US 20200291384A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2020/0291384 A1**

Kawaoka et al. (43) **Pub. Date: Sep. 17, 2020**

(54) **HUMANIZED CELL LINE**

(71) Applicants: **Wisconsin Alumin Research Foundation (WARF)**, Madison, WI (US); **National University Corporation, University of Tokyo**, Bunkyo-ku (JP)

(72) Inventors: **Yoshihiro Kawaoka**, Middleton, WI (US); **Kosuke Takada**, Tokyo (JP); **Masaki Imai**, Kawasaki-shi (JP)

(21) Appl. No.: **16/785,449**

(22) Filed: **Feb. 7, 2020**

Related U.S. Application Data

(60) Provisional application No. 62/803,266, filed on Feb. 8, 2019.

Publication Classification

(51) **Int. Cl.**

C12N 15/10	(2006.01)
C12N 9/10	(2006.01)
C12Q 1/04	(2006.01)
C12N 5/071	(2006.01)

(52) **U.S. Cl.**

CPC ***C12N 15/102*** (2013.01); ***C12N 9/10*** (2013.01); ***C12N 5/0602*** (2013.01); ***C12Q 1/04*** (2013.01); ***C12Y 204/99001*** (2013.01)

(57) **ABSTRACT**

A mammalian or avian cell line that expresses high levels of human influenza virus receptors is provided. In one embodiment, the cell line supports human influenza virus, e.g., human A/H3 influenza virus, isolation and growth much more effectively than corresponding conventional (unmodified) cells or in corresponding human virus receptor-over-expressing cells, and the propagated viruses may maintain higher genetic stability than in the corresponding cells.

Specification includes a Sequence Listing.

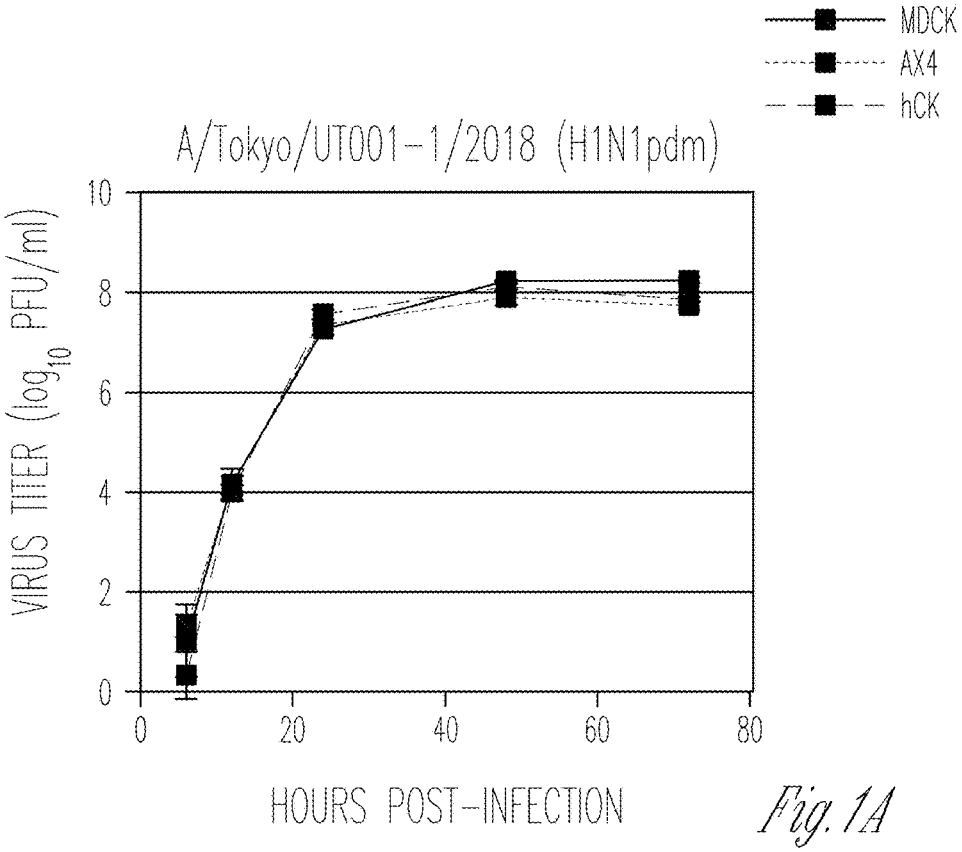
MDCK
AX4
hCK
A/Tokyo/UT001-1/2018 (H1N1pdm)
VIRUS TITER (log_{10} PFU/ml)
10
8
6
4
2
0
0
20
40
60
80
HOURS POST-INFECTION

Fig. 1A

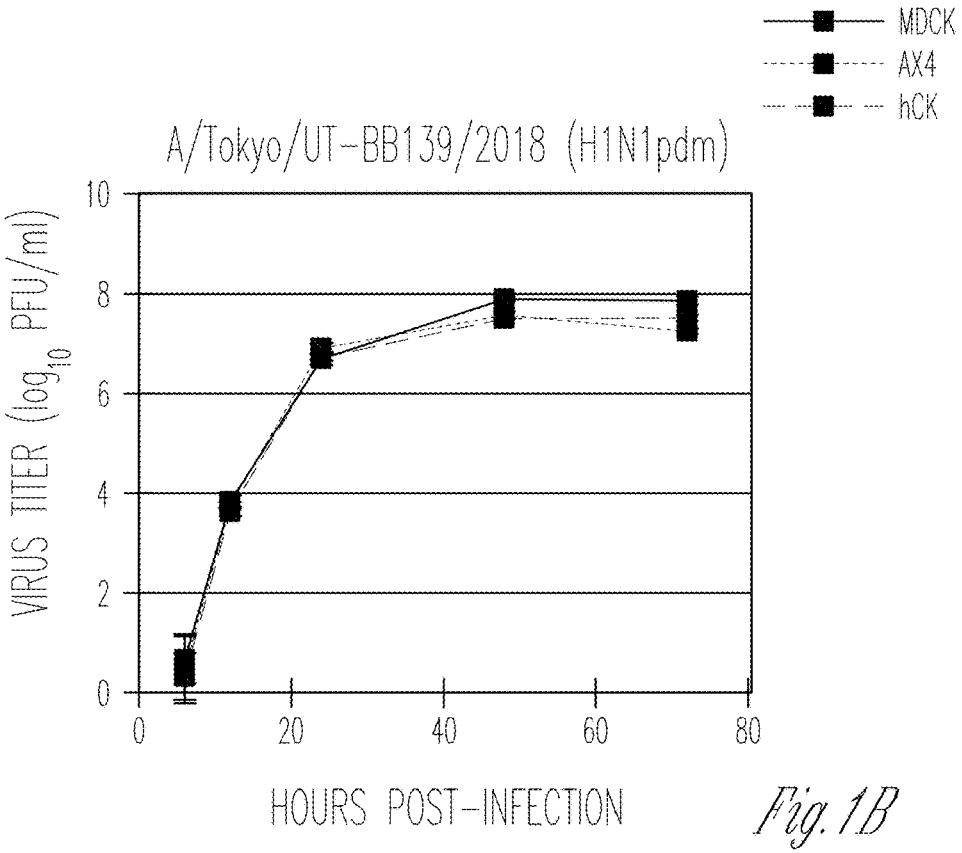
MDCK
AX4
hCK
A/Tokyo/UT-BB139/2018 (H1N1pdm)
VIRUS TITER (log_{10} PFU/ml)
10
8
6
4
2
0
0
20
40
60
80
HOURS POST-INFECTION

Fig. 1B

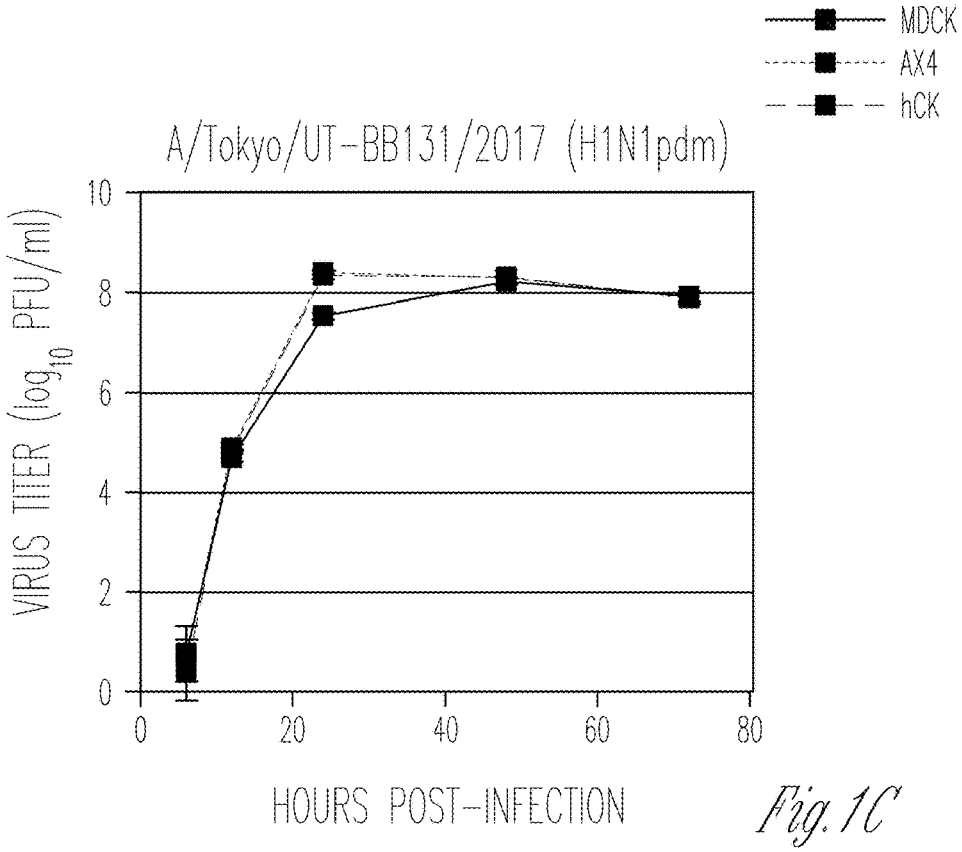
MDCK
AX4
hCK
A/Tokyo/UT-BB131/2017 (H1N1pdm)
VIRUS TITER (log_{10} PFU/ml)
10
8
6
4
2
0
0
20
40
60
80
HOURS POST-INFECTION

