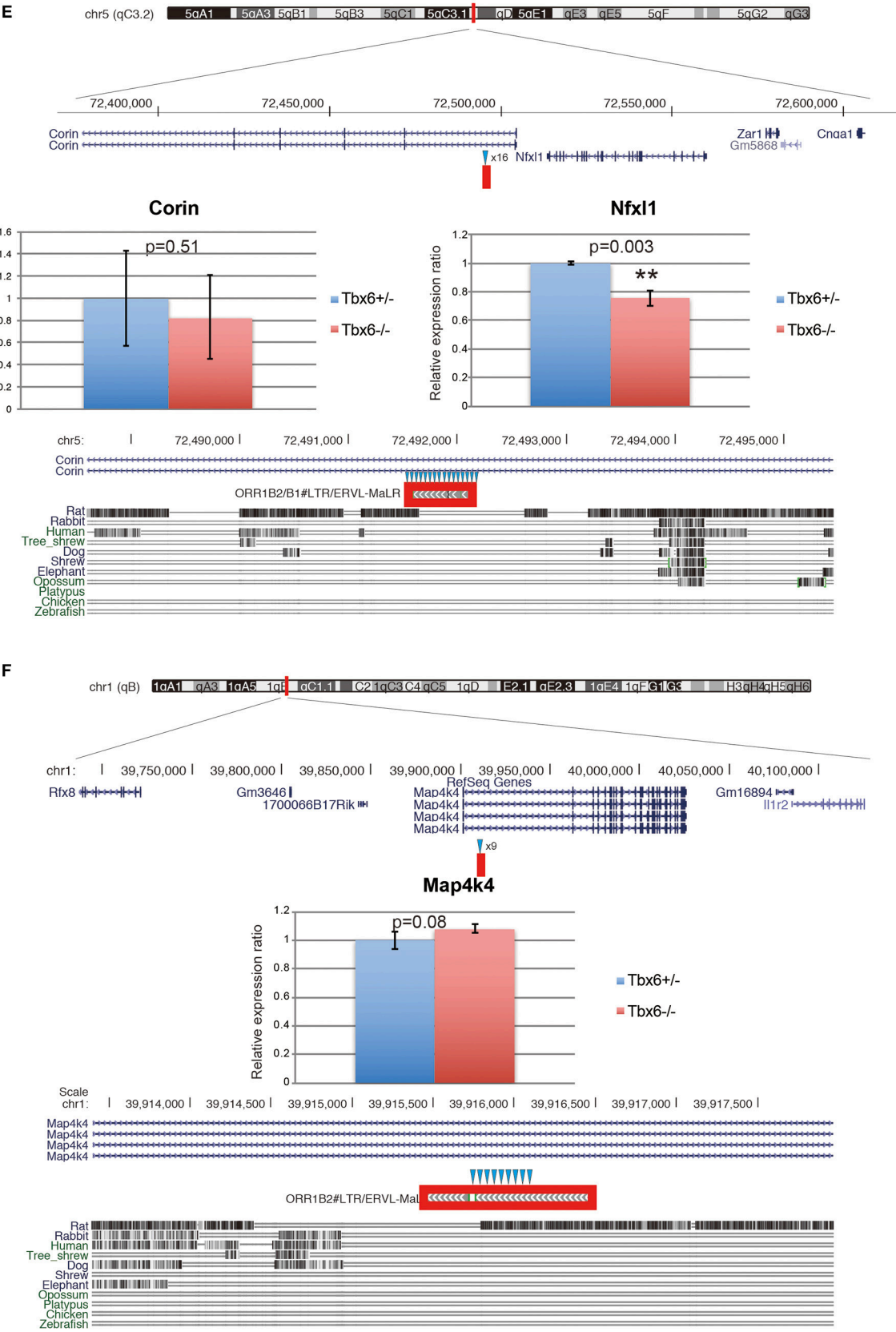


FIGURE 2 | Continued

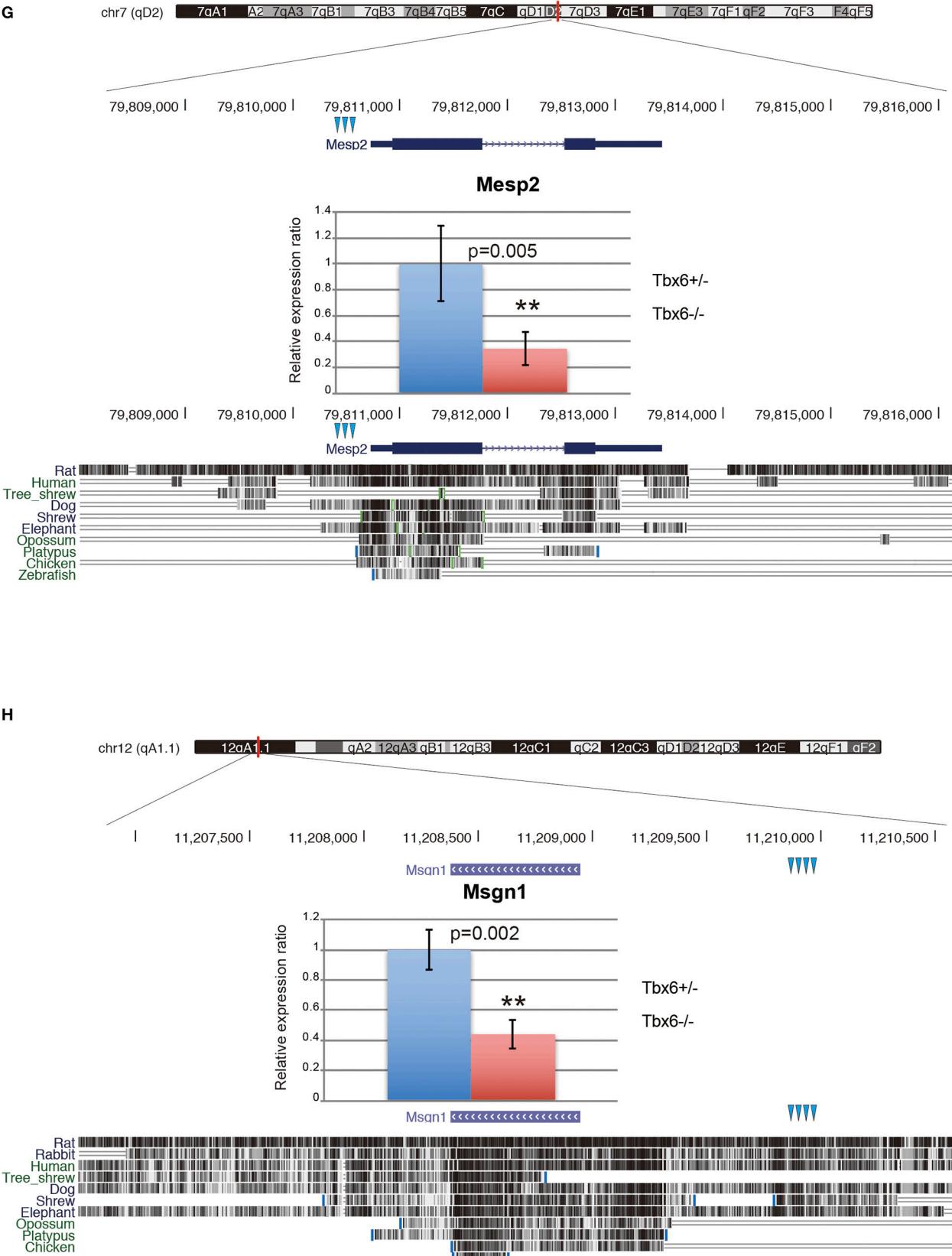
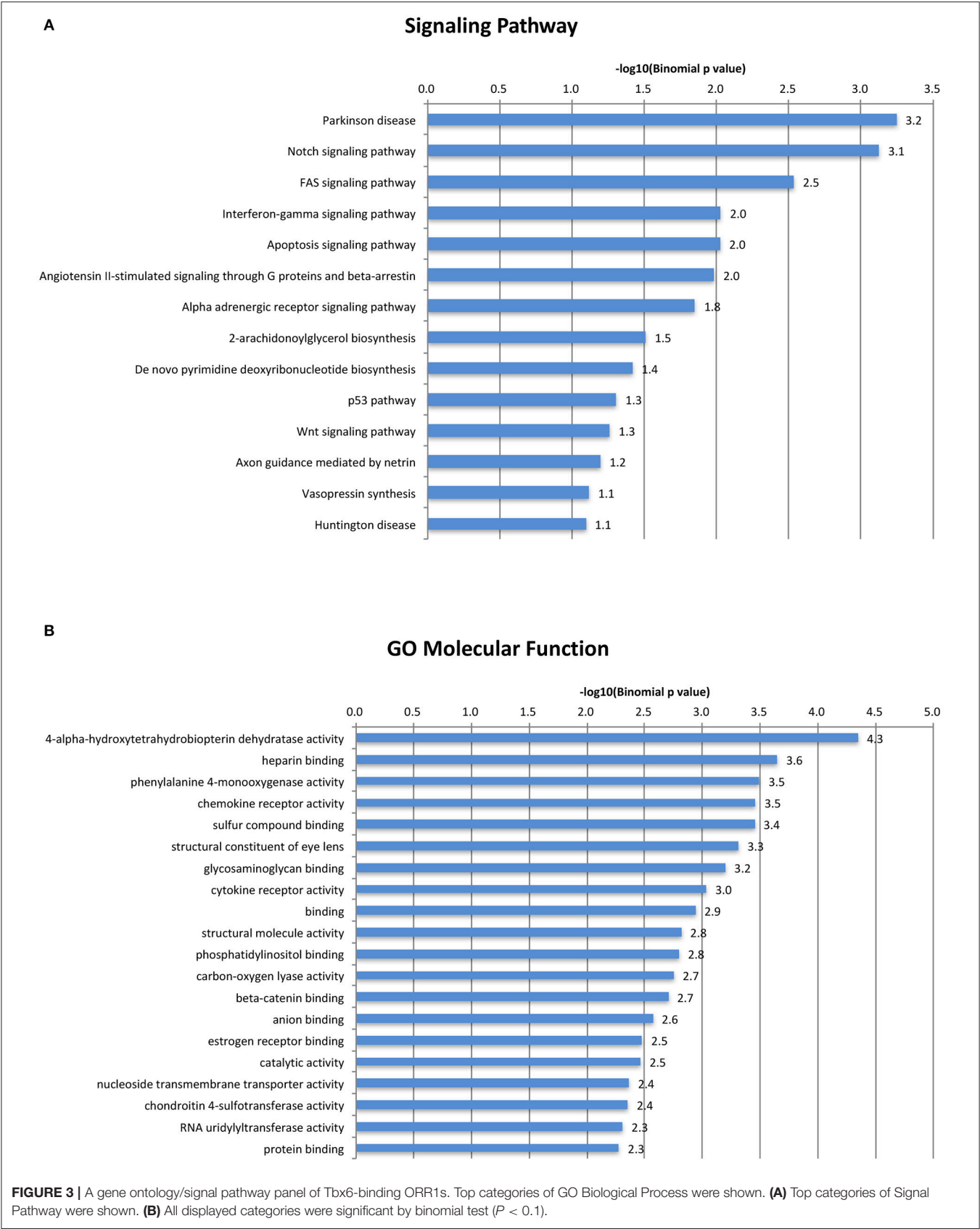


FIGURE 2 | Continued



protocol was approved by the animal welfare committee of National Institute of Health Sciences (NIHS; No.41). Animals had access to a standard chow diet and water *ad libitum* and were housed in a pathogen-free barrier facility with a 12L:12D cycle.

A *Tbx6* conditional knockout mouse (*Tbx6*^{flox}) was generated using the ES cell line TT2 and maintained in an ICR background (Yagi et al., 1993). Briefly, exon 3-5, encoding the T-box DNA binding domain of *Tbx6*, was flanked by a pair of loxP sites and knocked into the *Tbx6* locus by homologous recombination. The PGK-neo selection marker was removed by the FLP-FRT system to obtain *Tbx6*^{flox} mice. For cDNA preparation, embryos (8.0 days post-coitus) were obtained by crossing female CAG-*Cre*/*Tbx6*^{flox}/+ hybrid heterozygotes onto male *Tbx6*^{flox}/*flox* homozygotes. Embryos were genotyped by PCR using allantois genomic DNA, and total RNA were prepared using an RNeasy mini kit (QIAGEN). Total RNA was pooled from 5 (*Tbx6*+/-) and 4 (*Tbx6*-/-) 8.0dpc sibling embryos in the same litter. The sequences of primers for real-time RT-PCR were as follows:

*Twist2*_forward, 5'-TGTCCGCTCCCACTAGC-3';
*Twist2*_reverse, 5'-TGTCCAGGTGCCGAAAGTC-3';
*Pitx2*_forward, 5'-GGCAGTCACCCTGGGAAG-3';
*Pitx2*_reverse, 5'-GCCGACACTAGTTTGCGACA-3';
*Enpep*_forward, 5'-CCTGCTTTACGACCCCCTAC-3';
*Enpep*_reverse, 5'-TTAGCCACAAGTCGTCCCAC-3';
*Oscp1*_forward, 5'-GACTCTGCCGCTGCTCT-3';
*Oscp1*_reverse, 5'-TCGTCCATGAACCTTCCTGTTGA-3';
*Prdm2*_forward, 5'-GCTTCGAGGACTTCCAGAGG-3';
*Prdm2*_reverse, 5'-TGGTTTAGTGCCCCAGACAC-3';
*Pdpm*_forward, 5'-AGGTGCTACTGGAGGGCTTA-3';
*Pdpm*_reverse, 5'-GCTGAGGTGGACAGTTCCCTC-3';
*Nfxl1*_forward, 5'-AGAACCTCCTCAGTTGCTGC-3';
*Nfxl1*_reverse, 5'-AAGGGGCATTCACCAGGATG-3';
*Corin*_forward, 5'-GATATGTTACGAAACGGCCC-3';
*Corin*_reverse, 5'-CGCTCCTGTCTGCTCTCAAG-3';
*Map4k4*_forward, 5'-TTCCGGCCTCTCAAGCCT-3';
*Map4k4*_reverse, 5'-TCCCAGACTCCTCACTGGAG-3';
*Mesp2*_forward, 5'-ACCCTTACACCACTCCCTAGAAA-3';
*Mesp2*_reverse, 5'-GGTTCTGGAGACACAGAAAGACT-3';
*Msgn1*_forward, 5'-GCCAGAAAGGCAGCAAAGTC-3';
*Msgn1*_reverse, 5'-AGACAGGCGGCAGGTAATTC-3';
 β -actin_forward, 5'-CTGTGAGATCGCGTCCA-3';
 β -actin_reverse, 5'-ACGATGGAGGGGAATACAGC-3';

Primers were designed using Primer-BLAST (<https://www.ncbi.nlm.nih.gov/tools/primer-blast/>) tool to amplify 70–150 base pair (bp) fragment separated by at least one intron (>500 bp), except *Msgn1* (single exon gene). PCR reaction was performed using SYBR(R) Premix Ex Taq(TM) II (Takara RR820S) following the manufacturer's protocol, with PCR cycle as follows: 1 cycle of 95°C 30 s, 40 cycles of 95°C 5 s and 60°C 30 s.

