



Molecular genetic basis of the human Forssman glycolipid antigen negativity

Miyako Yamamoto*, Emili Cid* & Fumiichiro Yamamoto

From the ABO Histo-blood Groups and Cancer Laboratory, Cancer Genetics and Epigenetics Program, Institut de Medicina Predictiva i Personalitzada del Càncer (IMPPC), Badalona, Barcelona, Catalunya, 08916, Spain.

SUBJECT AREAS:
IMMUNOGENETICS
EVOLUTIONARY BIOLOGY
GLYCOBIOLOGY
EVOLUTIONARY GENETICS

Received
12 September 2012

Accepted
29 November 2012

Published
13 December 2012

Correspondence and
requests for materials
should be addressed to
F.Y. (fyamamoto@
imppc.org)

* These authors
contributed equally to
this work.

Forssman heterophilic glycolipid antigen has structural similarity to the histo-blood group A antigen, and the *GBGT1* gene encoding the Forssman glycolipid synthetase (FS) is evolutionarily related to the *ABO* gene. The antigen is present in various species, but not in others including humans. We have elucidated the molecular genetic basis of the Forssman antigen negativity in humans. In the human *GBGT1* gene, we identified two common inactivating missense mutations (c.688G>A [p.Gly230Ser] and c.887A>G [p.Gln296Arg]). The reversion of the two mutations fully restored the glycosyltransferase activity to synthesize the Forssman antigen *in vitro*. These glycine and glutamine residues are conserved among functional *GBGT1* genes in Forssman-positive species. Furthermore, the glycine and serine residues represent those at the corresponding position of the human blood group A and B transferases with GalNAc and galactose specificity, respectively, implicating the crucial role the glycine residue may play in the FS α 1,3-GalNAc transferase activity.

In 1911, Forssman discovered that rabbits injected with a suspension of kidney tissue from guinea pig or horse, but not from cow or rat, produced antibodies that were capable of hemolysing sheep erythrocytes in the presence of complement¹. The heterophilic antigen on sheep erythrocytes was later named as the Forssman antigen for its discoverer. This antigen is present in a variety of species, but not in others^{2,3}. Species were, therefore, categorized into Forssman-positive or negative species, depending on the antigen expression. Mouse and dog were also classified as Forssman-positive, and human and other anthropoid apes as Forssman-negative. The Forssman antigen was also found in species other than mammals. For example, chicken, turtles, and carp express the antigen, whereas goose, pigeon, and frog lack the antigen. Although the Forssman antigen was initially recognized on sheep erythrocytes, the expression on erythrocytes is rare. In many species its expression is restricted to tissues. However, in certain species it is observed in both erythrocytes and tissues (as in chicken) and in others only on erythrocytes (as in sheep). Its tissue distribution was studied in detail in mouse. Perivascular macrophages/dendritic cells located close to microvessels express this antigen, whereas fibroblasts, endothelial cells, pericytes, myocytes, and neural tissue elements do not⁴. This may explain the reason why anti-Forssman antibody *per se* fails to induce hyperacute rejection in a xenotransplantation model of mouse heart xenografts to rat⁵. Tissue-specific expression of the Forssman antigen was also observed in other Forssman-positive mammals^{3,6}. Its expression was shown to undergo changes during embryonic development and cellular differentiation as well⁷.

The Forssman antigen is a glycolipid with the structure GalNAc α 1->3GalNAc β 1->3Gal α 1->4Gal β 1->4Glc β 1->1Cer⁸. The terminal *N*-acetyl-D-galactosamine (GalNAc) residue bound by α 1-3 glycosidic linkage is the reason for immunological similarity to the blood group A antigen whose immunodominant structure is GalNAc α 1->3(Fuc α 1->2)Gal β 1-. As the oligosaccharide structure indicates, the Forssman antigen is synthesized through a series of glycosylation reactions. The enzyme responsible for the last step of its biosynthesis is Forssman glycolipid synthetase (FS; EC 2.4.1.88). This enzyme catalyzes the transfer of GalNAc from UDP-GalNAc nucleotide-sugar to the terminal GalNAc residue of globoside (GalNAc β 1->3Gal α 1->4Gal β 1->4Glc β 1->1Cer) by α 1-3 linkage. The FS activity was demonstrated to be present in tissues from various Forssman-positive species^{9,10}.

In 1996, the canine *GBGT1* gene cDNA encoding FS was cloned by the expression cloning method¹¹. The gene exhibited sequence homology to previously cloned bovine and murine *GGTA1* genes for α 1,3-galactosyltransferases that synthesize the α 1,3-galactosyl-epitope (Gal α 1->3Gal β 1->4GlcNAc-) and A and B alleles of the human *ABO* gene [MIM110300] encoding the blood group A and B transferases that catalyze the final reactions of biosynthesis of the histo-blood group A and B antigens, respectively¹²⁻¹⁵. In humans, the *GGTA1* gene became a pseudogene (*GGTA1P*) and the α 1,3-galactosyltransferase activity was lost [MIM104175]. Later, the same group



also cloned the rat *A3GALT2* gene cDNA encoding the isogloboside b3 synthetase (iGb3S), another enzyme with α 1,3-galactosyltransferase specificity, which acts on lactosylceramide to form iGb3 (Gal α 1->3Gal β 1->4Glc β 1->1Cer), initiating the biosynthesis of the isoglobo-series of glycosphingolipids¹⁶. The sequence homology among those enzymes illustrated that those genes have evolved from the same ancestral gene by duplications followed by divergence¹⁷. Those enzymes are categorized in the same GT6 family of α 1,3-Gal(NAC) transferases in the CAZy database that houses structurally-related catalytic and carbohydrate-binding modules of enzymes that create, modify, or degrade glycosidic bonds¹⁸. A full-length human *GBGT1* gene cDNA was later cloned, but no indel mutations were found [MIM 606074]¹⁹. The *GBGT1* mRNA expression was observed in a variety of human tissues ranging from heart, lung, skeletal muscle, kidney, pancreas, spleen, small and large intestines, peripheral leukocytes, prostate, ovary, to placenta. Because the human *GBGT1* cDNA construct failed to produce the Forssman antigen, a structural defect was suspected. However, the basis of the human Forssman negativity was unsolved. Commemorating the 100th anniversary of the discovery of Forssman antigen, we initiated the exploration to decipher the basis of Forssman antigen negativity, and we have finally decoded the mystery.