Fig. 1C

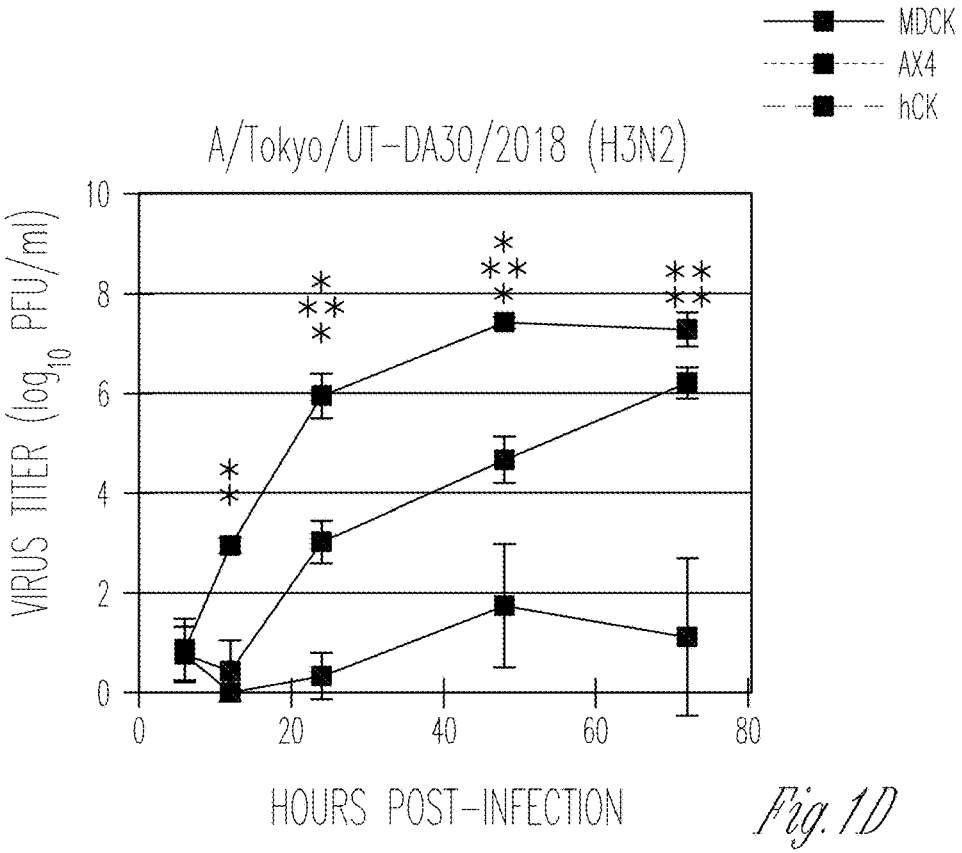
MDCK
AX4
hCK
A/Tokyo/UT-DA30/2018 (H3N2)
VIRUS TITER (log_{10} PFU/ml)
10
8
6
4
2
0
0
20
40
60
80
HOURS POST-INFECTION

Fig. 1D

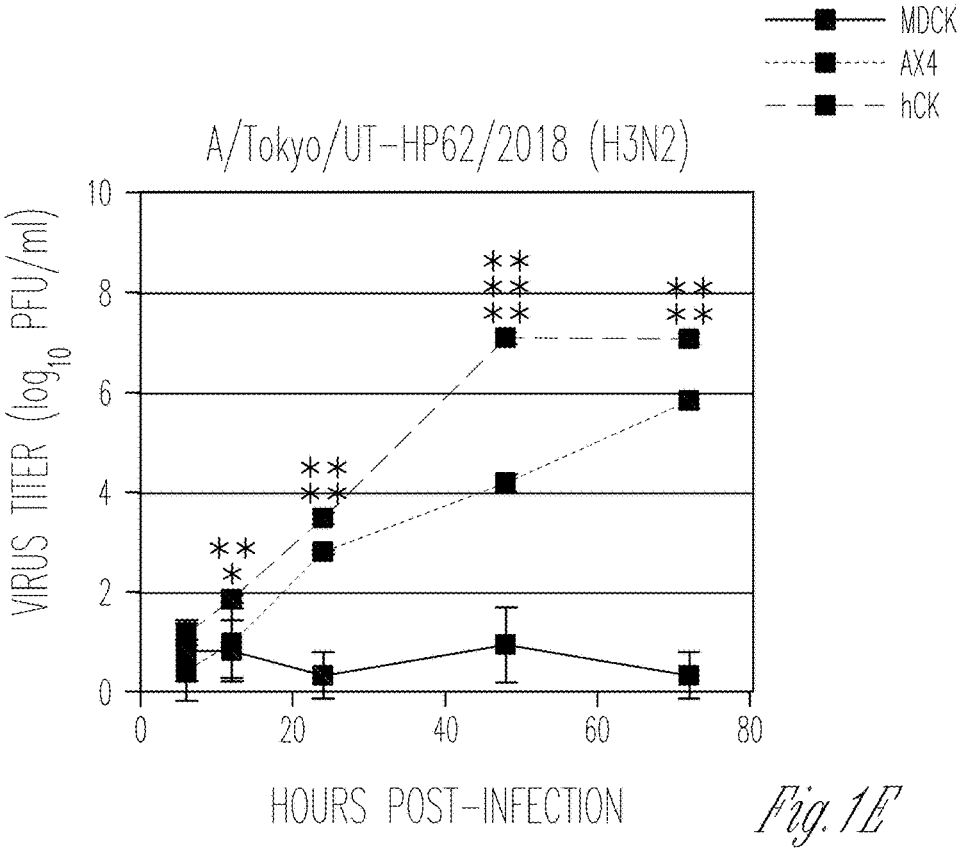
MDCK
AX4
hCK
A/Tokyo/UT-HP62/2018 (H3N2)
VIRUS TITER (log_{10} PFU/ml)
HOURS POST-INFECTION
0
2
4
6
8
10
0
20
40
60
80

Fig. 1E

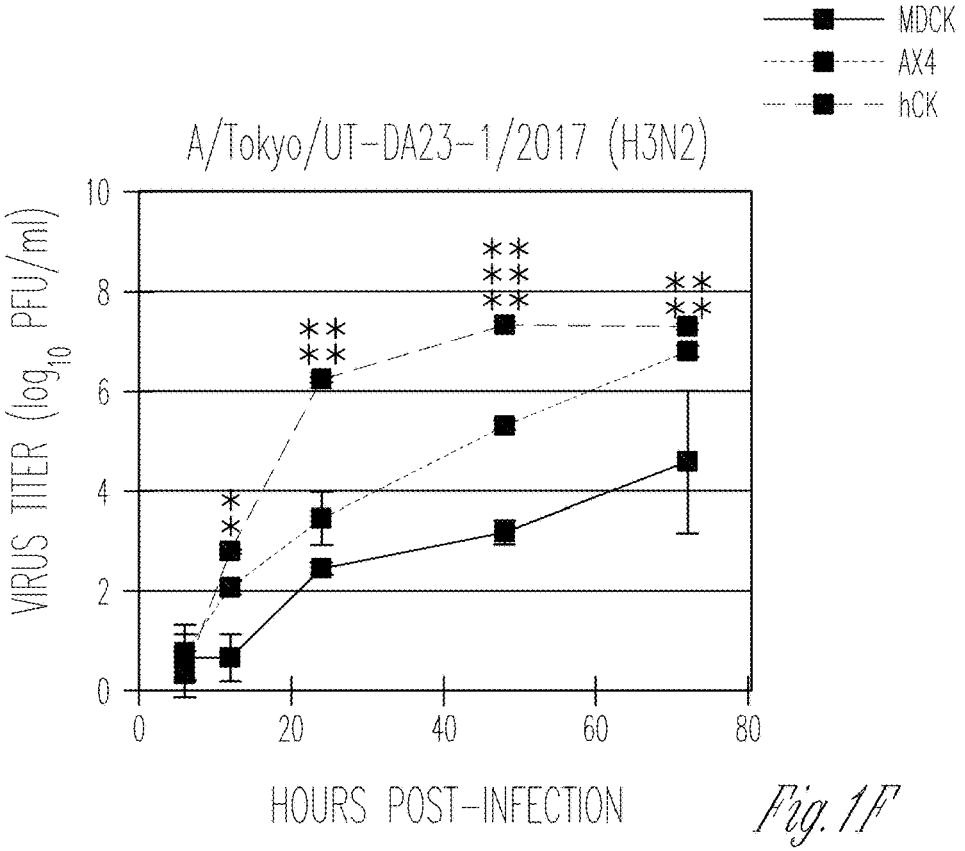
MDCK
AX4
hCK
A/Tokyo/UT-DA23-1/2017 (H3N2)
VIRUS TITER (log_{10} PFU/ml)
HOURS POST-INFECTION
0
2
4
6
8
10
0
20
40
60
80

Fig. 1F

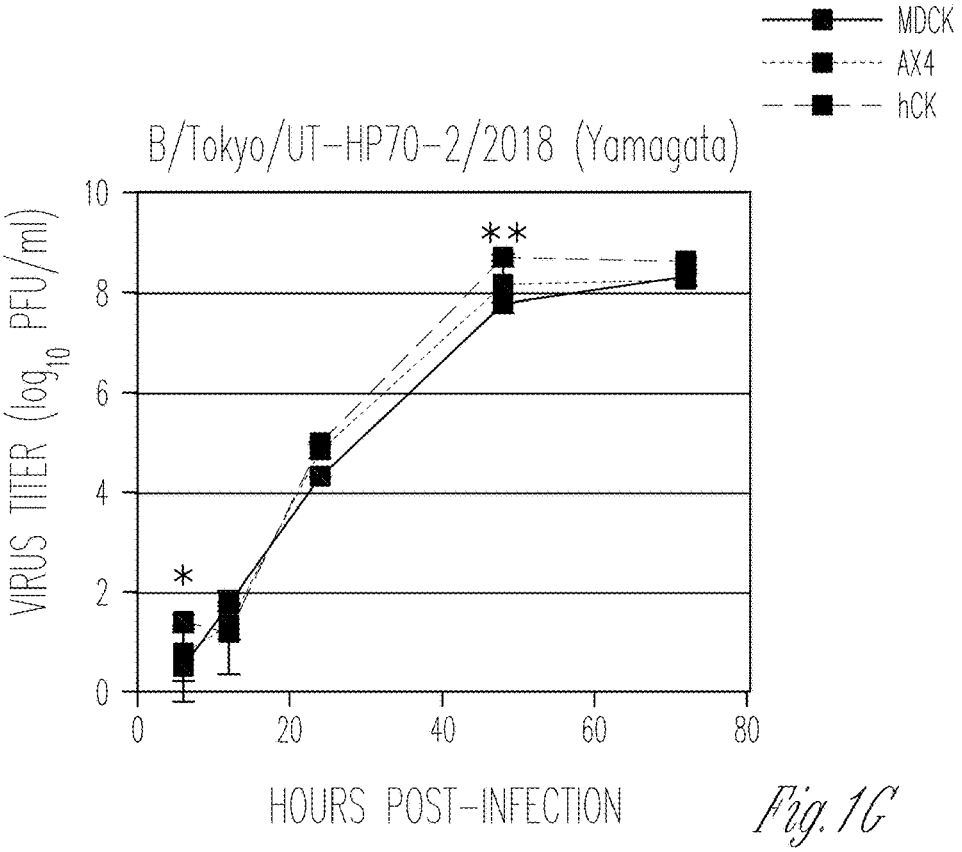
MDCK
AX4
hCK
B/Tokyo/UT-HP70-2/2018 (Yamagata)
VIRUS TITER (log_{10} PFU/ml)
HOURS POST-INFECTION
*
**
0
2
4
6
8
10
0
20
40
60
80