Statistical Analyses

Statistical significance for qPCR was assessed using a two-tailed unpaired Student's *t*-test with a threshold of $p < 0.1$.

ETHICS STATEMENT

The animal facility of the National Institute of Health Sciences was approved by the Japan Health Sciences Foundation since 2008. All animal studies were conducted in accordance with the guidelines approved by the animal welfare committee of the National Institute of Health Sciences (NIHS; No. 41).

AUTHOR CONTRIBUTIONS

RO conceived of the project. YY, YH, and RO participated in the experimental design. RO performed most analyses. YY produced *Tbx6* KO mice and performed EMSA and RT-PCR. RO wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from Grant-in-Aid for Scientific Research (C) (26430183) and the Research on Regulatory Science of Pharmaceuticals and Medical Devices from Japan Agency for Medical Research and Development, AMED to RO. We are grateful to Hiroyuki Takeda (University of Tokyo) for providing *zTbx6* cDNA clones, and to Satoshi Kitajima and Eriko Ikeno for technical assistance in generating *Tbx6* KO mouse.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fchem.2017.00034/full#supplementary-material>

REFERENCES

- Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6, 11. doi: 10.1186/s13100-015-0041-9
- Britten, R. J., and Davidson, E. H. (1969). Gene regulation for higher cells: a theory. *Science* 165, 349–357. doi: 10.1126/science.165.3891.349
- Chapman, D. L., and Papaioannou, V. E. (1998). Three neural tubes in mouse embryos with mutations in the T-box gene *Tbx6*. *Nature* 391, 695–697.
- Chuong, E. B., Elde, N. C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351, 1083–1087. doi: 10.1126/science.aad5497
- Criscione, S. W., Theodosakis, N., Micevic, G., Cornish, T. C., Burns, K. H., Neretti, N., et al. (2016). Genome-wide characterization of human L1 antisense promoter-driven transcripts. *BMC Genomics* 17:463. doi: 10.1186/s12864-016-2800-5
- Deininger, P. L., Moran, J. V., Batzer, M. A., and Kazazian, H. H. Jr. (2003). Mobile elements and mammalian genome evolution. *Curr. Opin. Genet. Dev.* 13, 651–658. doi: 10.1016/j.gde.2003.10.013
- de Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384
- Downen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387. doi: 10.1016/j.cell.2014.09.030
- Druker, R., Bruxner, T. J., Lehrbach, N. J., and Whitelaw, E. (2004). Complex patterns of transcription at the insertion site of a retrotransposon in the mouse. *Nucleic Acids Res.* 32, 5800–5808. doi: 10.1093/nar/gkh914

- Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., et al. (2009). Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl. Acad. Sci. U.S.A.* 106, 12127–12132. doi: 10.1073/pnas.0902925106
- Dupressoir, A., Vernochet, C., Harper, F., Guegan, J., Dessen, P., Pierron, G., et al. (2011). A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. *Proc. Natl. Acad. Sci. U.S.A.* 108, E1164–E1173. doi: 10.1073/pnas.1112304108
- Evsikov, A. V., de Vries, W. N., Peaston, A. E., Radford, E. E., Fancher, K. S., Chen, F. H., et al. (2004). Systems biology of the 2-cell mouse embryo. *Cytogenet. Genome Res.* 105, 240–250. doi: 10.1159/000078195
- Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405. doi: 10.1038/nrg2337
- Greer, E. L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* 13, 343–357. doi: 10.1038/nrg3173
- Hancks, D. C., and Kazazian, H. H. Jr. (2012). Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* 22, 191–203. doi: 10.1016/j.gde.2012.02.006
- Irie, M., Yoshikawa, M., Ono, R., Iwafune, H., Furuse, T., Yamada, I., et al. (2015). Cognitive Function Related to the Sirh11/Zcchc16 Gene Acquired from an LTR Retrotransposon in Eutherians. *PLoS Genet.* 11:e1005521. doi: 10.1371/journal.pgen.1005521
- Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* 68, 651–654. doi: 10.1095/biolreprod.102.007906
- 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
- Levin, H. L., and Moran, J. V. (2011). Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627. doi: 10.1038/nrg3030
- Lindblad-Toh, K., Wade, C. M., Mikkelsen, T. S., Karlsson, E. K., Jaffe, D. B., Kamal, M., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438, 803–819. doi: 10.1038/nature04338
- 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
- Matsui, T., Leung, D., Miyashita, H., Maksakova, I. A., Miyachi, H., Kimura, H., et al. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* 464, 927–931. doi: 10.1038/nature08858
- McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U.S.A.* 36, 344–355. doi: 10.1073/pnas.36.6.344
- Mi, S., Lee, X., Li, X., Veldman, G. M., Finnerty, H., Racine, L., et al. (2000). Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* 403, 785–789. doi: 10.1038/35001608
- Morgan, H. D., Sutherland, H. G., Martin, D. I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318. doi: 10.1038/15490
- Nakaya, Y., Koshi, K., Nakagawa, S., Hashizume, K., and Miyazawa, T. (2013). Fematrin-1 is involved in fetomaternal cell-to-cell fusion in Bovinae placenta and has contributed to diversity of ruminant placentation. *J. Virol.* 87, 10563–10572. doi: 10.1128/JVI.01398-13
- Naruse, M., Ono, R., Irie, M., Nakamura, K., Furuse, T., Hino, T., et al. (2014). Sirh7/Ldoci1 knockout mice exhibit placental P4 overproduction and delayed parturition. *Development* 141, 4763–4771. doi: 10.1242/dev.114520
- Ono, R., Ishii, M., Fujihara, Y., Kitazawa, M., Usami, T., Kaneko-Ishino, T., et al. (2015). Double strand break repair by capture of retrotransposon sequences and reverse-transcribed spliced mRNA sequences in mouse zygotes. *Sci. Rep.* 5:12281. doi: 10.1038/srep12281
- Ono, R., Kobayashi, S., Wagatsuma, H., Aisaka, K., Kohda, T., Kaneko-Ishino, T., et al. (2001). A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. *Genomics* 73, 232–237. doi: 10.1006/geno.2001.6494
- Ono, R., Kuroki, Y., Naruse, M., Ishii, M., Iwasaki, S., Toyoda, A., et al. (2011). Identification of tammar wallaby SIRH12, derived from a marsupial-specific retrotransposition event. *DNA Res.* 18, 211–219. doi: 10.1093/dnares/dsr012
- Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., et al. (2006). Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* 38, 101–106. doi: 10.1038/ng1699
- Ono, R., Shiura, H., Aburatani, H., Kohda, T., Kaneko-Ishino, T., and Ishino, F. (2003). Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. *Genome Res.* 13, 1696–1705. doi: 10.1101/gr.906803
- Roth, S. Y., Denu, J. M., and Allis, C. D. (2001). Histone acetyltransferases. *Annu. Rev. Biochem.* 70, 81–120. doi: 10.1146/annurev.biochem.70.1.81
- Rupp, R. A., Snider, L., Weintraub, H. (1994). Xenopus embryos regulate the nuclear localization of XMyoD. *Genes Dev.* 8, 1311–1323. doi: 10.1101/gad.8.11.1311
- Sasaki, T., Nishihara, H., Hirakawa, M., Fujimura, K., Tanaka, M., Kokubo, N., et al. (2008). Possible involvement of SINEs in mammalian-specific brain formation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4220–4225. doi: 10.1073/pnas.0709398105
- Schmidl, C., Klug, M., Boeld, T. J., Andreesen, R., Hoffmann, P., Edinger, M., et al. (2009). Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res.* 19, 1165–1174. doi: 10.1101/gr.091470.109
- Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., et al. (2008). Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. *Nat. Genet.* 40, 243–248. doi: 10.1038/ng.2007.51
- Smit, A. F. (1993). Identification of a new, abundant superfamily of mammalian LTR-transposons. *Nucleic Acids Res.* 21, 1863–1872. doi: 10.1093/nar/21.8.1863
- Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E., et al. (2011). Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells. *Nature* 470, 394–398. doi: 10.1038/nature09729
- Terasaki, H., Murakami, R., Yasuhiko, Y., Shin, I. T., Kohara, Y., Saga, Y., et al. (2006). Transgenic analysis of the medaka mesp-b enhancer in somitogenesis. *Dev. Growth Differ.* 48, 153–168. doi: 10.1111/j.1440-169X.2006.00853.x
- Vasicek, T. J., Zeng, L., Guan, X. J., Zhang, T., Costantini, F., and Tilghman, S. M. (1997). Two dominant mutations in the mouse fused gene are the result of transposon insertions. *Genetics* 147, 777–786.
- Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562. doi: 10.1038/nature01262
- White, P. H., and Chapman, D. L. (2005). Dll1 is a downstream target of Tbx6 in the paraxial mesoderm. *Genesis* 42, 193–202. doi: 10.1002/gene.20140
- Yagi, T., Tokunaga, T., Furuta, Y., Nada, S., Yoshida, M., Tsukada, T., et al. (1993). A novel ES cell line, TT2, with high germline-differentiating potency. *Anal. Biochem.* 214, 70–76. doi: 10.1006/abio.1993.1458
- Yasuhiko, Y., Haraguchi, S., Kitajima, S., Takahashi, Y., Kanno, J., and Saga, Y. (2006). Tbx6-mediated Notch signaling controls somite-specific Mesp2 expression. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3651–3656. doi: 10.1073/pnas.0508238103
- Yasuhiko, Y., Kitajima, S., Takahashi, Y., Oginuma, M., Kagiwada, H., Kanno, J., et al. (2008). Functional importance of evolutionally conserved Tbx6 binding sites in the presomitic mesoderm-specific enhancer of Mesp2. *Development* 135, 3511–3519. doi: 10.1242/dev.027144
- Yoder, J. A., Walsh, C. P., and Bestor, T. H. (1997). Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335–340. doi: 10.1016/S0168-9525(97)01181-5

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 © 2017 Yasuhiko, Hirabayashi and Ono. 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) or licensor 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.



Endogenous Retroviruses: With Us and against Us

Thomas J. Meyer^{1*}, Jimi L. Rosenkrantz^{2,3}, Lucia Carbone^{1,2,4} and Shawn L. Chavez^{3,5}

¹ Division of Bioinformatics and Computational Biology, Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA, ² Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA, ³ Division of Reproductive & Developmental Sciences, Oregon National Primate Research Center, Portland, OR, USA, ⁴ Department of Medicine, Knight Cardiovascular Institute, Oregon Health & Science University, Portland, OR, USA, ⁵ Departments of Obstetrics and Gynecology and Physiology and Pharmacology, Oregon Health & Science University School of Medicine, Portland, OR, USA

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Independent Investigator, Ann Arbor,
USA

Reviewed by:

Ahmet M. Denli,
Salk Institute for Biological Studies,
USA
Reiner Strick,
Universitätsklinikum Erlangen,
Germany

*Correspondence:

Thomas J. Meyer
thomas.joshua.meyer@gmail.com

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 22 December 2016

Accepted: 20 March 2017

Published: 07 April 2017

Citation:

Meyer TJ, Rosenkrantz JL, Carbone L
and Chavez SL (2017) Endogenous
Retroviruses: With Us and against Us.
Front. Chem. 5:23.
doi: 10.3389/fchem.2017.00023

Mammalian genomes are scattered with thousands of copies of endogenous retroviruses (ERVs), mobile genetic elements that are relics of ancient retroviral infections. After inserting copies into the germ line of a host, most ERVs accumulate mutations that prevent the normal assembly of infectious viral particles, becoming trapped in host genomes and unable to leave to infect other cells. While most copies of ERVs are inactive, some are transcribed and encode the proteins needed to generate new insertions at novel loci. In some cases, old copies are removed via recombination and other mechanisms. This creates a shifting landscape of ERV copies within host genomes. New insertions can disrupt normal expression of nearby genes via directly inserting into key regulatory elements or by containing regulatory motifs within their sequences. Further, the transcriptional silencing of ERVs via epigenetic modification may result in changes to the epigenetic regulation of adjacent genes. In these ways, ERVs can be potent sources of regulatory disruption as well as genetic innovation. Here, we provide a brief review of the association between ERVs and gene expression, especially as observed in pre-implantation development and placentation. Moreover, we will describe how disruption of the regulated mechanisms of ERVs may impact somatic tissues, mostly in the context of human disease, including cancer, neurodegenerative disorders, and schizophrenia. Lastly, we discuss the recent discovery that some ERVs may have been pressed into the service of their host genomes to aid in the innate immune response to exogenous viral infections.