Results

Potential molecular bases of the Forssman antigen negativity in non-human species. Figure 1 shows the *GBGT1* genes and their encoded proteins from various species. We have primarily examined them *in silico* and deduced the potential mechanisms of Forssman negativity in cow, macaque, pig, rat, rabbit, and frog species. In some cases we have also performed minor experiments of PCR and nucleotide sequencing. Important amino acid sequence changes and likely mechanisms are shown in the rightmost 2 columns. Compared with the FS protein sequences of the Forssman-positive mammals, the bovine (*Bos taurus*) protein lacks 11 amino acids due to an additional splicing. Furthermore, the second glycine position of the consensus FYYGGA sequence, which is conserved in all the Forssman-positive species whose sequences have been determined, is replaced by arginine (FYYGRA). We think that these alterations are the primary cause of Forssman negativity in ox/cow because the corresponding sequences of blood group A transferase (FYYLGG or FYYAGG), B transferase (FYYMGA), and α 1,3-galactosyltransferase (FYYHAA) were previously shown to be critical for the recognition and binding of the donor nucleotide-sugar substrates by specificity and activity analyses of the A/B transferase chimeras and amino acid substitution constructs^{20–22}, and separately

Species Name (Generic Name)	Forssman Antigen ^{a)}	Ensembl Gene Name	Ensembl Peptide Name	A.A. No. ^{b)}	Important A.A. Seq. ^{c)}	Remarks ^{d)}
<i>Homo sapiens</i> (Human)	-	ENSG00000148288	ENSP00000361110	347	FYYGGA	See Figures 2 & 3. Two inactivating missense mutations
<i>Pan troglodytes</i> (Chimpanzee)	-	ENSPTRG00000021510	ENSPTRP00000036809	347	FYYGGA	See Figure 2. Two inactivating missense mutations
<i>Gorilla gorilla</i> (Gorilla)	-	ENSGGOG00000003448	ENSGGOP00000003390	347	FYYGGA	See Figure 2. Two inactivating missense mutations
<i>Pongo abelii</i> (Orangutan)	-	ENSPPYG00000019724	ENSPYP000000022121	345	FYYGGA	
<i>Macaca mulatta</i> (Macaque)	-	ENSMUUG00000015277	ENSMMUP00000020033	343	FYYGGA	The sequence important for sugar binding is not conserved.
<i>Mus musculus</i> (Mouse)	+	ENSMUSG00000026829	ENSMUSP00000127071	347	FYYGGA	See Figures 2 & 3.
<i>Rattus norvegicus</i> (Rat)	-					A homologous sequence to the mouse <i>GBGT1</i> gene exists in the corresponding chromosomal region between the <i>Mps2</i> and <i>Ralgsd</i> genes. In addition to many nucleotide substitutions, there are several indels that make the encoded protein non-functional.
<i>Cavia porcellus</i> (Guinea pig)	+	ENSCPOG00000004874	ENSCPOP00000004388	347	FYYGGA	See Figure 2.
<i>Oryctolagus cuniculus</i> (Rabbit)	-					No <i>GBGT1</i> gene is annotated. The BLAST search did not identify any. The PCR using several pairs of degenerate oligonucleotide primers failed to amplify the equivalent sequences. Rabbits may lack the homologous sequence possibly due to the loss of chromosomal region containing the <i>GBGT1</i> gene.
<i>Bos taurus</i> (Cow)	-	ENSBTAG000000030319	ENSBTAP000000040428	335	FYYGRA	The sequence important for sugar binding is not conserved. Also 11 amino acids are missing due to an additional splicing.
<i>Ovis aries</i> (Sheep)	+			?	FYYGGA	No database is currently available for the BLAST search. However, PCR using degenerate primers amplified a homologous sequence, which was partially sequenced and confirmed to be the <i>GBGT1</i> gene.
<i>Sus scrofa</i> (Pig)	-				FYYGGA	No <i>GBGT1</i> gene is annotated. The BLAST search failed to identify any. However, PCR using degenerate primers amplified a homologous sequence, which was partially sequenced. Missense mutations are the likely cause of inactivity.
<i>Equus caballus</i> (Horse)	+	ENSECAG00000012442	ENSECAP00000010230	347	FYYGGA	See Figure 2.
<i>Felis catus</i> (Cat)	+	ENSFCAG00000013356	ENSFCAP00000012386	342	FYYGGA	
<i>Canis familiaris</i> (Dog)	+	ENSCAFG00000019864	ENSCAFP00000029424	347	FYYGGA	See Figure 2.
<i>Gallus gallus</i> (Chicken)	+	ENSGALG00000003340	ENSGALP00000005275	343	FYYGGA	
<i>Meleagris gallopavo</i> (Turkey)	?	ENSMGAG00000006307	ENSMGAP00000006329	344	FYYGGA	
<i>Taeniopygia guttata</i> (Zebra Finch)	?	ENSTGUG00000005430	ENSTGUP00000005586	344	FYYGGA	
<i>Xenopus tropicalis</i> (Frog)	-					No <i>GBGT1</i> gene is annotated. The BLAST search did not identify any frog <i>GBGT1</i> genes although several ABO genes were found. No <i>GBGT1</i> gene equivalent is present.
<i>Gadus morhua</i> (Cod)	?	ENSGMOG00000012375	ENSGMOP00000013247	275	YYYSA	Exons 1-4 missing?
<i>Gasterosteus aculeatus</i> (Stickleback)	?	ENSGACG00000000537 ENSGACG00000014813 ENSGACG00000016064	ENSGACP00000000689 ENSGACP00000019565 ENSGACP00000021194	275 346 275	YYTAA FYYCGA YYYTAA	Exons 1-4 missing? Exons 1-4 missing?
<i>Oryzias latipes</i> (Medaka fish)	?	ENSORLG00000010140 ENSORLG00000008616	ENSORLP00000012717 ENSORLP00000010810	358 220	YYTAA YYYSE	No ATG? Exons 1-4 correct? No ATG? A large deletion.
<i>Danio rerio</i> (Zebrafish)	?	ENSNDARG00000005257 ENSNDARG00000011283 ENSNDARG00000019207 ENSNDARG00000025275 ENSNDARG00000035555 ENSNDARG00000091936 ENSNDARG00000091944 ENSNDARG00000091969 ENSNDARG00000092718 ENSNDARG00000094600	ENSNDARP00000018621 ENSNDARP00000015549 ENSNDARP00000010651 ENSNDARP00000018900 ENSNDARP00000051540 ENSNDARP00000112590 ENSNDARP00000120433 ENSNDARP00000122748 ENSNDARP00000114822 ENSNDARP00000121282	277 302 304 316 342 137 342 275 301 140	YYTAA YYTAA YYCGA YYTAA YYGGA ----- YYGGA YYGGA YYGGA YY---	Exons 1-4 missing? Exons 1-4 missing? No ATG? Gene fragment? Exons 1-4 missing? Short N-terminal No ATG? Gene fragment?

Figure 1 | *GBGT1* genes and encoded proteins from Forssman-positive and negative species. The annotated entries of the *GBGT1* genes and their encoded proteins were extracted from the Ensembl database and categorized by species. a) The Forssman positivity and negativity are indicated by (+) and (-), respectively, when it is known for the species, whereas a question mark (?) shows unknown. b) In order to facilitate the identification of partial genes and pseudogenes, the numbers of the deduced amino acid residues of the individual proteins are indicated. c) The amino acid residues corresponding to the codons between 258 and 263 of the mouse FS, were also extracted, and the residues that are different from the consensus FYYGGA sequence are indicated in bold. d) Additional useful information is added in the Remarks. Potential mechanisms of Forssman negativity deduced from the *in silico* analysis are shown in bold.



by the determination of three-dimensional structures of those transferases^{23–25}. The same argument may also be applied to macaque (*Macaca mulatta*) with the FYYGK sequence. The replacement of the alanine by lysine may exert a devastating effect on the enzyme activity because the same substitution at the corresponding position of blood group B transferase abolished its activity²¹. In case of pig (*Sus scrofa*), the genome database is available, but no *GBGT1* genes have been annotated. We performed PCR using several pairs of degenerate oligonucleotide primers, which were designed to amplify the *GBGT1* gene sequences from diverse species. We could amplify DNA fragments, and the nucleotide sequencing confirmed that the sequences were from the porcine *GBGT1* gene. The FYYGGA sequence was conserved, however, there were several amino acid substitutions within the sequenced region (M.Y., unpublished data). We suspect that a few of the missense mutations are the probable cause of its Forssman negativity although there may be additional and more critical mutations in the unsequenced region of the gene.

