MINI REVIEW

A historical overview of advances in molecular genetic/genomic studies of the ABO blood group system

Fumiichiro Yamamoto[1]

Abstract

In 1990, 90 years after the discovery of ABO blood groups by Karl Landsteiner, my research team at the Molecular Biology Laboratory of the now-defunct Biomembrane Institute elucidated the molecular genetic basis of the ABO polymorphism. Henrik Clausen, Head of the Immunology Laboratory, initiated the project by isolating human group A transferase (AT), whose partial amino acid sequence was key to its success. Sen-itiroh Hakomori, the Scientific Director, provided all the institutional support. The characterization started from the 3 major alleles (A1, B, and O), and proceeded to the alleles of A2, A3, Ax and B3 subgroups and also to the *cis*-AB and B(A) alleles, which specify the expression of A and B antigens by single alleles. In addition to the identification of allele-specific single nucleotide polymorphism (SNP) variations, we also experimentally demonstrated their functional significance in glycosyltransferase activity and sugar specificity of the encoded proteins. Other scientists interested in blood group genes later characterized more than 250 ABO alleles. However, recent developments in next-generation sequencing have enabled the sequencing of millions of human genomes, transitioning from the era of genetics to the era of genomics. As a result, numerous SNP variations have been identified in the coding and noncoding regions of the ABO gene, making ABO one of the most studied loci for human polymorphism. As a tribute to Dr. Hakomori's scientific legacy, a historical overview in molecular genetic/genomic studies of the human ABO gene polymorphism is presented, with an emphasis on early discoveries made at his institute.

Keywords ABO blood group system · ABO genes · A and B glycosyltransferases (AT and BT) · A and B oligosaccharide antigens

Introduction

The discovery of ABO blood groups by Karl Landsteiner revolutionized risky blood transfusion into a substantially safer medical practice [1, 2]. In 1900, he separated the cellular component composed primarily of red blood cells (RBCs) and the liquid component of sera from the blood of himself and his colleagues at the University of Vienna Pathology Laboratory, and mixed them in combinations. He observed RBC agglutination in some combinations,

---

This mini review is dedicated to the memory of the mentor, Dr. Sen-itiroh Hakomori.

This article belongs to the Topical Collection: Tribute to Professor Sen-itiroh Hakomori

---

✉ Fumiichiro Yamamoto
fyamamoto@carrerasresearch.org

1 Laboratory of Immunohematology and Glycobiology, Josep Carreras Leukaemia Research Institute (IJC), Badalona, Spain

but not in others. Based on this discovery, he proposed a protocol to transfuse only compatible blood without RBC agglutination in a pre-transfusion test. Landsteiner well explained the potential mechanism of RBC agglutination, assuming two antigens, later named A and B, on RBCs and antibodies against these antigens in the sera of individuals lacking the expression of these antigens (Landsteiner's Law). The chemical nature of these antigens was revealed in the late 1950s when Elvin Kabat and separately Walter Morgan and Winifred Watkins demonstrated that these antigens are glycans with specific structures, $GalNAc\alpha1\text{-}>3(Fuc\alpha1\text{-}>2)$ $Gal\text{-}$ and $Gal\alpha1\text{-}>3(Fuc\alpha1\text{-}>2)Gal\text{-}$ for antigens A and B, respectively [3, 4]. Together with substance H ($Fuc\alpha1\text{-}>2Gal\text{-}$) found in individuals of group O, these structures led Morgan and Watkins [5], and Ruggero Ceppellini [6] separately, to propose a hypothesis on the biosynthesis of antigens A and B from the common precursor H substance by transfer of a GalNAc and a

galactose by group A and B glycosyltransferases (AT and BT) encoded by functional alleles A and B at the ABO genetic locus, respectively. Victor Ginsburg and Akira Kobata later showed the concordant presence/absence of AT and BT activity in tissues from individuals of different ABO groups [7]. However, the definitive demonstration of ABO's central dogma, namely that alleles A and B encode AT and BT, which synthesize glycan antigens A and B, respectively, had to wait until 1990 when we cloned A, B, and O allelic cDNAs and elucidated the molecular genetic basis of the ABO polymorphism [8, 9].

Cloning of human group A transferase cDNA

Once an enzymatic assay was designed to measure AT and BT activities *in vitro*, isolation of AT and BT proteins was attempted. In fact, isolation to homogeneity was claimed [10, 11], but no follow-up studies were published. Taking advantage of our expertise in DNA transfection techniques, we tried to clone human AT cDNA by expression cloning. We searched for appropriate cells to host transfection and found that HeLa cells of human uterine cancer origin express cell surface H substance, the acceptor substrate for the transfer of a GalNAc and a galactose by AT and BT, respectively. We made, in the eukaryotic expression plasmid vector pSG-5, a cDNA expression library of the stomach cancer cell line MKN45, which was previously shown to exhibit strong AT activity and express abundant A antigens. We obtained a library consisting of 2×10^5 independent plasmid clones. Because the size of the library was small, and because this number included clones containing a cDNA in the wrong orientation, we planned to scale up the library construction by 5 times before starting expression cloning using HeLa cells. Meanwhile, Henrik Clausen and his colleague Thayer White isolated two protein fractions from human lungs, which appeared to represent human AT [12]. These fractions did not show any AT activity because the final reverse phase chromatography isolation protocol abolished the activity. However, these fractions were only present in the lungs of group A individuals, but not in the lungs of group O individuals. Subsequently, Koiti Titani and Koji Takio determined partial amino acid sequences of trypsinized peptides of these proteins. We searched the sequence database for proteins that possessed sequence homology to these proteins. Although the number of the genes whose sequences were determined was quite small at the time, we found that one of the two proteins was identical to the cystic fibrosis transmembrane conductance regulator (CFTR) protein. Apparently, one or a few individuals in group A, whose lungs were used as a source of AT, had cystic fibrosis. On the other hand, the partial amino acid

sequence of the other protein did not share any homology with previously sequenced proteins. Encouraged by Henrik Clausen and Sen-itiroh Hakomori, we decided to clone the cDNA encoding this unknown protein.

For this, we prepared another MKN45 cDNA library, but in a lambda gt10 phage vector. The cDNA was synthesized with poly (A)$^+$ RNA prepared from MKN45 cells using random hexamers by the method of Gubler and Hoffman. After cDNA methylation using EcoRI methylase, phosphorylated EcoRI linkers were added to the ends of the cDNA fragments. The cDNA fragments were then digested with EcoRI, and electrophoresed through a 1% agarose gel. cDNA fragments larger than 1.3 kilo base pairs (kbp) were size-selected and ligated to the dephosphorylated EcoRI arms of the lambda gt10 vector and packaged *in vitro*. The library consisting of more than 1×10^6 independent phage clones was obtained. The cDNA library was screened, using ^{32}P-end-labelled degenerate oligonucleotides that were reverse-translated from the partial amino acid sequence of the isolated protein. After 3 rounds of selection, dozens of candidate clones were isolated. Phage DNA was prepared, digested with EcoRI, and ligated with dephosphorylated EcoRI arms of plasmid pT7T3 or phagemid Phagescript SK. After DNA transformation of XL-1 Blue strain of *Escherichia coli* bacteria, the clones with insert were identified by color selection with Isopropyl-β-D-thiogalactoside (IPTG) and 5-Bromo-4-Chloro-3-Indolyl β-D-Galactopyranoside (X-Gal). Next, the DNA sequences of the inserts were determined. However, none of the sequences of the identified phage clones showed any homology to the protein sequence.

