



OPEN

## Emergence of an erythroid cell-specific regulatory region in *ABO* intron 1 attributable to A- or B-antigen expression on erythrocytes in Hominoidea

Rie Sano<sup>1✉</sup>, Haruki Fukuda<sup>1</sup>, Rieko Kubo<sup>1</sup>, Takao Oishi<sup>2</sup>, Takako Miyabe-Nishiwaki<sup>2</sup>, Akihisa Kaneko<sup>2</sup>, Haruhisa Masato<sup>3</sup>, Yoichiro Takahashi<sup>4</sup>, Akira Hayakawa<sup>1</sup>, Shin Yazawa<sup>5</sup> & Yoshihiko Kominato<sup>1</sup>

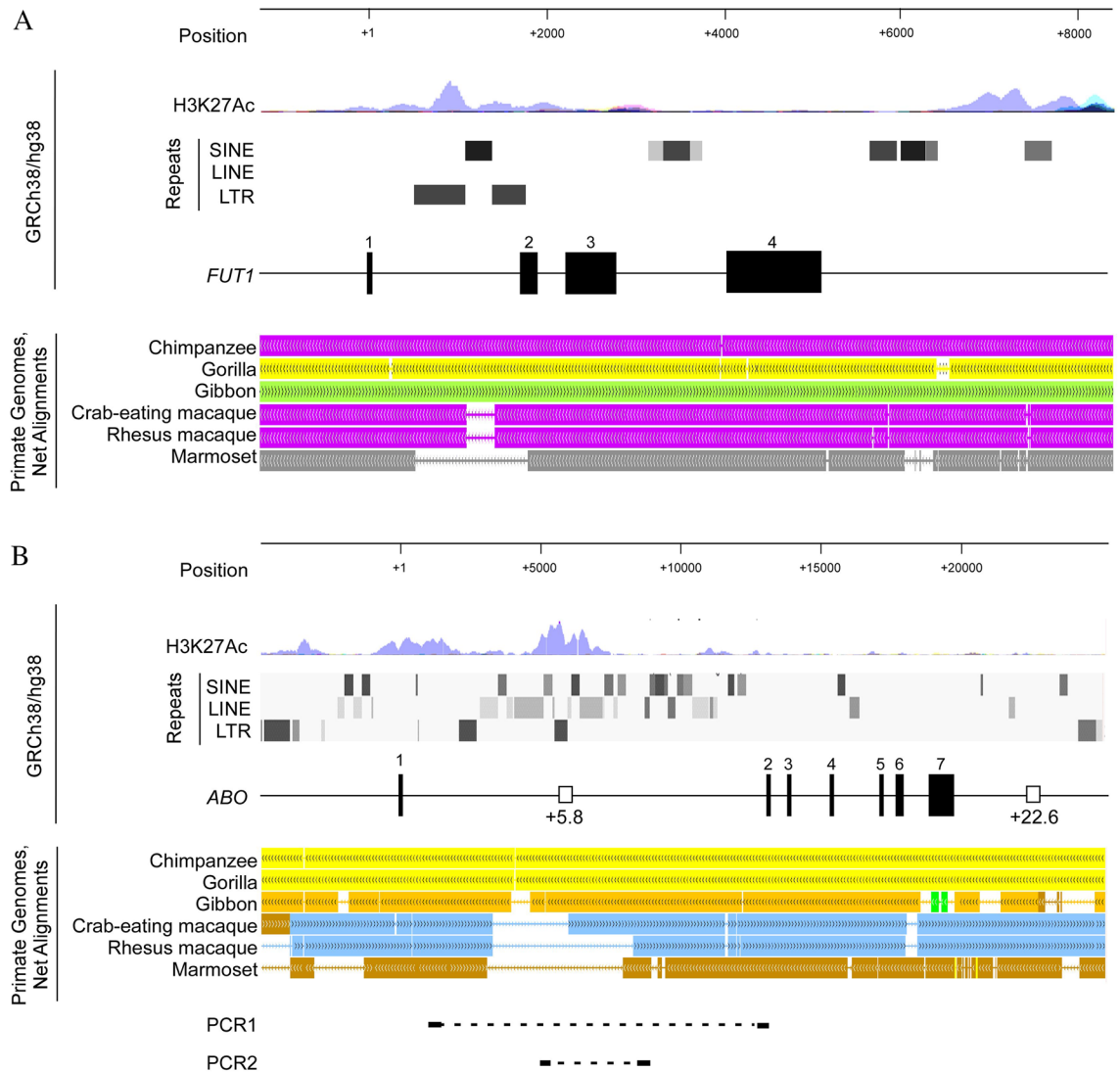
A- and B-antigens are present on red blood cells (RBCs) as well as other cells and secretions in Hominoidea including humans and apes such as chimpanzees and gibbons, whereas expression of these antigens on RBCs is subtle in monkeys such as Japanese macaques. Previous studies have indicated that H-antigen expression has not completely developed on RBCs in monkeys. Such antigen expression requires the presence of H-antigen and A- or B-transferase expression in cells of erythroid lineage, although whether or not ABO gene regulation is associated with the difference of A- or B-antigen expression between Hominoidea and monkeys has not been examined. Since it has been suggested that ABO expression on human erythrocytes is dependent upon an erythroid cell-specific regulatory region or the + 5.8-kb site in intron 1, we compared the sequences of *ABO* intron 1 among non-human primates, and demonstrated the presence of sites orthologous to the + 5.8-kb site in chimpanzees and gibbons, and their absence in Japanese macaques. In addition, luciferase assays revealed that the former orthologues enhanced promoter activity, whereas the corresponding site in the latter did not. These results suggested that the A- or B-antigens on RBCs might be ascribed to emergence of the + 5.8-kb site or the corresponding regions in *ABO* through genetic evolution.

The ABO blood group system was discovered by Karl Landsteiner in 1900<sup>1,2</sup>. Histo-blood group ABH(O) antigens are characterized as defined trisaccharide determinants GalNAc  $\alpha$ 1  $\rightarrow$  3 [Fuc $\alpha$ 1  $\rightarrow$  2]Gal $\beta$ 1  $\rightarrow$  R, Gal $\alpha$ 1  $\rightarrow$  3[Fuc $\alpha$ 1  $\rightarrow$  2]Gal $\beta$ 1  $\rightarrow$  R, and disaccharide determinants Fuc $\alpha$ 1  $\rightarrow$  2Gal  $\beta$ 1  $\rightarrow$  R for A, B, and H, respectively<sup>3,4</sup>. Gal  $\beta$ 1  $\rightarrow$  R denotes core chains such as Gal $\beta$ 1  $\rightarrow$  3GlcNAc and Gal $\beta$ 1  $\rightarrow$  4GlcNAc. These antigens are present on red blood cells (RBCs), in secretions, and on epithelial cells and vascular endothelial cells<sup>5-9</sup>. Their structures are synthesized from the precursor H substrate by  $\alpha$ 1  $\rightarrow$  3GalNAc transferase (A-transferase) and  $\alpha$ 1  $\rightarrow$  3Gal transferase (B-transferase), encoded by functional alleles at the ABO locus<sup>10,11</sup>. On the other hand, the precursor H substrate on RBCs is produced by  $\alpha$ 1  $\rightarrow$  2Fuc transferase (H-transferase) encoded by *FUT1*<sup>4</sup>. ABH-antigens are present at the non-reducing ends of carbohydrate chains, which are produced by sequential transfer of sugar by each glycosyltransferase specific to the non-reducing end of a precursor or immature carbohydrate chain. Thus, orchestrated expression of associated glycosyltransferases is needed for production of a carbohydrate chain, although the underlying mechanism responsible for regulation of the genes encoding the glycosyltransferases involved in carbohydrate chain production remains unclear.

