



Article

HERV-Derived *Ervpb1* Is Conserved in Simiiformes, Exhibiting Expression in Hematopoietic Cell Lineages Including Macrophages

Ayumi Matsuzawa ¹, Jiyong Lee ¹, So Nakagawa ² , Johbu Itoh ³, Mahoko Takahashi Ueda ⁴,
Satomi Mitsuhashi ⁴, Yuta Kochi ⁴, Tomoko Kaneko-Ishino ^{5,*} and Fumitoshi Ishino ^{1,*}

- ¹ Department of Epigenetics, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo 113-8510, Japan; amatepgn@tmd.ac.jp (A.M.); jlee.epgn@mri.tmd.ac.jp (J.L.)
² Department of Molecular Life Science, Tokai University School of Medicine, 143 Shimokasuya, Isehara 259-1193, Kanagawa, Japan; so@tokai.ac.jp
³ Department of Pathology, School of Medicine, Tokai University, 143 Shimokasuya, Isehara 259-1193, Kanagawa, Japan; itohj@is.icc.u-tokai.ac.jp
⁴ Department of Genomic Function and Diversity, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo 113-8510, Japan; mahoko.gfd@mri.tmd.ac.jp (M.T.U.); satomits.gfd@mri.tmd.ac.jp (S.M.); y-kochi.gfd@mri.tmd.ac.jp (Y.K.)
⁵ Department of Nursing, School of Medicine, Tokai University, 143 Shimokasuya, Isehara 259-1193, Kanagawa, Japan
* Correspondence: tkanekoi@is.icc.u-tokai.ac.jp (T.K.-I.); fishino.epgn@mri.tmd.ac.jp (F.I.)



Citation: Matsuzawa, A.; Lee, J.; Nakagawa, S.; Itoh, J.; Takahashi Ueda, M.; Mitsuhashi, S.; Kochi, Y.; Kaneko-Ishino, T.; Ishino, F. HERV-Derived *Ervpb1* Is Conserved in Simiiformes, Exhibiting Expression in Hematopoietic Cell Lineages Including Macrophages. *Int. J. Mol. Sci.* **2021**, *22*, 4504. <https://doi.org/10.3390/ijms22094504>

Academic Editor: Hee-Jae Cha

Received: 28 February 2021

Accepted: 23 April 2021

Published: 26 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: (1) Background: The *ERVpb1* gene in humans is derived from an envelope (*Env*) gene of a human endogenous retrovirus group, HERV-P(b). The *ERVpb1* gene reportedly has a conserved open reading frame (ORF) in Old World monkeys. Although its forced expression led to cell-fusion in an ex vivo cell culture system, like other *Env*-derived genes such as *syncytin-1* and *-2*, its mRNA expression is not placenta-specific, but almost ubiquitous, albeit being quite low in human tissues and organs, implying a distinct role for *ERVpb1*. (2) Methods: To elucidate the cell lineage(s) in which the *ERVpb1* protein is translated in human development, we developed a novel, highly sensitive system for detecting HERV-derived proteins/peptides expressed in the tissue differentiation process of human induced pluripotent stem cells (iPSCs). (3) Results: We first determined that *ERVpb1* is also conserved in New World monkeys. Then, we showed that the *ERVpb1* protein is translated from a uniquely spliced *ERVpb1* transcript in hematopoietic cell lineages, including a subset of macrophages, and further showed that its mRNA expression is upregulated by lipopolysaccharide (LPS) stimulation in primary human monocytes. (4) Conclusions: *ERVpb1* is unique to Simiiformes and actually translated in hematopoietic cell lineages, including a subset of macrophages.

Keywords: an *Env*-derived *ERVpb1* gene; a simiiforme-specific gene; retrovirus-derived genes (*RVDs*); human genome; hemangioblast cells; hematopoietic cells; macrophages; human iPSCs; organoid generation

1. Introduction

Approximately 8% of the human genome is comprised of long terminal repeat (LTR) retrotransposons/human endogenous retroviruses (HERVs) [1,2]. Although most of these have lost their original activity due to mutations, there still remain more than 10,000 HERVs that have open reading frames (ORFs) comprising >100 amino acids (aa) related to GAG-, POL-, and ENV-derived proteins/peptides [3,4]. Thus, these HERVs are potentially a large source of as yet unidentified retrovirus-derived (*RVD*) genes in the human genome, although it remains to be determined whether proteins/peptides are actually translated from such ORFs and thus play a role in human development.

For example, more than ten *RVD* genes have been identified so far in the human genome, such as paternally expressed 10 (*PEG10*) and eight other retrotransposon Gag-like

genes (*RTL*) (also called sushi-ichi-related retrotransposon homologues (*SIRHs*)) [5–7]. These genes play essential roles in the placenta and/or brain. There are also three *Env*-derived genes, *syncytin-1* and *-2* and *suppressyn*, that play essential roles in the placenta: two *syncytins* that play a role in the formation of the syncytial layer by inducing the fusion of syncytiotrophoblast cells, and *suppressyn*, which has a role in the inhibition of the activity of *syncytin-1* but not *syncytin-2* [8–10]. The latter *Env*-derived genes were captured (or domesticated) from three different HERVs, HERV-W, -FRD, and -Fb1, approximately 25, >40, and 25 million years ago (MYA), respectively. Therefore, *syncytin-1* and *suppressyn* are Catarrhini (Old World monkey)-specific, while *syncytin-2* is a Simiiforme-specific gene [9,11]. In contrast, the former genes encode GAG- (and sometimes also POL-) like proteins exhibiting the highest homology to the LTR sushi-ichi retrotransposon (each at most 20~30% homology). However, it is reasonable to think that these genes were derived from extinct retrovirus(es), because they were acquired (or domesticated) in the therian or eutherian ancestors approximately 168 and 148 MYA, respectively [12]. These data suggest that a number of as yet unidentified *RVD* genes may have been accumulated in a step-by-step manner in the human genome as mammalian-specific as well as primate-specific genes.

Among the *Env*-derived genes, *ERVpb1* (also called *ERVp(b)*) is reported to have an ORF comprising 665 amino acids in humans and retains several features typical of ENV: it is comprised of surface (SU) and transmembrane (TM) subunits, with a furin cleavage site (RXXR) between them [13,14]. In addition, there are two disulfide-bond isomerase motifs, Cys-X-X-Cys (CXXC) in SU and a CX6CC motif in TM, a putative receptor binding domain (RBD) in SU, an immunosuppressive domain (ISD) and two heptad repeats (HR1 and 2) in TM, all of which are well conserved. Importantly, *ERVpb1* is reported to be widely conserved in Catarrhini species, suggesting that it confers some evolutionary advantage. It is also reported that forced expression of the putative *ERVpb1* gene induces cell fusion in several types of cultured cells like *syncytin-1* and *-2*, even though it lacks a fusion peptide in TM [13,14]. *Syncytin-1* and *-2* exhibit placenta-specific expression, but *ERVpb1* mRNA expression is ubiquitous if also quite low in many tissues and organs, approximately 1/100 the level of *syncytin-1*, even in the placenta [13]. Therefore, it remains uncertain whether the *ERVpb1* protein is actually translated during human development and, if so, in which lineage of cells it is functional.