After repeated failures using different degenerate oligonucleotide probes, we changed the experimental strategy. Instead of oligodeoxynucleotides of low specificity due to degeneracy, we attempted to obtain a longer and unique DNA fragment probe by PCR using degenerate oligonucleotide primers [8]. In addition to genomic DNA, cDNA from MKN45 cells was also used as templates because intron sequences could be present between those 2 primers. When we analyzed the products of the PCR reaction by agarose gel electrophoresis, multiple bands of DNA, as well as a DNA smear, were observed in those amplifications of genomic and complementary DNA. The DNA was electrically transferred from the gel to a nylon membrane and fixed by UV irradiation. Subsequently, the membrane was hybridized with ^{32}P-end-labelled degenerate oligodeoxynucleotides corresponding to the internal amino acid sequence between the two amino acid sequences whose corresponding nucleotide sequences were used as PCR primers. After exposure to the X-ray film, a strong signal from a band located between 72 and 118 bp was observed. The gel fragment corresponding to the band was excised and immersed in distilled water. The eluate was used as a template for PCR using the same primer pair of oligodeoxynucleotides. The PCR amplified

products were then cloned into plasmid pT7T3 and used to transform frozen competent bacteria. Plasmid DNA was isolated from the transforming clones and subjected to DNA sequencing. Several clones contained the 98 bp DNA fragment, which corresponded to the expected size of the partial amino acid sequence of the protein. Bacteria were grown on a larger scale and plasmid DNA was prepared. After digestion, the 98 bp DNA fragment was excised and used to prepare the ^{32}P-labelled probe by PCR. With the help of Tsutomu Tsuji and John Marken, we then screened the MKN45 cDNA library prepared in the λgt10 vector, using the 98 bp probe. After 3 rounds of selection, we isolated several dozen positive clones. We then subcloned a DNA fragment containing the longest insert into a plasmid vector to prepare the ^{32}P-labelled probe by PCR to further screen the cDNA library. In this way, we were able to improve the probe specificity, starting with degenerate primers of 26-mer (a degeneracy of 576) and 29-mer (a degeneracy of 144) (triplet codons that are rarely used in humans were excluded to decrease degeneracy) to 98 bp with potentially mismatched pairs, and to a DNA fragment greater than 1 kbp of unique sequence. Indeed, this use of degenerate oligodeoxynucleotides for PCR primers and the selection of a single longer probe by hybridization with internal degenerate oligonucleotides was a novel strategy.

Elucidation of the molecular genetic basis of ABO polymorphism

After successful screening of cDNA clones encoding the isolated protein from the MKN45 library, we prepared 4 additional cDNA libraries using poly (A) + RNA isolated from colon cancer cell lines known to exhibit different ABO phenotypes [9]. We assumed that if the isolated protein was a true AT, the sequences of the cDNA clones to be isolated would reveal allele-specific single nucleotide polymorphism (SNP) variations. The cell lines used were: SW948 (phenotype O), SW48 (phenotype AB), COLO205 (phenotype O) and SW1417 (phenotype B). Among them, SW948, SW48, and SW1417 were known to be derived from individuals with the ABO blood types of O, AB, and B, respectively. However, the ABO blood type of the individual from whom the COLO205 cell line was derived was not known. From each screen, we obtained multiple cDNA clones that hybridized to the long specific cDNA probe isolated from the MKN45 library. Representative clones containing a long insert (> 1 kbp) were selected and the inserts were subcloned into a plasmid vector. Nucleotide sequencing was performed for individual clones. We found that all the clones had high sequence homology, but were different at the ends. We also observed some differences that could have resulted from premature splicing. The cDNA clones

were classified into groups, based on the sequence differences at the corresponding positions. The cDNA clones from the SW48 library were divided into 2 groups. The cDNAs whose sequences were more homologous to those of the MKN45 library were assumed to represent the A allele, while those with less homology represented the B allele. The cDNA clones from the SW948 library showed identical sequences although the ends of the fragments and the splicing patterns were diverse. Compared to the sequences assumed to represent the A allele of SW48, the corresponding sequences were nearly identical with the exception of a single nucleotide deletion of guanine at nucleotide 261 of the coding sequence, c.261delG, which causes a frameshift in codon reading. Because this deletion was located relatively close to the N-terminus of the protein, it explained well the inactivity of the O allele. Therefore, the sequences were assumed to represent the O allele. The cDNA clones from the COLO205 library showed the same sequences at corresponding positions, although splicing patterns and coverage regions varied. Compared to the O allele of the SW948 library, they also possessed the unique guanine deletion, but also contained several additional nucleotide substitutions. We assumed that they also represented the O allele due to the single nucleotide deletion. The SW1417 library cDNA clones were separated into 2 groups. The sequences of one group showed the same sequences corresponding to the B allele of the SW48 library, while those of the other group were identical to the sequences corresponding to the O allele of the SW948 library. This finding had an important implication. We previously knew that SW 1417 cells were derived from a colon cancer patient whose blood type was B. But with the results, we were able to deduce that the individual's genotype was BO and not BB. This actually became the first successful ABO genotyping. Furthermore, the results also demonstrated that the isolated protein was a true AT.

We found that AT and BT were comprised of 354 amino acids encoded by 1,065 nucleotides, including the termination codon (353 amino acids and 1,062 nucleotides in some, due to the loss of 3 nucleotides at the splice site). Allelic cDNAs A and B were highly homologous, but there were several nucleotide substitutions between them, among which 4 resulted in amino acid substitutions between AT and BT. They were arginine, glycine, leucine and glycine in codons 176, 235, 266 and 268 of the AT, while they were glycine, serine, methionine and alanine in these codons of the BT: c.526C > G (p.Arg176Gly), c.703G > A (p.Gly235Ser), c.796C > A (p.Leu266Met), and c.803G > C (p.Gly268Ala). We identified 2 types of O allelic cDNAs that were different by several nucleotide substitutions (c.297A > G, c.646 T > A, c.681G > A, c.771C > T, c.829G > A). Still, both had a single nucleotide deletion, c.261delG. Due to the codon reading frameshift (p.Thr88fs), these O allelic cDNAs were assumed to be translated into truncated proteins, possibly without AT

activity. To conclude, by cloning the human AT cDNA and B and O allelic cDNAs followed by determination of their nucleotide and deduced amino acid sequences, we were able to elucidate the molecular genetic basis of the ABO blood group polymorphism. Assuming that 99 alleles would be sufficient for some time, we later proposed an allele nomenclature composed of the phenotype plus a 2-digit number in the order of discovery. We named the A allele of SW48 as the A101 allele, and used the nucleotide and deduced amino acid sequences of this allele as standards. Similarly, the B alleles in SW48 and SW1417 were named B101 alleles because they possessed identical sequences at the corresponding positions. Furthermore, the O alleles of SW948 and SW1417 were designated O01 alleles, while the O alleles of COLO205 were designated O02 alleles. Compared to the nucleotide sequence of the A101 allele, the O01 and O02 alleles contained the same c.261delG, but the O02 allele possessed several additional nucleotide substitutions. The A alleles of MKN45 had 1 nucleotide substitution (c.467C > T) resulting in the amino acid substitution (p.Pro156Leu). Hence, we called it the A102 allele.