ABO blood groups are determined by these antigens on RBCs, as well as anti-A and anti-B antibodies in serum in accordance with Landsteiner's law. Weak expression of the ABH-antigen on RBCs is associated with ABO subgroups or variant types<sup>12</sup>. Similar reduced reactivity of RBCs with anti-A or anti-B antibody is observed

<sup>1</sup>Department of Legal Medicine, Gunma University Graduate School of Medicine, 3-39-22 Showa-Machi, Maebashi 371-8511, Japan. <sup>2</sup>Center for the Evolutionary Origins of Human Behavior, Kyoto University, Inuyama, Japan. <sup>3</sup>Hitachi City Kamine Zoo, Hitachi, Japan. <sup>4</sup>Department of Legal Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan. <sup>5</sup>Department of General Surgical Science, Graduate School of Medicine, Gunma University, Maebashi, Japan. ✉email: takagirie@gunma-u.ac.jp

in non-human primates. Studies using routine anti-A and anti-B reagents have shown that agglutination of RBCs in apes such as chimpanzees, gorillas, orangutans, and gibbons is similar to that of human subgroups such as  $A_{int}$  or  $B_{int}$ <sup>13,14</sup>.  $A_{int}$  erythrocytes have lower amounts of A-antigen than  $A_1$  erythrocytes and higher amounts than  $A_2$  erythrocytes. The  $B_{int}$  is comparable to  $A_{int}$ <sup>15,16</sup>. On the other hand, similar tests have shown that the RBCs of prosimians, New World monkeys and Old World monkeys appear to lack A- or B-antigens, although trace amounts have been demonstrated using very sensitive absorption-elution techniques or very potent anti-A or anti-B reagents<sup>13,14,17</sup>. Nakajima et al. have concluded that expression of H-antigen on RBCs has not completely developed in monkeys, because eluates from such RBCs obtained by absorption-elution tests with anti-H reagent were found to weakly agglutinate group O-RBCs<sup>14</sup>. One suggested potential mechanism leading to such increased expression of H-antigen on RBCs of human is insertion of a short interspersed nuclear element (SINE), one of several transposable elements, in the first intron of *FUT1* encoding H-transferase (Fig. 1A), since this element



**Figure 1.** Homology of nucleotide sequences from the upstream to downstream regions of *FUT1* or *ABO* between human and non-human primates. **(A)** Human *FUT1*, **(B)** Human *ABO*. Upper panel shows positions relative to the transcription start site of *FUT1* or *ABO*. The second panel from the top indicates the acetylation at lysine 27 of histone 3 in *FUT1* or *ABO*, often found near active regulatory elements, which was demonstrated by ENCODE Regulation Tracks. The third panel from the top denotes repeating elements including SINE, LINE, and LTR, which were revealed by RepeatMasker over the genomic structure of *FUT1* or *ABO*. The fourth panel shows the genomic structure from the upstream region through *FUT1* or *ABO* to the downstream region, including exons, as well as regulatory regions such as the +5.8- and +22.6-sites for erythroid cells and epithelial cells, respectively, on the basis of the human genome draft GRCh38/hg38. Exons are denoted by filled boxes, and the regulatory regions are indicated by empty boxes. The fifth panel shows a comparison of the genome sequences for human with their reference sequences for non-human primates including chimpanzee, gorilla, gibbon, crab-eating macaque, rhesus macaque, and marmoset using Primate Genomes, Chain and Net Alignments<sup>18</sup>. Nucleotides with high identity are highlighted. The bottom panel of **(B)** shows the locations of PCR-amplified fragments of PCR1 and PCR2.

is present in Hominoidea including humans and apes, whereas it is absent in monkeys or non-primate mammals, which lack ABH-antigens on RBCs (Fig. 1A)<sup>18–21</sup>. Either A- or B-antigen expression on RBCs requires the presence of H-antigen as well as expression of A- or B-transferase in erythroid lineage cells. However, whether or not ABO gene regulation is associated with this difference of A- or B-antigens on RBCs between Hominoidea and monkeys has not been examined.

Yamamoto et al. have demonstrated that human *ABO* is composed of seven exons spanning 20 kb<sup>22</sup>. Subsequently, a positive regulatory element, named the + 5.8-kb site, in intron 1 was implicated in erythroid cell-specific regulation of *ABO* transcription as a result of in vitro experiments and genetic analysis of individuals with B<sub>m</sub> harboring deletion of 5.8 kb or 3.0 kb including the site (Fig. 1B)<sup>23,24</sup>. As the site includes one transcription factor recognition motif for RUNX1 and two motifs for GATA1/2<sup>25,26</sup>, deletion and single-nucleotide substitutions in the RUNX motif in the + 5.8-kb site have been found in individuals with A<sub>m</sub>, B<sub>3</sub> or B<sub>w</sub><sup>26–28</sup>, while nucleotide substitutions in the downstream GATA motif of the site have been reported in individuals with B<sub>m</sub> and A<sub>m</sub><sup>25,29,30</sup>. Thus, erythroid cell-specific regulation of *ABO* transcription could be dependent upon the + 5.8-kb site and the constitutive *ABO* promoter<sup>23,31,32</sup>. In addition, in vitro experiments have shown that a positive regulatory element, named the + 22.6-kb site, downstream of *ABO* is associated with epithelial cell-specific regulation of *ABO* transcription (Fig. 1B)<sup>33</sup>. On the other hand, it remained to further investigate whether the *ABO* expression was dependent upon the + 5.8-kb site or the + 22.6-kb site in the vascular endothelium. However, given that hematopoietic stem cells and vascular endothelial cells are thought to have a common ancestral cell, the + 5.8-kb site might be involved in *ABO* transcriptional regulation in the vascular endothelium.

Here, we compared the sequences of the + 5.8-kb site in human *ABO* with those of the corresponding sites in primate species including chimpanzees, agile gibbons, and Japanese macaques, and carried out transient transfection experiments using reporter plasmids containing those sites. The results suggested that the appearance of A- and B-antigens on RBCs might be ascribable to emergence of the + 5.8-kb site or its orthologues through genomic evolution.