Recently, various in vitro differentiation methods using mouse and human induced pluripotent stem cells (iPSCs) and/or embryonic stem cells have been developed for the generation of several different tissue and organ types, including the liver, brain, kidney, and heart as well as yolk sac-derived erythroid and myeloid cells as well as macrophages [15–18]. Such human organoid generation systems provide excellent models for mimicking human development and/or tissue differentiation, and are thus essential tools for elucidating the biological functions of primate- or human-specific *RVD* genes by means of gene targeting instead of having to employ knockout mice, because these genes are not present in mice. In this paper, we report a novel system for the highly sensitive detection of HERV-derived proteins/peptides in the process of tissue differentiation from human iPSCs. We obtained a human iPSC line with an insertion of a fluorescent Venus protein just after the N-terminal region of the ORF that was examined using CRISPR/Cas9 genome editing, and performed in vitro differentiation according to tissue organoid generation protocols. Using this system, we detected the *ERVpb1*-Venus fusion protein under a condition of macrophage differentiation from iPSCs, demonstrating that the *ERVpb1* protein is actually translated in a hematopoietic cell lineage, including translation of a subset of macrophages.

2. Results

2.1. *ERVpb1* Is Unique to Simiiformes and Exhibits a Quite Low Level of Expression in Limited Tissues and Organs in Humans

Previous reports indicated that *ERVpb1* is unique to Catarrhini [13,14], but its presence in the owl monkey was suggested by PCR using locus-specific intra-envelop primers [13].

In this study, we found that the orthologous sequence is also conserved in New World monkeys (Platyrrhini), such as marmosets, capuchins, and owl monkeys, but is absent from the Philippine tarsier, a Tarsiiforme species, and the greater bamboo lemur, a Strepsirrhine species, demonstrating that *ERVpb1* is conserved in both Catarrhini and Platyrrhini, so it is unique to Simiiformes (Figure 1a,b). *ERVpb1* is located between exons 4 and 5 (intron 4) of the *Ras* and *Rab* interactor 3 (*RIN3*) gene except in two cases: in gibbons between exons 5 and 6 due to a split of exon 4 (exons 4 and 5) and in African green monkeys (AGMs) between exons 3 and 4 due to a deletion of exon 1. This location is also conserved in the Platyrrhini species and its ORF, comprising 666 aa (660 aa due to a 6 aa deletion in marmosets), has been conserved in all of the ENV-like motifs and domains in the Catarrhine *ERVpb1* orthologues (Figure 1b). Thus, it is probable that the domestication of *ERVpb1* from the original HERV10B3-like retrovirus would have occurred before the Platyrrhini-Catarrhini split 44.2 MYA [19].

It is also reported that *ERVpb1* mRNA expression is ubiquitous and quite low in the tissues and organs in which it is found [13,14], approximately 1/100 the level of *syncytin-1*, even in the placenta [13]. We also examined mRNA expression by RT-PCR using a commercially available human multiple tissue RNA panel and confirmed that the mRNA level was quite low in human tissues and organs: 35 cycles of the RT-PCR reaction were required for detection in the thymus, lung, placenta, fetal liver, and bone marrow, with no expression detected in other tissues and organs (Figure 1c). These results were supported by the short read RNA-seq data of 26 human tissue and organ samples registered in the SRA database (SRA, <https://www.ncbi.nlm.nih.gov/sra> (accessed on 21 January 2012–22 March 2019)). The values of the reads per million (RPM) mapped reads were very low (Table S1). There were a few brain samples (including the cerebrum, cerebellum, and hippocampus) and leukocyte cell samples (including B and T cells as well as macrophages) that exhibited > 0.1, but the placenta, heart, and most of the samples from the above tissues and cell types were nearly 0 (<0.1); the highest and second highest values were 0.76 and 0.33 from one hippocampus sample and one macrophage, respectively.

2.2. Generation of Human iPSCs with an *ERVpb1*–Venus Fusion Gene in Its Endogenous Locus

The above data indicate that the expression level of the putative *ERVpb1* protein is quite low, even in the tissues and organs in which its mRNA expression was confirmed. Therefore, to determine the role of this protein in human development, we developed a novel detection system by combining a highly sensitive method for detecting target proteins by means of a protein fused with a fluorescent Venus protein (Irie et al. submitted) together with in vitro organoid formation using human iPSCs. We introduced the *Venus* gene into the endogenous *ERVpb1* in human iPSCs (HPS0063 201B7) using the CRISPR/Cas9 genome editing process so that the Venus protein was translated just 48 amino acids after the N-terminal of the putative *ERVpb1* protein (Figure 2a). Therefore, it is expected that the *ERVpb1*–Venus fusion protein be translated by the same regulatory mechanism as the endogenous *ERVpb1* gene and thus detected in the tissues and organs with a high degree of sensitivity.