Keywords: endogenous retrovirus, genome, human disease, pre-implantation embryo, stem cells, placenta, innate immunity

BACKGROUND

A retroviral genome exists in different forms during its replication cycle. A viral particle, or virion, protects the RNA genome of the retrovirus during escape from the host cell and infection of new cells. A virion that enters a new host cell deploys its genomic payload, using its own reverse transcriptase to convert the RNA viral genome into a DNA copy which is integrated into the host genome, referred to as a provirus (**Figure 1**). Subsequently, a provirus can be transcribed into RNA again, and then translated by the host's ribosomal machinery to produce more virions. Ancient retroviral infections have occasionally resulted in such integrations into the germline of the



et al., 2015) and alternative repression (Thomas and Schneider, 2011; Manghara and Douville, 2013; Leung et al., 2014; Liu et al., 2014; Schlesinger and Goff, 2015; Wolf et al., 2015; Thompson et al., 2016) across the vast number and variety of ERVs within the genome during germ cell development and embryogenesis.

Despite the existence of elaborate mechanisms that mediate ERV inactivation within the genome, there is extensive evidence that some ERVs are still active and play an important role during gametogenesis and pre-implantation development. Upregulation of ERV proviral transcription and protein expression has been well documented in early human embryos and embryonic stem cells (hESCs). For example, elevated expression of the ERV-H family has been observed within both naïve-like and primed hESC sub-populations (Wang J. et al., 2014; Theunissen et al., 2016; Supplementary Table 1). Additional transcripts from the ERV-K (HML-2) family are also observed at high levels within hESCs and rapidly decrease upon differentiation (Fuchs et al., 2013). Expression of ERV-K begins at the 8-cell stage, concurrent with embryonic genome activation (EGA), and continues throughout pre-implantation development into the blastocyst stage. A majority of actively transcribed ERV-K loci during this time are associated with LTR5HS, a specific subclass of LTR, which is confined to human and chimpanzee and contains an OCT4 binding motif. The LTR5HS subclass requires both hypomethylation and OCT4 binding for transcriptional activation, which synergistically facilitated ERV-K expression (Grow et al., 2015; Supplementary Table 1). Based on the elevated activity of these ERVs within hESCs and pre-implantation embryos, as well as their known interactions with other cellular factors during this time, it is thought that these ERVs have been functionally incorporated into roles important for defining and maintaining pluripotent specific states.

The role of LTRs as regulatory regions for proviral DNA represents an additional function that can be utilized by or incorporated into host genomes. In particular, LTRs are known to be co-opted as promoters or enhancer elements of nearby genes important during embryonic development and maintenance of pluripotency (Friedli and Trono, 2015). Nearly, ~33% of all transcripts in human embryonic tissues are associated with repetitive elements, suggesting a clear pattern of embryonic cell specificity for viral promoters (Fort et al., 2014). Many transcripts detected in the totipotent blastomeres of mouse 2-cell embryos are initiated from LTRs upon EGA as well, indicating that these repeat sequences may help drive cell-fate regulation in mammals (MacFarlan et al., 2012). Regulatory activities of certain LTRs have also been shown to provide important functions not only in embryonic cells, but also within germ cells during gametogenesis. For example, germline-specific transactivating p63 (GTAp63), a member of the p53 family and a transcript important for maintaining genetic fidelity in the human male germline, is under the transcriptional control of ERV9 LTR (Ling et al., 2002; Beyer et al., 2011; Liu and Eiden, 2011; Supplementary Table 1). Transcriptionally active GTAp63 suppresses proliferation and induces apoptosis upon DNA

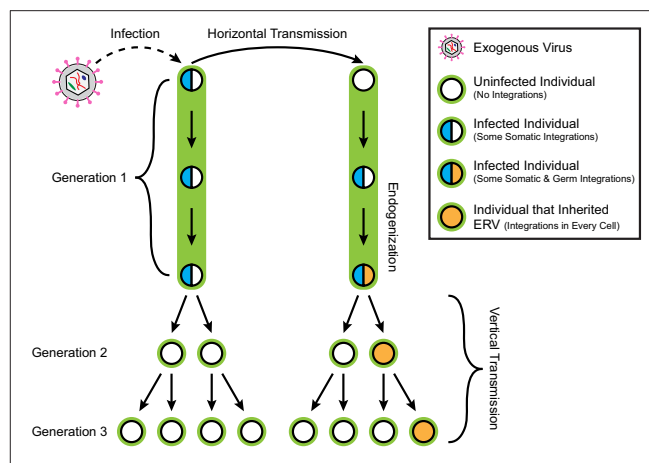


FIGURE 2 | Retroviral infection, horizontal transmission, endogenization, and vertical transmission.

An exogenous retrovirus infects an individual in generation 1, resulting in their accruing provirus integrations in some somatic cells. Horizontal transmission of the virus from the first individual to the second results in the second accruing somatic integrations as well. However, the second individual subsequently receives germline integrations. The descendants of the first individual do not inherit any retroviral integrations, while any germline integrations in the second individual are transmitted vertically to half of its descendants as endogenous retrovirus insertions present in every cell. Only half of the descendants of this second individual in Generation 2 inherit any given germline integration locus because any cell receiving a new integration does so on only one copy of the affected chromosome. This results in a heterozygous pattern of inheritance.

damage in healthy testis and is frequently lost in human testicular cancers. Restoration of GTAp63 expression levels in cancer cells was observed upon treatment with a histone deacetylase (HDAC) inhibitor, indicating possible epigenetic control of ERV9-mediated GTAp63 expression via activating histone acetylation marks. Thus, the ability of ERV9 regulatory regions to contribute to the maintenance of male germline stability is yet another example of how ERVs have evolved to serve an important function in their human hosts (Liu and Eiden, 2011).

ERVS IN THE PLACENTA

The placenta is a transient organ representing the maternal-fetal interface during pregnancy; it is derived from the outer trophoblast (TE) layer of blastocysts, and plays a critical role in the gas, nutrient, and waste exchange required for normal embryonic growth. It is well established that both mouse and human placentas are hypomethylated compared to other somatic cells derived from either *in vivo* or *in vitro* sources (Ehrlich et al., 1982; Fuke et al., 2004; Cotton et al., 2009; Popp et al., 2010; Hon et al., 2013). As such, the DNA methylation levels of LTRs within human placentas more closely resemble that observed in oocytes than in somatic tissues, averaging ~60% methylation across the genome (Schroeder et al., 2015). Given this hypomethylation of LTRs in placentas, it is not surprising that numerous sub-families of ERV proviruses are expressed within human placental tissues. More specifically,

there is evidence of proviral transcription from ERV-E (Yi and Kim, 2007), ERV3 (ERV-R; Boyd et al., 1993; Andersson et al., 2005), ERV-K (Kammerer et al., 2011), ERV-fb1 (Sugimoto et al., 2013), ERV-V1/2 (Esnault et al., 2013), ERV-W (Blond et al., 2000), and ERV-FRD (Blaise et al., 2003; Supplementary Tables 1, 2).

The most notable ERV families producing functional proteins during placentation are ERV-W and ERV-FRD, corresponding to Syncytin-1 and Syncytin-2, respectively, which are critical for the cellular fusion underlying human placental syncytia formation and maintenance (Blond et al., 2000; Mi et al., 2000; Blaise et al., 2003, 2005; Dunk et al., 2012; Supplementary Table 2). Cellular fusion is a relatively unique function in normal healthy tissues, with muscle, bone and placenta being the major exceptions. Since regulation of this highly specified function is of much interest, the precise mechanisms underlying the transcriptional control of the Syncytin-1 gene have been the topic of several investigations. Both DNA and histone H3K9 methylation have been reported to be important for inactivating ERV-W and thus repressing Syncytin-1 expression, resulting in pathological conditions such as exogenous viral infections and preeclampsia when repression does not occur (Matoušková et al., 2006; Gimenez et al., 2009; Li et al., 2014; Zhuang et al., 2014). It has been shown that transcriptional activation of the ERV-W locus and the promotion of cell fusion also requires the synergism of LTR promoter hypomethylation, along with the binding of several transcription factors such as GCM1, Sp1, and GATA family members (Yu et al., 2002; Cheng et al., 2004; Prudhomme et al., 2004; Cheng and Handwerger, 2005; Chang et al., 2011). Recently, another ERV-derived protein called suppressyn has been identified to alternatively regulate Syncytin-1, but not Syncytin-2-based cell fusion by inhibiting its interaction with the Syncytin-1 associated receptor, ASCT2 (Sugimoto et al., 2013; Supplementary Table 2). Suppressyn is a truncation product of the proviral *env* gene from the ERV-fb1 element and is transcribed within the placenta. Within normal human placentas, suppressyn is co-expressed with Syncytin-1 in the syncytiotrophoblast layer (Sugimoto et al., 2013), further supporting that these two factors are involved in cell-cell fusion regulation at the maternal-fetal interface *in utero*.

Notably, integration of ERV-W and ERV-FRD into the genome occurred prior to the divergence of Old World (Catarrhini; Cáceres et al., 2006) and New World (Platyrrhini) monkeys (Blaise et al., 2003), respectively, thus Syncytin-1 and Syncytin-2 are only present in higher-order primate (Haplorhini) species, although functionally similar yet distinct ERV proviral proteins have been discovered throughout most mammalian genomes, as reviewed in Imakawa et al. (2015). The ERV-V *env* gene present within Old World monkeys has also been implicated in trophoblast fusion activity, possibly alleviating the lack of functional Syncytin-1 within these species, while the ERV-V reiterations present within the human genome are not functional in this capacity (Esnault et al., 2013; Supplementary Table 2). Syncytin-A and Syncytin-B appear to function like human Syncytins within the mouse placenta and are known to have entered the murine (Muridae) lineages approximately 20 million years ago (Dupressoir et al.,

2005). Similarly, Syncytin-Ory1 has been discovered in rabbits and hares (Leporidae; Heidmann et al., 2009), Syncytin-Car1 within 26 different species of carnivorans (Carnivora; Cornelis et al., 2012), Syncytin-Mar1 within the squirrel-related clade (either Scuridae or Marmotini; Redelsperger et al., 2014), Syncytin-Ten1 within tenrec (Tenrecidae; Cornelis et al., 2014), Syncytin-Rum1 in ruminants (Ruminantia; Cornelis et al., 2013), and Syncytin-Opo1 within the short-lived placenta of opossum and kangaroo marsupials (Marsupialia; Cornelis et al., 2015).

Several ERV captured *env* genes have been proposed to have an immunosuppressive role that is important for preventing maternal rejection of the semi-allogenic fetus during pregnancy. In addition to fusogenic properties derived from the *env* gene of ERV-FRD, Syncytin-2 contains a classical Env retroviral immunosuppressive domain that has been shown to have immunosuppressive activity via *in vitro* tumor-rejection assay (Mangeney et al., 2007). Given observed protein expression within cytotrophoblasts cells of the human placenta, Syncytin-2 has been suggested to facilitate fetal tolerance by suppressing the maternal immune system. Other ERV-derived *env* proteins from ERV-V and ERV-K have also been proposed to possess an immunosuppressive role in controlling the maternal immune system during pregnancy. This is based on findings that both families have one or more proviral loci in the genome with intact *env* open reading frames (ORFs) and a corresponding immunosuppressive domain. Additionally, both ERV-V and ERV-K expression has been observed within placental trophoblast cells at the maternal-fetal interface, although corresponding *in vitro* functional assays have not yet been completed to directly support *in vivo* findings (Kammerer et al., 2011; Subramanian et al., 2011; Supplementary Table 1). Until these studies are undertaken, the exact function of ERV-V and ERV-K and whether *env* protein expression from these ERVs induce maternal immunosuppression within the placenta, will remain unknown.

ERVS AND HUMAN DISEASE

Through insertional mutagenesis, recombination between homologous copies, and the regulatory disruption that epigenetic suppression of ERV insertions can cause to nearby gene loci, there are many mechanisms by which these elements might cause disease. In particular, their association with various cancers has been well demonstrated, as reviewed in Katoh and Kurata (2013). For instance, ERV activity has been strongly associated with many breast cancers (Golan et al., 2008; Wang-Johanning et al., 2008; Salmons et al., 2014). While in melanoma tissues, ERV-K expression of both RNA and protein has been shown (Büscher et al., 2005), and one recent study identified 24 ERV-K (HML-2) loci transcribed (Schmitt et al., 2013). In another study of Hodgkin's lymphoma, all cancer patient samples were found to have alternative transcripts of the *CSF1R*, an important locus associated with this cancer, that initiate at the LTR of an ERV located ~6.2

kb upstream of the normal promoter (Lamprecht et al., 2010).

ERVs have been demonstrated to be associated with a variety of neurologic diseases, as reviewed in Douville and Nath (2014). One such disease is amyotrophic lateral sclerosis (ALS). Elevated ERV-K (HML-2) activity has been observed in the brain tissue of ALS patients (Douville et al., 2011), while transgenic animals expressing the ERV-K *env* gene in cortical and spinal neurons developed motor dysfunction, suggesting that these elements may contribute to neurodegeneration (Li et al., 2015). Additionally, the expression of ERV-W *env* and *gag* has been observed in samples of muscle from ALS patients (Oluwole et al., 2007). While the ERV-W findings may be due to the inflammatory response (Alfahad and Nath, 2013), the support for the involvement of ERV-K in ALS is mounting, though causality has yet to be demonstrated. Multiple sclerosis (MS) is another neurological disease in which ERVs have been strongly implicated. MSRV (multiple sclerosis-associated retrovirus), a subtype of ERV-W, as well as ERV-W1 and W2 and ERV-H/F have all been linked to MS (reviewed in Christensen, 2016). One study showed significantly elevated Env antigen in serum of MS patients relative to controls, while qPCR of ERV-W in mononuclear cells from blood (PBMC) showed association with MS relative to controls (Perron et al., 2012a). This same study demonstrated Env expression in eight well-characterized MS brains that had lesions throughout the parenchyma and in perivascular infiltrates, as well as at the rim of chronic active lesions. ERV association with schizophrenia and bipolar disorder has been demonstrated through the presence of biomarkers for ERV-K and ERV-W found in blood, cerebrospinal fluid, and the pre-frontal cortex (Karlsson et al., 2001; Huang et al., 2006, 2011; Perron et al., 2012b). In one study of schizophrenia, hypermethylation of a specific ERV-W LTR insertion located in the regulatory region of the *GABBR1* gene was associated with risk of schizophrenia (Hegyi, 2013). A nearly full-length ERV-K insertion near the *PRODH* gene, known to be associated with schizophrenia and other neuropsychiatric disorders, has been shown to work in concert with the internal *PRODH* CpG island to activate the gene. It is thought that aberrant DNA methylation of this locus may be a piece of the schizophrenia puzzle (Suntsova et al., 2013).