Regarding rat (*Rattus norvegicus*) and rabbit (*Oryctolagus cuniculus*) species, we failed to amplify DNA fragments using the same pairs of oligonucleotide primers that amplified the pig sequence. The reason for the failure soon became apparent for the rat case. There was, in fact, a nucleotide sequence homologous to the mouse *GBGT1* gene in the rat genome although the BLAST search using the amino acid

sequence inquiry did not return any homologous sequences. This sequence was located on the p12 band of rat chromosome 3 between *Mrps2* (mitochondrial ribosomal protein S2) and *GNDS_RAT*. Because the mouse *GBGT1* gene is mapped on the A3 band of chromosome 2 between *Mrps2* and *Ralgds* (ral guanine nucleotide dissociation stimulator) and because the rat *GNDS_RAT* is the *Ralgds* equivalent, the identified rat sequence seemed to be the *GBGT1* gene ortholog. However, due to numerous nucleotide sequence differences, including several indels, the protein sequence is not conserved in rat, which explains its loss of Forssman antigen expression. The OryCun2 (GCA_000003625.1) rabbit genome assembly used for the BLAST search had a limited coverage of 7×, and only 82% have been anchored to chromosomes. Neither *GBGT1* nor *Mrps2* genes have been annotated. BLAST searches were not successful, either. Considering the failed PCR attempts, the homologous sequence may not exist in rabbit possibly due to chromosomal region loss, or even if it exists, like the rat sequence, the sequence may be quite dissimilar to functional *GBGT1* gene sequences. For frog (*Xenopus tropicalis*), the BLAST search identified several *ABO* genes but no *GBGT1* genes, suggesting the absence of the *GBGT1* gene equivalent.

Identification of the missense mutations responsible for the human Forssman antigen negativity. For the loss of Forssman antigen expression in humans we assumed that missense mutations

<i>M. musculus</i>	MTRPRLAQLGAFLLGGTGLWVWLKFIKDWLLVSYIPYYLPCPEFFNMKLPFRKEKPLQPVTLQYQPQKLEHGPT ELLT LTLPWLAPIV	90
<i>C. porcellus</i>	MSRRRLVLSLGLIVLASTGLWALWVYIENWLPVSHVPPYLLPCPEIFNMKLQYKGEKPLPPVTAQYQPRLTQHRPT ELLT LTLPWLAPIV	90
<i>E. caballus</i>	MRRRLAIGLGVCLLVGTVLCTLWVYVEDWLPVSYVPPYLLPCPEIFNMKLQYKGEKLPVAVQSRYPQPKLLEQRPT ELLT LTLPWLAPIV	90
<i>C. familiaris</i>	MCRRLALGLGFSLLSGIALWLSLWIYMETWLPFVSYVPPYLLPCPEIFNMKLQYKGEKPFQPVTRSPHPQPKLLEQRPT ELLT LTLPWLAPIV	90
<i>H. sapiens</i>	MHRRRLALGLGFCLLAGTSLSVLWVYLENWLPVSYVPPYLLPCPEIFNMKLHYKREKPLQPVVWSQYQPQKLEHRPT QLLT LTLPWLAPIV	90
<i>P. troglodytes</i>	MHRRRLALGLGFCLLAGTSLSVLWVYLENWLPVSHVPPYLLPCPEIFNMKLHYKREKPLQPMVWSQYQPQKLEHRPT QLLT LTLPWLAPIV	90
<i>G. gorilla</i>	MHRRRLALGLGFCLLAGTSLSVLWVYLENWLPVSYVPPYLLPCPEIFNMKLHYKREKPLQPVVWSQYQPQKLEHRPT QLLT LTLPWLAPIV	90
<i>M. musculus</i>	SEGTFDPELLKSMYQPLNLTIGVTVFAVGKYTCFIQRFLESAAEFFMRGYQVHYVLFTHDPTAVPRVPLGPGRLLS I IPIQGSRWEE I S	180
<i>C. porcellus</i>	SEGTFDPELLQHIYQPLNLSIGVTVFAVGKYTRFVQHFLSAAEFFMRGFQVHYVVFTHNPAAI PRVLLGPRRLLD I IPIHGHTHWE I S	180
<i>E. caballus</i>	SEGTFAELLQHIYQPLNLTIGLTVFAVGKYTHFVQHFLSAAELFFMHGYRVCYVFTDDPTAIPQVPLGPGRRLL I IPIQRHSRWEE I S	180
<i>C. familiaris</i>	SEGTFNPELLQHIYQPLNLTIGLTVFAVGKYTRFVQHFLSAAEQFFMQGYQVYVYIFTNDPAAI PRVPLGPGRLLS I IPIQRHSRWEE I S	180
<i>H. sapiens</i>	SEGTFNPELLQHIYQPLNLTIGVTVFAVGKYTHFIQSFLSAAEFFMRGYRVHYVIFTNDPAAVPGVPLGPHRLLS S IPIQGHSHWEE T S	180
<i>P. troglodytes</i>	SEGTFNPELLQHIYQPLNLTIGVTVFAVGKYTHFIQSFLSAAEFFMRGYRVHYVIFTNDPAAI PGIPGLGPHRLLS S IPIQGHSHWEE T S	180
<i>G. gorilla</i>	SEGTFNPELLQHIYQPLNLTIGVTVFAVGKYTHFIQSFLSAAEFFMRGYRVHYVIFTNDPAAVPGVPLGPHRLLS S IPIQGHSHWEE T S	180
<i>M. musculus</i>	MRRMETINKHIAKRAHKEVDYLFCDVDMVFRNPWPGETLGDVLAIIHP G YFVAVPRRKFYERRQVSSAFVADNEGDFYGGALFVGG R V	270
<i>C. porcellus</i>	MRRMEAI SRHIAKKAHQEVDYLFCDVDMVFNHPWPGETLGDVLAIIHP G YFTVSRRQFPYERRQISTAFVAENEGDFYGGAVFVGG R V	270
<i>E. caballus</i>	TRRMEIISQHIKRAHREVDYLFCDVDMVFRNPWPGETLGDVLAIIHP G YAVVPRQFPYERRHVSTAFVADGEGDFYGGAVFVGG R V	270
<i>C. familiaris</i>	TRRMETISRHIAQRAHREVDYLFCDVDMVFRNPWPGETLGDVLAIIHP G YAVVPRQFPYERRHISTAFVAENEGDFYGGAVFVGG R V	270
<i>H. sapiens</i>	MRRMETISQHIKRAHREVDYLFCDVDMVFRNPWPGETLGDVLAIIHP S YAVVPRQFPYERRRVSTAFVADSEGDFYGGAVFVGG Q V	270
<i>P. troglodytes</i>	MRRMETISQHIKRAHREVDYLFCDVDMVFRNPWPGETLGDVLAIIHP S YAVSRQFPYERRRVSTAFVADSEGDFYGGAVFVGG Q V	270
<i>G. gorilla</i>	MRRMETISQHIKRAHQEVDYLFCDVDMVFRNPWPGETLGDVLAIIHP S YAVVPRQFPYERRRVSTAFVADSEGDFYGGAVFVGG Q V	270
<i>M. musculus</i>	RVYEFTRACHMAILADKANSIMAAW Q EESHNLNRHF I HKPSKVLSPPEYLWDERKPRPRSLKMIRFSSVKKAN W LRT	347
<i>C. porcellus</i>	NVYEFTRGCHMAILADKANGIMAAW Q EESHNLNRRL I HKPSKVLSPPEYLWDDRKPVPSSLKLIRFSTLLKDT N WLR	347
<i>E. caballus</i>	NVYEFTRGCHMAILADKANGIMAAW Q EESHNLNRFF I HKPSKVLSPPEYLWDDRKRQPPSLKLIRFSTLDKDT S WLR	347
<i>C. familiaris</i>	KVYEFRTGCHMAILADKANGIMAAW Q EESHNLNRFF I HKPSKVLSPPEYLWDDRKPQPPSLKLIRFSTLDKAT S WLR	347
<i>H. sapiens</i>	RVYEFTRGCHMAILADKANGIMAAW R EESHNLNRHF I SNKPSKVLSPPEYLWDDRKPQPPSLKLIRFSTLDKDIS C LRS	347
<i>P. troglodytes</i>	RVYEFTRGCHMAILADKANGIMAAW R EESHNLNCF I SNKPSKVLSPKYLWDDRKPQPPSLKLIRFSTLDKDIS C LRS	347
<i>G. gorilla</i>	RVYEFTRGCHMTILADKANGIMAAW R EESHNLNRFF I SNKPSKVLSPPEYLWDDRKPQPPSLKLIRFSTLDKDIS C LRS	347