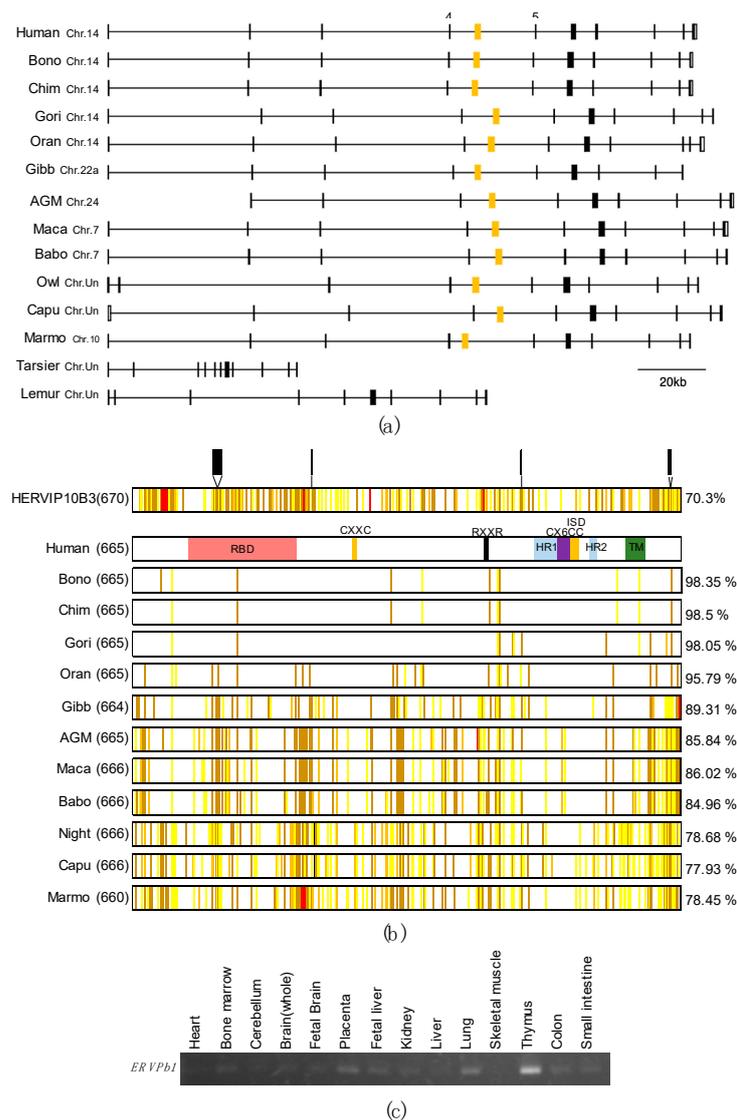


Figure 1. *ERVPb1* is conserved in Simiiformes and its mRNA expression is observed in human tissues. **(a)** Location of the conserved *ERVPb1* between *RIN3* exons 4 and 5 in Simiiformes including three Platyrrhini species. The orange and black colored boxes represent the *ERVPb1* locus and *RIN3* exons, respectively. Bono: Bonobo, Chim: Chimpanzee, Gori: Gorilla, Oran: Orangutan, Gibb: Gibbon, AGM: African green monkey, Maca: Macaca, Babo: Baboon, Owl: Ma's night monkey, Capu: Capuchin, Marmo: Marmoset, Tarsier: Philippine tarsier, Lemur: Greater bamboo lemur. **(b)** The putative *ERVPb1* ORFs found in Simiiformes and HERVIP10B3 ENV (upper line). The numbers in the parentheses and those on the right side indicate the number of amino acids and % identity to the human ORF, respectively. The black and red bars indicate insertion and deletion, respectively. The colored bars indicate conservation between groups of strongly similar (roughly equivalent to scoring > 0.5 in the Gonnet PAM 250 matrix: yellow) or weakly similar (roughly equivalent to scoring < 0.5 and > 0 in the Gonnet PAM 250 matrix: orange) properties, or that they are not similar (brown) to the human ORF. The colored boxes in the human ORF indicate *ERVPb1* motifs and domains, such as CXXC and CX6CC: disulfide-bond isomerase motifs, HR1 and HR2: heptad repeats, ISD: immunosuppressive domain, RBD: receptor binding domain, RXXR: furin cleavage site, and TM: transmembrane domain. **(c)** *ERVPb1* transcripts detected by RT-PCR (270 bp) in human tissues. Thirty-five cycles of the PCR reaction were performed.

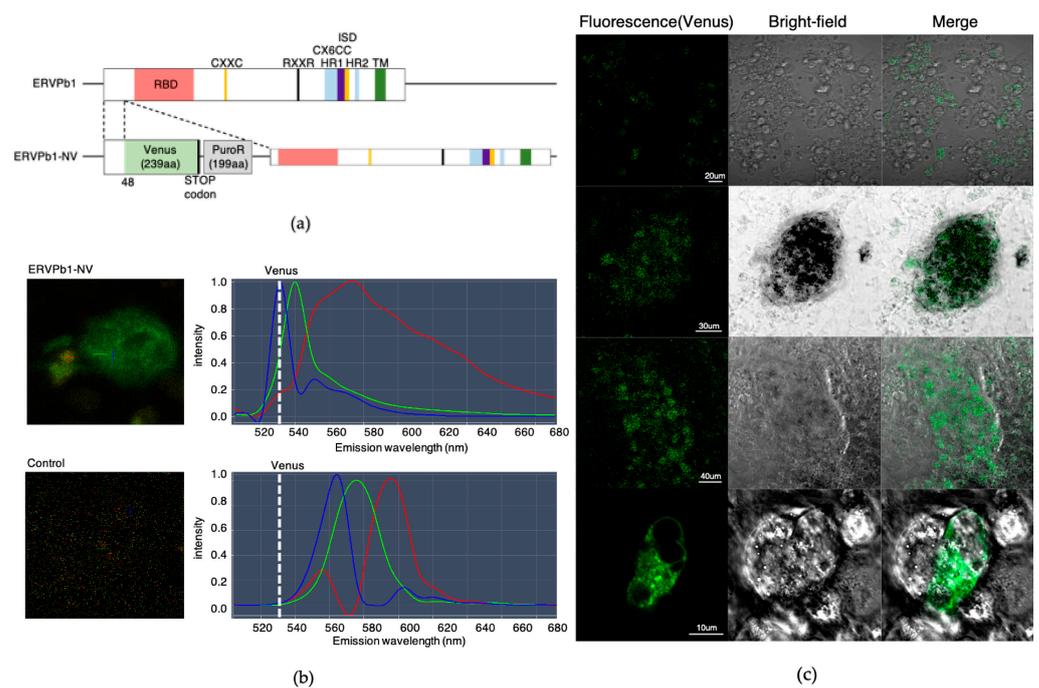


Figure 2. ERVPb1–Venus fusion protein detected under a condition of iPSC differentiation in the three germinal layers. (a) Structure of the ERVPb1–NV knock-in construct. The Venus gene was introduced into endogenous ERVPb1 so that the Venus protein was fused with ERVPb1 just after the 48 amino acids of its N-terminus (ERVPb1–Venus fusion protein). Most of the ENV features of this putative protein were removed in this construct except for the N-terminal signal peptide sequence. (b) Peak separation between the Venus-fusion protein (maximum peak emission fluorescence wavelength: 528 nm) and autofluorescence in human iPSC cells with or without Venus insertion. The white dotted line indicates the wavelength of Venus. The ERVPb1–Venus fusion protein was detected only in the ERVPb1–NV iPSCs, not in the control iPSCs. (c) Fluorescence (Venus), Bright-field, and the merged images of ERVPb1–NV iPSCs differentiated into the three germ layers.

2.3. Detection of the ERVPb1–Venus Fusion Protein in a Hematopoietic Cell Lineage

The Venus knock-in human iPSCs were differentiated to see whether the putative ERVPb1 ORF is actually translated during the *in vitro* differentiation process of human iPSC cells. First, we tested whether the Venus signal appeared in any of the cell types among the three germ layers differentiated from the embryonic bodies (EBs) of the Venus knock-in human iPSCs. We generated EBs in Cellartis DEF-CS 500 Culture System medium under feeder-free conditions for seven days, after which the EBs were differentiated into the three ectoderm, endoderm, and mesoderm germ cell lineages according to the Cellartis DEF-CS 500 Culture System *in vitro* differentiation protocol. After four days of differentiation, we detected the fluorescent signals emitted from the Venus fusion proteins. A strong Venus signal (with a maximal emission peak at 528 nm) was detected from one flattened cell as well as weak signals in a few other cell groups in a 3 cm diameter dish (Figure 2b,c). Although there were many other signals, they had peaks of 555, 565, 575, and 595 nm and were derived from the autofluorescence of the differentiated normal human iPSC cells without any Venus insertion (Figure 2b). Therefore, it is critically necessary to distinguish the Venus-derived signal from the other autofluorescence signals. For this purpose, it is essential to use confocal fluorescence microscopy equipped with the function of Multi-channel Unmixing or Automatic Composition Extraction (ACE) (Figure 2b). From the cell morphology of the one flattened cell, it was determined to be a kind of macrophage, along with groups of cells that were hematopoietic cell precursors, such as hemangioblast cells.