It should be noted that we were able to elucidate the allelic basis of the ABO polymorphism without using human RBCs. Instead, we used human cancer cell lines of the stomach and colon. This was possible only because those cells were known to express ABO differential phenotypes, exemplifying the importance of selecting the most suitable experimental system for scientific demonstration. It should also be emphasized that this elucidation was possible because the isolated protein was a true AT. Subsequently, we found, by PCR, the presence of the AT cDNA in the MKN cDNA library that we prepared in pSG-5. Therefore, it is possible that expression cloning would also have been successful. However, considering the competition with others, our timely cloning owed a lot to Henrik Clausen and Sen-itiroh Hakomori.

Our characterization of additional ABO alleles

As mentioned above, some cDNA clones isolated from cDNA cloning contained unspliced intron sequences. Taking advantage of the sequences, we showed that the sequences of the last two coding exons could be amplified by PCR of genomic DNA, using appropriate oligonucleotide primers. The portion of proteins encoded in these exons covered 90% of the coding sequence of the soluble form of AT [12]. Using PCR followed by cloning into a DNA sequencing vector, Patricia McNeill, Miyako Yamamoto and I determined the partial nucleotide sequences and deduced the amino acid sequences of the following A and B subgroup alleles (one A2 allele, one A3 allele, one Ax allele, and one B3 allele), in collaboration

with Teresa Harris (Pacific Northwest Regional Blood Services), W. John Judd and Robertson Davenport (The University of Michigan Hospital), and Imelda Bromilow and Jennifer Duguid (National Blood Transfusion Service, Liverpool). Compared to the corresponding sequences of the A101 allele, the A2 allele (A201) that we characterized [13] contained a single nucleotide deletion, c.1060delC, in one of the three cytosine residues (3Cs) at nucleotides 1059–1061, resulting in a frameshift at the C-terminal end of the coding sequence p.Pro354fs, in addition to the nucleotide/amino acid substitution (c.467C > T p.Pro156Leu) that was also found in the A102 allele. The A3 allele (A301) was identical to the A101 allele except for a single nucleotide substitution (c.871G > A) that resulted in an amino acid substitution (p.Asp291Asn) [14]. We analyzed several samples showing the A3 phenotype-specific mixed-field agglutination with anti-A antibodies. However, not all samples contained this mutation, indicating that there was heterogeneity among the A3 samples. The Ax01 allele that we characterized contained a single nucleotide substitution that resulted in an amino acid substitution at another position (c.646 T > A p.Phe216Ile) [15]. The B3 (B301) allele was found to possess a single nucleotide substitution (c.1054C > T) resulting in an amino acid substitution (p.Arg352Trp), compared to the sequences of the B101 allele [14]. In addition to these A and B subgroup alleles, we also molecularly characterized mutations in a *cis*-AB allele [16] and a B(A) allele [17], in collaboration with Yoshihiro Fujimura (Nara Medical University) and Teresa Harris. The *cis*-AB allele (*cis*-AB01) that we characterized was basically the A101 allele, except for the nucleotide/amino acid substitution (c.467C > T p.Pro156Leu) which was also found in the A102 and A201 alleles and an additional nucleotide/amino acid substitution (c.803G > C p.Gly268Ala) [18]. Since the latter substitution was also found in the B alleles, it was deduced that the *cis*-AB01 allele encoded a chimeric AT-BT protein. In contrast, the B(A)01 allele was basically the B101 allele except for a single nucleotide substitution (c.657 T > C) and a nucleotide/amino acid substitution (c.703A > G p.Ser235Gly) [15]. The latter was also found in the A alleles. Consequently, it was deduced that the B(A)01 allele encoded another type of chimeric AT-BT protein. While we were determining the nucleotide and deduced amino acid sequences of the B301 allele, we found a strange allele that appeared to be another type of O allele. Compared to the standard A101 sequences, this allele had a synonymous mutation (c.297A > G), and 2 nonsynonymous mutations (c.526C > G p.Arg176Gly and c.802G > A p.Gly268Arg). We called this allele O03. The inactivity of this allele was demonstrated by a functional analysis employing DNA transfection of the expression construct, which will be described later.

In total, we determined the complete nucleotide and deduced amino acid sequences of 2 A1 (A101 and A102),

one B (B101) and 2 O (O01 and O02) allelic cDNAs in all coding sequences. Furthermore, we also determined partial nucleotide and amino acid sequences of 1 A2 (A201), 1 A3 (A301), 1 Ax (Ax01), 1 B3 (B301), 1 *cis*-AB (*cis*-AB01), 1 B(A) (B(A)01), and 1 O (O03) alleles by sequencing the PCR-amplified genomic DNA fragments. The comparison of the nucleotide and deduced amino acid sequences of the ABO alleles that we characterized are shown in Fig. 1.

ABO genotyping

Once the allele-specific SNP variations were identified, we realized that several of them are located at the cleavage sites for certain restriction endonucleases [9]. For example, KpnI which recognizes the sequence 5'-GGT ACC and BstEII which recognizes the sequence 5'-GGT-NACC differentially cleave the sequences of ABO alleles with the c.261delG deletion and those without it. Similarly, at positions 1, 2 and 3 of 4 amino acid substitutions that discriminate between AT and BT, allele-specific SNP variations are cleaved by BssHII (5'-GCGCGC)/NarI (5'-GGCGCC), HpaII (5-CCGG)/AluI (5'-AGCT), and BstNI (5'-CCWGG)/NlaIII (5'-CATG) for A and O alleles versus B alleles, respectively. Consequently, we thought about designing ABO genotyping protocols, taking advantage of restriction fragment length polymorphism (RFLP). Allele-specific SNP variations used for original ABO genotyping by RFLP are shown in Fig. 2. Using Southern hybridization of BstEII or KpnI-digested genomic DNA

and a ^{32}P-labelled human A102 allelic cDNA probe, we first demonstrated that these differences in the nucleotide sequence were also present in genomic DNA, in addition to cDNA, of cell lines used for cDNA cloning. Unlike BstEII and KpnI, which are not frequent cutters, the other enzymes mentioned above cleave DNA very frequently, and thus Southern hybridization was difficult to perform due to the smaller sizes of the digested fragments. Therefore, we used PCR in combination with restriction enzyme digestion to determine the presence/absence of the cleavage sites for these enzymes. Practically, we amplified the DNA fragment containing the BssHII/NarI site or the DNA fragment containing the HpaII/AluI site, digested separately with these enzymes and analyzed by polyacrylamide gel electrophoresis. These sites were also confirmed to be present in the genomic DNA of those cell lines. In addition, we also examined genomic DNA isolated from 14 blood specimens from individuals with predetermined ABO phenotypes. No discrepancies were observed between blood ABO phenotypes and inferred ABO phenotypes from the RFLP results of ABO genotyping, indicating that these SNP variations are widely present in the human population. It should also be emphasized that thanks to the successful ABO genotyping, discrimination between genotypes A/A and A/O or between genotypes B/B and B/O was made possible, which was not achieved by immunohematological/serological methods. It should also be noted that we knew that simple PCR followed by DNA sequencing would also be useful for ABO genotyping, in addition to RFLP. In reality, a variety of