Figure 2 | Amino acid sequence comparison of the *GBGT1* genes between the Forssman-positive and negative species. The species compared are mouse, guinea pig, horse, and dog for Forssman-positive and human, chimpanzee, and gorilla for negative, respectively. The splicing junctions are indicated with the mark (∧). The amino acid residues that are different between the Forssman-positive and negative species are shown with the (+) mark. The marks and the amino acid residues are in bold when the amino acids are conserved in the 4 Forssman positive species. Additionally, they are underlined when conserved in cat and chicken (not shown). The restriction enzyme cleavage sites for the chimera constructions are also indicated.



might be responsible. To identify them, we took two different approaches: Sequence analysis of the *GBGT1* genes in the Forssman-positive and negative species and functional assays of FS chimeras between Forssman-positive mouse and Forssman-negative human and also of the *in vitro* mutagenized amino acid substitution constructs. We first aligned the amino acid sequences of the *GBGT1* genes from 4 Forssman-positive mammals and 3 Forssman-negative anthropoid apes, including humans (Figure 2). The locations of the splicing junctions are marked with the symbol ($\hat{\wedge}$), and the positions of the restriction enzyme cleavage sites used for the later construction of mouse-human FS chimeras are also indicated. The amino acid residues that are different between the Forssman-positive and negative species are shown with the (+) mark. When the amino acids are conserved in the 4 Forssman positive species, the marks, as well as the amino acid residues, are shown in bold. Additionally, when they are also conserved in Forssman-positive cat (*Felis catus*) and chicken (*Gallus gallus*) (not shown), the marks and the amino acids are also underlined (\pm). There are 3 such positions. The Forssman-negative human, chimpanzee, and gorilla share the c.536T>C [p.Ile179Thr], c.688G>A [p.Gly230Ser], and c.887A>G [p.Gln296Arg] substitutions.

The sequence homology between functional mouse and non-functional human FSs is shown in Figure 3. Both the mouse and human *GBGT1* genes encode a protein of 347 amino acid residues, of which 275 are identical (indicated with asterisk) and 72 are different. The known human SNP variations resulting in changes in the amino acid sequence are also shown. In the dbSNP-135 database, there is a nonsense codon termination SNP named rs35898523 (c.363C>A [p.Tyr121Ter]) with a minor allele frequency (MAF) of 0.048, which is marked with the symbol (=) in the figure. There are also 38 non-synonymous missense SNPs, among which one position is consistently different between the Forssman-positive and negative species. It is the SNP rs141041392 (c.886C>T [p.Arg296Trp]), which changes the arginine to tryptophan, but not to the glutamine of the mouse and other Forssman-positive species. This SNP is very rare as its MAF is listed as 0.00 (1/4,552) in the NHLBI-ESP: ESP Cohort Population (ss342287410).

We next constructed a variety of mouse-human FS chimeras, transfected DNA into Forssman-negative, but globoside-positive, African green monkey kidney fibroblast COS1 cells, and examined the Forssman antigen expression. Those cells transformed with an

origin-defective mutant of simian virus SV40 were previously used to clone the canine *GBGT1* gene cDNA encoding the functional FS¹¹. In Panel A of Figure 4, the proteins encoded by the mouse and human *GBGT1* genes are schematically shown in dark and light grey, respectively. The locations of restriction enzyme cleavage sites used for chimera constructions, and the positions of important conserved amino acid substitutions are indicated. The same dark/light grey code was also used to indicate the origins of sequences in the chimeras and amino acid substitution constructs. The results of immunocytochemistry are shown side-by-side with the chimeric constructs in Panel B. The results of immunostaining with anti-Forssman antibody were adjusted by transfection efficiency using co-transfected GFP-positive cell percentages. The values are shown in percentage, using the results of transfection with the original mouse FS construct and a negative control to be 100 and 0, respectively. The original human *GBGT1* gene construct, as well as several others, did not produce any Forssman antigen-positive cells (0%), whereas a decrease was observed with several other constructs. The presence of inactivating mutation(s) was evident within the *NcoI-BsiWI* region. Additional weakening mutations were suggested in the *KpnI-NcoI* and *BsiWI-C-terminal* regions.