ERVS MAY PLAY A ROLE IN THE INNATE IMMUNE RESPONSE

While the majority of ERV proviruses have acquired mutations, thereby preventing translation into protein, certain families have been especially well preserved and contain functional ORFs for one or more of the classical proviral genes. Within primates, ERV-K (HML-2) represents the best-preserved and most recently active ERV, containing a substantial number of loci that have predicted coding potential throughout different primate genomes. It has also been observed that ERV-K encodes a small accessory protein, Rec, in naïve ES cells and human blastocysts. Overexpression of Rec protein within

human pluripotent cells increases the innate antiviral response and can inhibit exogenous viral infections, suggesting an immunoprotective role of the ERV-K Rec protein during early embryonic development (Grow et al., 2015; Supplementary Table 1). An additional ERV-K proviral protein, gag, which makes up the core of viral particles in exogenous retroviruses, is also expressed within human blastocysts and pluripotent cells. Immunolabeling of ERV-K gag protein followed by confocal and transmission electron microscopy revealed ERV-K gag protein within structures of blastocysts resembling viral-like particle (VLPs). This suggested that some ERV proviral sequences within the human genome still retain the ability to code for viral proteins and form VLPs during normal human embryogenesis. Proteins produced from ERV *env* genes have also been demonstrated to function as restriction factors against exogenous retroviral infection (Malfavon-Borja and Feschotte, 2015).

Even ERV proviruses that do not contain functional ORFs can still harbor sequence motifs that serve to modulate the activity of nearby genes. For instance, interferon (IFN)-inducible enhancers have been dispersed via ERV insertions adjacent to IFN-inducible genes independently over mammalian evolution. This has resulted in regulatory networks of genes able to work in concert due to the presence of these ERV sequences. Further, CRISPR-Cas9 deletion of a MER41 insertion upstream of *AIM2* in HeLa cells disrupted the endogenous IFN γ -inducible regulation of this locus, demonstrating the utility that host genomes can obtain over time by harnessing ERV sequences (Chuong et al., 2016). In another example showing the variety of mechanisms by which ERVs are involved with innate immunity, Chiappinelli et al. (2015) demonstrated that induction of ERV expression, and especially bidirectional transcription of ERVs, activated a double-stranded RNA sensing pathway that triggers a type I interferon response and apoptosis.

CONCLUSIONS

The relationship between ERVs and the human genome is a diverse and complicated one, resulting from millions of years of co-evolution. ERVs are known to be involved in disease through insertional mutagenesis, as targets of epigenetic repression, and via recombination of sequences between the homologous copies of these elements scattered across the genome. Throughout mammalian evolution, the deleterious effects of ERVs have been balanced by the benefits gained from innovative co-option of their sequences and proteins by their host genomes. These innovations include the intimate relationship between ERV activity with embryonic and placental development, as well as a number of ERV-associated regulatory networks that have become important components of the normal function of our genome. An innate immune response to exogenous retroviral infection is likely only one of several ERV functional roles. Once thought to have been quiescent, dead residents of the human genome, we are only beginning to uncover the scope of how actively intertwined our biology is with these long-time genomic partners.

AUTHOR CONTRIBUTIONS

TM and JR drafted the manuscript and figures. SC and LC edited the manuscript.

FUNDING

TM was supported by the National Library of Medicine of the National Institutes of Health under Award Number T15LM007088. JR was supported by the Collins Medical Trust Foundation and Glenn/AFAR Scholarship for Research in the Biology of Aging. LC and SC were supported by NIH/NICHD R01HD086073-A1, National Centers for Translational Research in Reproduction and Infertility

REFERENCES

- Alfahad, T., and Nath, A. (2013). Retroviruses and amyotrophic lateral sclerosis. *Antiviral Res.* 99, 180–187. doi: 10.1016/j.antiviral.2013.05.006
- Andersson, A. C., Yun, Z., Sperber, G. O., Larsson, E., and Blomberg, J. (2005). ERV3 and related sequences in humans: structure and RNA expression. *J. Virol.* 79, 9270–9284. doi: 10.1128/JVI.79.14.9270-9284.2005
- Bénit, L., Dessen, P., and Heidmann, T. (2001). Identification, phylogeny, and evolution of retroviral elements based on their envelope genes. *J. Virol.* 75, 11709–11719. doi: 10.1128/JVI.75.23.11709-11719.2001
- Beyer, U., Moll-Rocce, J., Moll, U. M., and Döbelstein, M. (2011). Endogenous retrovirus drives hitherto unknown proapoptotic p63 isoforms in the male germ line of humans and great apes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3624–3629. doi: 10.1073/pnas.1016201108
- Blaise, S., de Parseval, N., and Heidmann, T. (2005). Functional characterization of two newly identified human endogenous retrovirus coding envelope genes. *Retrovirology* 2:19. doi: 10.1186/1742-4690-2-19
- Blaise, S., de Parseval, N., Bénit, L., and Heidmann, T. (2003). Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. U.S.A.* 100, 13013–13018. doi: 10.1073/pnas.2132646100
- Blond, J. L., Lavillette, D., Cheynet, V., Bouton, O., Oriol, G., Chapel-Fernandes, S., et al. (2000). 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.* 74, 3321–3329. doi: 10.1128/JVI.74.7.3321-3329.2000
- Boyd, M. T., Bax, C. M., Bax, B. E., Bloxam, D. L., and Weiss, R. A. (1993). The human endogenous retrovirus ERV-3 is upregulated in differentiating placental trophoblast cells. *Virology* 196, 905–909. doi: 10.1006/viro.1993.1556
- Büscher, K., Trefzer, U., Hofmann, M., Sterry, W., Kurth, R., and Denner, J. (2005). Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Res.* 65, 4172–4180. doi: 10.1158/0008-5472.CAN-04-2983
- Cáceres, M., NISC Comparative Sequencing Program, and Thomas, J. W. (2006). The gene of retroviral origin Syncytin 1 is specific to hominoids and is inactive in old world monkeys. *J. Hered.* 97, 100–106. doi: 10.1093/jhered/esj011
- Chang, C. W., Chang, G. D., and Chen, H. (2011). A novel cyclic AMP/Epac1/CaMKI signaling cascade promotes GCM1 desumoylation and placental cell fusion. *Mol. Cell. Biol.* 31, 3820–3831. doi: 10.1128/MCB.05582-11
- Cheng, Y. H., and Handwerger, S. (2005). A placenta-specific enhancer of the human syncytin gene. *Biol. Reprod.* 73, 500–509. doi: 10.1095/biolreprod.105.039941
- Cheng, Y. H., Richardson, B. D., Hubert, M. A., and Handwerger, S. (2004). Isolation and characterization of the human syncytin gene promoter. *Biol. Reprod.* 70, 694–701. doi: 10.1095/biolreprod.103.023473
- Chiappinelli, K. B., Strissel, P. L., Desrichard, A., Li, H., Henke, C., Akman, B., et al. (2015). Inhibiting DNA methylation causes an interferon response in cancer via dsRNA including endogenous retroviruses. *Cell* 162, 974–986. doi: 10.1016/j.cell.2015.07.011
- (NCTRI). Additional funding for SC came from the Georgeanna Jones Foundation for Reproductive Medicine, Medical Research Foundation of Oregon, and the Collins Medical Trust. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Collins Medical Trust Foundation, or Medical Research Foundation of Oregon.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fchem.2017.00023/full#supplementary-material>

- Christensen, T. (2016). Human endogenous retroviruses in neurologic disease. *APMIS* 124, 116–126. doi: 10.1111/apm.12486
- Chuong, E. B., Elde, N. C., and Feschotte, C. (2016). Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351, 1083–1087. doi: 10.1126/science.aad5497
- Cornelis, G., Heidmann, O., Bernard-Stoecklin, S., Reynaud, K., Véron, G., Mulot, B., et al. (2012). Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. *Proc. Natl. Acad. Sci. U.S.A.* 109, E432–E441. doi: 10.1073/pnas.1115346109
- Cornelis, G., Heidmann, O., Degrelle, S. A., Vernochet, C., Lavielle, C., Letzelter, C., et al. (2013). Captured retroviral envelope syncytin gene associated with the unique placental structure of higher ruminants. *Proc. Natl. Acad. Sci. U.S.A.* 110, E828–E837. doi: 10.1073/pnas.1215787110
- Cornelis, G., Vernochet, C., Carradec, Q., Souquere, S., Mulot, B., Catzeflis, F., et al. (2015). Retroviral envelope gene captures and syncytin exaptation for placentation in marsupials. *Proc. Natl. Acad. Sci. U.S.A.* 112, E487–E496. doi: 10.1073/pnas.1417000112
- Cornelis, G., Vernochet, C., Malicorne, S., Souquere, S., Tzika, A. C., Goodman, S. M., et al. (2014). Retroviral envelope syncytin capture in an ancestrally diverged mammalian clade for placentation in the primitive Afrotherian tenrecs. *Proc. Natl. Acad. Sci. U.S.A.* 111, E4332–E4341. doi: 10.1073/pnas.1412268111
- Cotton, A. M., Avila, L., Penaherrera, M. S., Affleck, J. G., Robinson, W. P., and Brown, C. J. (2009). Inactive X chromosome-specific reduction in placental DNA methylation. *Hum. Mol. Genet.* 18, 3544–3552. doi: 10.1093/hmg/ddp299
- Douville, R. N., and Nath, A. (2014). Human endogenous retroviruses and the nervous system. *Handb. Clin. Neurol.* 123, 465–485. doi: 10.1016/B978-0-444-53488-0.00022-5
- 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
- Dunk, C. E., Gellhaus, A., Drewlo, S., Baczyk, D., Pötgens, A. J., Winterhager, E., et al. (2012). The molecular role of connexin 43 in human trophoblast cell fusion. *Biol. Reprod.* 86, 115. doi: 10.1095/biolreprod.111.096925
- Dupressoir, A., Marceau, G., Vernochet, C., Bénit, L., Kanellopoulos, C., Sapin, V., et al. (2005). Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae. *Proc. Natl. Acad. Sci. U.S.A.* 102, 725–730. doi: 10.1073/pnas.0406509102
- Ehrlich, M., Gama-Sosa, M. A., Huang, L. H., Midgett, R. M., Kuo, K. C., McCune, R. A., et al. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res.* 10, 2709–2721. doi: 10.1093/nar/10.8.2709
- Esnault, C., Cornelis, G., Heidmann, O., and Heidmann, T. (2013). Differential evolutionary fate of an ancestral primate endogenous retrovirus envelope gene, the EnvV syncytin, captured for a function in placentation. *PLoS Genet.* 9:e1003400. doi: 10.1371/journal.pgen.1003400
- Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C. A., Saxena, A., et al. (2014). Deep transcriptome profiling of mammalian stem cells supports a