Fig. 1G

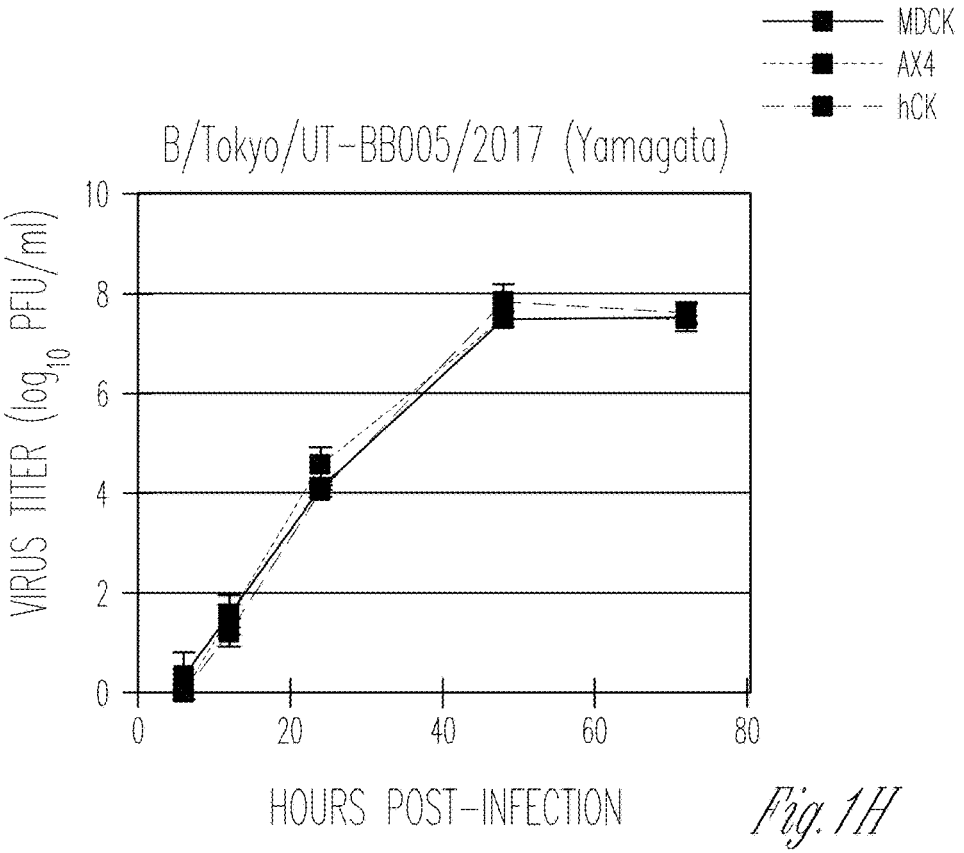
MDCK
AX4
hCK
B/Tokyo/UT-BB005/2017 (Yamagata)
VIRUS TITER (log_{10} PFU/ml)
HOURS POST-INFECTION
0
2
4
6
8
10
0
20
40
60
80

Fig. 1H

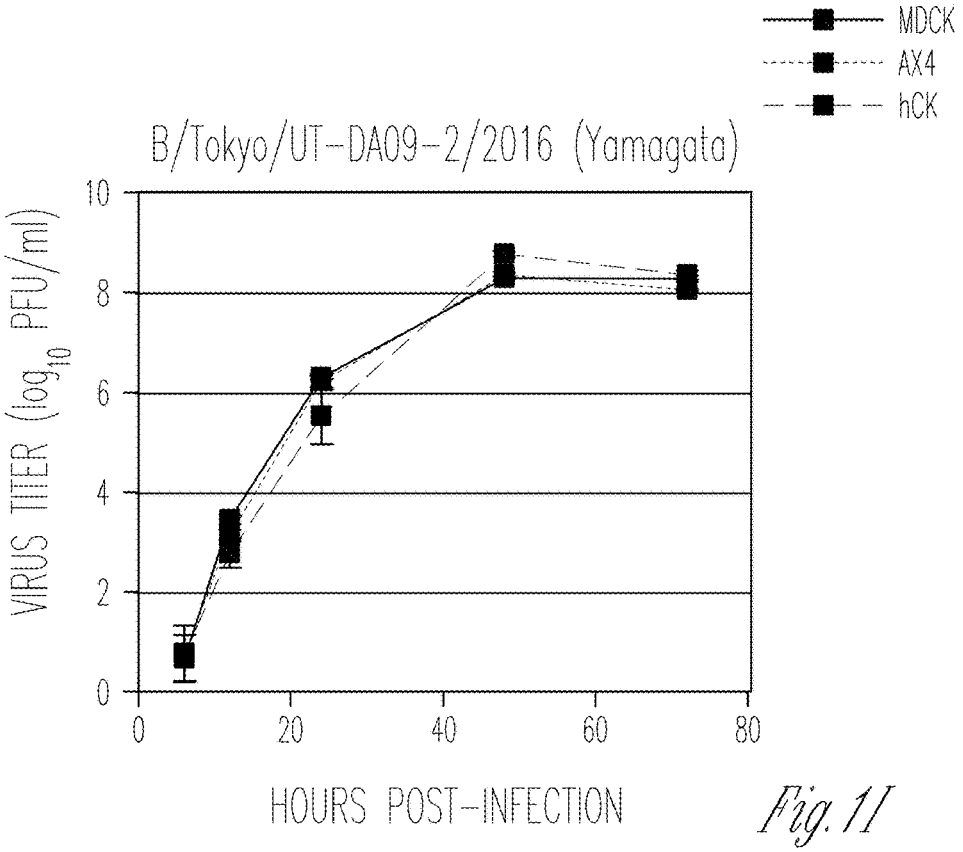
MDCK
AX4
hCK
B/Tokyo/UT-DA09-2/2016 (Yamagata)
VIRUS TITER (log_{10} PFU/ml)
0
2
4
6
8
10
0
20
40
60
80
HOURS POST-INFECTION

Fig. 1I

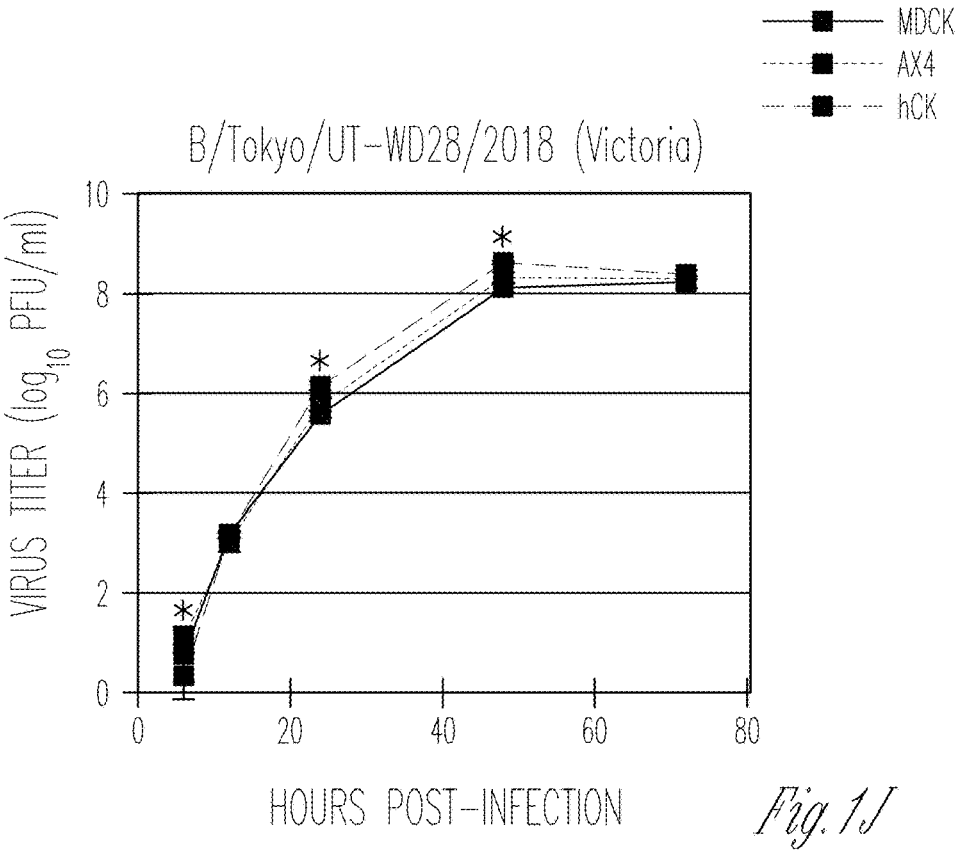
MDCK
AX4
hCK
B/Tokyo/UT-WD28/2018 (Victoria)
VIRUS TITER (log_{10} PFU/ml)
0
2
4
6
8
10
*
*
*
0
20
40
60
80
HOURS POST-INFECTION

Fig. 1J

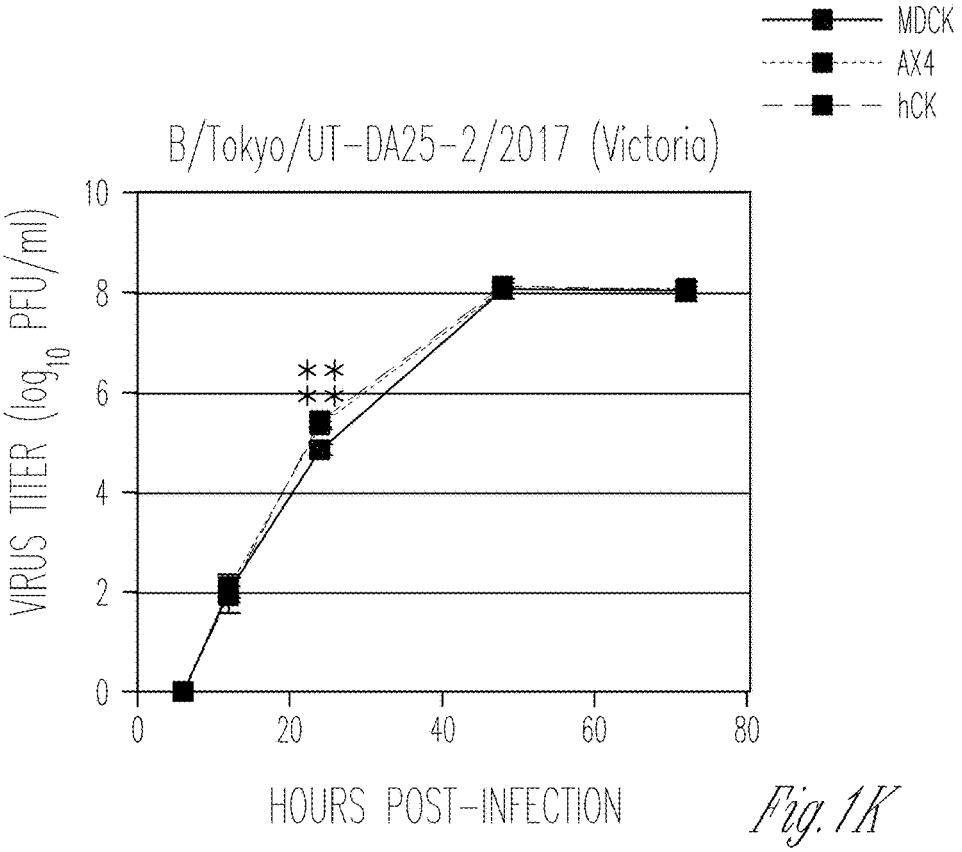
MDCK
AX4
hCK
B/Tokyo/UT-DA25-2/2017 (Victoria)
VIRUS TITER (log_{10} PFU/ml)
10
8
6
4
2
0
**
**
0
20
40
60
80
HOURS POST-INFECTION