We then introduced the Ile179Thr (I179T), Gly230Ser (G230S), or Gln296Arg (Q296R) substitutions individually into the mouse FS expression construct. We also introduced the Thr179Ile (T179I), Ser230Gly (S230G), Arg296Gln (R296Q), Asn308His (N308H), and Cys344Trp (C344W) substitution individually into the human *GBGT1* gene expression construct. The results of Forssman antigen expression after DNA transfection of the amino acid substitution constructs are shown in Panel C. Compared to the original mouse FS, the construct containing I179T lost some, but not much, FS activity. However, the Q296R substitution significantly decreased and the G230S substitution almost abolished the activity. None of the human constructs exhibited strong Forssman antigen expression. However, some cells, many dead or dying, were stained positive after transfection with either the S230G or R296Q substitution construct. In order to examine the effects of those two substitutions further, we prepared the double substitution constructs (the mouse G230S & Q296R construct and human S230G & R296Q construct). The results are also shown in the same panel. The introduction of both the G230S and Q296R substitutions into the mouse FS completely eliminated

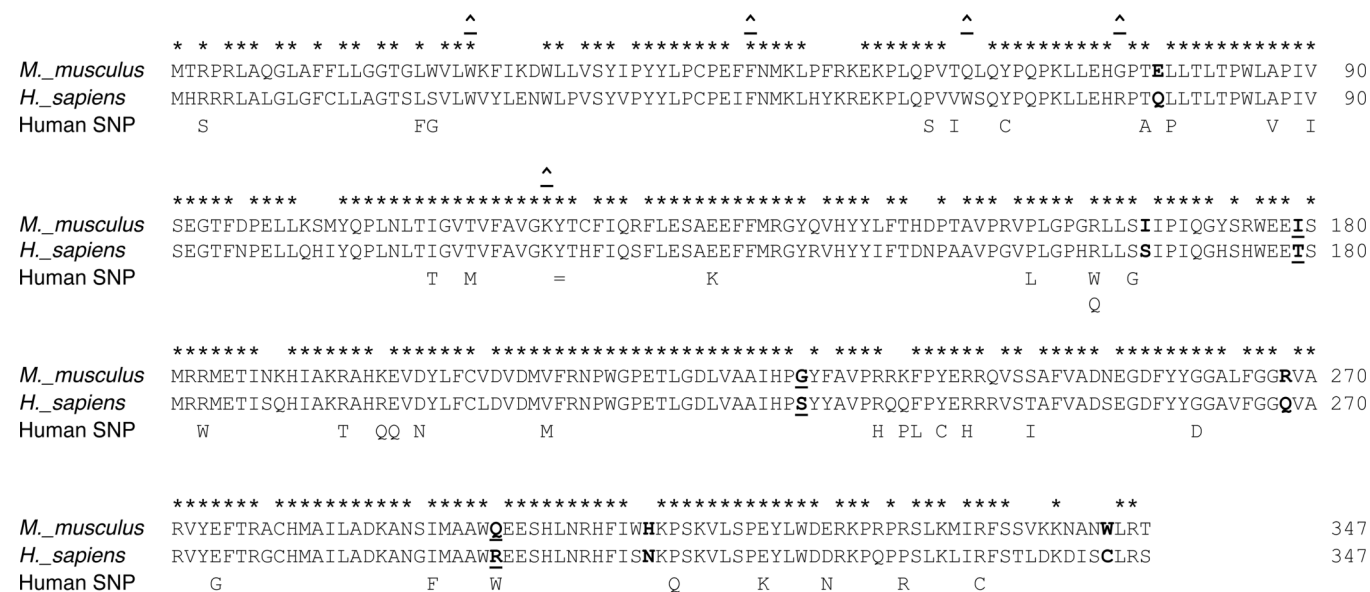


Figure 3 | Amino acid sequence homology between the murine and human *GBGT1*-encoded proteins and the human SNP variations. The amino acid residues identical in mice and humans are indicated with asterisk (*). The single SNP variation resulting in codon termination is marked with (=), and 38 SNPs resulting in amino acid substitutions are shown.

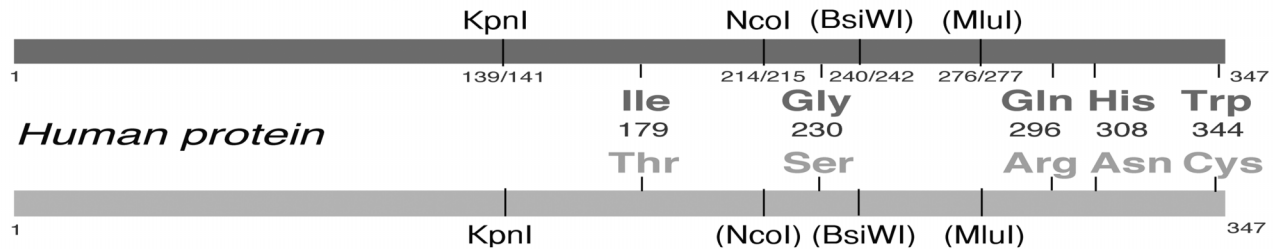
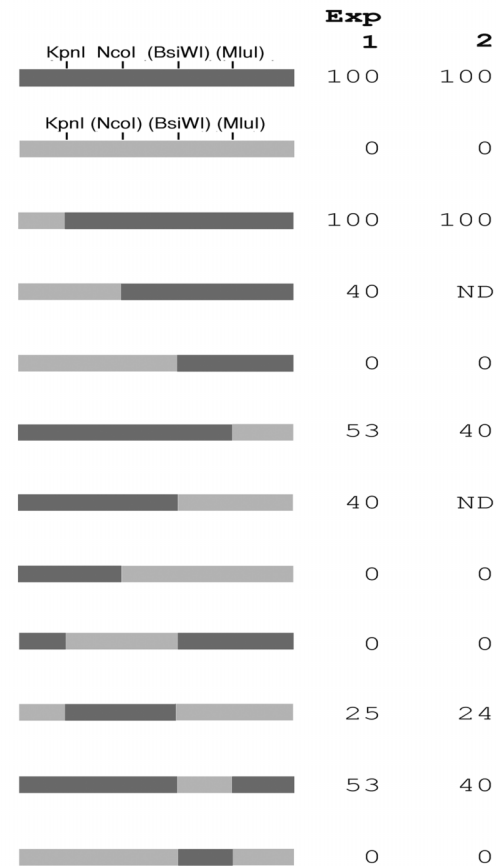
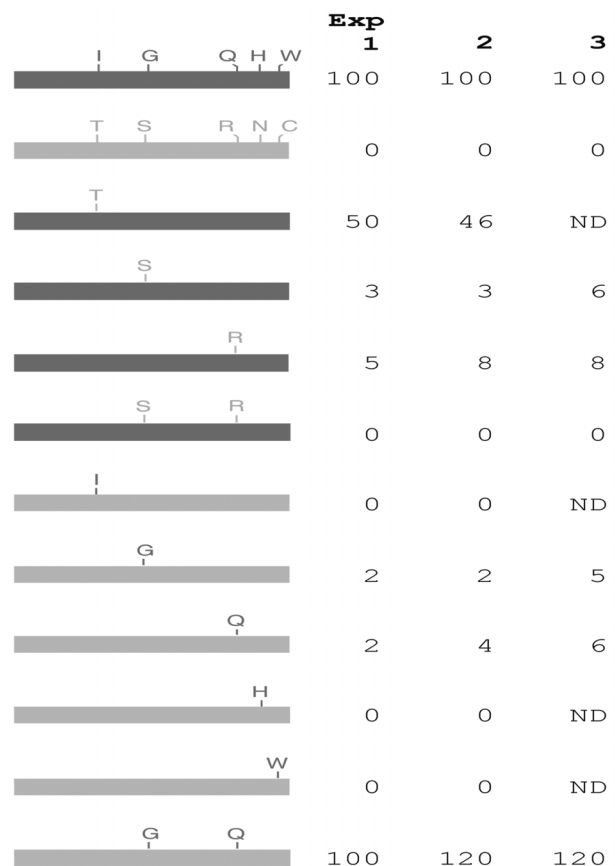
**A.****Mouse Forssman glycolipid synthetase****B.****C.**