## Materials and methods

**Blood and saliva.** Blood and saliva specimens were obtained at health check from several representative individuals of chimpanzee (*Pan troglodytes*), agile gibbon (*Hylobates agilis*), and Japanese macaque (*Macaca fuscata*) at the Center for the Evolutionary Origins of Human Behavior, Kyoto University. As mentioned briefly, the individual primates were anesthetized for a regular health examination. Anesthesia for chimpanzee was induced with intramuscular administration of the combination of medetomidine, midazolam, and ketamine and maintained with intravenous infusion of propofol<sup>34</sup>. Anesthesia for agile gibbon and Japanese macaque was induced with intramuscular administration of the combination of medetomidine, midazolam, and ketamine. Two or three ml of blood samples were obtained from the femoral vein and saliva were collected with a cotton swab. Experimental designs and procedures were approved under permit No.2021-037, by the Animal Welfare and Animal Care Committee at the Primate Research Institute, Kyoto University and all animal experiments were carried out in accordance with institutional guidelines for the care and use of nonhuman primates ([http://www.pri.kyotou.ac.jp/research/sisin2010/Guidelines\\_for\\_Care\\_and\\_Use\\_of\\_Nonhuman\\_Primates20100609.pdf](http://www.pri.kyotou.ac.jp/research/sisin2010/Guidelines_for_Care_and_Use_of_Nonhuman_Primates20100609.pdf)). All methods are reported in accordance with ARRIVE guidelines (<https://arriveguidelines.org>).

*ABO* blood groups of primates were determined using forward and reverse grouping of blood, and agglutination inhibition tests for ABH substances in saliva<sup>14,35</sup>. Briefly, for the forward grouping, one drop of 2% blood suspension was mixed with one drop of anti-A and anti-B antibodies (Wako) as well as *Ulex europaeus* lectin (J-Chemical) on a hole glass, and the agglutination reaction was observed 30 min after agitation at room temperature. For the reverse grouping, the plasma was mixed with human O blood cells overnight in a refrigerator to remove heterologous antibodies, and one drop each of the treated plasma and human A<sub>1</sub> or B blood cells was agitated on a hole glass, followed by observation of the agglutination reaction. For agglutination inhibition tests for ABH substances in saliva, a cotton swab wiped with fresh saliva was moistened with 300 µl of saline solution and heated to denature the proteins. Equal amount of anti-A, anti-B or anti-H reagents with an eightfold aggregation titer were mixed with the diluted saliva after heat treatment for two hours at room temperature. Human A<sub>1</sub>, B, and O blood cells were then added to the saliva-antibody mixture of the corresponding types, respectively, and inhibition of hemagglutination by the blood group substance in the saliva was observed.

Genomic DNA was extracted from blood using a QIAamp blood mini kit (Qiagen GmbH).

**PCR amplification and cycle sequencing.** PCR1 and PCR2 were carried out for PCR amplification of a portion of the first *ABO* intron in primate species (Fig. 1B). PCR1 used the primers hABOg997F and hABO-ex2R, whose sequences were 5'-AGAGGAGTGGAAAATTCATGAAGA-3' and 5'-CCAAACAAGACCAAGACAAGCAT-3', respectively. PCR amplification was performed in 50 µl of reaction mixture containing 200 ng of genomic DNA, 1 × KOD One PCR Master Mix (TOYOBO) with KOD DNA polymerase, and each primer at 0.1 µM. Amplification was carried out under the following conditions: initial denaturation at 98 °C for 3 min followed by 35 cycles of 98 °C for 10 s, 60 °C for 5 s and 68 °C for 10 min. PCR1 was followed by nested PCR2. PCR2 used the primers hABOc.28 + 5005 and hABOc.28 + 8800, whose sequences were 5'-TCGGCTCTTGCCAGGTGGAG-3' and 5'-CCACAATATCTCAGGACCCATA-3', respectively. PCR2 amplification was performed using the same reaction mixture as that for PCR1. Amplification was carried out under the following conditions: initial denaturation at 98 °C for 3 min followed by 35 cycles of 98 °C for 10 s and 68 °C for 2 min, followed by direct sequencing of PCR2 products with specific primers for the PCR2 target and the BigDye™ terminator kit version 1.1 (Thermo Fisher Scientific, Waltham, MA). The sequencing run was performed on a genetic analyzer (Thermo Fisher Scientific, Seqstudio). Also, the PCR2 products were ligated into the pUC 118 vector using a

Mighty Cloning Reagent set (Blunt End) (TaKaRa), followed by determination of the nucleotide sequences of the clones obtained using the primers M13M4 and M13RV.

**Plasmids.** Luciferase reporter plasmids SN, C(01)/SN, C(04)/SN, C(06)/SN, C(B<sub>m</sub>)/SN, have been described previously<sup>25,26</sup>. In reporter plasmid SN, the *ABO* proximal promoter located between −150 and −2 relative to the translation start site was subcloned upstream of the luciferase gene<sup>23</sup>, whereas the +5.8-kb site between c.28+5624 and c.28+6125 with haplotype *ABOInt1\*01*, *ABOInt1\*04*, or *ABOInt1\*06* was subcloned upstream of the *ABO* promoter in reporter plasmid C(01)/SN, C(04)/SN, or C(06)/SN, respectively<sup>25</sup>. The regions in the first intron 1 of *ABO* in non-human primates that corresponded to the human +5.8-kb site were generated by PCR using specific primers and the PCR2 products as templates, followed by insertion just upstream of the human *ABO* promoter in constructs C(Chi)/SN, C(Gib)/SN, and C(Mac)/SN. The directions of the inserts of all constructs used in this study were verified by DNA sequence analysis as described above. Plasmid DNA was purified using a plasmid purification kit (HiSpeed Plasmid Maxi kit, Qiagen GmbH).

**Transfection and luciferase assay.** Transient transfection into K562 cells was performed using Lipofectamine 3000 (Invitrogen Corp.) as reported previously<sup>23</sup>. For each analysis, 2.5 µg of firefly luciferase reporter plasmid and 0.01 µg of pRL-SV40 *Renilla* luciferase reporter vector were used. After the cells had been collected, cell lysis and luciferase assays were performed using the dual-luciferase reporter assay system (Promega) to measure the activities of the firefly and *Renilla* luciferases. Light emission was measured using an absorption spectrophotometer (Nivo F, Perkin Elmer). Variations in transfection efficiency were normalized to the activities of *Renilla* luciferase expressed from the cotransfected pRL-SV40 *Renilla* luciferase reporter.

## Results

**Comparison of genomic DNAs from the upstream to downstream region of *ABO* among humans and primate species.** Primate Genomes, Chain and Net Alignments<sup>18</sup> from the UCSC Genome Browser for humans and primate species have demonstrated high sequence conservation between the ATG translation start codon and the stop codon of *ABO*, except for a few regions including the +5.8-kb site (Fig. 1B). The +5.8-kb site is conserved among human, chimpanzee, gorilla, and gibbon, showing similar expression of A- and B-antigens on RBCs<sup>13</sup>, whereas it is not conserved in monkeys such as crab-eating macaque, rhesus macaque and marmoset, in which the A- and B-antigens are expressed only slightly on RBCs. On the other hand, the +22.6-kb site is conserved among humans and those primate species. Therefore, further study was carried out to examine whether the +5.8-kb site could contribute to the presence or absence of A- and B-antigens on RBCs using blood and saliva specimens from several representative individuals of chimpanzee, agile gibbon, and Japanese macaque without any anguineous relationship.