Fig. 1K

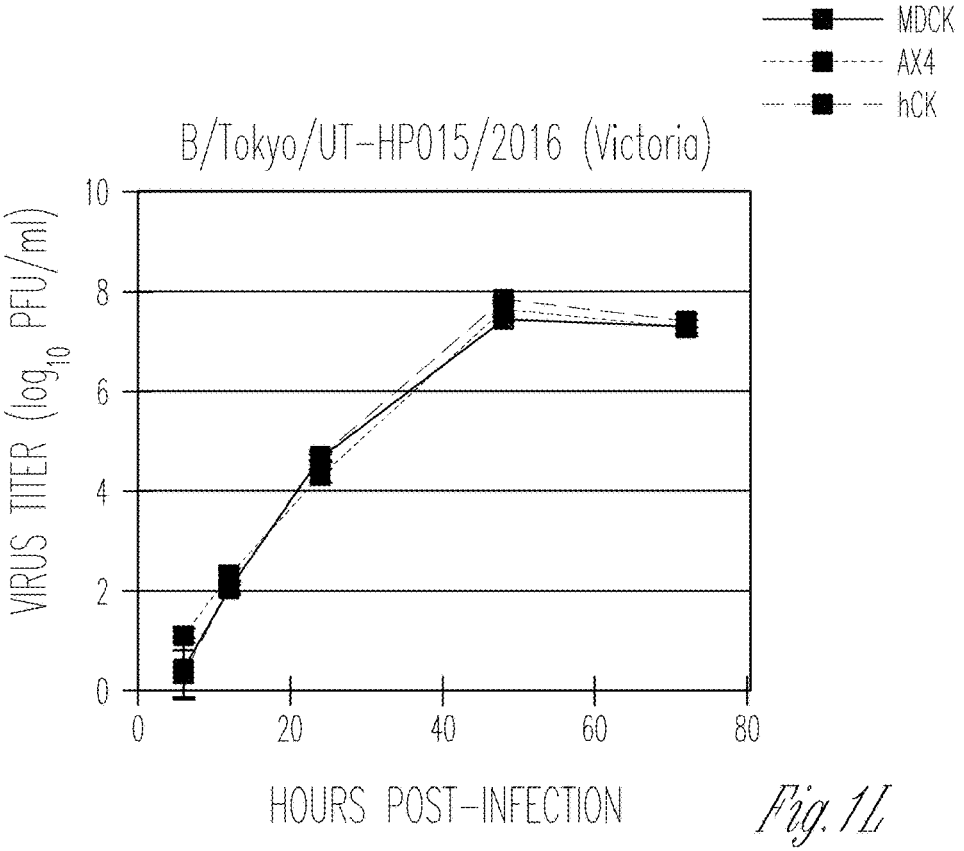
MDCK
AX4
hCK
B/Tokyo/UT-HP015/2016 (Victoria)
VIRUS TITER (log_{10} PFU/ml)
10
8
6
4
2
0
0
20
40
60
80
HOURS POST-INFECTION

Fig. 1L

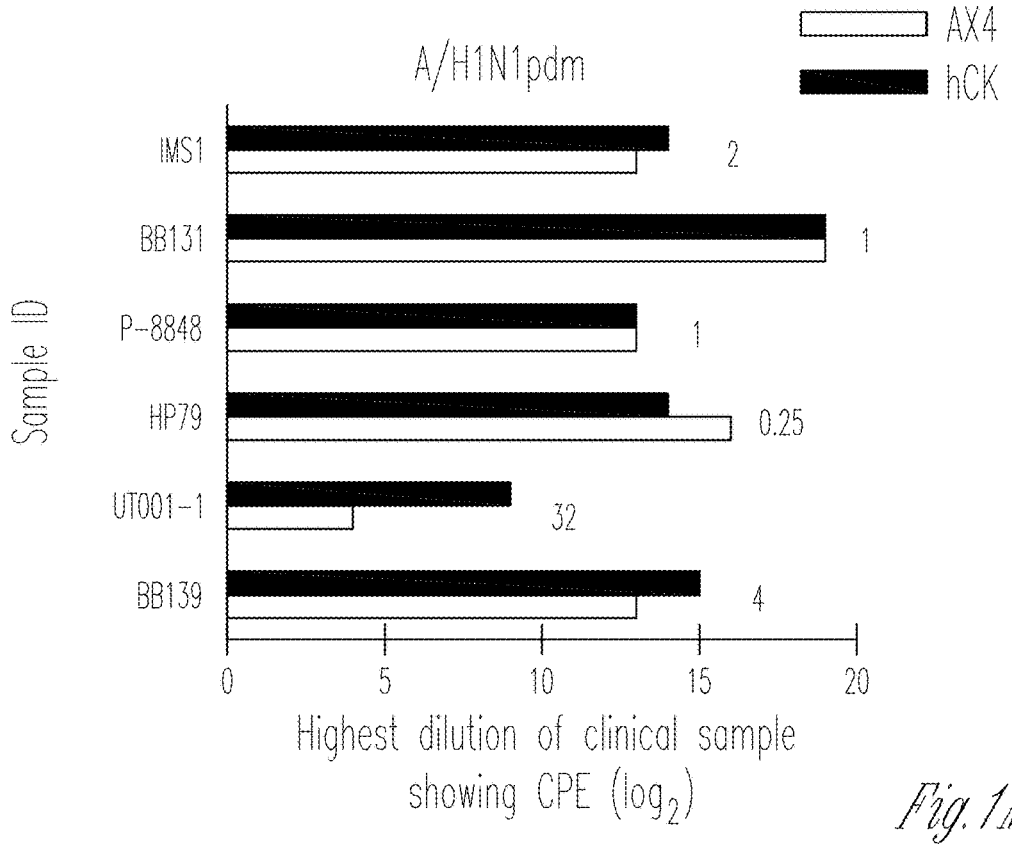
A/H1N1pdm
AX4
hCK
Sample ID
IMS1
BB131
P-8848
HP79
UT001-1
BB139
2
1
1
0.25
32
4
0
5
10
15
20
Highest dilution of clinical sample showing CPE (log_2)

Fig. 1M

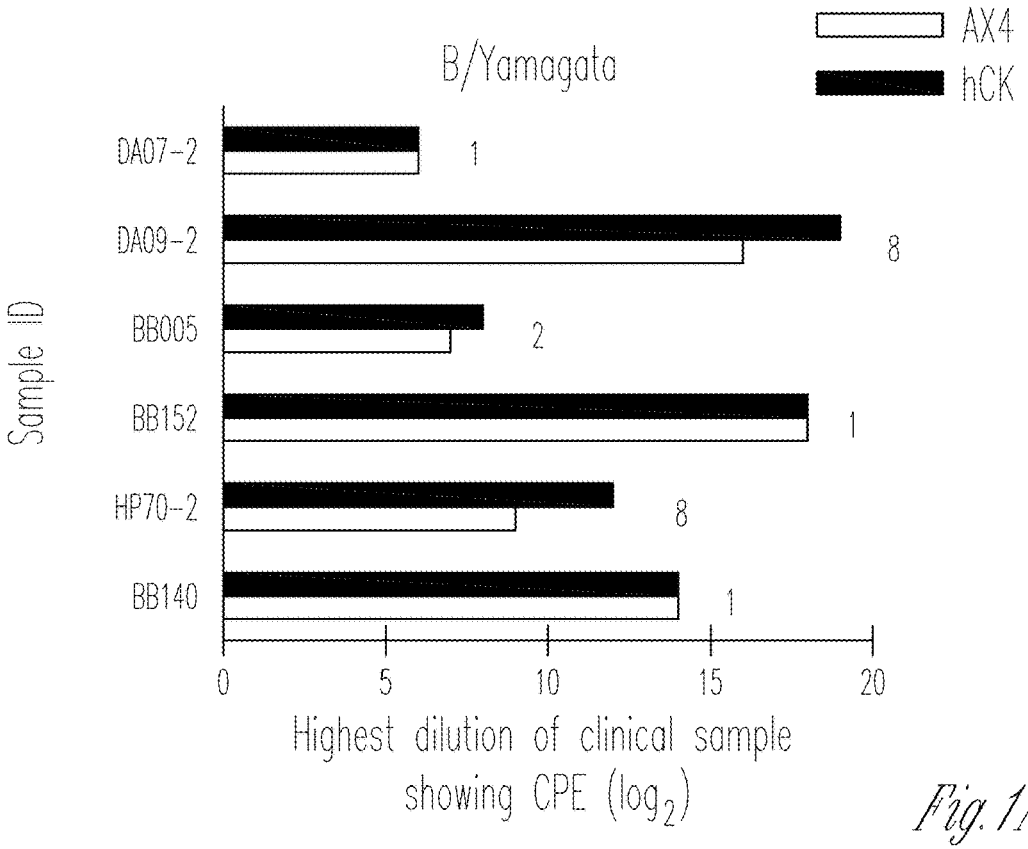
B/Yamagata
AX4
hCK
Sample ID
DA07-2
DA09-2
BB005
BB152
HP70-2
BB140
1
8
2
1
8
1
0
5
10
15
20
Highest dilution of clinical sample showing CPE (log_2)