Figure 4 | Forssman antigen expression after DNA transfection. (A) Schematic representation of the mouse and human *GBGT1* gene products. The mouse and human sequences are schematically shown in dark and light grey, respectively. The restriction enzyme cleavage sites used for the chimera constructions are indicated with and without parentheses for newly created artificial and pre-existed sites, respectively. The amino acid substitutions made are also shown. (B) The mouse-human FS chimeras and their FS activity. The constructs and the results of immunostaining with anti-Forssman antibody are shown. The values are shown in percentage with the expression of the original mouse FS construct and that of a negative control to be 100 and 0, respectively. The experiments were repeated with some constructs. The “ND” stands for “Not determined”. (C) The *in vitro* mutagenized amino acid substitution constructs and their FS activity. The results of the amino acid substitution constructs are shown.

Forssman antigen expression. On the other hand, the introduction of both the S230G and R296Q substitutions enabled the human *GBGT1* protein to synthesize Forssman glycolipid at the equivalent level of mouse FS. The pictures of cells transfected with these constructs are shown in Figure 5, as well as those transfected with either the original mouse construct, the original human construct, or the human R296Q construct (to be discussed later). The pictures at left and right represent the expression of co-transfected GFP protein and Forssman glycolipid, respectively.

These results also showed that none of the remaining 70 amino acid substitutions in the human *GBGT1* gene were responsible for Forssman negativity. The codon 230 corresponds to the codon 235 of the human A/B transferases, and the glycine and serine residues are present in A and B transferases with GalNAc and galactose

specificity, respectively^{15,20}. Therefore, we concluded that the glycine residue is crucial for the α 1,3-GalNAc transferase activity of FS. The fact that only 2 amino acid changes reinstated FS activity may suggest that the inactivation of the *GBGT1* genes in anthropoid apes was recent in evolution.

Discussion

The biosynthesis of Forssman glycolipid is complex and requires a multi-step glycosylation. Deficiency at any step may result in the failure to synthesize this antigen. For instance, non-functional alleles are known to exist for the *B3GALNT1* gene encoding β 1,3-N-acetyl-D-galactosaminyltransferase 1, which catalyzes the biosynthesis of globoside, the acceptor substrate of FS. Those individuals who are homozygous of the non-functional alleles cannot produce globoside

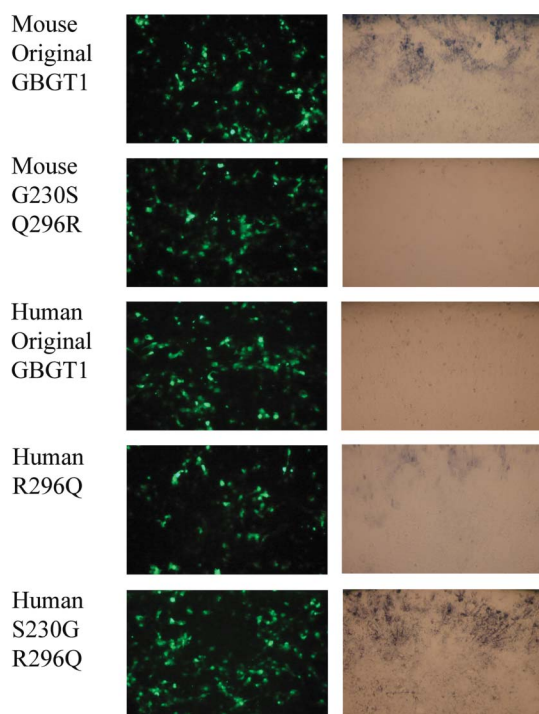


Figure 5 | GFP and Forssman glycolipid antigen expression of COS1 cells transfected with key constructs. The left and right panels show the pictures of GFP and Forssman antigen expression, respectively.

and exhibit the P^k phenotype of the P blood group system^{26,27}. Therefore, the inactivating mutations of the *B3GALNT1* gene can also be regarded as a cause of Forssman glycolipid antigen negativity. However, the combined frequency of such non-functional alleles is extremely low (<0.001). As shown in Figure 3, a nonsense mutation (Tyr121Ter) is known to exist for the *GBGT1* gene as an SNP (rs35898523). With a MAF of 0.048, this mutation causing early termination cannot be ignored. However, its frequency is also too low to explain the human Forssman antigen negativity because the chance of having non-functional recessive alleles in a homozygous manner is small. In addition to the nonsense SNP, there are additional 38 non-synonymous SNPs, 14 synonymous SNPs, and 3 and 16 SNPs in the 5' upstream and 3' downstream, untranslated regions of mRNA, respectively, of the human *GBGT1* gene. It is possible that some of these SNPs may also affect the functionality of human FS. However, the two inactivating missense mutations we identified are very common and present in almost 100% of individuals, which makes these the most likely and sufficient causes of Forssman antigen negativity in humans.

Categorizing species into Forssman-positive and negative is solid because Forssman's results were confirmed and expanded by others. Our results presented in this paper also support this classification. However, this categorization is not absolute. Like ABO, intra-species polymorphism of the Forssman antigen expression exists. An example has recently been reported of the individuals with rare ABO subgroup A_{pae} ²⁸. Those individuals possess the *GBGT1* gene with the c.887G>A [p.Arg296Gln] substitution and weakly express Forssman antigen on their erythrocytes. Because the frequency of the A_{pae} subgroup is exceptionally low, no cases are deposited in the SNP database. In our DNA transfection experiments the human R296Q construct having the same mutation resulted in the appearance of some cells with Forssman antibody reactivity (Figures 4C and 5), corroborating the presumption that the germline mutation confers weak FS activity to the *GBGT1* protein with the substitution. Our results have also proven that the mutant allele is dominant over the

regular allele in its functionality. The Forssman antigen on erythrocytes in those A_{pae} individuals may be produced by those cells. Alternatively, like the Lewis antigens the glycolipid produced in other tissues may be adsorbed onto their erythrocytes.