**Serological tests and phenotypes.** Serological tests were performed for two individual chimpanzees (Table 1). Their RBCs were agglutinated by anti-A antibody, but not by anti-B antibody. Their sera contained antibodies that agglutinated B-RBCs, and lacked antibodies that agglutinated A-RBCs. Consistently, their saliva was shown to contain A- and H-substances by saliva tests, based on its ability to inhibit anti-A and anti-B reagents. Thus, their blood types were determined as A. Serological tests were also carried out for three individual agile gibbons (Table 1). The RBCs from two individuals were agglutinated by anti-A and anti-H antibodies, while RBCs from the other individual were agglutinated by anti-B antibody, but not by anti-A antibody. The sera from the former individuals lacked antibodies that agglutinated A- and B-RBCs, while the serum from the latter individual contained antibody that agglutinated A-RBCs and lacked antibodies that agglutinated B-RBCs. Consistently, the saliva from the two individuals was shown to contain A-, B- and H-substances, while that from the other contained B- and H-substances. Thus, the blood type of the former individuals was determined as

Individual	Species	Blood				*ABH-antigen in saliva	ABO phenotype
		Agglutinability of red cells		Reaction of serum			
		Anti-A	Anti-B	Human A-RBC	Human B-RBC		
#1	Chimpanzee	‡4+	–	–	+	H, A	A
#2	Chimpanzee	4+	–	–	+	H, A	A
#3	Agile gibbon	4+	4+	–	–	H, A, B	AB
#4	Agile gibbon	4+	4+	–	–	H, A, B	AB
#5	Agile gibbon	–	4+	+	–	H, B	B
#6	Japanese Macaque	–	–	+	+	H	O
#7	Japanese Macaque	–	–	+	–	H, B	B
#8	Japanese Macaque	–	–	+	+	H	O

**Table 1.** Summary of serological results obtained using samples from individual primates. \*Presence of ABH-antigens was examined by testing saliva for its ability to inhibit anti-A, anti-B and anti-H reagents. ‡4+ denotes complete agglutination observed.

AB, whereas that of the latter individual was determined as B. When serological tests were carried out on three individual Japanese macaques, their RBCs were not agglutinated with anti-A and anti-B antibodies (Table 1). The serum from one individual contained antibody that agglutinated A-RBCs and lacked antibodies that agglutinated B-RBCs, while that from the others contained antibody that agglutinated both A-RBCs and B-RBCs. Consistently, the saliva from one individual was shown to contain B- and H-substances, while that from the others contained only H-substances. Thus, the blood type of the one individual was determined as B, whereas that of the other two individuals was determined as O. In conclusion, the presence of A- or B-antigens on RBCs was concordant with that in saliva in chimpanzees and agile gibbons. On the other hand, B-antigens were not detectable from RBCs of a Japanese macaque, but were found in saliva. This phenotype seemed to correspond to the weak human phenotype B<sub>m</sub>.

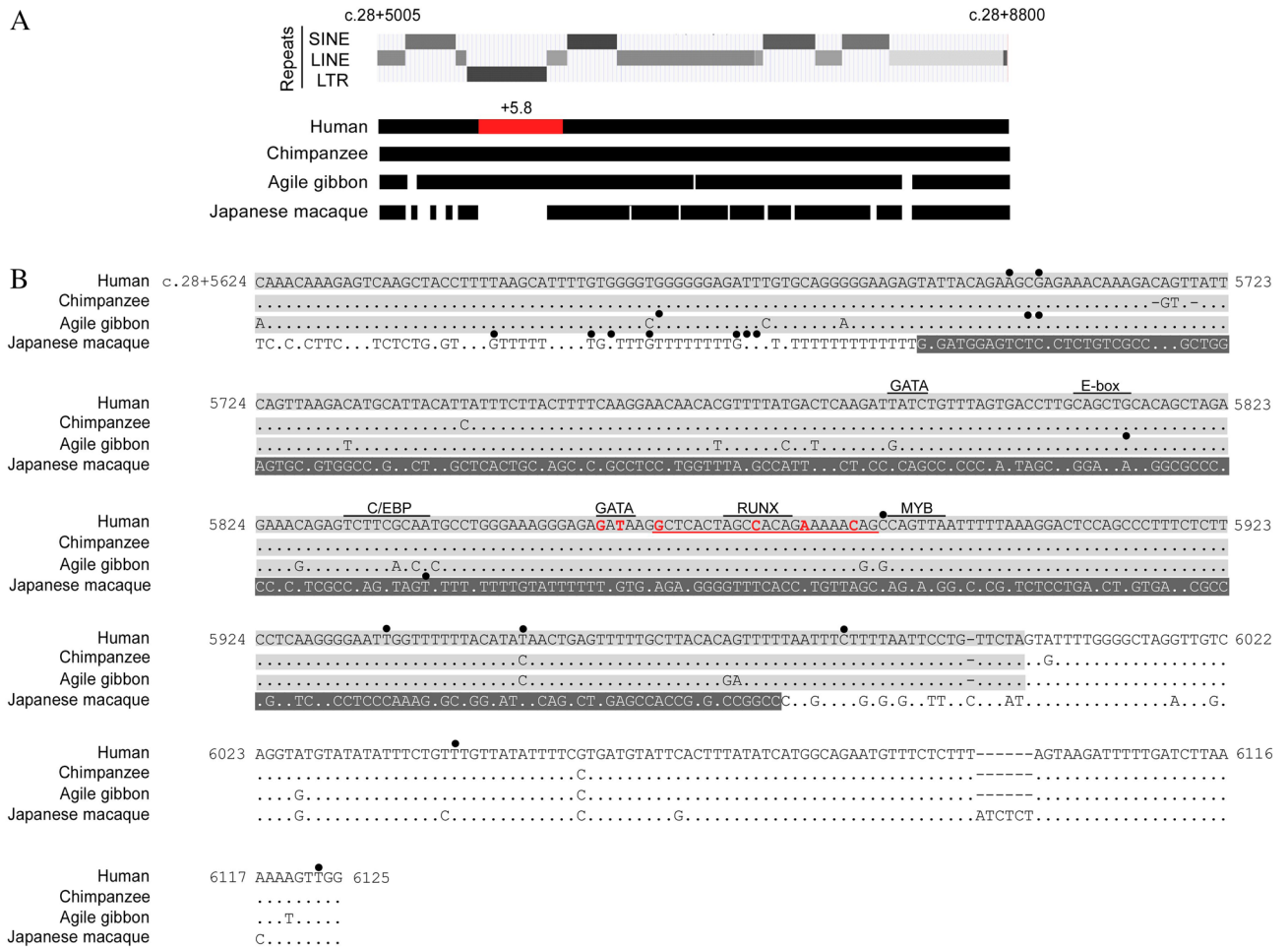
**LTRs at sites corresponding to the human +5.8-kb site was preserved in chimpanzees and agile gibbons, and absent in Japanese macaques.** Further investigation was carried out to clarify the mechanism responsible for loss or reduction in B-antigen expression in the above Japanese macaque. Nested PCR amplification was performed to determine the sequences of 3.8-kb in the first intron, which corresponded to the regions including the upstream sites corresponding to the human +5.8-kb site to the downstream stretch using the genomic DNAs obtained from two individual chimpanzees, three individual agile gibbons, and three individual Japanese macaques. When the nucleotide sequences from c.28 + 5005 to c.28 + 8800 in the first intron of human *ABO* were compared with those among the chimpanzees, gibbons, and Japanese macaques, most of the chimpanzee and gibbon nucleotide sequences were similar to those of the human +5.8-kb site, although in Japanese macaques several regions were not conserved (Fig. 2A). Specifically, regarding the +5.8-kb site, the human nucleotide sequences from c.28 + 6002 to c.28 + 6125 were conserved among the four primates (Fig. 2B). In contrast, the human sequences from c.28 + 5624 to c.28 + 6001 were conserved in the chimpanzees and agile gibbons, but not in Japanese macaques. The human sequences from c.28 + 5624 to c.28 + 6001 were shown to belong to long terminal repeats (LTR) by RepeatMasker<sup>36</sup>, and the LTR was located between long interspersed nuclear elements (LINE) that were classified as L1MB7 (Fig. 2A). However, the site corresponding to the +5.8-kb site included SINE in Japanese macaques (Fig. 2B). Therefore, since the +5.8-kb site plays a role in erythroid cell-specific expression of human *ABO*, it seemed likely that the LTR of the human +5.8-kb site and its orthologues could be involved in erythroid cell-specific regulation of *ABO* expression in human, chimpanzee and agile gibbon.