We then differentiated the Venus knock-in human iPSCs according to the macrophage differentiation protocol using Cellartis DEF-CS 500 Culture System medium containing 80 ng/mL BMP4 without bFGF. Under this condition, many Venus-positive cells were

induced after four days of culture (Figure 3a). We were also able to detect the Venus mRNA by RT-PCR, while it was hardly detected under the first differentiation condition (Figure 3b). As a control, human iPSCs without the Venus insertion were then differentiated using this protocol, and we were able to detect the *ERVpb1* mRNA by real-time PCR. The *ERVpb1* mRNA was detected in the period of days 3–5, with the peak on day 4 (Figure 3c,d), like the Venus signals from the Venus knock-in iPSCs (Figure 3b). Under this condition, the mesoderm marker *TBXT* was induced from day 1 of the differentiation process and peaked on day 2. The hemangioblast marker *CD235a* was detected on days 2 and 3 (Figure 3d), so the differentiation via the mesoderm into primitive hemangioblast cells occurred on these three days of differentiation. *ERVpb1* mRNA expression peaked on day 4 when the macrophage-like cells appeared in the differentiated Venus knock-in iPSC culture. Its expression was subsequently reduced, indicating that *ERVpb1* mRNA expression was temporary, that is, limited to the early stage of hematopoietic cell differentiation. These results clearly demonstrate that *ERVpb1* is a real, protein coding gene that manifests in hematopoietic cell lineages, including a subpopulation of macrophages, in human development. When we carried out the RT-PCR experiment using three primer sets, one designed within the *ERVpb1* ORF and the other designed 28, 551 or 1487 bp upstream of the ORF, there was a band detected using the first primer set but not the last two primer sets, suggesting there is no full length mRNA (Supplementary Figure S1). We then carried out the 5'-RACE experiment and found that the *ERVpb1* mRNA is transcribed from an unexpected promoter located approximately 2.5 kb upstream from the ORF, but not from the presumable 5' LTR approximately 3.0 kb upstream, and then spliced to the 172 bp upstream of the first ATG codon of the *ERVpb1* protein (Figure 3e).

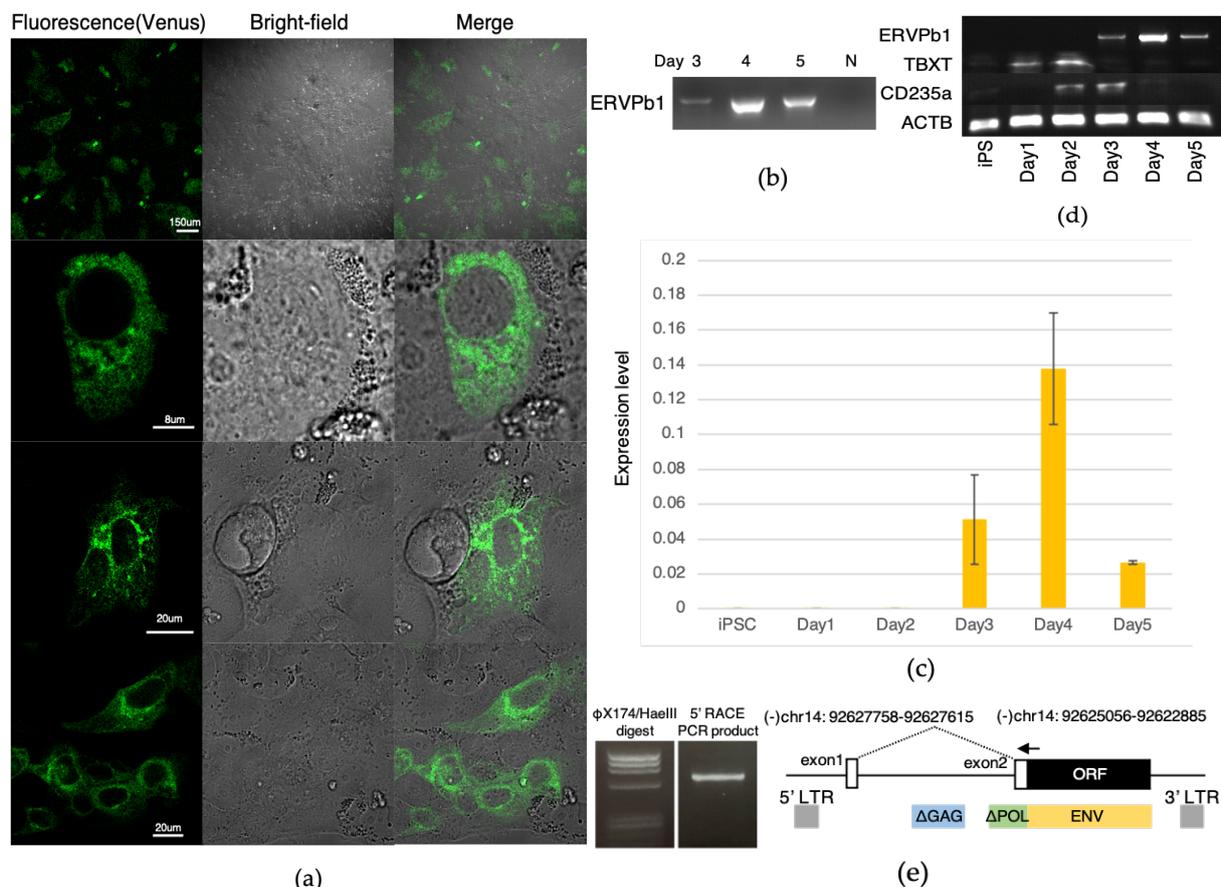


Figure 3. ERVPb1–Venus fusion protein detected under an iPSC differentiation condition induced in macrophages cells. (a) ERVP1-NV iPSCs were differentiated under a condition for macrophage differentiation containing BMP4 without bFGF in day 4 cultures. (b) Analysis of Venus expression by RT-PCR. (c) Detection of *ERVpb1* by real-time PCR in normal human iPSCs in the differentiation medium containing BMP4 on days 1–5. The values were normalized to β -actin expression and

are shown in a bar graph. Each PCR was run in triplicate. (d) Detection of *TBXT* (a mesoderm marker) and *CD235a* (a hemangioblast marker) in addition to *ERVpb1* in normal human iPSC differentiation, the same as (c). (e) A spliced form of *ERVpb1* from an unusual promoter was detected in the differentiated human iPSCs using 5'-RACE. The arrow indicates PCR primer using 5'-RACE. The PCR product (535 bp) contained the known 219 bp sequence in the *ERVpb1* ORF and upstream 316 bp sequences, which included the 144 bp exon 1 sequence approximately 2.5 kb upstream and a 172 bp sequence upstream of the first ATG in exon 2.