Exon	6		7														
Nucleotide Position	261	297	467	526	646	657	681	703	771	796	802	803	829	871	930	1054	1060
A alleles																	
A101	G	A	C	C	T	C	G	G	C	C	G	G	G	G	G	C	C
A102	*	*	T	*	*	*	*	*	*	*	*	*	*	*	*	*	*
A201	*	*	T	*	*	*	*	*	*	*	*	*	*	*	*	*	Δ
A301	*	*	*	*	*	*	*	*	*	*	*	*	*	A	*	*	*
Ax01	*	*	*	*	A	*	*	*	*	*	*	*	*	*	*	*	*
B alleles																	
B101	*	G	*	G	*	T	*	A	*	A	*	C	*	*	A	*	*
B301	*	G	*	G	*	T	*	A	*	A	*	C	*	*	A	T	*
AB alleles																	
cis-AB01	*	*	T	*	*	*	*	*	*	*	*	C	*	*	*	*	*
B(A)01	*	G	*	G	*	*	*	*	*	A	*	C	*	*	A	*	*
O alleles																	
O01	Δ	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
O02	Δ	G	*	*	A	*	A	*	T	*	*	*	A	*	*	*	*
O03	*	G	*	G	*	*	*	*	*	*	A	*	*	*	*	*	*
Possible Amino Acid Change	Frameshift (p.Thr88fs)	No change	p.Pro156Leu	p.Arg176Gly	p.Phe216Ile	No change	No change	p.Gly235Ser	No change	p.Leu266Met	p.Gly268Arg	p.Gly268Ala	No change	p.Asp291Asn	No change	p.Arg352Trp	Frameshift (p.Pro354fs)

Fig. 1 Comparison of the nucleotide and deduced amino acid sequences of the twelve ABO alleles that we molecular genetically characterized. Only the differences of the nucleotide and deduced amino acid sequences from the A101 allele are shown. Similar tables have been published previously [20, 34, 41, 42]

Fig. 2 ABO genotyping by allele-specific restriction fragment length polymorphism. ABO alleles with c.261delG and those without it can be discriminated by digestion with restriction endonucleases, KpnI and BstEII, respectively. Similarly, at positions 1, 2 and 3 of the 4 amino acid substitutions that discriminate AT and BT, A/O alleles and B alleles can be discriminated by cleavage by BssHII and NarI, HpaII and AluI, or BstNI and NlaIII, respectively

A/B allele

```
                              BstEII
            GTCCTCGTGGT  G  ACCCCTTGGCTGG
                        261
            GTCCTCGTGGT  -  ACCCCTTGGCTGG
                         KpnI
```

O allele

A/O allele

```
                              BssHII
            TGCTGGAGGTG  C  GCGCCTACAAGCG
                        526
            TGCTGGAGGTG  G  GCGCCTACAAGCG
                         NarI
```

B allele

A/O allele

```
                              HpaII
            CCCTGCACCCC  G  GCTTCTACGGAAG
                        703
            CCCTGCACCCC  A  GCTTCTACGGAAG
                         AluI
```

B allele

A/O allele

```
                              BstNI
            ATTTCTACTAC  C  TGGGGG  G  GTTCTTCGGG
                        796          803
            ATTTCTACTAC  A  TGGGGG  C  GTTCTTCGGG
                         NlaIII
```

B allele

modifications have been applied to ABO genotyping over the past 30 years, and ABO genotyping is now routinely performed in immunohematology reference laboratories.

Demonstration of the functional significance of the identified mutations

As mentioned above, we initially planned to clone human AT cDNA by expression cloning. Because the correlation between the putative ABO alleles and their unique nucleotide and amino acid sequences clearly demonstrated the identity of the isolated protein as human AT, we did not restart the backup plan of expression cloning. However, the

experimental system that we established for this purpose turned out to be quite useful in examining the functionality of the SNPs/mutations that we identified from various ABO alleles. HeLa cells expressing cell surface H substance were thought to provide a useful host for DNA transfection. However, the possibility existed that the A and/or B antigens were not expressed in these cells due to the absence of transcription of the ABO gene, rather than a structural deficiency of the encoded protein. In that case, the results can be confused to interpret, because the DNA transfection procedure alone could activate gene expression. Therefore, we first determined the ABO genotype of HeLa cells [19]. We took advantage of the KpnI RFLP specific to the O01/O02 alleles. Although the KpnI digestion was partial, no

Construction of eukaryotic expression constructs
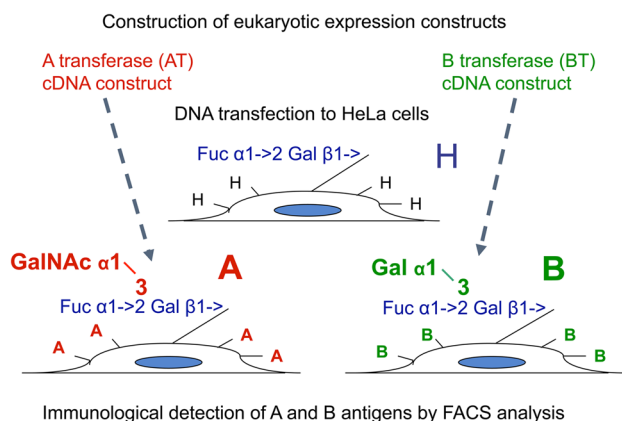
Fig. 3 The functional assay protocols are shown schematically. Expression constructs are prepared in pSG-5 of the A101 and B101 allelic cDNAs encoding AT and BT, respectively. DNA is transfected into HeLa cells, and the appearance of A and B antigens on the cell surface is immunologically detected, using anti-A and anti-B antibodies, respectively. Once the system is established, expression constructs possessing mutations/SNPs are then analyzed

BstEII digestion was observed. Therefore, we concluded that HeLa cells are homozygous for the O alleles with c.261delG. With this assurance, we chose these cells to examine the functionality of the SNPs/mutations that we identified in 12 ABO alleles.

The protocols of the functional assay we designed are shown schematically in Fig. 3. We first elaborated the expression constructs of the A101 and B101 allelic cDNAs that contained the complete coding sequences of AT and BT, respectively, in the correct orientation in pSG-5. DNA transfection resulted in the appearance of A and B antigens on the cell surface, which were immunologically detectable using anti-A and anti-B antibodies, respectively [19]. Next, we introduced the single nucleotide c.261delG deletion specific for alleles O01 and O02 in construct A101. The DNA from these constructs was co-transfected with a fixed amount of pSG-5 vector DNA to express green fluorescent protein (GFP) to normalize differences in transfection efficiency among different constructs. A antigens were not detected, probably because the deletion caused a codon reading frameshift in AT translation and abrogated transferase activity [20]. When we introduced the 2 nucleotide substitutions that resulted in amino acid substitutions of the O03 allele (c.526C > G p.Arg176Gly and c.802G > A p.Gly268Arg), A antigens were also not detected [21]. We later showed that p.Gly268Arg, and not p.Arg176Gly, was responsible for the inactivation of AT activity [22]. When we introduced the specific mutation of the A201 or A301 allele, A antigens were still expressed, but decreased in number [13, 20]. The A201 allele had an amino acid substitution (c.467C > T p.Pro156Leu) and a single nucleotide deletion (c.1060delC) at the C-terminus of the encoded protein,

causing a reading frameshift of codons and an extension of 21 additional amino acids p.Pro354fs. We later showed that c.1060delC was responsible for the decrease in AT activity, possibly due to the steric hindrance of the additional domain that partially inhibited AT access to the reaction substrates [20]. The results are summarized in Fig. 4.