**Significance of other LTRs similar to the LTR in the +5.8-kb site as a positive regulatory element in humans.** To examine whether other LTRs similar to that in the +5.8-kb site might play a role in gene regulation in humans, a BLAST search was performed using the query sequence of the LTR in the human +5.8-kb site, as the LTR was classified group *HERVER1* according to Vargiu et al.<sup>37</sup> or LTR15 by RepeatMasker. Above an identity of 92%, 30 LTRs are listed in Table 2, 19 of them belonging to LTR15. ENCODE Candidates Cis-regulatory Elements<sup>38</sup> indicated 10 LTRs with a distal enhancer-like signature, a promoter-like signature or DNase-H3K4me3, although GeneHancer Regulatory Elements and Gene Interactions<sup>39</sup> showed no interaction configuration in the LTRs other than that in the +5.8-kb site which had been shown previously to interact with the transcription start sites of genes including *ABO*, *SURF6*, *MED22*, *RPL7A*, *SNORD24*, *SNORD36*, *SURF1*, *SURF2*, and *DBH* in humans<sup>40</sup>. Thus, it was unclear whether other LTRs similar to that in the +5.8-kb site could be involved in gene expression in humans, and heterogeneity for positive regulatory potential of the LTRs might be dependent upon local genomic circumstances or positional effects.

**Examination of the *trans*-activation potential of sites corresponding to the human +5.8-kb site in chimpanzee, agile gibbon, and Japanese macaque.** To examine whether the sites corresponding to the human +5.8-kb site in chimpanzee, agile gibbon, or Japanese macaque have potential for *trans*-activating the *ABO* promoter, we prepared reporter plasmids C(Chi)/SN, C(Gib)/SN, or C(Mac)/SN containing sites corresponding to the human +5.8-kb site in chimpanzee, agile gibbon, or Japanese macaque, respectively, inserted upstream of the human *ABO* promoter. When these reporters were transiently transfected into K562 cells, the sites corresponding to the human +5.8-kb site in chimpanzee and agile gibbon enhanced the promoter activity, while that in Japanese macaque did not (Fig. 3). The orthologous site in chimpanzee increased the promoter activity similarly to the human +5.8-kb site of haplotypes *ABOInt1\*01*, *ABOInt1\*04* and *ABOInt1\*06*, with an activity higher than that of the site in the agile gibbon. Thus, it was likely that the sites orthologous to the human +5.8-kb site in chimpanzee and agile gibbon had the potential to enhance the promoter activity, whereas the enhancing activity of the site corresponding to the human +5.8-kb site in Japanese macaque was deficient.

## Discussion

Here, we compared the genomic sequences of the +5.8-kb site in human *ABO* with the corresponding sequences in primate species including the chimpanzee, agile gibbon, and Japanese macaque. We found that the sequences of the human +5.8-kb site and its orthologues including the LTR were conserved among human, chimpanzee, and agile gibbon, whereas in Japanese macaque the site corresponding to the +5.8-kb site included SINE. In addition, transient transfection experiments using reporter *luciferase* plasmids including the human +5.8-kb site and its corresponding sites from the three primate species showed that the sequences from chimpanzee and agile gibbon enhanced the transcriptional activity of the *ABO* promoter, whereas that in Japanese macaque did not. These results suggested that the +5.8-kb site or its orthologues could play a role in *ABO* expression in erythroid lineage cells of human, chimpanzee and agile gibbon. On the other hand, the epithelial cell-specific regulatory region or the +22.6-kb site appeared to be conserved among primates including human, chimpanzee,



**Figure 2.** Comparison of nucleotide sequences in the first intron of *ABO* involving the human + 5.8-kb site and the corresponding regions in chimpanzee, gibbon, and Japanese macaque. **(A)** Comparison of nucleotide sequences in the first intron of *ABO* among human, chimpanzee, gibbon, and Japanese macaque. The top line shows the nucleotide sequence of the first intron in human *ABO* from c.28 + 5005 to c.28 + 8800, and the + 5.8-kb site in human *ABO* is denoted in red. The other lines represent the nucleotide sequences in chimpanzee, gibbon, and Japanese macaque corresponding to the human genome. Retrotransposons in the human genome are denoted over those lines for human and non-human primates. **(B)** Comparison of nucleotide sequences between the human + 5.8-kb site and the corresponding sites found in non-human primates. The top line shows the nucleotide sequence of the + 5.8-kb site in human *ABO* from c.28 + 5624 to c.28 + 6125, and the other lines represent the nucleotide sequences of the sites corresponding to the human + 5.8-kb site from chimpanzee, gibbon, and Japanese macaque. The sequences were deposited in GenBank with accession numbers OP775662 (chimpanzee), OP775663 (gibbon), and OP775664 (Japanese macaque). Nucleotides that are red or underlined indicate single nucleotide variants (SNVs) or deletion in individuals with weak phenotypes such as  $A_m$ ,  $A_3$ ,  $B_m$  and  $B_3$ . The motifs for transcription factor recognition sites are indicated by overbars. For comparison with the nucleotides in human *ABO*, the substituted nucleotides are shown and horizontal bars denote deletion of nucleotides in those monkeys. Solid circles over nucleotides in the human and monkey sequences indicate single nucleotide polymorphisms (SNPs) or variants. Nucleotide sequences belonging to LTR are highlighted in gray, and those belonging to SINE are highlighted in dark gray.

gorilla, gibbon, crab-eating macaque, rhesus macaque and marmoset (Fig. 1). Since saliva tests were valid for *ABO* typing in three primate species, the + 22.6-kb site or its orthologues might be involved in epithelial cell-specific regulation of *ABO* transcription in non-human primates including Japanese macaque, chimpanzee and agile gibbon. Thus, it seemed plausible that the presence of A- and B-antigens on RBCs might be ascribed to emergence of the + 5.8-kb site or its orthologues in intron 1 of *ABO* through genetic evolution in the common ancestral lineage of Hominoidea including human, chimpanzee and agile gibbon.