- regulatory role for retrotransposons in pluripotency maintenance. *Nat. Genet.* 46, 558–566. doi: 10.1038/ng.2965
- Friedli, M., and Trono, D. (2015). The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* 31, 429–451. doi: 10.1146/annurev-cellbio-100814-125514
- Fuchs, N. V., Loewer, S., Daley, G. Q., Izsvak, Z., Löwer, J., and Löwer, R. (2013). Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. *Retrovirology* 10:115. doi: 10.1186/1742-4690-10-115
- Fuke, C., Shimabukuro, M., Petronis, A., Sugimoto, J., Oda, T., Miura, K., et al. (2004). Age related changes in 5-methylcytosine content in human peripheral leukocytes and placentas: an HPLC-based study. *Ann. Hum. Genet.* 68(Pt 3), 196–204. doi: 10.1046/j.1529-8817.2004.00081.x
- Gimenez, J., Montgiraud, C., Oriol, G., Pichon, J. P., Ruel, K., Tsatsaris, V., et al. (2009). Comparative methylation of ERVWE1/syncytin-1 and other human endogenous retrovirus LTRs in placenta tissues. *DNA Res.* 16, 195–211. doi: 10.1093/dnares/dsp011
- Golan, M., Hizi, A., Resau, J. H., Yaal-Hahoshen, N., Reichman, H., Keydar, I., et al. (2008). Human endogenous retrovirus (HERV-K) reverse transcriptase as a breast cancer prognostic marker. *Neoplasia* 10, 521–533. doi: 10.1593/neo.07986
- 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
- Guo, H., Zhu, P., Yan, L., Li, R., Hu, B., Lian, Y., et al. (2014). The DNA methylation landscape of human early embryos. *Nature* 511, 606–610. doi: 10.1038/nature13544
- Hegyí, H. (2013). GABBR1 has a HERV-W LTR in its regulatory region—a possible implication for schizophrenia. *Biol. Direct* 8:5. doi: 10.1186/1745-6150-8-5
- Heidmann, O., Vernochet, C., Dupressoir, A., and Heidmann, T. (2009). Identification of an endogenous retroviral envelope gene with fusogenic activity and placenta-specific expression in the rabbit: a new “syncytin” in a third order of mammals. *Retrovirology* 6:107. doi: 10.1186/1742-4690-6-107
- Hon, G. C., Rajagopal, N., Shen, Y., McCleary, D. F., Yue, F., Dang, M. D., et al. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat. Genet.* 45, 1198–1206. doi: 10.1038/ng.2746
- Huang, W. J., Liu, Z. C., Wei, W., Wang, G. H., Wu, J. G., and Zhu, F. (2006). Human endogenous retroviral pol RNA and protein detected and identified in the blood of individuals with schizophrenia. *Schizophr. Res.* 83, 193–199. doi: 10.1016/j.schres.2006.01.007
- Huang, W., Li, S., Hu, Y., Yu, H., Luo, F., Zhang, Q., et al. (2011). Implication of the env gene of the human endogenous retrovirus W family in the expression of BDNF and DRD3 and development of recent-onset schizophrenia. *Schizophr. Bull.* 37, 988–1000. doi: 10.1093/schbul/sbp166
- Imakawa, K., Nakagawa, S., and Miyazawa, T. (2015). Baton pass hypothesis: successive incorporation of unconserved endogenous retroviral genes for placental during mammalian evolution. *Genes Cells* 20, 771–788. doi: 10.1111/gtc.12278
- Kammerer, U., Germeyer, A., Stengel, S., Kapp, M., and Denner, J. (2011). Human endogenous retrovirus K (HERV-K) is expressed in villous and extravillous cytotrophoblast cells of the human placenta. *J. Reprod. Immunol.* 91, 1–8. doi: 10.1016/j.jri.2011.06.102
- Karlsson, H., Bachmann, S., Schröder, J., McArthur, J., Torrey, E. F., and Yolken, R. H. (2001). Retroviral RNA identified in the cerebrospinal fluids and brains of individuals with schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* 98, 4634–4639. doi: 10.1073/pnas.061021998
- Katoh, I., and Kurata, S. (2013). Association of endogenous retroviruses and long terminal repeats with human disorders. *Front. Oncol.* 3:234. doi: 10.3389/fonc.2013.00234
- Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., et al. (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* 8:e1002440. doi: 10.1371/journal.pgen.1002440
- Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., et al. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* 16, 571–579. doi: 10.1038/nm.2129
- Lee, H. J., Hore, T. A., and Reik, W. (2014). Reprogramming the methylome: erasing memory and creating diversity. *Cell Stem Cell* 14, 710–719. doi: 10.1016/j.stem.2014.05.008
- Leung, D., Du, T., Wagner, U., Xie, W., Lee, A. Y., Goyal, P., et al. (2014). Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6690–6695. doi: 10.1073/pnas.1322273111
- Li, F., Nellaker, C., Sabuncian, S., Yolken, R. H., Jones-Brando, L., Johansson, A. S., et al. (2014). Transcriptional derepression of the ERVWE1 locus following influenza A virus infection. *J. Virol.* 88, 4328–4337. doi: 10.1128/JVI.03628-13
- 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
- Ling, J., Pi, W., Bollag, R., Zeng, S., Keskinetepe, M., Saliman, H., et al. (2002). The solitary long terminal repeats of ERV-9 endogenous retrovirus are conserved during primate evolution and possess enhancer activities in embryonic and hematopoietic cells. *J. Virol.* 76, 2410–2423. doi: 10.1128/jvi.76.5.2410-2423.2002
- Liu, M., and Eiden, M. V. (2011). Role of human endogenous retroviral long terminal repeats (LTRs) in maintaining the integrity of the human germ line. *Viruses* 3, 901–905. doi: 10.3390/v3060901
- Liu, S., Brind'Amour, J., Karimi, M. M., Shirane, K., Bogutz, A., Lefebvre, L., et al. (2014). Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev.* 28, 2041–2055. doi: 10.1101/gad.244848.114
- 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
- Malfavon-Borja, R., and Feschotte, C. (2015). Fighting fire with fire: endogenous retrovirus envelopes as restriction factors. *J. Virol.* 89, 4047–4050. doi: 10.1128/JVI.03653-14
- Mangeney, M., Renard, M., Schlecht-Louf, G., Bouallaga, I., Heidmann, O., Letzelter, C., et al. (2007). Placental syncytins: genetic disjunction between the fusogenic and immunosuppressive activity of retroviral envelope proteins. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20534–20539. doi: 10.1073/pnas.0707873105
- Manghera, M., and Douville, R. N. (2013). Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology* 10:16. doi: 10.1186/1742-4690-10-16
- Matoušková, M., Blazková, J., Pajer, P., Pavlíček, A., and Hejnar, J. (2006). CpG methylation suppresses transcriptional activity of human syncytin-1 in non-placental tissues. *Exp. Cell Res.* 312, 1011–1020. doi: 10.1016/j.yexcr.2005.12.010
- 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
- Okazaki, H., Chiba, H., Hiura, H., Hamada, H., Sato, A., Utsunomiya, T., et al. (2014). Genome-wide analysis of DNA methylation dynamics during early human development. *PLoS Genet.* 10:e1004868. doi: 10.1371/journal.pgen.1004868
- Oluwole, S. O., Yao, Y., Conradi, S., Kristensson, K., and Karlsson, H. (2007). Elevated levels of transcripts encoding a human retroviral envelope protein (syncytin) in muscles from patients with motor neuron disease. *Amyotroph. Lateral Scler.* 8, 67–72. doi: 10.1080/17482960600864207
- Perron, H., Germin, R., Bernard, C., Garcia-Montojo, M., Deluen, C., Farinelli, L., et al. (2012a). Human endogenous retrovirus type W envelope expression in blood and brain cells provides new insights into multiple sclerosis disease. *Mult. Scler.* 18, 1721–1736. doi: 10.1177/1352458512441381
- Perron, H., Hamdani, N., Faucard, R., Lajnef, M., Jamain, S., Daban-Huard, C., et al. (2012b). Molecular characteristics of human endogenous retrovirus type-W in schizophrenia and bipolar disorder. *Transl. Psychiatry* 2:e201. doi: 10.1038/tp.2012.125
- Popp, C., Dean, W., Feng, S., Cokus, S. J., Andrews, S., Pellegrini, M., et al. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463, 1101–1105. doi: 10.1038/nature08829
- Prudhomme, S., Oriol, G., and Mallet, F. (2004). A retroviral promoter and a cellular enhancer define a bipartite element which controls env ERVWE1 placental expression. *J. Virol.* 78, 12157–12168. doi: 10.1128/JVI.78.22.12157-12168.2004

- Redelsperger, F., Cornelis, G., Vernochet, C., Tennant, B. C., Catzeflis, F., Mulot, B., et al. (2014). Capture of syncytin-Mar1, a fusogenic endogenous retroviral envelope gene involved in placentation in the Rodentia squirrel-related clade. *J. Virol.* 88, 7915–7928. doi: 10.1128/JVI.00141-14
- Salmons, B., Lawson, J. S., and Günzburg, W. H. (2014). Recent developments linking retroviruses to human breast cancer: infectious agent, enemy within or both? *J. Gen. Virol.* 95(Pt 12), 2589–2593. doi: 10.1099/vir.0.070631-0
- Schlesinger, S., and Goff, S. P. (2015). Retroviral transcriptional regulation and embryonic stem cells: war and peace. *Mol. Cell. Biol.* 35, 770–777. doi: 10.1128/MCB.01293-14
- Schmitt, K., Reichrath, J., Roesch, A., Meese, E., and Mayer, J. (2013). Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. *Genome Biol. Evol.* 5, 307–328. doi: 10.1093/gbe/evt010
- Schroeder, D. I., Jayashankar, K., Douglas, K. C., Thirkill, T. L., York, D., Dickinson, P. J., et al. (2015). Early developmental and evolutionary origins of gene body DNA methylation patterns in mammalian placentas. *PLoS Genet.* 11:e1005442. doi: 10.1371/journal.pgen.1005442
- Subramanian, R. P., Wildschutte, J. H., Russo, C., and Coffin, J. M. (2011). Identification, characterization, and comparative genomic distribution of the HERV-K (HML-2) group of human endogenous retroviruses. *Retrovirology* 8:90. doi: 10.1186/1742-4690-8-90
- Sugimoto, J., Sugimoto, M., Bernstein, H., Jinno, Y., and Schust, D. (2013). A novel human endogenous retroviral protein inhibits cell-cell fusion. *Sci. Rep.* 3:1462. doi: 10.1038/srep01462
- Suntsova, M., Gogvadze, E. V., Salozhin, S., Gaifullin, N., Eroshkin, F., Dmitriev, S. E., et al. (2013). Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19472–19477. doi: 10.1073/pnas.1318172110
- Theunissen, T. W., Friedli, M., He, Y., Planet, E., O'Neil, R. C., Markoulaki, S., et al. (2016). Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* 19, 502–515. doi: 10.1016/j.stem.2016.06.011
- Thomas, J. H., and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. *Genome Res.* 21, 1800–1812. doi: 10.1101/gr.121749.111
- Thompson, P. J., MacFarlan, T. S., and Lorincz, M. C. (2016). Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* 62, 766–776. doi: 10.1016/j.molcel.2016.03.029
- Tristem, M. (2000). Identification and characterization of novel human endogenous retrovirus families by phylogenetic screening of the human genome mapping project database. *J. Virol.* 74, 3715–3730. doi: 10.1128/JVI.74.8.3715-3730.2000
- van der Laan, L. J., Lockey, C., Griffith, B. C., Frasier, F. S., Wilson, C. A., Onions, D. E., et al. (2000). Infection by porcine endogenous retrovirus after islet xenotransplantation in SCID mice. *Nature* 407, 90–94. doi: 10.1038/35024089
- Walsh, C. P., Chaillet, J. R., and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* 20, 116–117. doi: 10.1038/2413
- Wang, J., Xie, G., Singh, M., Ghanbarian, A. T., Rasko, T., Szvetnik, A., et al. (2014). Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* 516, 405–409. doi: 10.1038/nature13804
- Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., et al. (2014). Programming and inheritance of parental DNA methylomes in mammals. *Cell* 157, 979–991. doi: 10.1016/j.cell.2014.04.017
- Wang-Johanning, F., Radvanyi, L., Rycaj, K., Plummer, J. B., Yan, P., Sastry, K. J., et al. (2008). Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients. *Cancer Res.* 68, 5869–5877. doi: 10.1158/0008-5472.CAN-07-6838
- Wolf, G., Yang, P., Fuchtbauer, A. C., Fuchtbauer, E. M., Silva, A. M., Park, C., et al. (2015). The KRAB zinc finger protein ZFP809 is required to initiate epigenetic silencing of endogenous retroviruses. *Genes Dev.* 29, 538–554. doi: 10.1101/gad.252767.114
- Yi, J. M., and Kim, H. S. (2007). Molecular phylogenetic analysis of the human endogenous retrovirus E (HERV-E) family in human tissues and human cancers. *Genes Genet. Syst.* 82, 89–98. doi: 10.1266/ggs.82.89
- Yu, C., Shen, K., Lin, M., Chen, P., Lin, C., Chang, G. D., et al. (2002). GCMA regulates the syncytin-mediated trophoblastic fusion. *J. Biol. Chem.* 277, 50062–50068. doi: 10.1074/jbc.M209316200
- Zhuang, X. W., Li, J., Brost, B. C., Xia, X. Y., Chen, H. B., Wang, C. X., et al. (2014). Decreased expression and altered methylation of syncytin-1 gene in human placentas associated with preeclampsia. *Curr. Pharm. Des.* 20, 1796–1802. doi: 10.2174/13816128113199990541

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 © 2017 Meyer, Rosenkrantz, Carbone and Chavez. 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) or licensor 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.



Friends-Enemies: Endogenous Retroviruses Are Major Transcriptional Regulators of Human DNA

Anton A. Buzdin^{1,2*}, Vladimir Prassolov¹ and Andrew V. Garazha^{3,4}

¹ Department of Cell Biology, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia, ² Centre for Convergence of Nano-, Bio-, Information and Cognitive Sciences and Technologies, National Research Centre "Kurchatov Institute," Moscow, Russia, ³ Group for Genomic Regulation of Cell Signaling Systems, Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia, ⁴ Department of Biomedicine, Moscow Institute of Physics and Technology, Moscow, Russia

OPEN ACCESS

Edited by:

Tammy A. Morrish,
Independent Investigator,
United States

Reviewed by:

Artem Babaian,
University of British Columbia, Canada
Avindra Nath,
National Institute of Neurological
Disorders and Stroke, United States

*Correspondence:

Anton A. Buzdin
buzdin@ponkc.com

Specialty section:

This article was submitted to
Cellular Biochemistry,
a section of the journal
Frontiers in Chemistry

Received: 17 December 2016

Accepted: 24 May 2017

Published: 08 June 2017

Citation:

Buzdin AA, Prassolov V and
Garazha AV (2017) Friends-Enemies:
Endogenous Retroviruses Are Major
Transcriptional Regulators of Human
DNA. *Front. Chem.* 5:35.
doi: 10.3389/fchem.2017.00035

Endogenous retroviruses are mobile genetic elements hardly distinguishable from infectious, or "exogenous," retroviruses at the time of insertion in the host DNA. Human endogenous retroviruses (HERVs) are not rare. They gave rise to multiple families of closely related mobile elements that occupy ~8% of the human genome. Together, they shape genomic regulatory landscape by providing at least ~320,000 human transcription factor binding sites (TFBS) located on ~110,000 individual HERV elements. The HERVs host as many as 155,000 mapped DNaseI hypersensitivity sites, which denote loci active in the regulation of gene expression or chromatin structure. The contemporary view of the HERVs evolutionary dynamics suggests that at the early stages after insertion, the HERV is treated by the host cells as a foreign genetic element, and is likely to be suppressed by the targeted methylation and mutations. However, at the later stages, when significant number of mutations has been already accumulated and when the retroviral genes are broken, the regulatory potential of a HERV may be released and recruited to modify the genomic balance of transcription factor binding sites. This process goes together with further accumulation and selection of mutations, which reshape the regulatory landscape of the human DNA. However, developmental reprogramming, stress or pathological conditions like cancer, inflammation and infectious diseases, can remove the blocks limiting expression and HERV-mediated host gene regulation. This, in turn, can dramatically alter the gene expression equilibrium and shift it to a newer state, thus further amplifying instability and exacerbating the stressful situation.

Keywords: retrovirus, gene expression regulation, pathology, cancer, inflammation, stress, stability, infection

Human endogenous retroviruses (HERVs) and related genetic elements occupy ~8% of human genome. They are thought to be remnants of multiple ancient retroviral infections (Sverdlov, 2000; Belshaw et al., 2004; Buzdin, 2007). HERV insertions occurred in the ancestral germ cell lineage, fixed in the genome and became inheritable (Buzdin et al., 2003; Dewannieux and Heidmann, 2013). In the human DNA, HERVs are represented by 504 groups including 717,778 individual fragments (RepeatMasker, hg19). The individual HERV copies are frequently interrupted by other sequences, such as transposable elements, and may each represent two or more genomic fragments.