Critical amino acids for differential sugar specificity between AT and BT were investigated using functional assays [19]. As mentioned above, we found 4 amino acid substitutions between these two enzymes. They are p.Arg176Gly, p.Gly235Ser, p.Leu266Met and p.Gly268Ala (AT versus BT). To examine which amino acid substitutions are important in the difference in specificity of GalNAc versus galactose, we prepared 14 AT-BT chimeras that possessed AT or BT amino acids at those 4 positions. The original plasmid constructs of AT and BT were named pAAAA and pBBBB, respectively. The chimeras were: pAAAB, pAABA, pAABB, pABAA, pABAB, pABBA, pABBB, pBAAA, pBAAB, pBABA, pBABB, pBBAA, pBBAB, and pBBBA. The DNA of these constructs was transfected separately into HeLa cells, and cell surface expression of A and B antigens was quantified by immunocytometry using a fluorescence-activated cell sorter (FACS) after the first staining with anti-A and anti-B antibodies, respectively, and later with FITC-conjugated secondary antibody. The results are shown in Fig. 5. They can be summarized as follows. When the constructs had AT amino acids in the last two positions (pAAAA, pABAA, pBAAA, and pBBAA), the constructs showed a strong AT activity. Similarly, when the constructs had BT amino acids in the last two positions (pAABB, pABBB, pBABB, and pBBBB), the constructs showed a strong BT activity. When the constructs had the AT and BT amino acids in this order in the last two positions (pAAAB, pABAB, pBAAB, and pBBAB), the activity depended on the amino acid in the 2nd position of the 4 amino acid

Construct	Possible Gene	Expression	
		A antigen	B antigen
A1 construct	A101	++++	-
A1 (p.Pro156Leu)	A102	++++	-
A1 (c.1060delC)	A2	++	-
A1 (p.Pro156Leu, c.1060delC)	A201	++	-
A1 (p.Asp291Asn)	A301	++	-
A1 (p.Phe216Ile)	Ax01	ND	ND
A1 (c.261delG)	O01	-	-
A1 (p.Arg176Gly, p.Gly268Arg)	O03	-	-
B construct	B101	-	++++
B (p.Arg352Trp)	B301	ND	ND

Fig. 4 Functional significance of the mutations (SNP variations) that we identified in the ABO alleles. DNA from the A1 construct (AT), B construct (BT), and their derivative constructs possessing the mutations/SNPs was transfected separately into HeLa cells, and cell surface expression of A and B antigens was quantified by FACS. The antigen expression levels are schematically shown: + + + + (very strong); + + (strong);—(no); and ND (not determined). Similar tables have been published previously [20, 41]

Plasmid DNA	Experiment 1		Experiment 2		Experiment 3		Activity
	A	B	A	B	A	B	
pAAAA	41.5	0.0	3.8	0.0	14.0	0.1	A
pAAAB	17.5	0.3	1.2	0.0	7.4	0.1	A
pAABA	NT	NT	1.0	0.6	2.4	1.1	AB
pAABB	0.2	26.3	0.1	1.4	0.1	5.6	B
pABAA	27.5	0.2	5.7	0.1	11.6	0.1	A
pABAB	21.3	3.0	1.8	0.1	5.8	0.2	A(B)
pABBA	17.0	22.1	0.8	1.4	2.1	2.9	AB
pABBB	0.1	31.1	0.1	1.6	0.1	5.5	B
pBAAA	29.1	0.1	2.9	0.1	10.3	0.0	A
pBAAB	10.0	0.1	0.5	0.0	4.8	0.1	A
pBABA	NT	NT	0.5	0.4	3.1	1.3	AB
pBABB	0.1	20.7	0.0	1.0	0.0	5.4	B
pBBAA	12.7	0.1	4.8	0.0	12.3	0.0	A
pBBAB	29.5	2.9	1.4	0.0	8.0	0.4	A(B)
pBBBA	NT	NT	1.0	0.6	3.2	2.0	AB
pBBBB	0.1	30.6	0.0	2.5	0.1	3.3	B
no DNA	0.0	0.1	0.0	0.0	0.0	0.0	

Fig. 5 Functional significance of amino acid substitutions between AT and BT. The DNA from the AT, BT, and 14 AT-BT chimeric expression constructs was transfected separately into HeLa cells, and cell surface expression of A and B antigens was quantified by FACS. The numbers indicate the percentages of cells expressing A/B antigens above background levels. The symbol "NT" indicates "not tested". Similar tables have been published previously [19, 41]

substitutions. When it was from AT, the constructs (pAAAB and pBAAB) showed only a strong AT activity. On the other hand, when it was from BT, the constructs (pABAB and pBBAB) showed a weak BT activity, in addition to a strong AT activity. And finally, when the constructs had the BT and AT amino acids in this order in the last two positions (pAABA, pABBA, pBABA, and pBBBA), strong activities of both AT and BT were observed. Based on these results, we reached the following conclusion: The amino acid substitution in the 1st position (arginine versus glycine) is not important for the differential specificity of sugars between AT and BT. Substitution of amino acids in the 2nd position (glycine versus serine) plays a role, but not much. The most important are the amino acid substitutions at positions 3 and 4. They are leucine and glycine in AT and methionine and alanine in BT, and these amino acids determined the sugar specificity of AT and BT. Subsequently, others determined the three-dimensional structures of the crystalized catalytic domains of human AT and BT, with and without substrates [23]. The results showed that codon 176 did not physically interact with the substrates, but codon 235 was in proximity, and codons 266 and 268 showed direct contact, which confirmed our results obtained from functional analysis.

Later, we expanded this line of enzymological analysis even further, preparing *in vitro* mutagenized derivative constructs that have any amino acid at codon 268 with the backbone of human AT or BT [22]. In these experiments, African green monkey kidney COS 1 cells were also used as host for DNA transfection, in addition to HeLa cells. The pSG-5 vector contains the SV40 promoter and the SV40 origin of replication. Because COS 1 cells express SV40 T antigens, the cDNA in pSG-5 was highly expressed by transient transfection. However, COS 1 cells do not express H substances and therefore AT and BT activities had to be examined using an *in vitro* enzymatic assay system. The DNA from the constructs was co-transfected with DNA from the control vector pSV-β-galactosidase to normalize the differences in transfection efficiency among the constructs. For COS 1 cells, cell extracts were prepared and used as sources of enzymes (AT, BT and β-galactosidase). For the detection of AT and BT activities, 2'-fucosyllactose was used as the acceptor substrate and ^{14}C-labelled UDP-GalNAc and UDP-galactose were used as donor substrates. After enzymatic reactions with appropriate cacodylate buffer and manganese ions in the presence of reaction substrates, neutral reaction products were separated from acidic nucleotide-sugars by AG1-X8 anion exchange column chromatography and quantified using a γ-scintillation counter. The activity of β-galactosidase was measured by hydrolysis of o-nitrophenyl-β-D-galactopyranoside, a substrate for β-galactosidase. For HeLa cells, cells were fixed after DNA transfection, and A and B antigens were detected by cytometry. The percentages of cells stained for β-galactosidase activity *in situ* were used to normalize the efficiency of transfection. Although experiments using HeLa and COS 1 cells showed similar results, the FACS results showed somewhat greater sensitivity due to higher background in *in vitro* enzymatic assays potentially caused by enzymes endogenous to COS 1 cells. The results confirmed our earlier finding that amino acids at codon 268 of AT and BT are critical for the determination of GalNAc/galactose specificity.