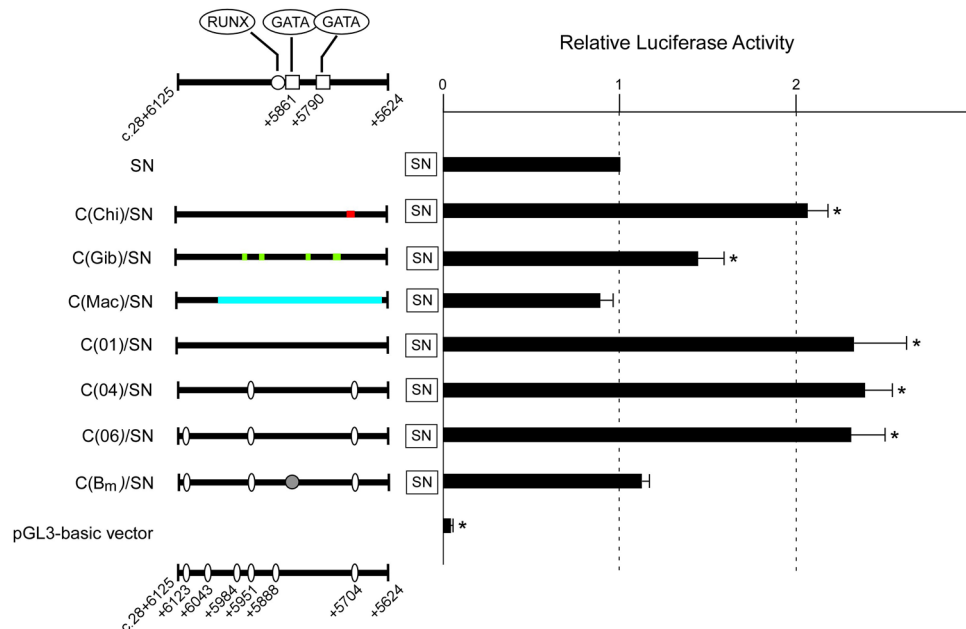
Almost half of the human genome is derived from transposable elements (TEs)<sup>41</sup>. The majority of TEs consist of retrotransposons including LTRs, SINEs, and long interspersed nuclear elements (LINEs). Retrotransposons are divided according to whether they contain LTRs. Human LTR retrotransposons containing endogenous retroviruses (ERV) and related sequences represent ~ 8% of the human genome whereas non-LTR retrotransposon elements including SINEs and LINEs account for 34%. Most ERV open reading frames have been degraded by

Position*	Identity %	Cis-regulatory elements <sup>§</sup>	Location	
chr9:133269164–133269541 <sup>†</sup>	99.5	Distal enhancer-like signature	<i>ABO</i>	Intron 1
chr1:154291653–154292344 <sup>†</sup>	94.2	Distal enhancer-like signature	Intergenic region	
chr14:94084570–94084951 <sup>†</sup>	93.9	None	<i>IFI27L1</i>	Intron 1
chr3:145546392–145546770 <sup>‡§</sup>	93.6	DNase-H3K4me3	Intergenic region	
chr10:25279470–25279845 <sup>†</sup>	93.5	None	<i>GPR158</i>	Intron 3
chr19:19826275–19826655	93.3	None	<i>ZNF56P</i>	Intron 1
chr9:36304059–36304746 <sup>†</sup>	93.3	None	Intergenic region	
chr6:150365509–150365886 <sup>†</sup>	92.9	None	Intergenic region	
chr1:197097013–197097382 <sup>†</sup>	92.9	None	<i>ASPM</i>	Intron 17 or 18
chr3:146922327–146922695	92.9	None	Intergenic region	
chr17:39334919–39335299 <sup>†</sup>	92.8	Distal enhancer-like signature	<i>FBXL20</i>	Intron 2
chr12:86129189–86129890 <sup>†</sup>	92.8	Distal enhancer-like signature	<i>MGAT4C</i>	Intron 1 or 4
chr13:57436874–57437242 <sup>‡§</sup>	92.6	Distal enhancer-like signature	Intergenic region	
chr14:88819677–88820049 <sup>†</sup>	92.5	None	ENSG00000274492	Exon 1
chr2:149240139–149240517 <sup>†</sup>	92.4	None	Intergenic region	
chr16:3085288–3085666 <sup>†</sup>	92.3	Promoter-like signature	ENSG00000262370	Intron 1
chr4:44032311–44032686	92.3	None	Intergenic region	
chr19:49124931–49125639	92.3	None	<i>SPHK2</i>	Intron 2
chr14:19776986–19777376	92.3	None	Intergenic region	
<sup>†</sup> chr4:156063554–156063924	92.2	Distal enhancer-like signature	Intergenic region	
chr4:156984706–156985076	92.2	None	Intergenic region	
chr11:24047586–24047954	92.2	None	Intergenic region	
chr8:522284–522666 <sup>†</sup>	92.1	None	<i>TDRP</i>	Intron 1
chr16:31241396–31241760	92.1	None	Intergenic region	
chr4:99080845–99081222	92.0	None	<i>ADH5</i>	Intron 4
chr12:9392179–9392557 <sup>†</sup>	92.0	None	<i>LINC02367</i>	Intron 2
chr1:67141852–67142227 <sup>†</sup>	92.0	Distal enhancer-like signature	<i>C1orf141, IL23R</i>	Intron 1
chr19:44573691–44574071 <sup>†</sup>	92.0	Distal enhancer-like signature	Intergenic region	
chr7:106630097–106630468	92.0	None	Intergenic region	
chr22:16065595–16065985	92.0	None	Intergenic region	

**Table 2.** Sequences homologous to the human + 5.8-kb site demonstrated by a BLAST search with candidates for estimated cis-regulatory elements. \*Positions are described according to Human GRCh38/hg38. For the BLAST, a portion of the + 5.8-kb site of haplotype *ABOInt1\*01* belonging to LTR15 was used as the query sequence. Since the + 5.8-kb site in human GRCh38/hg38 belonged to haplotype *ABOInt1\*04*, there was a difference of two nucleotides between these + 5.8-kb sites. †Descriptions are according to ENCODE Candidates Cis-regulatory Elements (cCREs). ‡Sequences indicated to belong to LTR15 by RepeatMasker. §The regions chr3:145546392–145546770 and chr13:57436874–57437242 belonging to LTR15 were located between LINEs of L1M1 and L1Pba1, respectively. Similarly, the human + 5.8-kb site belonging to LTR15 was located between L1MB7s.

mutation. The majority of ERVs are devoid of all coding regions, having undergone recombination between the two flanking LTRs to produce the elements known as solitary LTRs. LTRs including ERV and solitary LTRs may be located at any region of the adjacent genes, including the 5'UTR, intron, exon and 3'UTR. These distributions provide favorable conditions for LTRs regulating the expression of their neighboring genes in different ways, such as promoters, enhancers, selective splicer sites and polyadenylation sites<sup>41–45</sup>. Consistently, the LTR of the + 5.8-kb site and its orthologues are suggested to be involved in the regulation of *ABO* expression in erythroid lineage cells.