Fig. 1N

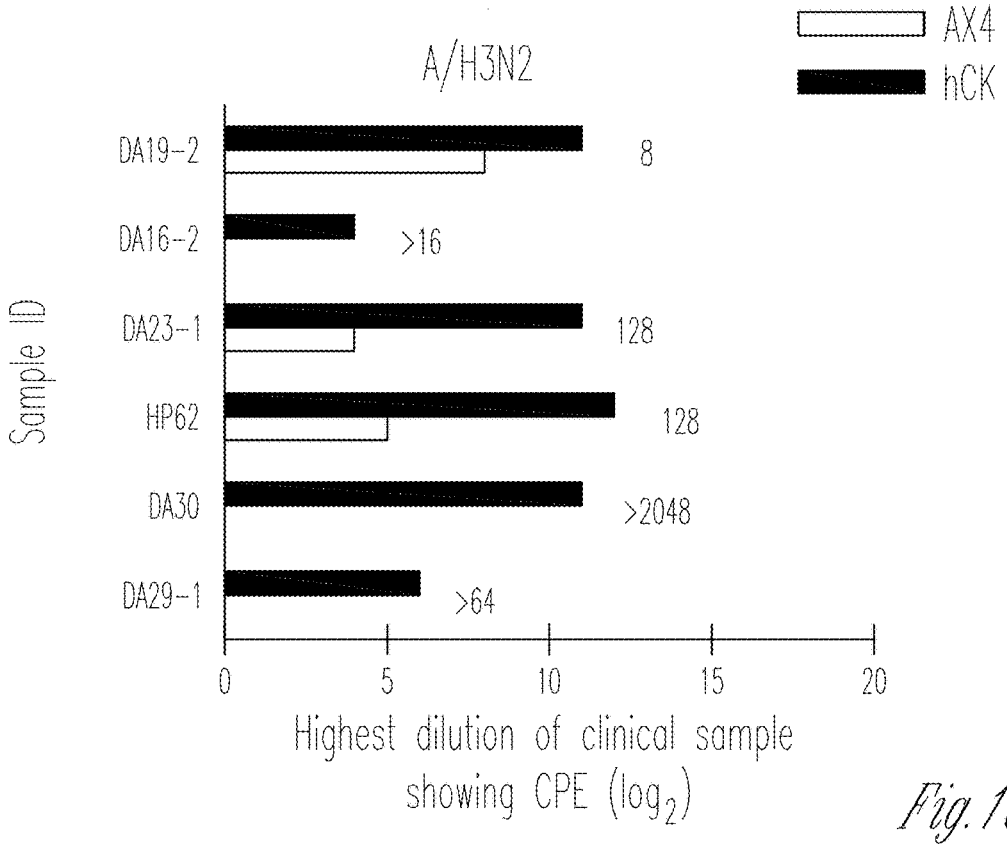
A/H3N2
AX4
hCK
Sample ID
DA19-2
DA16-2
DA23-1
HP62
DA30
DA29-1
8
>16
128
128
>2048
>64
0
5
10
15
20
Highest dilution of clinical sample showing CPE (log2)

Fig. 10

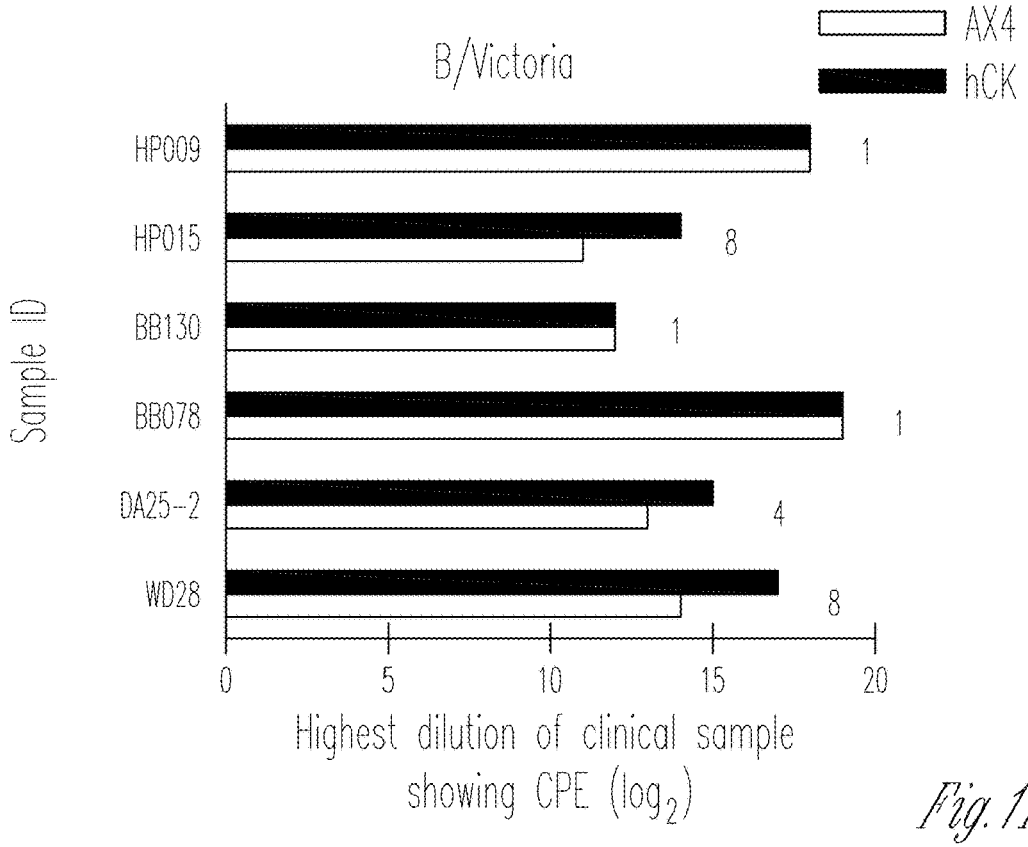
B/Victoria
AX4
hCK
Sample ID
HP009
HP015
BB130
BB078
DA25-2
WD28
1
8
1
1
4
8
0
5
10
15
20
Highest dilution of clinical sample showing CPE (log2)

Fig. 1P

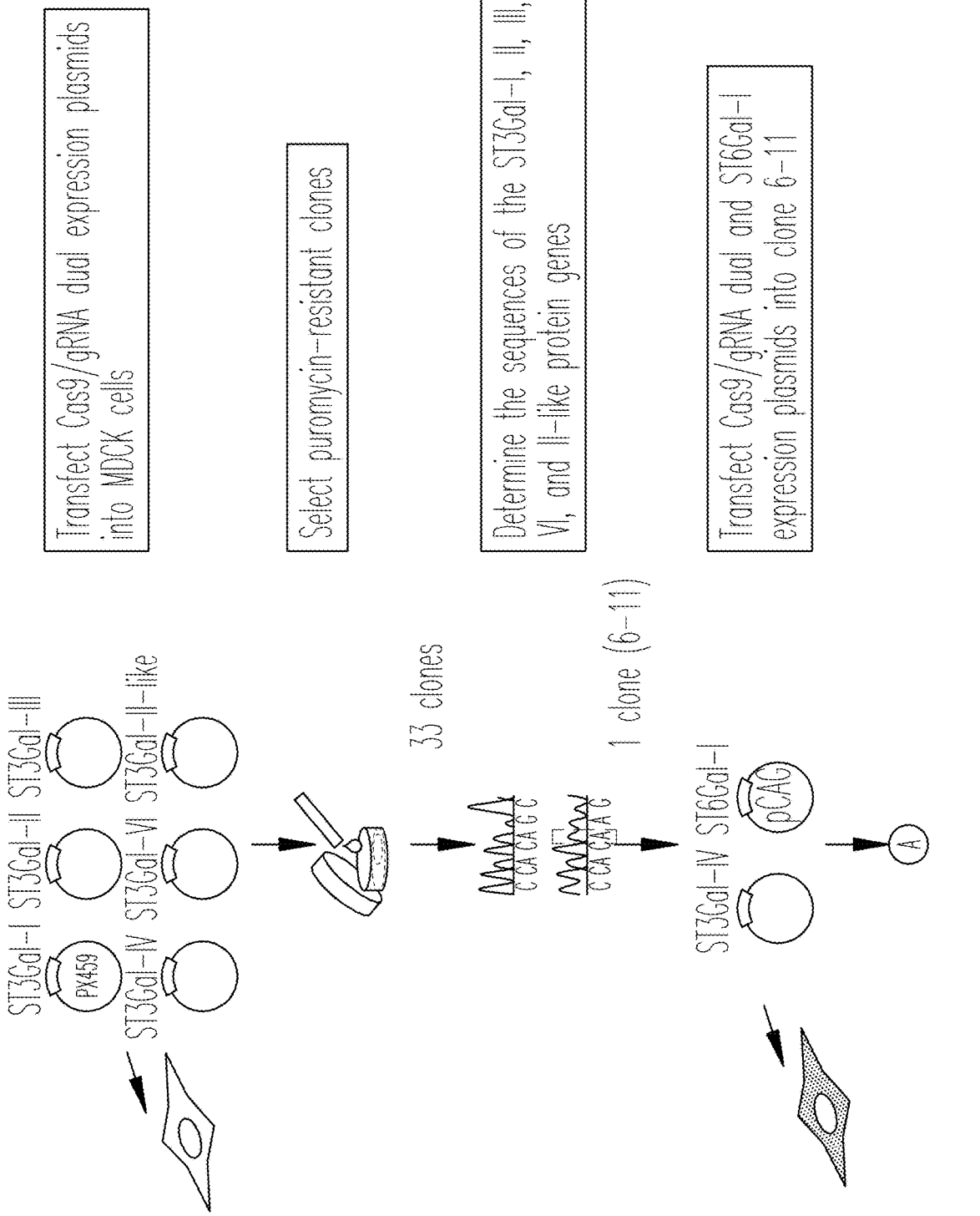
Transfect Cas9/gRNA dual expression plasmids into MDCK cells
Select puromycin-resistant clones
Determine the sequences of the ST3Gal-I, II, III, IV, VI, and II-like protein genes
Transfect Cas9/gRNA dual and ST6Gal-I expression plasmids into clone 6-11
ST3Gal-I ST3Gal-II ST3Gal-III
PX459
ST3Gal-IV ST3Gal-VI ST3Gal-II-like
33 clones
1 clone (6-11)
ST3Gal-IV ST6Gal-I
pCAG
A