Finally, we analyzed *ERVpb1* expression in primary human monocytes using the short-read RNA-seq data of 183 LPS-treated and 200 control samples registered in the European Genome-Phenome Archive (EGA: accession code EGA: EGAS00001001895) [20]. We mapped the single-end short-read RNA-seq data on the *ERVpb1* locus (Supplementary Figure S2) and counted the data at normalized level. Interestingly, the *ERVpb1* expression significantly increased from 12.5099 to 16.2717104 reads with lipopolysaccharide (LPS) stimulation (Games-Howell test, significant, $p = 8.71 \times 10^{-8}$) (Figure 4). These data clearly demonstrate that *ERVpb1* is upregulated by LPS administration in primary human monocytes.

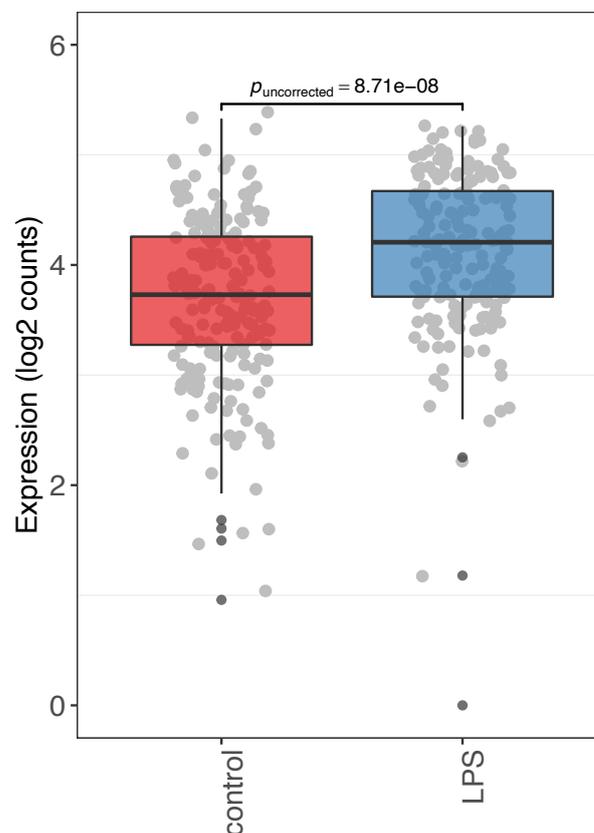


Figure 4. Increment of *ERVpb1* mRNA expression in primary monocytes by LPS administration. Comparison of RNA-Seq data from healthy primary and LPS-treated monocytes is presented.

ERVpb1 expression in primary human monocytes using short-read RNA-seq data of 183 LPS-treated and 200 control samples registered in the European Genome-Phenome Archive (EGA: accession code EGA: EGAS00001001895). Plots were constructed using the values of $\log_2(\text{count} + 1)$. The p -value was calculated by two-sided Games-Howell test, which is parametric, assuming non-equal variance, and based on Welch's statistics.

3. Discussion

It was demonstrated in an in vitro differentiation system using human iPSCs that *ERVpb1* is a bona fide protein-coding gene that is expressed in human hematopoietic cell

lineages, including macrophages, and also that its mRNA is induced by LPS stimulation in primary human monocytes. Importantly, there is no full length mRNA, and only a spliced form of the *ERVpb1* mRNA is expressed in differentiated human iPSCs. Interestingly, it is not a typical spliced transcript from 5' LTR as reported in the cases of the usual *ENV* and *ERVWE/syncytin-1* transcripts in seminomas [21,22], suggesting that it has a unique promoter that resides between the 5' LTR and degenerated GAG sequence (Figure 3e). Thus, further analysis of the features of this unique promoter will be required in the future. It should be noted that the hemangioblast and hematopoietic cells induced in the in vitro human iPSC differentiation experiments have yolk sac-derived properties (Figures 2 and 3). Thus, it is highly probable that *ERVpb1* is expressed in a subpopulation of both yolk-sac- and bone-marrow-derived hematopoietic cells, including macrophages. This explains why *ERVpb1* exhibits such a wide expression profile range in human tissues and organs. Yolk sac-derived macrophages become tissue-resident macrophages in certain tissues and organs, such as the lung, liver, kidney, epidermis, and brain (where they are known as microglia), and it is also possible that *ERVpb1* is expressed in other lymphoid cell populations, such as B and T cells, based on the RNA-seq analysis data shown in Table S1. Under normal conditions, *ERVpb1* expression is maintained at a low level, however, and importantly, its expression is induced by LPS stimulation in primary monocytes, and possibly by the administration of various interferons in other lymphoid cell lineages. Nevertheless, it remains elusive whether the cell fusion activity induced by its forced expression in a variety of cell lines is the actual function of the *ERVpb1* in the course of normal human development [23]. The *ERVpb1*-Venus fusion protein lacks all of the other *ENV*-like motifs, but retains the N-terminal signal peptide motif, so the Venus signals were clearly detected in the Golgi body, a multi-layered structure in the proximity of the nucleus, in macrophages (Figure 3a). We expect that another type of fusion protein with a Venus protein in the C-terminus of *ERVpb1* would prove useful for determining the precise cellular location of *ERVpb1* in the macrophage cells, and would also be helpful in establishing the actual function of the *ERVpb1* protein.

The presence of *ERVpb1* in Platyrrhini species as well as its absence from the Tarsiiforme as well as Strepsirrhini species demonstrates that its domestication from the original HERVIP10B3-like retrovirus must have occurred after the Tarsiiforme-Simiiforme split 71.1 MYA and before the Platyrrhini-Catarrhini split 44.2 MYA [19]. Its intact ORF has been conserved in the Simiiforme species, suggesting that it confers some evolutionary advantage. Human *syncytin-2* is also conserved in Simiiformes, but is derived from a different HERV (HERV-FRD) [9]. Thus, while the number of Simiiforme-specific *RVDs* is currently two, it is possible that more Simiiforme-specific *RVDs* may be found in the human genome if the HERV-derived sequences are searched with an ORF of less than 500 aa. Similarly, there may be more Catarrhini-specific genes in the human genome in addition to *syncytin-1* and *suppressyn*.