Older HERVs have accumulated more mutations, including indels, and thus are more fragmented than the evolutionary young elements. For example, the MER-41-int element located at the position chr1:26,952,949-26,962,938 (hg19 assembly) is broken into four fragments in the genome, but biologically this was a single HERV.

Many families of HERVs are highly transcriptionally active in human tissues (Buzdin et al., 2006a; Maliniemi et al., 2013). Genomic copies of HERVs are of particular interest because in addition to viral genes they also have various regulatory sequences concentrated in their *long terminal repeats* (LTRs)—about 1 kb long fragments of DNA flanking the “body” of an element (**Figure 1**). The LTRs serve as promoters (Buzdin et al., 2006a), enhancers (Chuong et al., 2013; Suntsova et al., 2013), polyadenylation signals (Suntsova et al., 2015), chromatin folding reshapers (Schumann et al., 2010), and binding sites for various nuclear proteins (Young et al., 2013). Importantly, most of HERVs reside in the human genome as solitary LTRs arisen due to homologous recombinations between the two 5′- and 3′-flanking LTRs of the same full-length element (Hughes and Coffin, 2004). In turn, further recombinations between the different HERVs may cause genomic instability (Trombetta et al., 2016). For example, this mechanism may be responsible for at least 78 copy number variation cases encompassing known human genes (Campbell et al., 2014).

Most of the newly inserted HERVs harbor functional retroviral genes, such as those encoding for the reverse transcriptase/integrase, the structural polyprotein Gag and the envelope polyprotein Env, and the canonical function of an LTR is the regulation of retroviral expression. However, the LTRs may also drive the transcription of closely located genomic sequences and human genes (Buzdin et al., 2006a). In this minireview, we pay attention to the regulatory function of HERVs which donated multitude of functional sequences to the human genome.

STRUCTURE OF LTR AND BINDING OF NUCLEAR PROTEINS

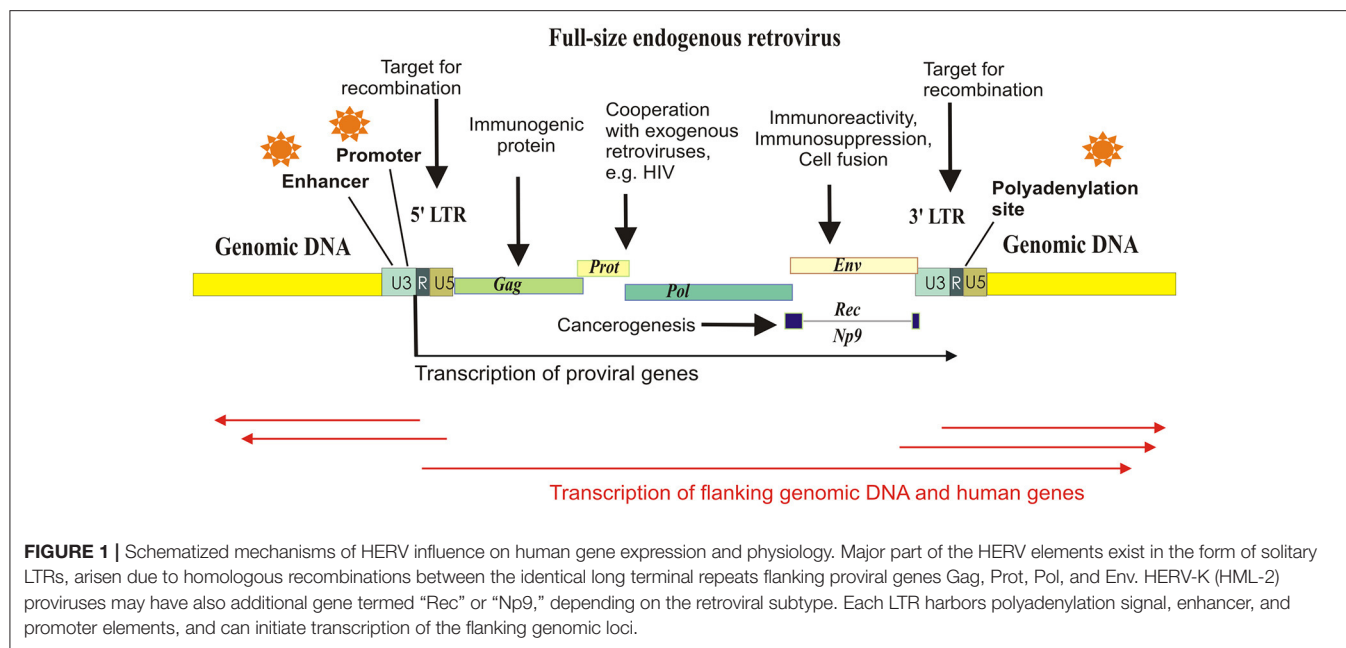
Most of the HERVs exist in the form of solitary LTRs. The LTRs include promoter elements, enhancers, transcriptional factor binding sites, splice sites, and polyadenylation signals, and are thought to serve as the major transcriptional regulators of HERVs. LTRs specifically bind host cell nuclear proteins (Trubetskoy et al., 2002) and serve in the following five pathways of human transcriptional regulation: (i) LTRs may have enhancer/repressor activities (Domansky et al., 2000; Hughes and Coffin, 2004; Ruda et al., 2004; Suntsova et al., 2013); (ii) LTRs may be promoter active; (iii) LTR may provide polyadenylation sites to terminate read-through transcripts; (iv) LTRs may provide splice sites; (v) LTRs may regulate host genes by RNA interference (Gogvadze et al., 2009).

Mapping DNaseI hypersensitivity sites (DHS) is the method of choice for the high-throughput identification of the regulatory genomic regions. Similarly, transcription factor binding sites (TFBS) denote fragments of DNA with nuclear protein binding

capacities (Ho et al., 2012). We combined investigation of both DHS and TFBS content of HERVs on a genomic scale (Garazha et al., 2015). To this end, we annotated all the genomic copies of HERVs and devised a bioinformatic algorithm mapping relevant TFBS and DHS features. For the entire set of HERVs, ~140,000 individual inserts (~19%) had at least one DHS and ~110,000 inserts (~15%)—at least one TFBS. Totally, there were identified ~155,000 and ~320,000 HERV-related DHS and TFBS, respectively (Garazha et al., 2015). This directly evidences potential implication of HERVs in the regulation of thousands of human genes. This is also in line with the previous finding that ~30% of all p53 binding sites localized by chromatin immunoprecipitation approach in the human genome fall within the HERV elements (Wang et al., 2007). Finally, as much as ~31.4% of all human transcription start sites were mapped within various transposable elements, including the HERVs (Faulkner et al., 2009).

All the 504 known HERV groups were characterized with regard to their TFBS content and showed very different results (available at <http://herv.pparser.net/TotalStatistic.php>). The families differed in their copy number, ranging from several copies as for the HERV-F, to more than 22,000 members as for the THE1B family. The total number of TFBS was also strikingly different—from zero (LTR5, LTR7A) to ~13,000 (MLT1K). The densities of TFBS also varied among the families. This is also important to quantitate absolute numbers of TFBS in each family. For example, the LTR12 family had the biggest proportion of TFBS-positive members and donated ~1,300 TFBS to the human DNA, whereas the family MLT1K contributed the greatest number of TFBS (~13,000), but had a small occurrence of TFBS-positive members. Interestingly, the TFBS and DHS tended to commonly appear in the same HERV elements. The probability that a particular element had DHS, was proportionate to the number of TFBS mapped herein (Garazha et al., 2015). Although, these findings provide clues for identification and functional annotation of multiple previously unknown human regulatory sequences, they are most likely still an underestimation of the HERV-generated TFBS pool. The repetitive nature of HERVs in many cases did not allow to directly attribute TFBS or DHS to any particular HERV element (Garazha et al., 2015).

Importantly, all the interrogated transcriptional factors had TFBS in the HERVs. This can explain extremely diverse and sometimes strongly tissue-specific influence of the different HERVs on the gene expression. For example, the LTR of the most recent HERV family HERV-K (HML-2) containing many human-specific and even polymorphic members, shows very high promoter and enhancer activities in the human germ cells and the corresponding tumors (seminoma), being transcriptionally silent in the other tissues (Domansky et al., 2000; Ruda et al., 2004). The promoter activity of the HERV-K (HML-2) inserts also provided the first evidence for the human specific antisense regulation of gene expression (Gogvadze et al., 2009). The human-specific LTRs located in the introns of genes *SLC4A8* and *IFT172* (for sodium bicarbonate cotransporter and intraflagellar transport protein



172, respectively) can *in vivo* generate transcripts that are reverse-complementary to the exons of those genes. Overexpression of the antisense transcripts resulted in approximately three- to four-fold decrease in mRNA levels for these genes (Gogvadze et al., 2009).

The HERVs can also provide polyadenylation signals for the regulation of gene expression. For example, mRNA for 8-kDa human protein similar to transcription factor GON4L is polyadenylated using the HERV-K (HML-2) LTR sequence (Baust et al., 2000). Another human transcription factor gene, *ZNF195*, utilizes the HERV-F LTR as the alternative polyadenylation site (Kjellman et al., 1999).

FUNCTIONAL INTERPLAY OF HERVs AND HUMAN GENOME

Expression of HERVs is tightly controlled by the host cell because it may be deleterious. Even the physical presence of the repetitive sequences in the genome can generate genomic instability due to homologous recombination between the HERV elements. HERVs can bias normal gene regulatory networks (Suntsova et al., 2015; reviewed by Rebollo et al., 2012). Expression of HERV proteins may result in dangerous inflammatory or immunosuppressive effects (Cho et al., 2008). In mammals, endogenous retroviruses are transcriptionally repressed using the KRAB domain zinc finger proteins and their cofactor TRIM28, which recruit methylation machinery to HERV copies (Turelli et al., 2014). In embryonic cells, a zinc finger protein Yin Yang 1 may serve as another repressor of HERV transcription by suppressing promoter activities of the LTRs (Schlesinger et al., 2013). Besides DNA methylation, histone modification is considered an alternative mechanism of endogenous retroviral repression in embryonal stem cells with the proteins SETDB1

(methyltransferase responsible for H3K9 trimethylation) and H3K4 demethylase LSD1/KDM1A involved (reviewed by Rebollo et al., 2012).

APOBEC3 protein family has another function in suppression of HERVs and retroviruses. APOBEC3G (hA3G) inhibits the retroviruses by entering viral particles and inducing hypermutation of viral genome during reverse transcription, leading to G to A substitutions (Bae and Jung, 2014). In concert, the protein hA3F induces viral hypermutation by deaminating minus-strand of viral cDNA during reverse transcription (Bae and Jung, 2014). Taken together, these factors induce epigenetic silencing and hypermutation of HERVs. Indeed, the LTRs have a bigger mutation rate than the rest of non-coding fraction of the human genome (Romano et al., 2006).

Conversely, the content of TFBS among the HERVs decreases with their evolutionary age (Garazha et al., 2015). For the heavily mutated, highly diverged (>20%) HERV elements, this content is approximately six-fold lower compared to the top evolutionary young elements. This observation may suggest that genomic "domestication" of HERVs involved reformatting of the active TFBS profiles and their further "standardization" upon accumulation of mutations, until they get equilibrated with the rest of non-coding DNA (Garazha et al., 2015). However, this type of analysis can be biased by the higher fragmentation in the evolutionary older HERVs, because each fragment is considered as an independent element. Further studies are, therefore, needed to explore the TFBS accumulation trends in linkage to the evolutionary dynamics of the human genome.

Sometimes co-evolution with the human genome resulted in a recruitment of certain HERV regulatory modules by the host organism (Table 1). The best-known example is the acquisition of salivary expression of the carbohydrate digestive enzyme amylase

TABLE 1 | Implication of HERV transcriptional regulation in human physiology and pathology (selected examples).

HERV element	Function	Mechanism	References
NORMAL PHYSIOLOGY			
HERV-E	Expression of amylase genes <i>AMY1A</i> , <i>AMY1B</i> , <i>AMY1C</i> in salivary glands	Creates tissue-specific enhancer	Ting et al., 1992
HERV-K (HML-2)	Expression of proline dehydrogenase gene <i>PRODH</i> in hippocampus	Creates tissue-specific enhancer	Suntsova et al., 2013
HERV-H	Maintaining pluripotency in stem cells	Recruits transcriptional activators by initiating transcription of intergenic RNAs	Ohnuki et al., 2014
HUERS-P1	Maintaining pluripotency in stem cells	Promotes transcription of a non-coding RNA serving as a molecular sponge for miR let-7 microRNAs	Durruthy-Durruthy et al., 2015
MER39	Expression of Prolactin during pregnancy	Creates tissue-specific promoter	Emera et al., 2012
HERV9	Control of fetal and adult expression of globin locus	Recruits transcriptional factors to the downstream Beta-globin promoter	Tuan and Pi, 2014
HERV-K (HML-2)	Control of <i>SLC4A8</i> and <i>IFT172</i> gene expression	Promotes negative regulator antisense RNAs	Gogvadze et al., 2009
HERV-W	Fusion of trophoblast cells in placenta	Encodes protein Syncytin	Frendo et al., 2003
CANCER			
MaLR LTR	Survival of Hodgkin's lymphoma cells by upregulation of <i>CSF1R</i> gene	Creates alternative promoter	Lamprecht et al., 2010
HERV-K (HML-2)	Survival of Chronic Lymphocytic Leukemia cells	Encodes protein NP9 with possible oncogenic functions	Fischer et al., 2014
HERV-W	Tumor growth and metastasis via immunosuppression	Encodes protein Syncytin	Kassiotis, 2014
INFECTIOUS DISEASES			
HERV-L/HERV16	Suppression of Varicella zoster virus and HIV infection	Creates HLA Complex P5 gene	Crosslin et al., 2015
HERV-K	dUTPase activity for HIV life cycle	Encodes endogenous retroviral dUTPase	Mayer and Meese, 2003
AUTOIMMUNE DISEASES			
HERV-E	Possible role in promotion of systemic lupus erythematosus (SLE)	Encodes potentially immunogenic retroviral proteins	Wu et al., 2015
HERV-K10	Possible role in promotion of rheumatoid arthritis	Encodes potentially immunogenic retroviral protein HERV-K10 Gag	Nelson et al., 2014
HERV-W	Possible role in promotion of osteoarthritis	Encodes potentially immunogenic retroviral protein Syncytin	Bendiksen et al., 2014
HERV-K18	Possible role in promotion of osteoarthritis	Encodes potentially immunogenic retroviral proteins	Garcia-Montojo et al., 2013
NEUROLOGICAL DISORDERS			
Multiple HERVs	Possible role in promotion of multiple sclerosis	Encode potentially immunogenic proteins and induce autoimmunoreactivity	Libbey et al., 2014; Manghera et al., 2015
HERV-W	Possible role in promotion of schizophrenia and bipolar disorder	Encodes potentially immunogenic retroviral protein Syncytin	Diern et al., 2012
HERV-K (HML-2)	Possible role in regulation of proline dehydrogenase in schizophrenia	Creates tissue-specific enhancer for gene <i>PRODH</i>	Suntsova et al., 2013
HERV-H	Induction of hypotonia and motor, language, and cognitive delays	Due to recombinations, mediate 3q13.2-q13.31 deletions	Shuvarikov et al., 2013

from a HERV element inserted in the common ancestor of great apes (Ting et al., 1992).