Fig. 2A

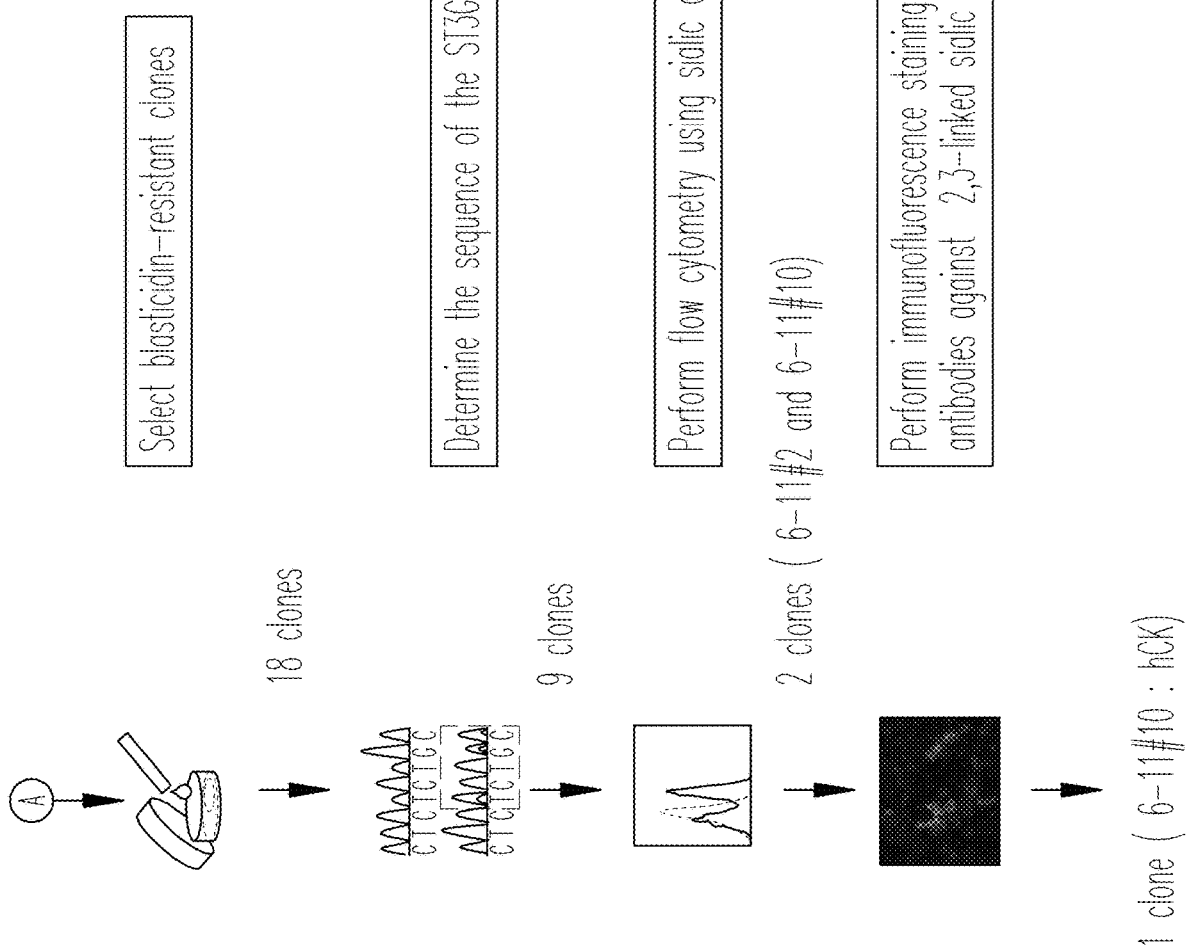
A
Select blasticidin-resistant clones
18 clones
Determine the sequence of the ST3Gal-V gene
9 clones
Perform flow cytometry using sialic acid-specific lectins
2 clones (6-11#2 and 6-11#10)
Perform immunofluorescence staining using antibodies against 2,3-linked sialic acid
1 clone (6-11#10 : hCK)

Fig. 2B

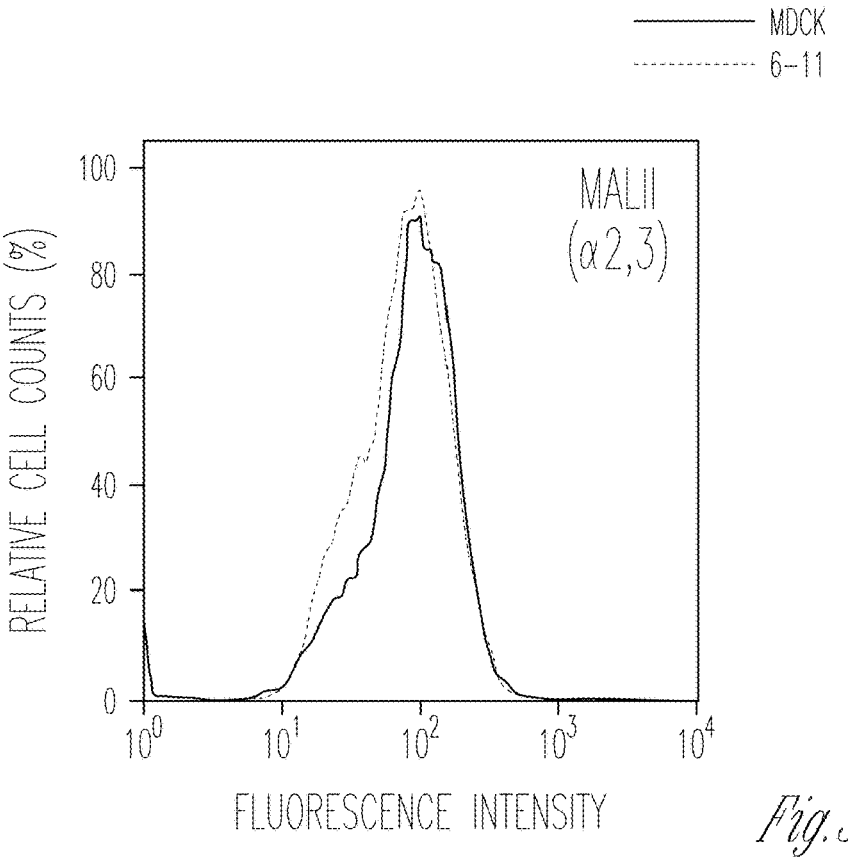
MDCK
6-11
MALII
(α2,3)
RELATIVE CELL COUNTS (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FLUORESCENCE INTENSITY

Fig. 3A

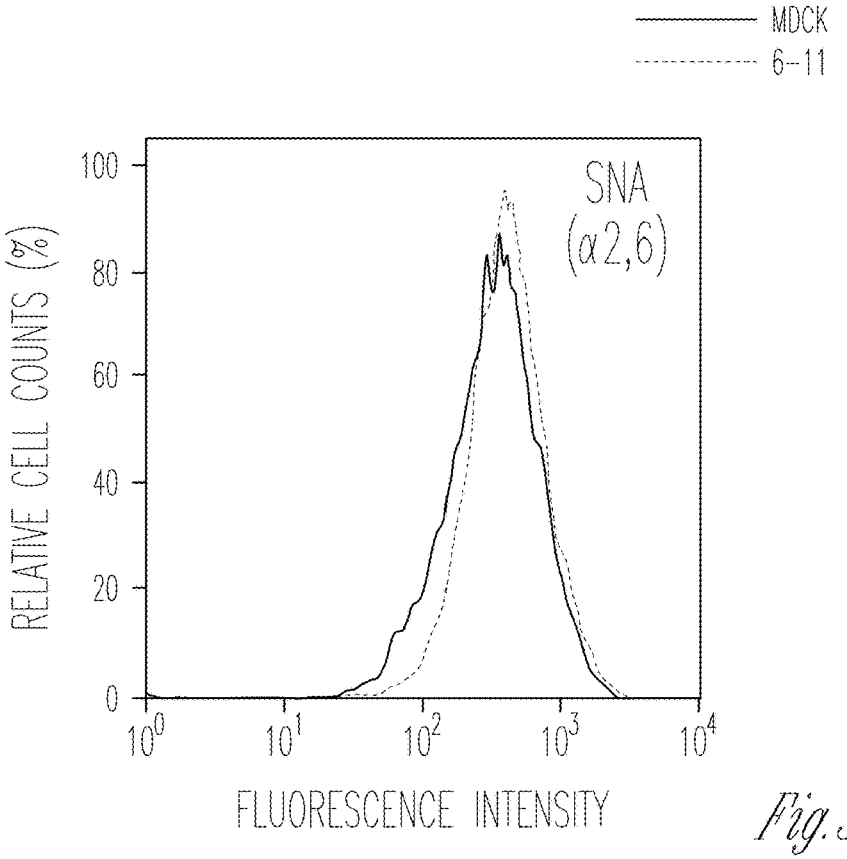
MDCK
6-11
SNA
(α2,6)
RELATIVE CELL COUNTS (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FLUORESCENCE INTENSITY

Fig. 3B

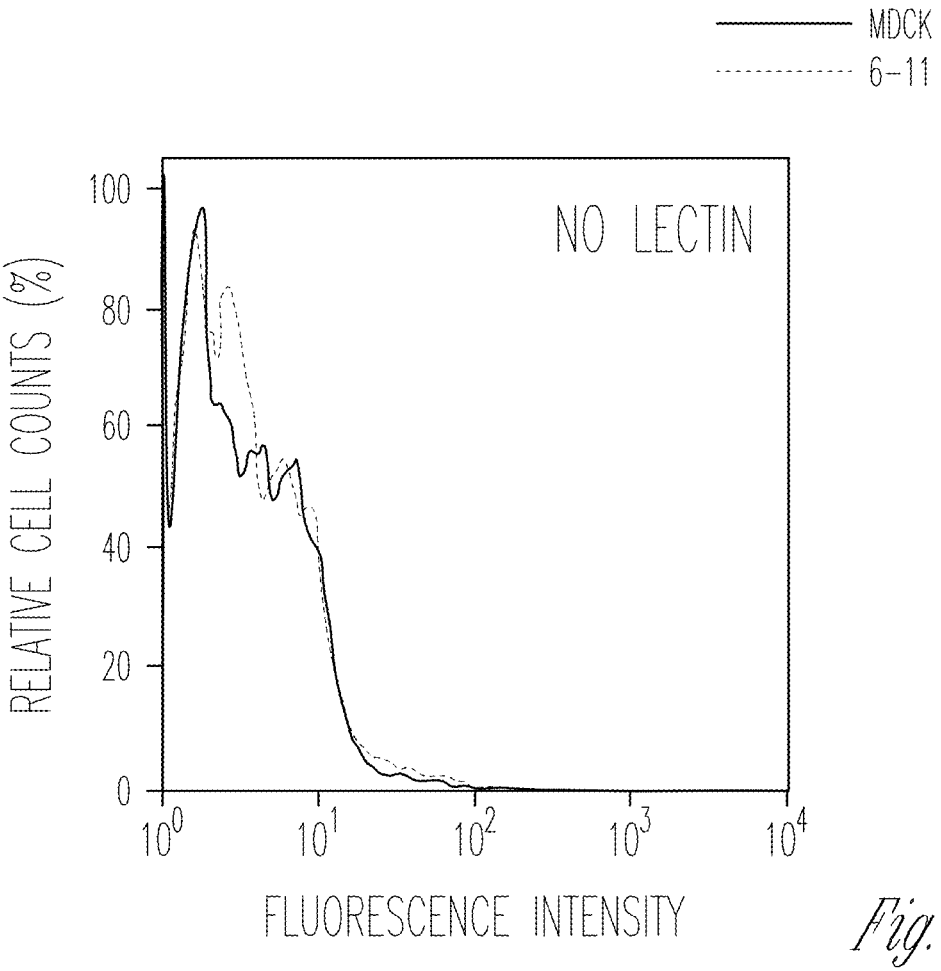
MDCK
6-11
NO LECTIN
RELATIVE CELL COUNTS (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FLUORESCENCE INTENSITY