Recently, we used the same *in vitro* mutagenesis strategy to produce AT derivatives that contained various amino acid substitutions around codons 266–268, and codons 263–268 in some, which were found in the ABO genes in different species of organisms. We then determined AT and BT activities by DNA transfection to HeLa cells and generated a code table that associated amino acid sequences with AT and BT activities/specificities [24]. We combined the data in the table with the results of phylogenetic analysis of ABO genes in a variety of species and demonstrated that there are vertebrate species that possess multiple ABO genes and that several species even possess non-allelic A and B genes. These findings were not entirely new because rats had previously been shown to possess non-allelic A and B alleles. However, our results showed that multiple ABO genes in the genome are not confined to rats, but are observed widely. We also demonstrated that both horizontal and vertical transmissions have occurred during the evolution of bacterial ABO genes [24].

Genomic organization of the human ABO gene

To determine the organization of the ABO gene and better understand the molecular mechanisms that control gene transcription, we cloned genomic DNA that encompassed the human ABO gene [25]. For this purpose, we screened one million independent phage clones from a human genomic DNA library containing partially MboI-digested male leukocyte genomic DNA at the unique BamHl site of the λEMBL-3 vector, using a ^{32}P-radiolabelled AT cDNA probe primed with random hexamers. We obtained a positive clone, which contained most of the coding sequence except for the latter part of the last coding exon. We then screened 5 million clones from a human placenta genomic DNA library that contained partially Sau3AI-digested genomic DNA at the unique XhoI site of the λFIX™ II vector. We identified several positive clones. Preliminary analysis showed that they all covered most of the 3' end of the gene. We then re-screened 1 million additional phage clones from the λEMBL-3 library and isolated another positive clone covering the 5' region upstream of Exon 1, potentially containing the promoter and enhancer sequences. Restriction enzyme cleavage sites for EcoRI, SstI and SalI were determined, and a map was constructed. In total, the genomic DNA sequence spanning more than 30 kbp was cloned. Using Southern hybridization of the digested DNA fragments of inserts with a ^{32}P-labelled human AT cDNA probe, the exons were mapped. The nucleotide sequences of the genomic DNA around possible exon–intron boundaries were determined and compared with those of the cDNA. And finally, the precise exon–intron boundaries were determined. The coding sequences are dispersed in 7 coding exons (Exon 1–7). The hydrophobic region, which appears to be a possible transmembrane domain, was found mainly in Exon 3. The N-terminal end of the soluble form of AT (codon 54) was mapped into Exon 4. And most of the coding sequence was located in the last 2 exons (Exon 6 and 7). The single nucleotide deletion of c.261delG present in the O01/O02 alleles was located in Exon 6, while the 4 amino acid substitutions that discriminate AT and BT were in Exon 7, the largest exon of all. Furthermore, we used cloned genomic DNA to determine the transcription start sites of the ABO gene mRNAs and to characterize the transcriptional regulatory elements of the ABO gene. Yoshihiko Kominato, a former postdoctoral fellow, initiated work to define epithelial and erythroid cell-specific regulatory elements in ABO gene expression, in collaboration with us [26–28], and has later expanded independently [29, 30]. In addition, his research team also identified mutations in several alleles of A and B subgroups that caused a deficiency in gene expression (see Kominato's recent review of the ABO gene expression [31] for more details).

# Characterization of additional ABO alleles in the era of genetics

Millions of blood samples are determined for ABO phenotypes each year primarily to test transfusion compatibility. There are two kinds of tests. Forward typing examines RBCs for the presence or absence of A and B antigens, using anti-A and anti-B antibodies. Anti-A,B antibodies and anti-A1 *Dolichos biflorus* plant lectin can also be used. Reverse typing examines plasma/sera for the presence or absence of anti-A and anti-B antibodies, using reference type A1 and type B RBCs. Additionally, a cross-match is also performed to further ensure transfusion compatibility. The ABO typing of so many samples has led to the identification of discrepancies between the forward and reverse typing results in some, mainly due to A/B subgroups or diseases. Once we determined the nucleotide and amino acid sequences of the 12 ABO alleles and demonstrated that mutations in ABO alleles could be easily identified by genomic DNA PCR followed by DNA sequencing of the amplified fragments, some researchers took advantage of the sequence information that was available from our studies and characterized the ABO alleles in their samples showing such discrepancies. As the number of ABO alleles that were molecularly characterized increased, confusion ensued because different investigators sometimes identified the same alleles without knowing that others had previously characterized the same alleles due to the delay in publication. To house the information on the alleles of the blood group systems and avoid such a problem, Olga Blumenfeld established the Blood group antigen Gene MUTation (BGMUT) database in 1999 [32]. She assigned me as a curator of the ABO system. Following the guidelines of the Human Variation Genome Society, we used an allele nomenclature based on the ABO phenotype followed by the 2-digit number in the order of discovery. In 2006, BGMUT became part of the NCBI's dbRBC (database Red Blood Cells) resource at National Institute of Health. Over 250 ABO alleles were deposited into the database, including alleles specifying A1, A2, A3, Ael, Aint, Am, Aw, Ax, *cis*-AB, B(A), B (B1), B3, Bel, Bw, Bx, and O phenotypes. Many missense mutations, nonsense mutations, and frameshift mutations (insertions and deletions) identified in the coding sequence of the ABO gene have been extensively reviewed and therefore are not duplicated in this review (see review articles [33, 34] for detailed information on individual alleles and their references). Mutations have also been discovered that changed the mRNA splicing pattern of the ABO gene [35–37] or that altered the promoter/enhancer activity in the transcription of the ABO gene [31, 38, 39]. In 2016, BGMUT was closed due to the unfortunate death of Dr. Blumenfeld. However, data previously on the site became available on the website (ftp://ftp.ncbi.nlm.nih.

gov/pub/mhc/rbc/Final Archive). The International Society of Blood Transfusion (ISBT) subsequently organized the Working Party on red cell immunogenetics and blood group terminology and used an allele nomenclature similar to that of BGMUT.

Identification of numerous single nucleotide polymorphism (SNP) variations in the ABO gene in the era of genomics

Over the past decade, millions of human genomes have been sequenced, thanks to recent advances in next-generation sequencing technology. And many of the sequences are available in public databases, such as Genbank and Ensembl. As a result, thousands of SNP variations have been identified in the ABO gene. The nucleotide and deduced amino acid