On the other hand, HERV-H is a family expressed preferentially in human embryonal stem cells. Surprisingly, these are the HERV-H LTRs that appeared to be the primary mediators of cell fate reprogramming using famous “Yamanaka cocktail” (by overexpressing OCT3/4, SOX2, and KLF4 proteins), due to regulatory HERV-H-driven intergenic non-coding RNAs that help to recruit the transcriptional activator genes by serving as the scaffold (Ohnuki et al., 2014). Another human long non-coding RNA (human pluripotency-associated transcript 5, *HPAT5*) derived from both a HERV element HUERS-P1 and an Alu retrotransposon, was shown to promote pluripotency by functioning as a molecular sponge for the let-7 family of

microRNAs (Durruthy-Durruthy et al., 2015; Chuong et al., 2017).

The element MER39 forms an endometrium-specific promoter that regulates expression of Prolactin during pregnancy (Emera et al., 2012). The developmental switch from fetal to adult beta-globin gene expression in human is controlled by a copy of HERV9 element (Tuan and Pi, 2014). In hippocampus, transcription of gene *PRODH* is regulated human-specifically by a HERV-K (HML-2) LTR (Suntsova et al., 2013). *PRODH* metabolizes neuromediator molecules and has a strong implication in higher nervous activity and neurological disorders, and its deregulation might have an important impact on human evolution (Suntsova et al., 2013).

HERV-MEDIATED REGULATION OF GENE EXPRESSION IN PATHOLOGY

Proliferative Disorders

Recent findings indicate that HERV-mediated control of gene expression may be involved in various human diseases including cancer (Kassiotis, 2014). The role of HERVs in cancer is most likely limited to regulation of gene expression (Hohn et al., 2013). The data from cancer genome sequencing identified over 180 somatic insertions caused by LINE-1 retrotransposon activity, vs. only a single integration of a short HERV fragment, most likely replicated due to microhomology-mediated DNA repair mechanism (Lee et al., 2012). Many HERVs are abnormally expressed in cancer. For instance, HERV-K (HML-2) elements are up to ~3,000 times overexpressed in germ cell tumors and in melanoma (Buzdin et al., 2006b; Schmitt et al., 2013). Upregulation of HERVs can be mediated by either biased content of the specific transcription factors or by disruption of the anti-retroviral suppression mechanisms, such as aberrant demethylation (Conti et al., 2016) and decreased expression of APOBEC3 proteins (Shepelin et al., 2016). HERVs, in turn, may promote cellular transformation by regulating downstream human genes. For example, a demethylated copy of MaLR LTR can act as an alternative promoter to transcriptionally derepress the gene *CSF1R*, encoding colony stimulating factor-1 receptor, which is linked with survival of the Hodgkin's lymphoma cells (Lamprecht et al., 2010). More examples can be found in the other specific reviews (Babaian and Mager, 2016; Gonzalez-Cao et al., 2016; Anwar et al., 2017).

Infectious Diseases

The evolution of human pathogens might generate mechanisms involving transcriptional interactions of endogenous and exogenous retroviruses. For example, in HIV-infected patients, the HERV-K (HML-2) proviruses are expressed in peripheral blood mononuclear cells at higher levels compared to the non-infected individuals (Bhardwaj et al., 2014). The antibodies against HERV-K (HML-2) Env protein in blood were proposed as the new biomarker of HIV-1 infection, because HIV-1 can upregulate expression of a fully N-glycosylated HERV-K (HML-2) envelope protein on the cell surface (Michaud et al., 2014). Moreover, the HERV-K (HML-2)-specific T-cells from the HIV-1 infected patients *in vitro* completely eliminated the human cells infected with a panel of globally diverse HIV isolates. The mechanism of HIV-1 induced activation of human transposable elements possibly involves the activity of an HIV-1 Tat protein (Jones et al., 2013). Recent studies showed that out of 91 annotated HERV-K (HML-2) proviruses, Tat could activate expression of 26 proviruses, silenced 12, and did not change the expression of the others (Gonzalez-Hernandez et al., 2014). In addition, HIV infection may cause transactivation of HERV-W elements with their Env genes and Syncytin (Uleri et al., 2014). However, a controversial data were reported on the presence of HERV-K (HML-2) viral particles in the plasma of HIV-infected patients—higher levels of HERV-K (HML-2) RNA were detected in the HIV patients from Uganda, but not from the USA (Li et al., 2013). Of note, the recent association

study showed that susceptibility to infection with varicella zoster virus is linked with the non-coding gene HLA Complex P5 in the major histocompatibility complex. This gene is a copy of an endogenous retrovirus that may have a potential to suppress viral activity through indirect regulatory mechanisms. In previous studies, particular genetic variants of this region were associated with delay in development of AIDS in HIV-infected individuals (Crosslin et al., 2015).

Autoimmunity

The biased expression of HERVs is considered as one of the triggers of autoimmune disorders (Suntsova et al., 2015), which is evidenced by increased proviral RNA levels (Ehlhardt et al., 2006) and anti-HERV protein antibodies in sera from several types of patients (Bannert and Kurth, 2004). Immune reactivity against ERV proteins can be experimentally induced in mice and non-human primates, evidencing that immunological tolerance to endogenous retroviral products is not complete (Kassiotis, 2014). The HERV overexpression may be linked with massive DNA hypomethylation as seen for T-cells in systemic lupus erythematosus (SLE) patients (Wu et al., 2015).

Compared to the normal controls, in the patients with rheumatoid arthritis, increased antibody response was detected against the HERV-K10 Gag protein (Nelson et al., 2014). HERV-W transcripts and protein isoforms of Syncytin were overexpressed in cartilage of osteoarthritis patients (Bendiksen et al., 2014). In osteoarthritis, the patient's individual disease severity index was correlated with the expression of HERV-K18 provirus (Garcia-Montojo et al., 2013). However, inflammatory diseases may be also associated with the decreased expression of HERVs (Table 1).

Neurological Diseases

Expression of HERVs may serve as the biomarker for various neurological diseases (Table 1). For example, the HERV expression may be inducible in human astrocytes and neurons under inflammatory conditions in an IFN γ -dependent manner (Manghera et al., 2015). For multiple sclerosis (MS), a hypothesis was proposed that HERV-encoded proteins can act as the powerful immune stimulators inducing disease progression following neurodegeneration (Libbey et al., 2014). Indeed, genetic variants in some genes restricting retroviral infections were statistically linked with the risk of getting MS, as shown for the TRIM5, TRIM22, and BST2 genes (Nexo et al., 2013).

The abnormally high levels of the HERV-W Env gene product were detected in the plasma of the patients with schizophrenia and bipolar disorder (Diem et al., 2012), and in the active lesions in multiple sclerosis (van Horssen et al., 2016) and in the biopsies from the chronic inflammatory demyelinating polyradiculoneuropathies (Faucard et al., 2016). The increased expression of endogenous HERV-K (HML-2) proviral Env gene, in turn, may contribute to the development of amyotrophic lateral sclerosis by inducing neurodegeneration (Li et al., 2015). Finally, HERVs may also cause neurological disorders due to HERV-linked genomic rearrangements (Table 1). The human-specific enhancer activity of a HERV-K (HML-2) provirus on schizophrenia-associated gene *PRODH* may be another

active mechanism of HERV involvement in schizophrenia (Suntsova et al., 2013). Recently, a link was discovered between schizophrenia risk and the complement C4 system (Sekar et al., 2016). The individuals having a polymorphic HERV intronic insertion have elevated C4 expression, which in turn may cause neuronal synapse over-pruning, a phenotype that is associated with schizophrenia. Although this evidence is still indirect, this case is intriguing in light of previous observations of an association between schizophrenia and elevated ERV transcriptional activity (Chuong et al., 2017).

CONCLUSIONS

Taken together, these findings suggest that at the early stages after insertion, the HERV is treated by the host cells as a foreign genetic element, and is likely to be suppressed by the targeted methylation and mutations. However, at the later stages, when significant number of mutations has been already accumulated and when the retroviral genes are broken, the regulatory potential of a HERV may be released and recruited

to modify the genomic balance of transcription factor binding sites. This process goes together with further accumulation and selection of mutations, which reshape the regulatory landscape of the human DNA. However, developmental reprogramming, stress or pathological conditions like cancer, inflammation and infectious diseases, can remove the blocks limiting expression and HERV-mediated host gene regulation. This, in turn, can dramatically alter the gene expression equilibrium and shift it to a newer state, thus further exacerbating the stressful or unstable situation.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct and intellectual contributions to the work, and approved it for publication.

FUNDING

This work was supported by the the Russian Scientific Foundation grant 14-14-00060.

REFERENCES

- Anwar, S. L., Wulaningsih, W., and Lehmann, U. (2017). Transposable elements in human cancer: causes and consequences of deregulation. *Int. J. Mol. Sci.* 18:E974. doi: 10.3390/ijms18050974
- Babaian, A., and Mager, D. L. (2016). Endogenous retroviral promoter exaptation in human cancer. *Mob. DNA* 7:24. doi: 10.1186/s13100-016-0080-x
- Bae, E. H., and Jung, Y.-T. (2014). Comparison of the effects of retroviral restriction factors involved in resistance to porcine endogenous retrovirus. *J. Microbiol. Biotechnol.* 24, 577–583. doi: 10.4014/jmb.1312.12079
- 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, 14572–14579. doi: 10.1073/pnas.0404838101
- Baust, C., Seifarth, W., Germaier, H., Hehlmann, R., and Leib-Mösch, C. (2000). HERV-K-T47D-Related long terminal repeats mediate polyadenylation of cellular transcripts. *Genomics* 66, 98–103. doi: 10.1006/geno.2000.6175
- Belshaw, R., Pereira, V., Katzourakis, A., Talbot, G., Paces, J., Burt, A., et al. (2004). Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4894–4899. doi: 10.1073/pnas.0307800101
- Bendiksen, S., Martinez-Zubiavrra, I., Tummler, C., Knutsen, G., Elvenes, J., Olsen, E., et al. (2014). Human endogenous retrovirus W activity in cartilage of osteoarthritis patients. *Biomed. Res. Int.* 2014:698609. doi: 10.1155/2014/698609
- Bhardwaj, N., Maldarelli, F., Mellors, J., and Coffin, J. M. (2014). 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.* 88, 11108–11120. doi: 10.1128/JVI.01623-14
- Buzdin, A. (2007). Human-specific endogenous retroviruses. *ScientificWorldJournal* 7, 1848–1868. doi: 10.1100/tsw.2007.270
- Buzdin, A., Kovalskaya-Alexandrova, E., Gogvadze, E., and Sverdlov, E. (2006a). At least 50% of human-specific HERV-K (HML-2) long terminal repeats serve *in vivo* as active promoters for host nonrepetitive DNA transcription. *J. Virol.* 80, 10752–10762. doi: 10.1128/JVI.00871-06
- Buzdin, A., Kovalskaya-Alexandrova, E., Gogvadze, E., and Sverdlov, E. (2006b). GREM, a technique for genome-wide isolation and quantitative analysis of promoter active repeats. *Nucleic Acids Res.* 34:e67. doi: 10.1093/nar/gkl335
- Buzdin, A., Ustyugova, S., Khodosevich, K., Mamedov, I., Lebedev, Y., Hunsmann, G., et al. (2003). Human-specific subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active simultaneously during branching of hominoid lineages. *Genomics* 81, 149–156. doi: 10.1016/S0888-7543(02)00027-7
- Campbell, I. M., Gambin, T., Dittwald, P., Beck, C. R., Shuvarikov, A., Hixson, P., et al. (2014). Human endogenous retroviral elements promote genome instability via non-allelic homologous recombination. *BMC Biol.* 12:74. doi: 10.1186/s12915-014-0074-4
- Cho, K., Lee, Y.-K., and Greenhalgh, D. G. (2008). Endogenous retroviruses in systemic response to stress signals. *Shock* 30, 105–116. doi: 10.1097/SHK.0b013e31816a363f
- Chuong, E. B., Elde, N. C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86. doi: 10.1038/nrg.2016.139
- Chuong, E. B., Rumi, M. A. K., Soares, M. J., and Baker, J. C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat. Genet.* 45, 325–329. doi: 10.1038/ng.2553
- Conti, A., Rota, F., Ragni, E., Favero, C., Motta, V., Lazzari, L., et al. (2016). Hydroquinone induces DNA hypomethylation-independent overexpression of retroelements in human leukemia and hematopoietic stem cells. *Biochem. Biophys. Res. Commun.* 474, 691–695. doi: 10.1016/j.bbrc.2016.05.010
- Crosslin, D. R., Carrell, D. S., Burt, A., Kim, D. S., Underwood, J. G., Hanna, D. S., et al. (2015). Genetic variation in the HLA region is associated with susceptibility to herpes zoster. *Genes Immun.* 16, 1–7. doi: 10.1038/gene.2014.51
- Dewannieux, M., and Heidmann, T. (2013). Endogenous retroviruses: acquisition, amplification and taming of genome invaders. *Curr. Opin. Virol.* 3, 646–656. doi: 10.1016/j.coviro.2013.08.005
- Diem, O., Schaffner, M., Seifarth, W., and Leib-Mosch, C. (2012). Influence of antipsychotic drugs on human endogenous retrovirus (HERV) transcription in brain cells. *PLoS ONE* 7:e30054. doi: 10.1371/journal.pone.0030054
- Domansky, A. N., Kopantzev, E. P., Snezhkov, E. V., Lebedev, Y. B., Leib-Mosch, C., and Sverdlov, E. D. (2000). Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. *FEBS Lett.* 472, 191–195. doi: 10.1016/S0014-5793(00)01460-5
- Durruthy-Durruthy, J., Sebastiano, V., Wossidlo, M., Cepeda, D., Cui, J., Grow, E. J., et al. (2015). The primate-specific noncoding RNA HPAT5 regulates pluripotency during human preimplantation development and nuclear reprogramming. *Nat. Genet.* 48, 44–52. doi: 10.1038/ng.3449
- Ehlhardt, S., Seifert, M., Schneider, J., Ojak, A., Zang, K. D., and Mehraein, Y. (2006). Human endogenous retrovirus HERV-K (HML-2) Rec expression and transcriptional activities in normal and rheumatoid arthritis synovia. *J. Rheumatol.* 3, 316–323.
- Emera, D., Casola, C., Lynch, V. J., Wildman, D. E., Agnew, D., and Wagner, G. P. (2012). Convergent evolution of endometrial prolactin expression in primates,