A series of KO mouse studies of mammalian-specific *RVD* genes, including *Peg10*, *Rtl1/Peg11*, and *Ldoc1/Sirh7/Rtl7*, as well as other *Rtl/Sirh* genes, has demonstrated that they exert important functions in the placenta and/or brain in mice and presumably humans [24–30]. Thus, the concept of mammalian-specific genes acquired from retroviruses is evidently well-supported, providing a useful model for approaching a variety of mammalian-specific features related to viviparous reproduction and the highly developed central nervous system in terms of genomic function. These results strongly indicate that these *RTL/SIRH* genes derived from retroviruses may be major players in the macroevolution of mammals [12,31,32]. It should be noted that most of them encode only a portion of the differently sized GAG-derived proteins, for example only 112 and 113 aa for *RTL8C*, *A*, *B/SIRH4*, *5*, and *6*; 151 aa for *RTL7/SIRH7*; 243 aa for *RTL6/SIRH3*; 304 aa for *RTL4/SIRH11*; and 599 aa for *RTL5/SIRH8*. Thus, not only *ENV* but also GAG and POL-derived ORFs may be functional in human development. There still remain more than 10,000 HERVs that have ORFs comprising >100 amino acids (aa) for GAG, POL, and *ENV*-derived proteins/peptides [3,4,33]. We have now reached a stage at which it is

important to ask how many primate- and human-specific *RVD* genes exist in the human genome. It is critically important for understanding the evolutionary emergence of human beings to elucidate the biological functions of these genes during the course of human development and growth. Thus, speeding up the efforts to carry out a comprehensive screening of the primate- and human-specific *RVD* genes is of considerable importance.

4. Materials and Methods

4.1. Comparison of ERVPb1 ORF in Simiiformes

All of the HERV and HERV-derived ORFs as well as viral sequences were obtained from the Ensembl (hg38, version 103.38), Repbase (version 23.02), and gEVE databases (version 1.1), respectively [4,34]. Using the human ERVPb1 ORF sequence as a query, we conducted tblastn searches against the genome sequences of eight *Catarrhini* primates (bonobo, chimpanzee, gorilla, orangutan, gibbon, African green monkey, rhesus macaque, and baboon), three *Platyrrhini* primates (Ma's night monkey, Capuchin and Marmoset), a *Tarsiiforme* species (Philippine tarsier), and a *Strepsirrhini* species (greater bamboo lemur). The amino acid identity in each of the ERVPb1 ORF sequences was calculated using the Clustal Omega program (Clustal Omega, <https://www.ebi.ac.uk/Tools/msa/clustalo/> (accessed on 6 January 2021–10 April 2021)) in the default mode.

4.2. Analysis of Short Read RNA-Seq Data

The data in Table S1 are summarized for each SRA ID of the RNA-seq data obtained from the SRA database (SRA, <http://www.ncbi.nlm.nih.gov/sra/> (accessed on 21 January 2012–22 March 2019)). They were aligned to HERV-derived ORFs in the human genome (hg38) obtained from the gEVE database (gEVE, <http://geve.med.u-tokai.ac.jp> (accessed on 18 February 2019)) using bowtie2 version 2.2.6 and the default parameter settings [35]. The expression level of each ORF was calculated by the total mapped reads as the reads per million (RPM) of the mapped reads. The RNA-Seq data of human primary monocytes of the 200 healthy individuals shown in Figure 4 were obtained from the European Genome-Phenome Archive (EGA), accession code EGA: EGAS00001001895, mapped onto the hg38 using STAR (version 2.73a) with the option “–outFilterMultimapNmax 1” and assembled using Scallop (version 0.10.5) [19,36,37]. The uniquely mapped read counts were normalized using size factors calculated in the DESeq2 package version 1.26.0 [38] and R version 3.6.1 [39].

4.3. Generation of Human iPSCs with an ERVPb1-NV Construct

Human ERVPb1-NV (Venus knock-in) iPSCs were generated with the CRISPR editing tool. A gRNA was designed to target the ERVPb1 N-terminal using the Benchling online tool (Benchling, <https://www.benchling.com> (accessed on 11 November 2019)), then cloned into the pGuide-it-ZsGreen1 Vector (Guide-it™ CRISPR/Cas9 System, Takara Bio, Shiga, Japan). The gRNA oligonucleotide sequence was 5'-AGGTGGCTAGCAAGAGCACA. The ERVPb1-Venus NV donor vector was assembled with a 5' homologous arm (1631 bp), Venus-Puro, and 3' homologous arm (1651 bp) by means of a cloning enzyme (In-Fusion® HD Cloning Kit, Takara Bio, Shiga, Japan). The plasmid containing the gRNA and the donor vector was introduced into human iPSCs by electroporation. Human iPSCs in which homologous recombination had occurred were selected by puromycin. ERVPb1-NV human iPSCs were confirmed by the DNA sequence from 5'HA to 3'HA.

4.4. Culture and Differentiation of Human iPSCs

Human iPSCs (HPS0063 201B7) were cultured in Cellartis DEF-CS 500 Culture System medium (Takara Bio, Shiga, Japan) under feeder-free conditions. Human iPSCs were induced into the embryoid body and differentiated into the three germ layers based on the in vitro differentiation protocol (Takara Bio). For macrophage differentiation, human iPSCs was induced in Cellartis DEF-CS 500 Culture System medium containing 80 ng/mL BMP4 (Recombinant human BMP-4, R&D Systems, Minneapolis, MN, USA) without bFGF [17].

4.5. Gene Expression Analysis by Real-Time PCR

The total RNA of human tissues (Human Total RNA Master Panel II, Takara Bio) was used for RT-PCR. The RT-PCR primer used were *ERVpb1*-forward: 5' GCTAGTAGGCCCTAATTGTTTCAG, *ERVpb1*-reverse: 5' AAGAAACCACAACGCCGTAA, *Venus*-forward: 5' GTGAGCAAGGGCGAGGAGCTGTT, *Venus*-reverse: 5' CTACTTGTACAGCTCGTCCA, *TBXT*-forward: 5' GCACACCAGAATTTGTAACTCC, *TBXT*-reverse: 5' TGTCCTTAACA GCTCAACTC, *CD235a*-forward: 5' TGTCCTTAACAGCTCAACTC, and *CD235a*-reverse: 5' AAATATTCACCTGCATGTCC. The total RNA derived from human iPSCs was extracted using an RNeasy Plus Kit (QIAGEN, Hilden, Germany) and used to carry out first-strand cDNA synthesis using Superscript III (SuperScript™ III First-Strand Synthesis System, Thermo Fisher Scientific, Waltham, MA, USA). Real-time PCR was performed using PowerTrack SYBR Green Master Mix (Thermo Fisher Scientific) in a StepOne machine (StepOne Real-Time PCR System laptop, Thermo Fisher Scientific), technical triplicates were analyzed. The *ERVpb1*-forward and -reverse primers were used for real-time PCR. The PCR conditions were one cycle of 95 °C for 2 min, 40 cycles of 95 °C for 5 s, and 60 °C for 30 s. The relative expression level was determined by the standard curve method and normalized to the expression of β -Actin.