Fig. 3C

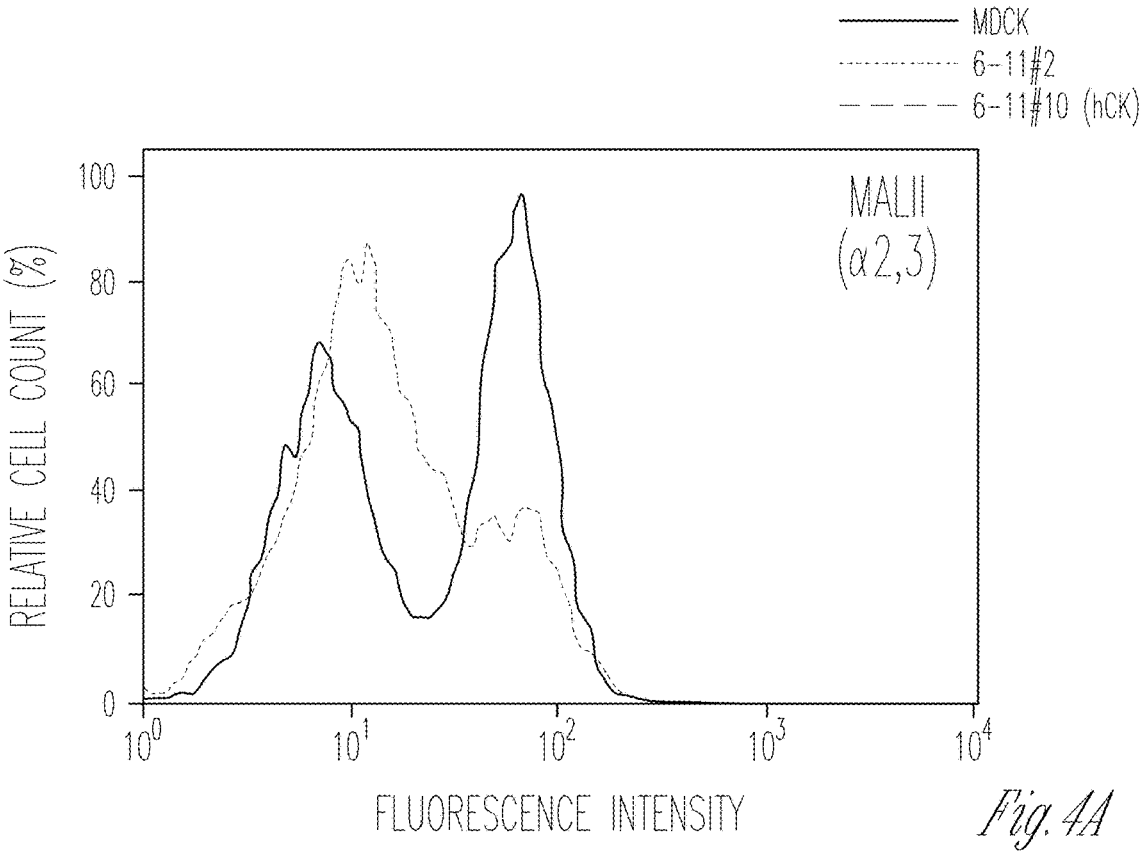
MDCK
6-11#2
6-11#10 (hCK)
MALII
(α2,3)
RELATIVE CELL COUNT (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FLUORESCENCE INTENSITY

Fig. 4A

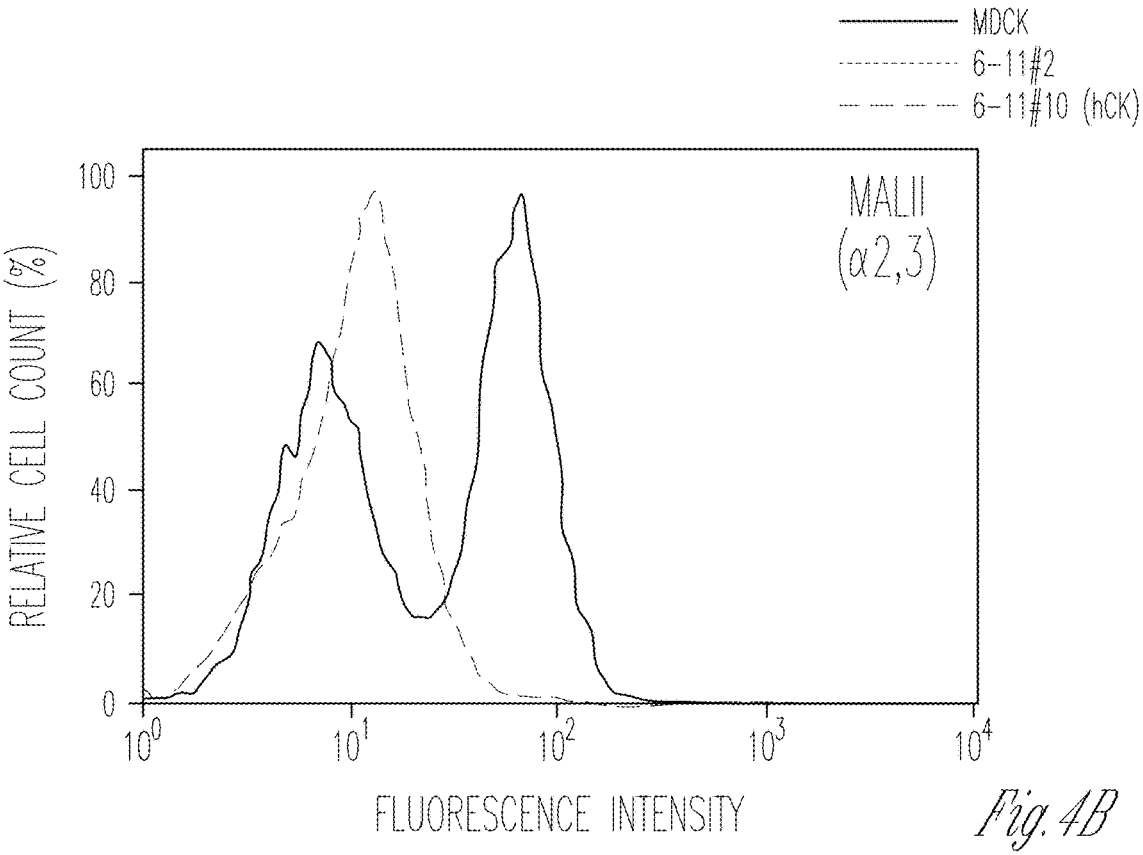
MDCK
6-11#2
6-11#10 (hCK)
MALII
(α2,3)
RELATIVE CELL COUNT (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FLUORESCENCE INTENSITY

Fig. 4B

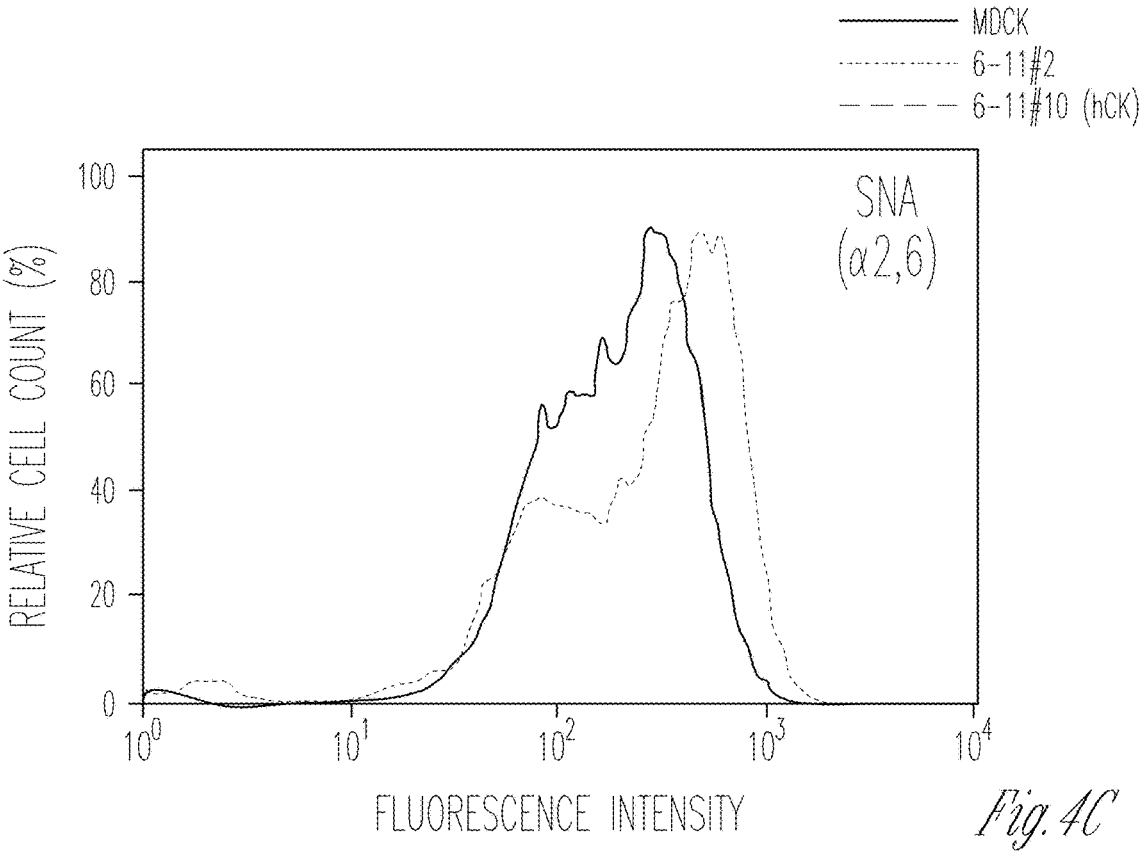
MDCK
6-11#2
6-11#10 (hCK)
SNA
(α2,6)
RELATIVE CELL COUNT (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FLUORESCENCE INTENSITY