Since TEs have repeat sequences spread throughout the genome, and their repeat sequences involve transcription factor binding sites (TFBSs) that act as TE-derived cis-regulatory elements (CREs), it has long been speculated that specific TE families could contribute to dispersion of TFBSs and CREs through the genome, thus facilitating the emergence of a regulatory network controlling a series of cooperative genes<sup>46,47</sup>. It is intriguing to speculate that repeat sequences of TEs play a role in the orchestrated expression of genes encoding glycosyltransferases involved in the production of carbohydrate chains. Indeed, it has been reported that an LTR is a dominant promoter for the gene coding human  $\beta 1 \rightarrow 3$  galactosyltransferase 5, which is involved in synthesizing the core chain of ABH-antigens<sup>48</sup>. In the first intron of *FUT1*, ENCODE Candidates Cis-regulatory Elements and RepeatMasker indicated distal enhancer-like signatures in the LTRs flanking the SINE, which is assumed to be responsible for H-antigen expression in erythroid lineage cells (Fig. 1A)<sup>19–21</sup>. In addition, it was indicated that the + 5.8-kb site and its orthologues including the LTR are involved in the regulation of *ABO* expression in erythroid lineage cells. Thus, it seems that LTRs could contribute to the cis-regulatory network facilitating cooperative expression of the genes involved in the production of carbohydrate chains with A- and B-antigens at



**Figure 3.** Transcriptional regulatory activity of the human +5.8-kb site was lost by change from its homologs in apes to the corresponding site in Japanese macaque. K562 cells were transiently transfected with *ABO* enhancer-luciferase gene constructs including SN, C(01)/SN, C(04)/SN, C(06)/SN, and C(B<sub>m</sub>)/SN as well as C(Chi)/SN, C(Gib)/SN, and C(Mac)/SN. Reporter plasmid SN contained the human *ABO* promoter, and reporter plasmids C(01)/SN, C(04)/SN, and C(06) comprised the promoter and +5.8-kb site of haplotypes *ABOInt1\*01*, *ABOInt1\*04*, and *ABOInt1\*06*, respectively. Haplotypes *ABOInt1\*01*, *ABOInt1\*04*, and *ABOInt1\*06* were linked to the *O* allele, *A* allele and *B* allele, respectively, with a few exceptions attributable to genetic recombination<sup>49</sup>. Locations of the single nucleotide polymorphisms are shown by ovals. Reporter plasmid C(B<sub>m</sub>)/SN included the promoter and +5.8-kb site harboring a mutation of the GATA recognition site in the background of haplotype *ABOInt1\*01*<sup>25</sup>. Location of the mutation is denoted by a solid circle. Reporter plasmids C(Chi)/SN, C(Gib)/SN, or C(Mac)/SN contained the promoter and the sites corresponding to the human +5.8-kb site in chimpanzee, agile gibbon, or Japanese macaque, respectively. The parts in red, green and blue in chimpanzee, agile gibbon, and Japanese macaque, respectively, represent differences from the human sequences. The results are expressed as the average fold of the activity observed for the promoter. The standard deviations indicated are for three independent experiments. The significance of differences was determined by Student's *t* test at a significance level of  $p < 0.05$  (\*) compared to the relative luciferase activity of SN.

the non-reducing end. Future studies will need to investigate the relationship between the LTR at the +5.8-kb site of *ABO* and the LTRs flanking the SINE in intron 1 of *FUT1* to clarify the mechanism underlying the expression network of genes involved in ABH-antigen production on human RBCs.

### Data availability

The datasets generated and/or analysed during the current study are available in the Genbank repository, <https://ncbi.nlm.nih.gov/nuccore/OP775662> (chimpanzee), <https://ncbi.nlm.nih.gov/nuccore/OP775663> (gibbon), <https://ncbi.nlm.nih.gov/nuccore/OP775664> (Japanese macaque).

Received: 1 November 2022; Accepted: 20 March 2023

Published online: 27 March 2023

### References

- Landsteiner, K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden wirkungen des blutserums und der lympe. *Zentralbl. Bakteriol.* **27**, 357–362 (1900).
- Landsteiner, K. Über Agglutinationserscheinungen normalen menschlichen Blutes. *Wien Klin Wochenschr* **14**, 1132–1134 (1901).
- Kabat, E. A. *Blood Group Substances: Their Chemistry and Immunochemistry* (Springer, 1956).
- Watkins, W. M. Biochemistry and genetics of the ABO, Lewis, and P blood group systems. *Adv. Hum. Genet.* **10**, 379–385 (1980).
- Sarafian, V. *et al.* Differential expression of ABH histo-blood group antigens and LAMPs in infantile hemangioma. *J. Mol. Histol.* **36**, 455–460 (2005).
- Nosaka, M. *et al.* Aberrant expression of histo-blood group A type 3 antigens in vascular endothelial cells in inflammatory sites. *J. Histochem Cytochem.* **56**, 223–231 (2008).
- Ravn, V. & Dabelsteen, E. Tissue distribution of histo-blood group antigens (review). *APMIS* **108**, 1–28 (2000).
- Oriol, R. *et al.* Genetic regulation of the expression of ABH and Lewis antigens in tissues. *APMIS Suppl.* **27**, 28–38 (1992).
- Bentall, A. *et al.* Characterization of ABH-subtype donor-specific antibodies in ABO-A-incompatible kidney transplantation. *Am. J. Transplant.* **21**, 3649–3662 (2021).