In addition of intra-species polymorphism, somatic changes may also enable human cells to produce Forssman glycolipid. The expression of Forssman antigen has been reported of cancer cells and tissues, including metastatic tumor of biliary adenocarcinoma in liver²⁹, gastric and colonic mucosa and tumors of these tissues^{30–32}, lung and lung carcinoma³³, and MRK-nu-1 breast carcinoma cell line, PC10 lung carcinoma cell line, and KATOIII, MKN45, MKN74 gastric carcinoma cell lines³⁴. Questionable specificity of the antibodies used in the immunological detection of Forssman antigen may be responsible in some studies. However, Forssman glycolipid is expressed undoubtedly in certain human cells and tissues because its structure was chemically determined³⁰. Together with the fact that R296Q polymorphism confers Forssman antigen positivity in the A_{pae} subgroup individuals, the results of our *in vitro* mutagenesis study suggest that either this or S230G substitution alone may grant the human *GBGT1* gene protein with weak FS activity. However, no such somatic mutations have been reported of the gene in the Catalogue of Somatic Mutations in Cancer – COSMIC database. Therefore, other mechanisms may be responsible if changes in the *GBGT1* gene cause the appearance of Forssman antigen. The potential mechanisms may include overexpression of the *GBGT1* gene mRNA, splicing variations, increased stability of mRNA and protein, and post-translational protein modifications. Additionally, alterations in specificity and activity of other glycosyltransferases, such as blood group A transferase, may allow aberrant glycosylation to synthesize Forssman glycolipid under certain conditions. Similarly, vestigial human FS may be activated to synthesize Forssman antigen if extraordinarily higher concentrations of UDP-GalNAc and globoside substrates than threshold levels become available in the Golgi apparatus. This argument may be supported by previous findings that certain tumor tissues exhibit higher concentrations of UDP-GalNAc and UDP-GlcNAc³⁵, and that an increased level of intracellular UDP-GlcNAc activates O-GlcNAc transferase (OGT) and leads to enhanced O-GlcNAcylation of target proteins³⁶. The Forssman antigen expression also seems to be influenced by cell density because more Forssman-positive cells appeared after DNA transfection on the dish edges (top of each picture in Figure 5), where cells are more crowded. Now that the basis of the Forssman antigen negativity in humans is clarified, the elucidation of molecular mechanisms leading to Forssman antigen expression in Forssman-negative human species may not be too far.

Methods

Sequence analysis. The nucleotide and deduced amino acid sequences of the annotated *GBGT1* genes from various species were extracted from the Ensembl database. Additional homologous sequences were BLAST searched among the sequences deposited in the Ensembl and GenBank databases using the nucleotide and deduced amino acid sequences of known *GBGT1* genes in the evolutionarily related species as query sequences. We used the 5.05 version of MEGA5 software for sequence alignment (<http://www.megasoftware.net/>)³⁷.

Preparations of mouse and human *GBGT1* gene expression constructs, their chimeric constructs, and *in vitro* mutagenized amino acid substitution constructs. The plasmids containing the human *GBGT1* cDNA and the mouse *GBGT1* cDNA were purchased from the Open Biosystem (Lafayette, CO). Compared with the reference sequence (*GBGT1*-001: ENST00000372040), the human *GBGT1* cDNA sequence contained 2 SNPs: T residues were found in place of Cs at the nucleotides 58 and 987 of the coding sequence (rs2073924 and rs17853056), respectively. The former resulted in the Leu20Phe substitution although the latter did not change the amino acid sequence. After PCR amplification using forward and reverse oligonucleotide primers that were tagged with the *EcoRI* and *BamHI* restriction enzyme cleavage sites, respectively, the coding regions of the *GBGT1* cDNAs from the two species were cleaved with those enzymes and ligated with the *EcoRI*-*BamHI* fragment of the eukaryotic expression vector pSG5 (Stratagene, CA). The unique cleavage site for the restriction enzyme *KpnI* pre-existed at equivalent locations of both human and mouse cDNAs whereas the *NcoI* site was only present in the mouse sequence. We created an *NcoI* site at the corresponding position of the human gene without



modifying the amino acid sequence, by oligonucleotide primer-mediated *in vitro* mutagenesis through two rounds of PCR reactions as previously described²¹. Similarly, the cleavage sites for *BsiWI* and *MluI* enzymes were also crafted at the corresponding sequences of the human and mouse genes without changes in their amino acid sequences. Next, a variety of chimeras possessing different portions of functional mouse FS and non-functional human FS were produced, by shuffling the equivalent restriction fragments between those two species. Later, the amino acid substitution constructs were prepared by the oligonucleotide primer-mediated *in vitro* mutagenesis. The nucleotide sequence of the *GBGT1* gene was determined for all of the constructs. Only those without any unintended non-synonymous mutations were used for the subsequent DNA transfection experiments.

DNA transfection and immunostaining. DNA from the *GBGT1* gene expression constructs and their derivatives was used for transfection. Forssman-negative but globoside-positive COS1 cells from African green monkey kidney were used as a recipient. These cells were previously used to clone the canine *GBGT1* gene cDNA encoding the functional FS, together with the MDCK II dog kidney cell cDNA expression library¹¹. The LipofectAMINE 2000 reagent (Life Technologies, Carlsbad, CA) was used to mediate DNA transfection. The cells were plated in a 96-well plate, and DNA transfection experiments were performed following the protocol provided by the manufacturer. In order to determine the efficiency of DNA transfection and the capacity of the transfected cells to express the transgenes, an equal amount of the pEGFP plasmid that expresses a red-shifted variant of GFP was co-transfected. Three days after transfection, the cells were fixed with 4% paraformaldehyde, washed with PBS, and dried. Immunostaining was later performed to detect the expression of Forssman antigen using a rat monoclonal antibody against murine Forssman glycolipid antigen (Clone FOM-1 from BMA Biomedicals, Switzerland). Biotin-SP-conjugated AffinitPure Goat AntiRat IgG+IgM(H+L) secondary antibody from Jackson ImmunoResearch Laboratories (West Grove, PA) and Vectastain ABC kit and DAB Peroxidase Substrate kit from Vector Laboratories (Burlingame, CA) were used. GFP-positive cells were counted under a fluorescence microscope. Forssman antigen-expressing cells were also counted visually after immunostaining. The ratios of the cells positive with Forssman antigen versus the cells positive with GFP were calculated, and they were adjusted so that the values become 100 for the cells transfected with the original mouse *GBGT1* gene expression construct. For several crucial constructs, DNA transfection experiments were repeated using independent preparations of DNA from different clones, in order to exclude the possibility of unintended mutation(s) in the vector sequence. Cell pictures were taken at 80× magnification, using a Leica DMIL Fluorescence Microscope and a Leica D-Lux3 digital camera, before (for GFP expression) and after (for Forssman antigen expression) the treatment with the DAB staining reagent for color development.