sequences of the A101 allelic cDNA are shown in Fig. 6, along with the known SNP variations listed in the Ensembl database. Different types of SNP variations are highlighted in different colors. In this figure, the original data (ABO-203 ENST00000611156.4) of the O01 allele was modified to encode functional AT. Of 1,065 nucleotides in the coding region, at more than 500 positions, SNP variations have already been identified. Because the ABO gene is close to 25 kbp in length, the number of SNP variations that have been identified throughout the gene is much higher because the noncoding sequence occupies most of the genetic locus. Furthermore, this number increases rapidly as additional genomes are sequenced. If alleles are defined as genes that occupy the same genetic locus but have different sequences, the number of ABO alleles can be in the millions, considering SNP combinations. In 2012, predicting this situation, we proposed a new allele nomenclature based on the assortment
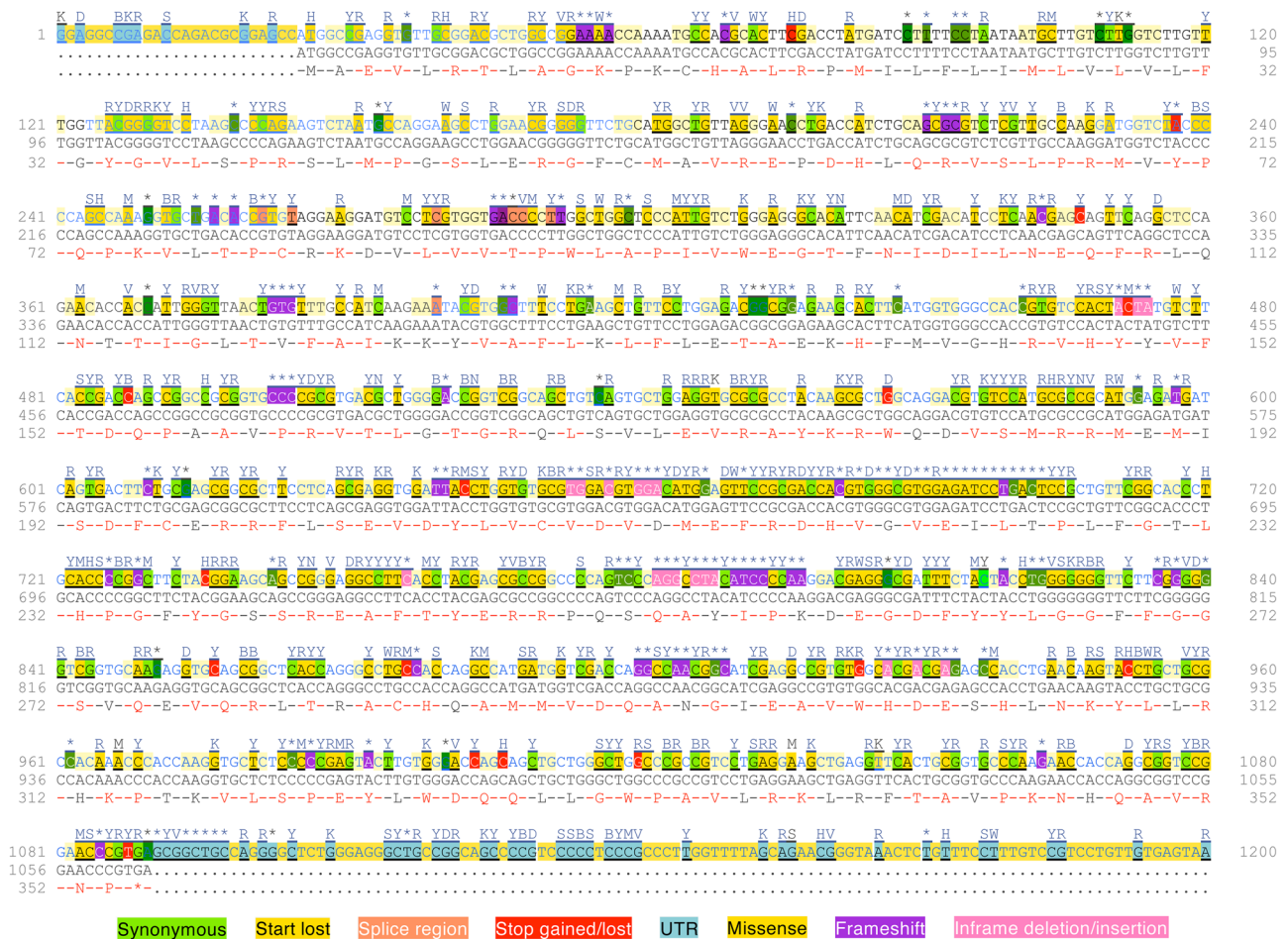
Fig. 6 SNP variations identified in the human ABO gene. Recent advances in next-generation sequencing technology have enabled the identification of thousands of SNP variations in the coding region of the ABO gene, in addition to previously characterized SNPs in ABO alleles that specify a variety of ABO phenotypes. Together with the

transcript sequence of the A101 allele, the different types of SNP variations are highlighted in different colors. This figure was modified from the transcript of the O01 allele (ABO-203 ENST00000611156.4) (retrieved on July 15, 2021). Similar figures have been published previously [41, 42].

of SNP variations, using the nucleotide and amino acid sequences of allele A101 as standards [40]. Although ISBT has not yet used this nomenclature, we anticipate that time will come soon. It is obvious that the ISBT numbering system currently in use cannot keep up with the rapid increase in new alleles, and that members of the task force will soon tire of curating them. What is important are the SNP variations, especially allele-specific SNP variations with functional significance, and not the allele names. The transition from the age of genetics to the age of genomics has revolutionized ABO research, although most of the SNP variations identified through genome sequencing may have little or no functional relevance, unlike those that were shown responsible for specific ABO phenotypic variations.

Since the end of 2019, the world has been suffering from the SARS-CoV-2 infection outbreak and the COVID-19 pandemic. The ABO polymorphism appears to affect the infectivity of SARS-CoV-2 through modified glycosylation of Spike and other viral proteins and the inhibition/neutralization by natural antibodies against A/B antigens. It may also influence the progression and severity of COVID-19 disease due to differential susceptibility to blood clotting (see our reviews [41, 43] for additional information). It will be exciting to find out what the next breakthrough discoveries in ABO's research field would be.

Declarations

Ethics approval Does not apply.

Consent to participate Does not apply.

Consent for publication Does not apply.

Conflicts of interest Does not apply.

References

1. Landsteiner, K.: Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. Zentralblatt Bakteriologie **27**, 357–362 (1900)

2. Landsteiner, K.: Über Agglutinationserscheinungen normalen menschlichen Blutes. Wien Klin Woschenschr **14**, 1132–1134 (1901)

3. Kabat, E.A.: Blood Group substances; Their Chemistry and Immunochemistry, (1956)

4. Watkins, W.M.: Biochemistry and genetics of the ABO, Lewis, and P blood group systems. Adv Hum Genet **10**(1–136), 379–385 (1980)

5. Watkins, W.M.: Some genetical aspects of the biosynthesis of human blood group substances in: Ciba Foundation Symposium on Biochemistry of Human Genetics, Churchill, London (1959)

6. Ceppellini, R.: Physiological genetics of human blood factors in: Ciba Foundation Symposium on Biochemistry of Human Genetics, Churchill, London (1959)

7. Ginsburg, V.: Enzymatic basis for blood groups in man. Adv Enzymol Relat Areas Mol Biol **36**, 131–149 (1972)

8. Yamamoto, F., Marken, J., Tsuji, T., White, T., Clausen, H., Hakomori, S.: 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**(2), 1146–1151 (1990)

9. Yamamoto, F., Clausen, H., White, T., Marken, J., Hakomori, S.: Molecular genetic basis of the histo-blood group ABO system. Nature **345**, 229–233 (1990)

10. Nagai, M., Dave, V., Kaplan, B.E., Yoshida, A.: Human blood group glycosyltransferases. I. Purification of n-acetylgalactosaminyltransferase. J Biol Chem **253**(2), 377–379 (1978)

11. Nagai, M., Dave, V., Muensch, H., Yoshida, A.: Human blood group glycosyltransferase. II. Purification of galactosyltransferase. J Biol Chem **253**(2), 380–381 (1978)

12. Clausen, H., White, T., Takio, K., Titani, K., Stroud, M., Holmes, E., Karkov, J., Thim, L., Hakomori, S.: Isolation to homogeneity and partial characterization of a histo-blood group A defined Fuc alpha 1–2Gal alpha 1—3-N-acetylgalactosaminyltransferase from human lung tissue. J Biol Chem. **265**, 1139–1145 (1990)