10. Yamamoto, F. *et al.* Cloning and characterization of DNA complementary to human UDP-GalNAc: Fuc alpha 1->2Gal alpha 1->3GalNAc transferase (histo-blood group A transferase) mRNA. *J. Biol. Chem.* **265**, 1146–1151 (1990).
11. Yamamoto, F. *et al.* Molecular genetic basis of the histo-blood group ABO system. *Nature* **345**, 229–233 (1990).
12. Ogasawara, K. *et al.* Molecular genetic analysis of variant phenotypes of the ABO blood group system. *Blood* **88**, 2732–2737 (1996).
13. Nakajima, T., Miyazaki, S. & Yazawa, S. ABH antigens in red blood cells and digestive organs of non-human primates (in Japanese). *Primate Res.* **5**, 36–45 (1989).
14. Nakajima, T., Furukawa, K. & Takenaka, O. Blood group A and B glycosyltransferases in nonhuman primate plasma. *Exp. Clin. Immunogenet.* **10**, 21–30 (1993).
15. Landsteiner, K. & Levine, P. The differentiation of a type of human blood by means of normal animal serum. *J. Immunol.* **20**(2), 179–185 (1931).
16. Akira, Y. *et al.* An enzyme basis for blood type A intermediate status. *Am. J. Hum. Genet.* **34**(6), 919–924 (1982).
17. Blancher, A., Klein, J. & Socha, W. W. *Molecular Biology and Evolution of Blood Group and MHC Antigens in Primates* (Springer, 2012).
18. [https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1471358881\\_MMEQqLMEDZ3GILBMnlu8gNmV8pwA&db=hg19&c=chr19&g=primateChainNet](https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1471358881_MMEQqLMEDZ3GILBMnlu8gNmV8pwA&db=hg19&c=chr19&g=primateChainNet). Accessed 12 Oct 2022.
19. Koda, Y., Soejima, M. & Kimura, H. Structure and expression of H-type GDP-L-fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase gene (FUT1): Two transcription start sites and alternative splicing generate several forms of FUT1 mRNA. *J. Biol. Chem.* **272**, 7501–7505 (1997).
20. Koda, Y., Soejima, M. & Kimura, H. Changing transcription start sites in H-type alpha(1,2)fucosyltransferase gene (FUT1) during differentiation of the human erythroid lineage. *Eur. J. Biochem.* **256**, 379–387 (1998).
21. Apoil, P. A. *et al.* Evolution of alpha 2-fucosyltransferase genes in primates: Relation between an intronic Alu-Y element and red cell expression of ABH antigens. *Mol. Biol. Evol.* **17**, 337–351 (2000).
22. Yamamoto, F. Molecular genetics of ABO. *Vox Sang.* **78**, 91–103 (2000).
23. Sano, R. *et al.* Expression of ABO blood-group genes is dependent upon an erythroid cell-specific regulatory element that is deleted in persons with the B<sub>m</sub> phenotype. *Blood* **119**, 5301–5310 (2012).
24. Sano, R. *et al.* A 3.0-kb deletion including an erythroid cell-specific regulatory element in intron 1 of the ABO blood group gene in an individual with the B<sub>m</sub> phenotype. *Vox Sang.* **108**, 310–313 (2015).
25. Nakajima, T. *et al.* Mutation of the GATA site in the erythroid cell-specific regulatory element of the ABO gene in a B<sub>m</sub> subgroup individual. *Transfusion* **53**, 2917–2927 (2013).
26. Takahashi, Y. *et al.* Deletion of the RUNX1 binding site in the erythroid cell-specific regulatory element of the ABO gene in two individuals with the A<sub>m</sub> phenotype. *Vox Sang.* **106**, 167–175 (2014).
27. Ying, Y. *et al.* A novel mutation +5904 C>T of RUNX1 site in the erythroid cell-specific regulatory element decreases the ABO antigen expression in Chinese population. *Vox Sang.* **113**, 594–600 (2018).
28. Hult, A. K. *et al.* Disrupted RUNX1 motifs in the ABO gene explain samples with A<sub>3</sub> and B<sub>3</sub> phenotypes [abstract 3A-S04-03]. *Vox Sang.* **115**, 15 (2020).
29. Oda, A. *et al.* A novel mutation of the GATA site in the erythroid cell-specific regulatory element of the ABO gene in a blood donor with the A<sub>m</sub>B phenotype. *Vox Sang.* **108**, 425–427 (2015).
30. Fennell, K. *et al.* Effect on gene expression of three allelic variants in GATA motifs of ABO, RHD, and RHCE regulatory elements. *Transfusion* **57**, 2804–2808 (2017).
31. Kominato, Y. *et al.* Alternative promoter identified between a hypermethylated upstream region of repetitive elements and a CpG island in human ABO histo-blood group genes. *J. Biol. Chem.* **277**, 37936–37948 (2002).
32. Hata, Y. *et al.* Characterization of the human ABO gene promoter in erythroid cell lineage. *Vox Sang.* **82**, 39–46 (2002).
33. Sano, R. *et al.* Epithelial expression of human ABO blood-group genes is dependent upon a downstream regulatory element functioning through an epithelial cell-specific transcription factor, Elf5. *J. Biol. Chem.* **291**, 22594–22606 (2016).
34. Miyabe-Nishiwaki, T. *et al.* Propofol infusions using a human target controlled infusion (TCI) pump in chimpanzees (*Pan troglodytes*). *Sci. Rep.* **11**, 1214 (2021).
35. Schenkel-Brunner, H. *Human Blood Groups* 2nd edn. (Springer, 2000).
36. [https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1471358881\\_MMEQqLMEDZ3GILBMnlu8gNmV8pwA&db=hg19&c=chr19&g=rmsk](https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1471358881_MMEQqLMEDZ3GILBMnlu8gNmV8pwA&db=hg19&c=chr19&g=rmsk). Accessed 12 Oct 2022.
37. Vargiu, L. *et al.* Classification and characterization of human endogenous retroviruses; mosaic forms are common. *Retrovirology* **13**, 7 (2016).
38. [https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1471381413\\_vLgN55RboFbWBu8Fbd5ubUaijRT&db=hg38&c=chrX&g=encodeCcreCombined](https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1471381413_vLgN55RboFbWBu8Fbd5ubUaijRT&db=hg38&c=chrX&g=encodeCcreCombined). Accessed 12 Oct 2022.
39. [https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1471358881\\_MMEQqLMEDZ3GILBMnlu8gNmV8pwA&db=hg19&c=chr19&g=geneHancer](https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=1471358881_MMEQqLMEDZ3GILBMnlu8gNmV8pwA&db=hg19&c=chr19&g=geneHancer). Accessed 12 Oct 2022.
40. Sano, R. *et al.* A cell-specific regulatory region of the human ABO blood group gene regulates the neighbourhood gene encoding odorant binding protein 2B. *Sci. Rep.* **11**, 7325 (2021).
41. Yu, H. L., Zhao, Z. K. & Zhu, F. The role of human endogenous retroviral long terminal repeat sequences in human cancer (review). *Int. J. Mol. Med.* **32**, 755–762 (2013).
42. Durnaoglu, S., Lee, S. K. & Ahnn, J. Human endogenous retroviruses as gene expression regulators: Insights from animal models into human diseases. *Mol. Cells.* **44**, 861–878 (2021).
43. Medstrand, P., Landry, J. R. & Mager, D. L. Long terminal repeats are used as alternative promoters for the endothelin B receptor and apolipoprotein C-I genes in humans. *J. Biol. Chem.* **276**, 1896–1903 (2001).
44. Yu, X. *et al.* The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active LTR enhancer complex NF-Y/MZF1/GATA-2. *J. Biol. Chem.* **280**, 35184–35194 (2005).
45. Ruda, V. M. *et al.* Tissue specificity of enhancer and promoter activities of a HERV-K (HML-2) LTR. *Virus Res.* **104**, 11–16 (2004).
46. Medstrand, P. *et al.* Impact of transposable elements on the evolution of mammalian gene regulation. *Cytogenet. Genome Res.* **110**, 342–352 (2005).
47. Fueyo, R., Judd, J., Feschotte, C. & Wysocka, J. Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* **23**, 481–497 (2022).
48. Dunn, C. A., Medstrand, P. & Mager, D. L. An endogenous retroviral long terminal repeat is the dominant promoter for human beta1,3-galactosyltransferase 5 in the colon. *Proc. Natl. Acad. Sci. USA* **100**, 12841–12846 (2003).
49. Nakajima, T. *et al.* ABO alleles are linked with haplotypes of an erythroid cell-specific regulatory element in intron 1 with a few exceptions attributable to genetic recombination. *Vox Sang.* **110**, 90–92 (2016).

## Acknowledgements

This work was supported by the Cooperative Research Program of the Primate Research Institute, Kyoto University (2021-B-7). This work was also supported by the Fostering Health Professionals for Changing Needs of Cancer by MEXT of Japan and Gunma University Initiative for Advanced Research (GIAR). Moreover, this work was supported in part by JSPS KAKENHI Grant Numbers 20K10551 to R.S. and 19H03916 to Y.K.

### Author contributions

R.S. conceived, designed, coordinated, performed research, analyzed data, and wrote the paper; H.F., R.K., Y.T., A.H., and S.Y. performed research; T.O., T.M.-N., A.K., and H.M. collected sample of monkeys; Y.K. conceived, designed, coordinated, analyzed data, and wrote the paper. All authors reviewed the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to R.S.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023