4.6. Rapid Amplification of cDNA Ends (RACE)

For the 5'-RACE experiment, the 5'-Full RACE Core Set (Takara Bio, Shiga, Japan) was used to extend the *ERVpb1* mRNA extracted in 4.5. The first strand cDNA synthesis was carried out with 1 µg of total RNA using *ERVpb1* 5'-RACE GSP: 5' TTATACATGTTCTGC (5'-phosphorylated). The 5'-RACE PCR was performed with PrimeSTAR GXL DNA Polymerase (Takara Bio, Shiga, Japan) using the following primers: *ERVpb1*-race-F1: 5' CAGGATGGTTAGGGCAATGG and *ERVpb1*-race-R1: 5' GGTAATGGTCTGGGGTGAGT. The 5'-RACE PCR products (first 5'-RACE products) were confirmed by 1.5% agarose gel electrophoresis. Nested PCR was performed with PrimeSTAR GXL DNA Polymerase using the following primers: *ERVpb1*-race-F2: 5' CGTGAACGAACACGAAATCCGAA and *ERVpb1*-race-R2: 5' AGGTGGCTAGCAAGAGCACAA. The nested PCR products were purified by Agencourt AMPure XP (Beckman Coulter, Brea, CA, USA) and were determined by Sanger sequencing.

5. Conclusions

We have demonstrated that *ERVpb1* is a novel Simiiforme-specific gene and actually translated in hematopoietic cell lineages, including a subset of macrophages, in the differentiated human iPSCs. In addition, its expression is up-regulated by LPS administration in human primary monocytes.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms22094504/s1>: Figure S1: IGV browser views of *ERVpb1* locus. Figure S2: IGV browser views of *ERVpb1* locus. Table S1: RNA-seq data from several different tissues and organs.

Author Contributions: Conceptualization: T.K.-I. and F.I.; methodology: A.M., J.L., J.I., T.K.-I. and F.I.; software: A.M., S.N., S.M., M.T.U. and Y.K.; validation: S.N., Y.K. and F.I.; formal analysis: A.M.; investigation: A.M.; data curation: S.N., M.T.U., S.M. and Y.K.; writing—original draft preparation: A.M., T.K.-I. and F.I.; writing—review and editing: S.N. and F.I.; visualization: J.I.; supervision: Y.K., T.K.-I. and F.I.; project administration: T.K.-I. and F.I.; funding acquisition: F.I., T.K.-I. and Y.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the funding program for Grants-in-Aid for Scientific Research (A) (16H02478 and 19H00978) from Japan Society for the Promotion of Science (JSPS) to F.I. and T.K.-I. and Scientific Research (C) (20K06775) to S.N., Joint Usage/Research Program of Medical Research Institute (MRI) Tokyo Medical and Dental University (TMDU) grants to T.K.-I. and F.I. and Trans-Omics Project of MRI, TMDU to A.M., F.I. and Y.K.

Institutional Review Board Statement: The expression analysis of human monocyte data was approved by the ethics committee of Medical Research Institute, Tokyo Medical and Dental University (02019-005-04).

Informed Consent Statement: Written informed consent had been obtained for the collection of monocyte samples as described in the original study [20].

Data Availability Statement: Data can be available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bock, M.; Stoye, J.P. Endogenous retroviruses and the human germline. *Curr. Opin. Genet. Dev.* **2000**, *10*, 651–655. [[CrossRef](#)]
2. Belshaw, R.; Pereira, V.; Katzourakis, A.; Talbot, G.; Paces, J.; Burt, A.; Tristem, M. Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 4894–4899. [[CrossRef](#)]
3. Villesen, P.; Aagaard, L.; Wiuf, C.; Pedersen, F.S. Identification of endogenous retroviral reading frames in the human genome. *Retrovirology* **2004**, *1*, 32. [[CrossRef](#)]
4. Nakagawa, S.; Takahashi, M.U. gEVE: A genome-based endogenous viral element database provides comprehensive viral protein-coding sequences in mammalian genomes. *Database* **2016**, *2016*, 1–8. [[CrossRef](#)] [[PubMed](#)]
5. Ono, R.; Kobayashi, S.; Wagatsuma, H.; Aisaka, K.; Kohda, T.; Kaneko-Ishino, T.; Ishino, F. A retrotransposon-derived gene, PEG10, is a novel imprinted gene located on human chromosome 7q21. *Genomics* **2001**, *73*, 232–237. [[CrossRef](#)]
6. Ono, R.; Shiura, H.; Aburatani, H.; Kohda, T.; Kaneko-Ishino, T.; Ishino, F. Identification of a large novel imprinted gene cluster on mouse proximal chromosome 6. *Genome Res.* **2003**, *13*, 1696–1705. [[CrossRef](#)]
7. Ono, R.; Nakamura, K.; Inoue, K.; Naruse, M.; Usami, T.; Wakisaka-Saito, N.; Hino, T.; Suzuki-Migishima, R.; Ogonuki, N.; Miki, H.; et al. Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. *Nat. Genet.* **2006**, *38*, 101–106. [[CrossRef](#)]
8. Mi, S.; Lee, X.; Li, X.; Veldman, G.M.; Finnerty, H.; Racie, L.; LaVallie, E.; Tang, X.Y.; Edouard, P.; Howes, S.; et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature* **2000**, *403*, 785–789. [[CrossRef](#)]
9. Blaise, S.; de Parseval, N.; B nit, L.; Heidmann, T. Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 13013–13018. [[CrossRef](#)] [[PubMed](#)]
10. Sugimoto, J.; Sugimoto, M.; Bernstein, H.; Jinno, Y.; Schust, D. A novel human endogenous retroviral protein inhibits cell-cell fusion. *Sci. Rep.* **2013**, *3*, 1462. [[CrossRef](#)] [[PubMed](#)]
11. Kim, H.S.; Takenaka, O.; Crow, T.J. Isolation and phylogeny of endogenous retrovirus sequences belonging to the HERV-W family in primates. *J. Gen. Virol.* **1999**, *80 Pt 10*, 2613–2619. [[CrossRef](#)]
12. Kaneko-Ishino, T.; Ishino, F. Mammalian-specific genomic functions: Newly acquired traits generated by genomic imprinting and LTR retrotransposon-derived genes in mammals. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* **2015**, *91*, 511–538. [[CrossRef](#)]
13. Aagaard, L.; Villesen, P.; Kjeldbjerg, A.L.; Pedersen, F.S. The approximately 30-million-year-old ERVPb1 envelope gene is evolutionarily conserved among hominoids and Old World monkeys. *Genomics* **2005**, *86*, 685–691. [[CrossRef](#)]
14. Blaise, S.; de Parseval, N.; Heidmann, T. Functional characterization of two newly identified Human Endogenous Retrovirus coding envelope genes. *Retrovirology* **2005**, *2*, 19. [[CrossRef](#)]
15. Lee, J.; Sutani, A.; Kaneko, R.; Takeuchi, J.; Sasano, T.; Kohda, T.; Ihara, K.; Takahashi, K.; Yamazoe, M.; Morio, T.; et al. In vitro generation of functional murine heart organoids via FGF4 and extracellular matrix. *Nat. Commun.* **2020**, *11*, 4283. [[CrossRef](#)]
16. Huch, M.; Koo, B.K. Modeling mouse and human development using organoid cultures. *Development* **2015**, *142*, 3113–3125. [[CrossRef](#)]
17. Douvaras, P.; Sun, B.; Wang, M.; Kruglikov, I.; Lallo, G.; Zimmer, M.; Terrenoire, C.; Zhang, B.; Gandy, S.; Schadt, E.; et al. Directed Differentiation of Human Pluripotent Stem Cells to Microglia. *Stem. Cell Rep.* **2017**, *8*, 1516–1524. [[CrossRef](#)] [[PubMed](#)]
18. Takata, K.; Kozaki, T.; Lee, C.Z.W.; Thion, M.S.; Otsuka, M.; Lim, S.; Utami, K.H.; Fidan, K.; Park, D.S.; Malleret, B.; et al. Induced-Pluripotent-Stem-Cell-Derived Primitive Macrophages Provide a Platform for Modeling Tissue-Resident Macrophage Differentiation and Function. *Immunity* **2017**, *47*, 183–198. [[CrossRef](#)]
19. Steiper, M.E.; Young, N.M. *The Timetree of Life*; Oxford University Press: Oxford, UK, 2009; pp. 482–486.
20. Quach, H.; Rotival, M.; Pothlichet, J.; Loh, Y.E.; Dannemann, M.; Zidane, N.; Laval, G.; Patin, E.; Harmant, C.; Lopez, M.; et al. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell* **2016**, *167*, 643–656. [[CrossRef](#)]
21. Machinaga, A.; Ishihara, S.; Shirai, A.; Takase-Yoden, S. Splicing of Friend Murine Leukemia Virus env-mRNA Enhances Its Ability to Form Polysomes. *Front. Microbiol.* **2016**, *7*, 160. [[CrossRef](#)] [[PubMed](#)]
22. Benešová, M.; Trejbalová, K.; Kovářová, D.; Vernerová, Z.; Hron, T.; Kučerová, D.; Hejnar, J. DNA hypomethylation and aberrant expression of the human endogenous retrovirus ERVWE1/syncytin-1 in seminomas. *Retrovirology* **2017**, *14*, 20. [[CrossRef](#)]
23. Vargas, A.; Thiery, M.; Lafond, J.; Barbeau, B. Transcriptional and functional studies of Human Endogenous Retrovirus envelope EnvP(b) and EnvV genes in human trophoblasts. *Virology* **2012**, *425*, 1–10. [[CrossRef](#)]