13. Yamamoto, F., McNeill, P.D., Hakomori, S.: Human histo-blood group A2 transferase coded by A2 allele, one of the A subtypes, is characterized by a single base deletion in the coding sequence, which results in an additional domain at the carboxyl terminal. Biochem Biophys Res Commun **187**(1), 366–374 (1992)

14. Yamamoto, F., McNeill, P.D., Yamamoto, M., Hakomori, S., Harris, T., Judd, W.J., Davenport, R.D.: Molecular genetic analysis of the ABO blood group system: 1. Weak subgroups: A3 and B3 alleles. Vox Sang **64**, 116–119 (1993)

15. Yamamoto, F., McNeill, P.D., Yamamoto, M., Hakomori, S., Harris, T.: Molecular genetic analysis of the ABO blood group system: 3. A(X) and B(A) alleles. Vox Sang **64**(3), 171–174 (1993)

16. Yamaguchi, H.: A review of cis AB blood. Jinrui Idengaku Zasshi **18**(1), 1–9 (1973)

17. Treacy, M., Stroup, M.: A scientific forum on blood grouping serum Anti-A (Murine Monoclonal Blend) BioClone*. Raritan, NJ: Ortho Diagnostic Systems, (1987)

18. Yamamoto, F., McNeill, P.D., Kominato, Y., Yamamoto, M., Hakomori, S., Ishimoto, S., Nishida, S., Shima, M., Fujimura, Y.: Molecular genetic analysis of the ABO blood group system: 2. *cis*-AB alleles. Vox Sang **64**(2), 120–123 (1993)

19. Yamamoto, F., Hakomori, S.: Sugar-nucleotide donor specificity of histo-blood group A and B transferases is based on amino acid substitutions. J Biol Chem **265**(31), 19257–19262 (1990)

20. Yamamoto, F.: Molecular genetics of the ABO histo-blood group system. Vox Sang **69**(1), 1–7 (1995)

21. Yamamoto, F., McNeill, P.D., Yamamoto, M., Hakomori, S., Bromilow, I.M., Duguid, J.K.: Molecular genetic analysis of the ABO blood group system: 4. Another type of O allele. Vox Sang **64**(3), 175–178 (1993)

22. Yamamoto, F., McNeill, P.D.: Amino acid residue at codon 268 determines both activity and nucleotide-sugar donor substrate specificity of human histo-blood group A and B transferases: *In vitro* mutagenesis study. J Biol Chem **271**(18), 10515–10520 (1996)

23. Patenaude, S.I., Seto, N.O., Borisova, S.N., Szpacenko, A., Marcus, S.L., Palcic, M.M., Evans, S.V.: The structural basis for specificity in human ABO(H) blood group biosynthesis. Nat Struct Biol **9**(9), 685–690 (2002)

24. Yamamoto, F., Cid, E., Yamamoto, M., Saitou, N., Bertranpetit, J., Blancher, A.: An integrative evolution theory of histo-blood group ABO and related genes. Sci Rep **4**, 6601 (2014)

25. Yamamoto, F., McNeill, P.D., Hakomori, S.: Genomic organization of human histo-blood group ABO genes. Glycobiology **5**(1), 51–58 (1995)

26. Kominato, Y., Tsuchiya, T., Hata, N., Takizawa, H., Yamamoto, F.: Transcription of human ABO histo-blood group genes is dependent upon binding of transcription factor CBF/NF-Y to minisatellite sequence. J Biol Chem **272**(41), 25890–25898 (1997)

27. Kominato, Y., Hata, Y., Takizawa, H., Tsuchiya, T., Tsukada, J., Yamamoto, F.: Expression of human histo-blood group ABO genes is dependent upon DNA methylation of the promoter region. J Biol Chem **274**(52), 37240–37250 (1999)

28. Kominato, Y., Hata, Y., Takizawa, H., Matsumoto, K., Yasui, K., Tsukada, J., Yamamoto, F.: 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**(40), 37936–37948 (2002)

29. Hata, Y., Kominato, Y., Yamamoto, F.I., Takizawa, H.: Characterization of the human ABO gene promoter in erythroid cell lineage. Vox Sang **82**(1), 39–46 (2002)

30. Hata, Y., Kominato, Y., Takizawa, H., Tabata, S., Michino, J., Nishino, K., Yasumura, S., Yamamoto, F.: Transcription starting from an alternative promoter leads to the expression of the human ABO histo-blood group antigen. Transfusion **43**(5), 656–662 (2003)

31. Kominato, Y., Sano, R., Takahashi, Y., Hayakawa, A., Ogasawara, K.: Human ABO gene transcription regulation. Transfusion **60**(4), 860–869 (2020)

32. Blumenfeld, O.O.: Mutation databases and other online sites as a resource for transfusion medicine: history and attributes. Transf Med Rev **16**(2), 103–114 (2002)

33. Yip, S.P.: Sequence variation at the human ABO locus. Ann Human Genet **66**(1), 1–27 (2002)

34. Yamamoto, F.: Review: ABO blood group system–ABH oligosaccharide antigens, anti-A and anti-B, A and B glycosyltransferases, and ABO genes. Immunohematology **20**(1), 3–22 (2004)

35. Yu, L.C., Twu, Y.C., Chou, M.L., Chang, C.Y., Wu, C.Y., Lin, M.: Molecular genetic analysis for the B(3) allele. Blood **100**(4), 1490–1492 (2002)

36. Hosseini-Maaf, B., Smart, E., Chester, M.A., Olsson, M.L.: The Abantu phenotype in the ABO blood group system is due to a splice-site mutation in a hybrid between a new O1-like allelic lineage and the A2 allele. Vox Sang **88**(4), 256–264 (2005)

37. Cai, X., Jin, S., Liu, X., Fan, L., Lu, Q., Wang, J., Shen, W., Gong, S., Qiu, L., Xiang, D.: Molecular genetic analysis of ABO blood group variations reveals 29 novel ABO subgroup alleles. Transfusion **53**(11 Suppl 2), 2910–2916 (2013)

38. Sano, R., Kuboya, E., Nakajima, T., Takahashi, Y., Takahashi, K., Kubo, R., Kominato, Y., Takeshita, H., Yamao, H., Kishida, T., Isa, K., Ogasawara, K., Uchikawa, M.: 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 Bm phenotype. Vox Sang **108**(3), 310–313 (2015)

39. Ying, Y., Hong, X., Xu, X., Ma, K., He, J., Zhu, F.: 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**(6), 594–600 (2018)

40. Yamamoto, F., Cid, E., Yamamoto, M., Blancher, A.: ABO research in the modern era of genomics. Transfus Med Rev **26**(2), 103–118 (2012)

41. Yamamoto, F.: Molecular Genetics and Genomics of the ABO Blood Group Ann Blood **6**, 25 (2021). https://doi.org/10.21037/aob-20-71

42. Yamamoto, F.: Molecular genetics to genomics – historical overview of the ABO research. ISBT Sci. Ser. **15**(1), 174–184 (2019)

43. Yamamoto, F., Yamamoto, M., Muniz-Diaz, E.: Blood group ABO polymorphism inhibits SARS-CoV-2 infection and affects COVID-19 progression. Vox Sang **116**(1), 15–17 (2021)

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