- mice, and elephants through the independent recruitment of transposable elements. *Mol. Biol. Evol.* 29, 239–247. doi: 10.1093/molbev/msr189
- Faucard, R., Madeira, A., Gehin, N., Authier, F. J., Panaite, P. A., and Lesage, C. (2016). Human endogenous retrovirus and neuroinflammation in chronic inflammatory demyelinating polyradiculoneuropathy. *EBioMedicine* 6, 190–198. doi: 10.1016/j.ebiom.2016.03.001
- Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 41, 563–571. doi: 10.1038/ng.368
- Fischer, S., Echeverria, N., Moratorio, G., Landoni, A. I., Dighiero, G., Cristina, J., et al. (2014). Human endogenous retrovirus np9 gene is over expressed in chronic lymphocytic leukemia patients. *Leuk. Res. Rep.* 3, 70–72.
- Frendo, J. L., Olivier, D., Cheynet, V., Blond, J. L., Bouton, O., Vidaud, M., et al. (2003). Direct involvement of HERV-W Env glycoprotein in human trophoblast cell fusion and differentiation. *Mol. Cell. Biol.* 23, 3566–3574. doi: 10.1016/j.lrr.2014.06.005
- Garazha, A., Ivanova, A., Suntsova, M., Malakhova, G., Roumiantsev, S., Zhavoronkov, A., et al. (2015). New bioinformatic tool for quick identification of functionally relevant endogenous retroviral inserts in human genome. *Cell Cycle* 14, 1476–1484. doi: 10.1080/15384101.2015.1022696
- Garcia-Montojo, M., Varade, J., Villafuertes, E., de La Hera, B., Hoyas-Fernandez, J., Dominguez-Mozo, M. I., et al. (2013). Expression of human endogenous retrovirus HERV-K18 is associated with clinical severity in osteoarthritis patients. *Scand. J. Rheumatol.* 42, 498–504. doi: 10.3109/03009742.2013.779021
- Gogvadze, E., Stukacheva, E., Buzdin, A., and Sverdlov, E. (2009). Human-specific modulation of transcriptional activity provided by endogenous retroviral insertions. *J. Virol.* 83, 6098–6105. doi: 10.1128/JVI.00123-09
- Gonzalez-Cao, M., Iduma, P., Karachaliou, N., Santarpia, M., Blanco, J., and Rosell, R. (2016). Human endogenous retroviruses and cancer. *Cancer Biol. Med.* 13, 483–488. doi: 10.20892/j.issn.2095-3941.2016.0080
- Gonzalez-Hernandez, M. J., Cavalcoti, J. D., Sartor, M. A., Contreras-Galindo, R., Meng, F., Dai, M., et al. (2014). Regulation of the human endogenous retrovirus K (HML-2) transcriptome by the HIV-1 Tat protein. *J. Virol.* 88, 8924–8935. doi: 10.1128/JVI.00556-14
- Ho, B., Baker, P. M., Singh, S., Shih, S.-J., and Vaughan, A. T. (2012). Localized DNA cleavage secondary to genotoxic exposure adjacent to an Alu inverted repeat. *Genes Chromosomes Cancer* 51, 501–509. doi: 10.1002/gcc.21938
- Hohn, O., Hanke, K., and Bannert, N. (2013). HERV-K (HML-2), the best preserved family of HERVs: endogenization, expression, and implications in health and disease. *Front. Oncol.* 3:246. doi: 10.3389/fonc.2013.00246
- Hughes, J. F., and Coffin, J. M. (2004). Human endogenous retrovirus K solo-LTR formation and insertional polymorphisms: implications for human and viral evolution. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1668–1672. doi: 10.1073/pnas.0307885100
- Jones, R. B., Song, H., Xu, Y., Garrison, K. E., Buzdin, A. A., Anwar, N., et al. (2013). LINE-1 retrotransposable element DNA accumulates in HIV-1-infected cells. *J. Virol.* 87, 13307–13320. doi: 10.1128/JVI.02257-13
- Kassiotis, G. (2014). Endogenous retroviruses and the development of cancer. *J. Immunol.* 192, 1343–1349. doi: 10.4049/jimmunol.1302972
- Kjellman, C., Sjögren, H. O., Salford, L. G., and Widegren, B. (1999). HERV-F (XA34) is a full-length human endogenous retrovirus expressed in placental and fetal tissues. *Gene* 239, 99–107. doi: 10.1016/S0378-1119(99)00372-8
- Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., et al. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. *Nat. Med.* 16, 571–579. doi: 10.1038/nm.2129
- Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971. doi: 10.1126/science.1222077
- Li, L., Deng, X., Linsuwanon, P., Bangsberg, D., Bwana, M. B., Hunt, P., et al. (2013). AIDS alters the commensal plasma virome. *J. Virol.* 87, 10912–10915. doi: 10.1128/JVI.01839-13
- 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
- Libbey, J. E., Cusick, M. F., and Fujinami, R. S. (2014). Role of pathogens in multiple sclerosis. *Int. Rev. Immunol.* 33, 266–283. doi: 10.3109/08830185.2013.823422
- Maliniemi, P., Vincendeau, M., Mayer, J., Frank, O., Hahtola, S., Karenko, L., et al. (2013). Expression of human endogenous retrovirus-w including syncytin-1 in cutaneous T-cell lymphoma. *PLoS ONE* 8:e76281. doi: 10.1371/journal.pone.0076281
- 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., and Meese, E. U. (2003). Presence of dUTPase in the various human endogenous retrovirus K (HERV-K) families. *J. Mol. Evol.* 57, 642–649. doi: 10.1007/s00239-003-2514-6
- 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
- Nelson, P. N., Roden, D., Nevill, A., Freimanis, G. L., Trela, M., Ejtehadi, H. D., et al. (2014). Rheumatoid arthritis is associated with IgG antibodies to human endogenous retrovirus gag matrix: a potential pathogenic mechanism of disease? *J. Rheumatol.* 41, 1952–1960. doi: 10.3899/jrheum.130502
- Nexo, B. A., Hansen, B., Nissen, K. K., Gundestrup, L., Terkelsen, T., Villesen, P., et al. (2013). Restriction genes for retroviruses influence the risk of multiple sclerosis. *PLoS ONE* 8:e74063. doi: 10.1371/journal.pone.0074063
- Ohnuki, M., Tanabe, K., Sutou, K., Teramoto, I., Sawamura, Y., Narita, M., et al. (2014). Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12426–12431. doi: 10.1073/pnas.1413299111
- Rebollo, R., Romanish, M. T., and Mager, D. L. (2012). Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu. Rev. Genet.* 46, 21–42. doi: 10.1146/annurev-genet-110711-155621
- Romano, C. M., Ramalho, R. F., and Zanotto, P. M., de A. (2006). Tempo and mode of ERV-K evolution in human and chimpanzee genomes. *Arch. Virol.* 151, 2215–2228. doi: 10.1007/s00705-006-0792-1
- 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
- Schlesinger, S., Lee, A. H., Wang, G. Z., Green, L., and Goff, S. P. (2013). Proviral silencing in embryonic cells is regulated by Yin Yang 1. *Cell Rep.* 4, 50–58. doi: 10.1016/j.celrep.2013.06.003
- Schmitt, K., Reichrath, J., Roesch, A., Meese, E., and Mayer, J. (2013). Transcriptional profiling of human endogenous retrovirus group HERV-K (HML-2) loci in melanoma. *Genome Biol. Evol.* 5, 307–328. doi: 10.1093/gbe/evt010
- Schumann, G. G., Gogvadze, E. V., Osanai-Futahashi, M., Kuroki, A., Münk, C., Fujiwara, H., et al. (2010). Unique functions of repetitive transcriptomes. *Int. Rev. Cell Mol. Biol.* 115–188. doi: 10.1016/B978-0-12-381047-2.00003-7
- Sekar, A., Bialas, A. R., de Rivera, H., Davis, A., Hammond, T. R., Kamitaki, N., et al. (2016). Schizophrenia risk from complex variation of complement component 4. *Nature* 530, 177–183. doi: 10.1038/nature16549
- Shepelin, D., Korzinkin, M., Vanyushina, A., Aliper, A., Borisov, N., Vasilov, R., et al. (2016). Molecular pathway activation features linked with transition from normal skin to primary and metastatic melanomas in human. *Oncotarget* 7, 656–670. doi: 10.18632/oncotarget.6394
- Shuvarikov, A., Campbell, I. M., Dittwald, P., Neill, N. J., Bialer, M. G., Moore, C., et al. (2013). Recurrent HERV-H-mediated 3q13.2–q13.31 deletions cause a syndrome of hypotonia and motor, language, and cognitive delays. *Hum. Mutat.* 34, 1415–1423. doi: 10.1002/humu.22384
- Suntsova, M., Garazha, A., Ivanova, A., Kaminsky, D., Zhavoronkov, A., and Buzdin, A. (2015). Molecular functions of human endogenous retroviruses in health and disease. *Cell. Mol. Life Sci.* 72, 3653–3675. doi: 10.1007/s00018-015-1947-6
- Suntsova, M., Gogvadze, E. V., Salozhin, S., Gaifullin, N., Eroshkin, F., Dmitriev, S. E., et al. (2013). Human-specific endogenous retroviral insert serves as an enhancer for the schizophrenia-linked gene PRODH. *Proc. Natl. Acad. Sci. U.S.A.* 110, 19472–19477. doi: 10.1073/pnas.1318172110

- Sverdlov, E. D. (2000). Retroviruses and primate evolution. *Bioessays* 22, 161–171. doi: 10.1002/(SICI)1521-1878(200002)22:2<161::AID-BIES7>3.0.CO;2-X
- Ting, C. N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C., and Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes Dev.* 6, 1457–1465. doi: 10.1101/gad.6.8.1457
- Trombetta, B., Fantini, G., D'Atanasio, E., Sellitto, D., and Cruciani, F. (2016). Evidence of extensive non-allelic gene conversion among LTR elements in the human genome. *Sci. Rep.* 6:28710. doi: 10.1038/srep28710
- Trubetskoy, D. O., Zavalova, L. L., Akopov, S. B., and Nikolaev, L. G. (2002). Purification of proteins specifically binding human endogenous retrovirus K long terminal repeat by affinity elution chromatography. *J. Chromatogr. A* 976, 95–101. doi: 10.1016/S0021-9673(02)01236-0
- Tuan, D., and Pi, W. (2014). In human β -globin gene locus, ERV-9 LTR retrotransposon interacts with and activates β -but not γ -globin gene. *Blood*. 124:2686.
- Turelli, P., Castro-Diaz, N., Marzetta, F., Kapopoulou, A., Raclot, C., Duc, J., et al. (2014). Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. *Genome Res.* 24, 1260–1270. doi: 10.1101/gr.172833.114
- Uleri, E., Mei, A., Mameli, G., Poddighe, L., Serra, C., and Dolei, A. (2014). HIV Tat acts on endogenous retroviruses of the W family and this occurs via Toll-like receptor 4: inference for neuroAIDS. *AIDS* 28, 2659–2670. doi: 10.1097/QAD.0000000000000477
- van Horssen, J., van der Pol, S., Nijland, P., Amor, S., and Perron, H. (2016). Human endogenous retrovirus W in brain lesions: rationale for targeted therapy in multiple sclerosis. *Mult. Scler. Relat. Disord.* 8, 11–18. doi: 10.1016/j.msard.2016.04.006
- Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., et al. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18613–18618. doi: 10.1073/pnas.0703637104
- Wu, Z., Mei, X., Zhao, D., Sun, Y., Song, J., Pan, W., et al. (2015). DNA methylation modulates HERV-E expression in CD4+ T cells from systemic lupus erythematosus patients. *J. Dermatol. Sci.* 77, 110–116. doi: 10.1016/j.jdermsci.2014.12.004
- Young, J. M., Whiddon, J. L., Yao, Z., Kasinathan, B., Snider, L., Geng, L. N., et al. (2013). DUX4 binding to retroelements creates promoters that are active in FSHD muscle and testis. *PLoS Genet.* 9:e1003947. doi: 10.1371/journal.pgen.1003947

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 © 2017 Buzdin, Prassolov and Garazha. 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) or licensor 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



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