24. Sekita, Y.; Wagatsuma, H.; Nakamura, K.; Ono, R.; Kagami, M.; Wakisaka, N.; Hino, T.; Suzuki-Migishima, R.; Kohda, T.; Ogura, A.; et al. Role of retrotransposon-derived imprinted gene, *Rtl1*, in the feto-maternal interface of mouse placenta. *Nat. Genet.* **2008**, *40*, 243–248. [[CrossRef](#)]
25. Naruse, M.; Ono, R.; Irie, M.; Nakamura, K.; Furuse, T.; Hino, T.; Oda, K.; Kashimura, M.; Yamada, I.; Wakana, S.; et al. *Sirh7/Ldoc1* knockout mice exhibit placental P4 overproduction and delayed parturition. *Development* **2014**, *141*, 4763–4771. [[CrossRef](#)]
26. Irie, M.; Yoshikawa, M.; Ono, R.; Iwafune, H.; Furuse, T.; Yamada, I.; Wakana, S.; Yamashita, Y.; Abe, T.; Ishino, F.; et al. Cognitive Function Related to the *Sirh11/Zcchc16* Gene Acquired from an LTR Retrotransposon in Eutherians. *PLoS Genet.* **2015**, *11*, e1005521. [[CrossRef](#)]
27. Irie, M.; Koga, A.; Kaneko-Ishino, T.; Ishino, F. An LTR Retrotransposon-Derived Gene Displays Lineage-Specific Structural and Putative Species-Specific Functional Variations in Eutherians. *Front. Chem.* **2016**, *4*, 26. [[CrossRef](#)] [[PubMed](#)]
28. Kitazawa, M.; Tamura, M.; Kaneko-Ishino, T.; Ishino, F. Severe damage to the placental fetal capillary network causes mid- to late fetal lethality and reduction in placental size in *Peg11/Rtl1* KO mice. *Genes Cells* **2017**, *22*, 174–188. [[CrossRef](#)] [[PubMed](#)]
29. Kitazawa, M.; Hayashi, S.; Imamura, M.; Takeda, S.; Oishi, Y.; Kaneko-Ishino, T.; Ishino, F. Deficiency and overexpression of *Rtl1* in the mouse cause distinct muscle abnormalities related to Temple and Kagami-Ogata syndromes. *Development* **2020**, *147*. [[CrossRef](#)] [[PubMed](#)]
30. Kitazawa, M.; Sutani, A.; Kaneko-Ishino, T.; Ishino, F. The role of eutherian-specific *RTL1* in the nervous system and its implications for the Kagami-Ogata and Temple syndromes. *Genes Cells* **2021**, *26*, 165–179. [[CrossRef](#)]
31. Kaneko-Ishino, T.; Ishino, F. The role of genes domesticated from LTR retrotransposons and retroviruses in mammals. *Front. Microbiol.* **2012**, *3*, 262. [[CrossRef](#)]
32. Kaneko-Ishino, T.; Ishino, F. Cooperation and Competition in Mammalian Evolution –Gene Domestication from LTR Retrotransposons. In *Evolution, Origin of Life, Concepts and Methods*; Pontarotti, P., Ed.; Springer Nature: Berlin/Heidelberg, Germany, 2019; Chapter 15; pp. 317–333.
33. Ueda, M.T.; Kryukov, K.; Mitsunashi, S.; Mitsunashi, H.; Imanishi, T.; Nakagawa, S. Comprehensive genomic analysis reveals dynamic evolution of endogenous retroviruses that code for retroviral-like protein domains. *Mob. DNA* **2020**, *11*, 29. [[CrossRef](#)] [[PubMed](#)]
34. Bao, W.; Kojima, K.K.; Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [[CrossRef](#)] [[PubMed](#)]
35. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [[CrossRef](#)] [[PubMed](#)]
36. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [[CrossRef](#)] [[PubMed](#)]
37. Shao, M.; Kingsford, C. Accurate assembly of transcripts through phase-preserving graph decomposition. *Nat. Biotechnol.* **2017**, *35*, 1167–1169. [[CrossRef](#)]
38. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
39. Statisticat LLC. LaplacesDemon: Complete Environment for Bayesian Inference. Bayesian-Inference.com. R Package Version 16.1.1. 2018. Available online: <https://cran.r-project.org/web/packages/LaplacesDemon/citation.html> (accessed on 22 March 2019).