- Forssman, J. Die Herstellung hochwertiger spezifischer Schafhämonysine ohne Verwendung von Schafblut. Ein Beitrag zur Lehre von heterologer Antikörperbildung. *Biochem. Z.* **37**, 78–115 (1911).
- Bushbinder, L. Heterophile phenomena in immunology. *Arch. Path.* **19**, 841–80 (1935).
- Tanaka, N. & Leduc, E. H. A study of the cellular distribution of Forssman antigen in various species. *J. Immunol.* **77**, 198–212 (1956).
- Gustavsson, M. L. *et al.* Characterization of Forssman and other antigen/antibody systems in vascularized mouse heart to rat xenotransplantation. *Scand. J. Immunol.* **53**, 121–31 (2001).
- Wu, G. D. *et al.* Failure of anti-Forssman antibodies to induce rejection of mouse heart xenografts. *Xenotransplantation* **6**, 90–7 (1999).
- Willison, K. R. & Stern, P. L. Expression of a Forssman antigenic specificity in the preimplantation mouse embryo. *Cell* **14**, 785–93 (1978).
- Fenderson, B. A., Eddy, E. M. & Hakomori, S. Glycoconjugate expression during embryogenesis and its biological significance. *Bioessays* **12**, 173–9 (1990).
- Siddiqui, B. & Hakomori, S. A revised structure for the Forssman glycolipid hapten. *J. Biol. Chem.* **246**, 5766–9 (1971).
- Kijimoto, S., Ishibashi, T. & Makita, A. Biosynthesis of Forssman hapten from globoside by alpha-N-acetylgalactosaminyltransferase of guinea pig tissues. *Biochem. Biophys. Res. Commun.* **56**, 177–84 (1974).
- Das, K. K., Basu, M., Basu, S. & Evans, C. H. Biosynthesis *in vitro* of a globoside containing a 2-acetamido-2-deoxy-beta-D-galactopyranosyl group (1->3)-linked and Forssman glycolipid by two N-acetylgalactosaminyltransferases from chemically transformed guinea pig cells. *Carbohydr. Res.* **149**, 119–35 (1986).
- Haslam, D. B. & Baenziger, J. U. Expression cloning of Forssman glycolipid synthetase: a novel member of the histo-blood group ABO gene family. *Proc. Natl. Acad. Sci. USA* **93**, 10697–702 (1996).
- Joziasse, D. H., Shaper, J. H., Van den Eijnden, D. H., Van Tunen, A. J. & Shaper, N. L. Bovine alpha 1->3-galactosyltransferase: isolation and characterization of a cDNA clone. Identification of homologous sequences in human genomic DNA. *J. Biol. Chem.* **264**, 14290–7 (1989).
- Larsen, R. D. *et al.* Isolation of a cDNA encoding a murine UDPgalactose: beta-D-galactosyl-1,4-N-acetyl-D-glucosaminide alpha-1,3-galactosyltransferase: expression cloning by gene transfer. *Proc. Natl. Acad. Sci. USA* **86**, 8227–8231 (1989).
- 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).

- 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).
- Keusch, J. J., Manzella, S. M., Nyame, K. A., Cummings, R. D. & Baenziger, J. U. Expression cloning of a new member of the ABO blood group glycosyltransferases, iGb3 synthase, that directs the synthesis of isogloboglycosphingolipids. *J. Biol. Chem.* **275**, 25308–14 (2000).
- Turcot-Dubois, A. L. *et al.* Long-term evolution of the CAZY glycosyltransferase 6 (ABO) gene family from fishes to mammals—a birth-and-death evolution model. *Glycobiology* **17**, 516–28 (2007).
- Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–8 (2009).
- Xu, H., Storch, T., Yu, M., Elliott, S. P. & Haslam, D. B. Characterization of the human Forssman synthetase gene. An evolving association between glycolipid synthesis and host-microbial interactions. *J. Biol. Chem.* **274**, 29390–8 (1999).
- 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**, 19257–19262 (1990).
- 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**, 10515–10520 (1996).
- Yamamoto, F., Yamamoto, M. & Blancher, A. Generation of histo-blood group B transferase by replacing the N-acetyl-D-galactosamine recognition domain of human A transferase with the galactose-recognition domain of evolutionarily related murine alpha1,3-galactosyltransferase. *Transfusion* **50**, 622–30 (2010).
- Boix, E. *et al.* Structure of UDP complex of UDP-galactose:beta-galactoside-alpha-1,3-galactosyltransferase at 1.53-Å resolution reveals a conformational change in the catalytically important C terminus. *J. Biol. Chem.* **276**, 48608–14 (2001).
- Gastinel, L. N. *et al.* Bovine alpha1,3-galactosyltransferase catalytic domain structure and its relationship with ABO histo-blood group and glycosphingolipid glycosyltransferases. *EMBO J.* **20**, 638–49 (2001).
- Patenaude, S. I. *et al.* The structural basis for specificity in human ABO(H) blood group biosynthesis. *Nat. Struct. Biol.* **9**, 685–90 (2002).
- Hellberg, A., Poole, J. & Olsson, M. L. Molecular basis of the globoside-deficient P(k) blood group phenotype. Identification of four inactivating mutations in the UDP-N-acetylgalactosamine: globotriaosylceramide 3-beta-N-acetylgalactosaminyltransferase gene. *J. Biol. Chem.* **277**, 29455–9 (2002).
- Hellberg, A. *et al.* Genetic heterogeneity at the glycosyltransferase loci underlying the GLOB blood group system and collection. *Br. J. Haematol.* **125**, 528–36 (2004).
- Hult, A. K. *et al.* Forssman expression of human red cells; Biochemical and genetic basis of a novel histo-blood group system candidate (Abstract of presentation P2-020A at the AABB Annual Meeting). *Transfusion* **51**, 1A (2011).
- Kawanami, J. The appearance of Forssman hapten in human tumor. *J. Biochem.* **72**, 783–5 (1972).
- Hakomori, S., Wang, S. M. & Young, W. W. Jr. Isoantigenic expression of Forssman glycolipid in human gastric and colonic mucosa: its possible identity with "A-like antigen" in human cancer. *Proc. Natl. Acad. Sci. USA* **74**, 3023–7 (1977).
- Uemura, K., Hattori, H., Ono, K., Ogata, H. & Taketomi, T. Expression of Forssman glycolipid and blood group-related antigens A, Le(x), and Le(y) in human gastric cancer and in fetal tissues. *Jpn. J. Exp. Med.* **59**, 239–49 (1989).
- Ono, K. *et al.* Expression of Forssman antigen in human large intestine. *J. Histochem. Cytochem.* **42**, 659–65 (1994).
- Yoda, Y., Ishibashi, T. & Makita, A. Isolation, characterization, and biosynthesis of Forssman antigen in human lung and lung carcinoma. *J. Biochem.* **88**, 1887–90 (1980).
- Mori, E., Mori, T., Sanai, Y. & Nagai, Y. Radioimmuno-thin-layer chromatographic detection of Forssman antigen in human carcinoma cell lines. *Biochem. Biophys. Res. Commun.* **108**, 926–32 (1982).
- Gribbestad, I. S., Petersen, S. B., Fjosne, H. E., Kvinnsland, S. & Krane, J. 1H NMR spectroscopic characterization of perchloric acid extracts from breast carcinomas and non-involved breast tissue. *NMR Biomed.* **7**, 181–94 (1994).
- Hart, G. W., Housley, M. P. & Slawson, C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* **446**, 1017–22 (2007).
- Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–9 (2011).

Acknowledgments

We are grateful to Patricia Barrero and Marta Soler for their technical assistance. This work was supported by the Instituto de Salud Carlos III grant (PI11/00454) and the fund from IMPPC to F.Y.

Author contributions

F.Y. conceived and designed research. M.Y., E.C. and F.Y. performed experiments. F.Y. wrote manuscript and all the authors contributed to the editing of the manuscript.



Additional information

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

How to cite this article: Yamamoto, M., Cid, E. & Yamamoto, F. Molecular genetic basis of the human Forssman glycolipid antigen negativity. *Sci. Rep.* 2, 975; DOI:10.1038/srep00975 (2012).