Fig. 4C

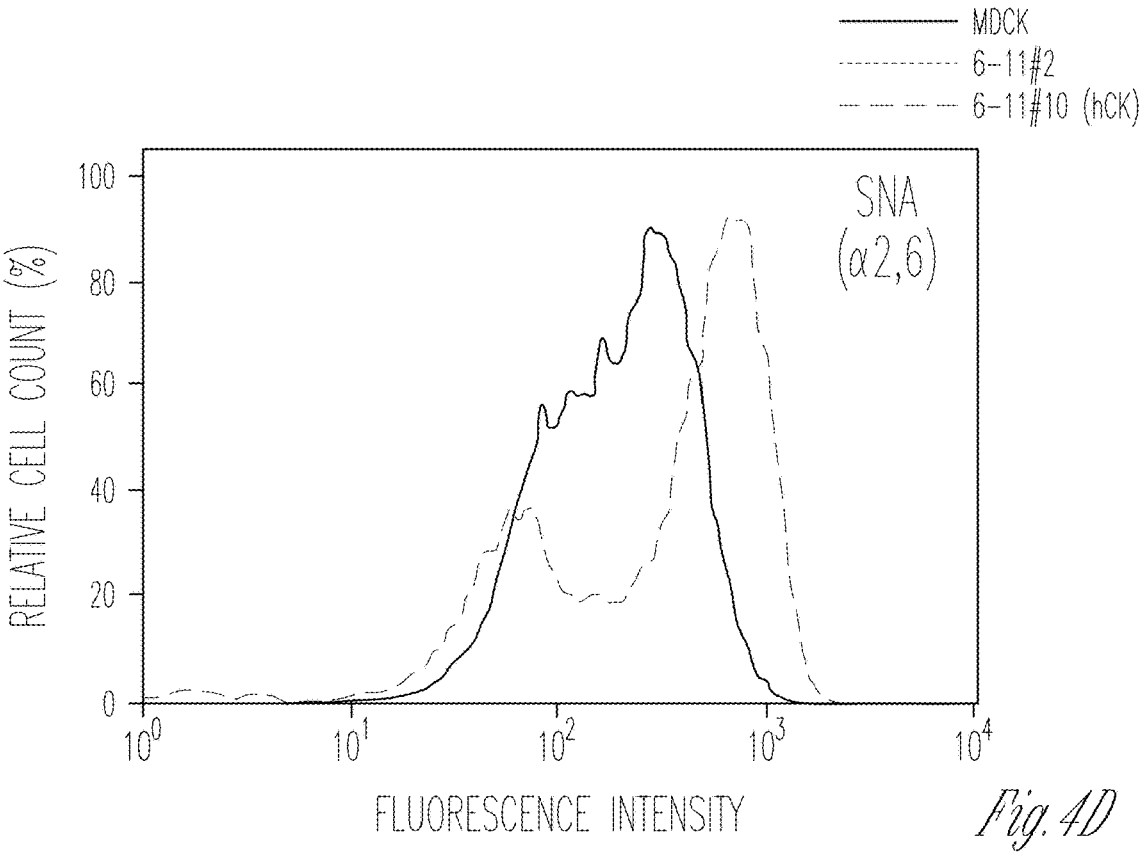
MDCK
6-11#2
6-11#10 (hCK)
SNA
(α2,6)
RELATIVE CELL COUNT (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FLUORESCENCE INTENSITY

Fig. 4D

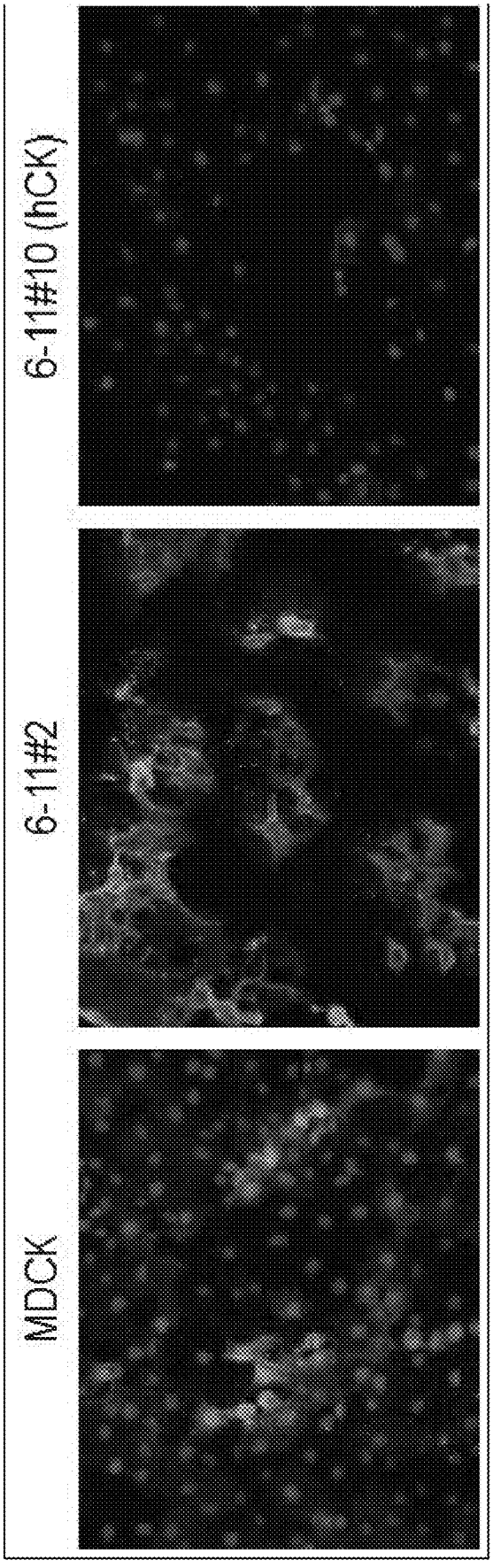

Fig. 4G

Fig. 4F

Fig. 4E

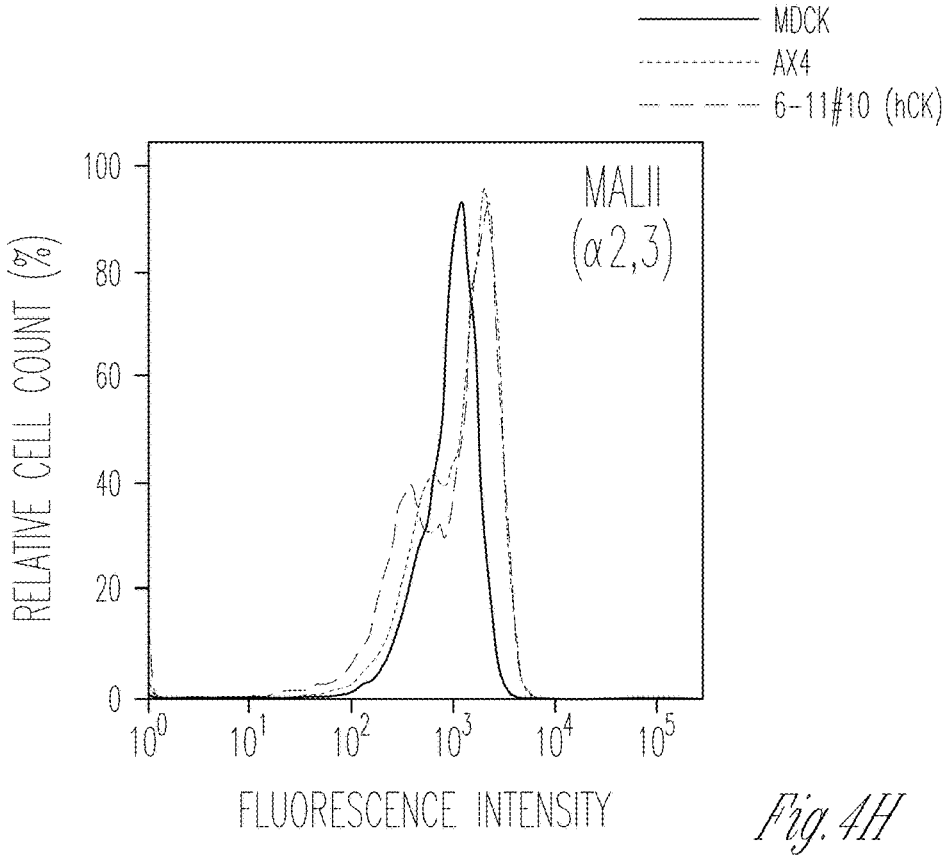
MDCK
AX4
6-11#10 (hCK)
MALII
(α2,3)
RELATIVE CELL COUNT (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
10^5
FLUORESCENCE INTENSITY

Fig. 4H

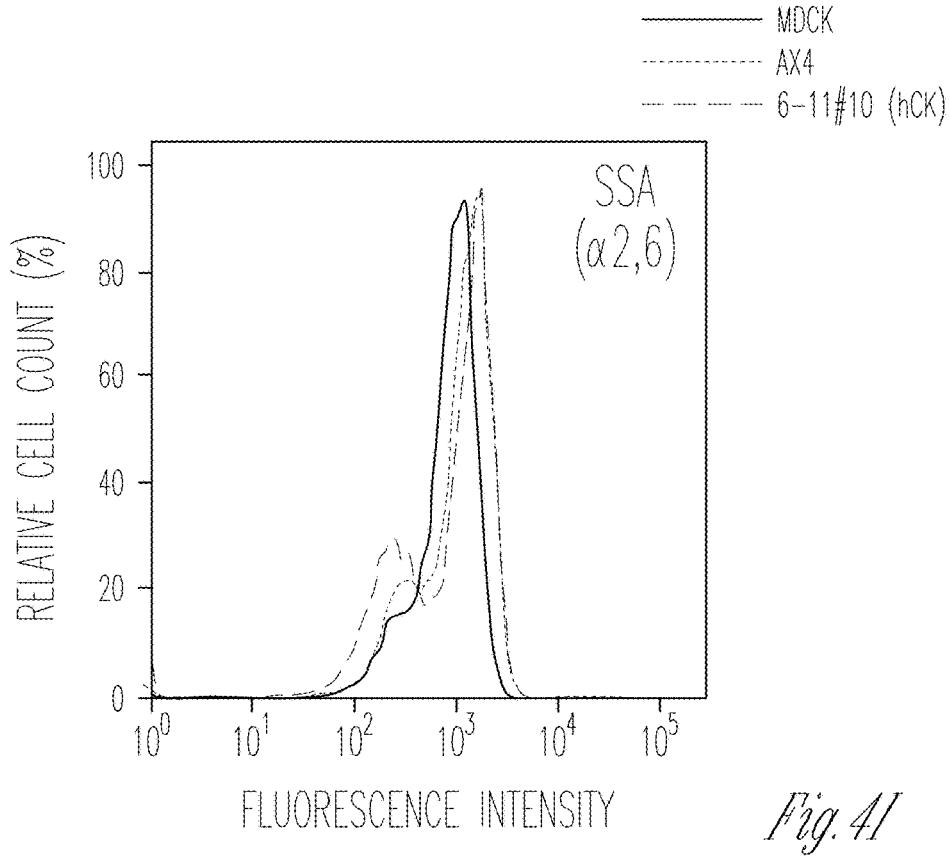
MDCK
AX4
6-11#10 (hCK)
SSA
(α2,6)
RELATIVE CELL COUNT (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
10^5
FLUORESCENCE INTENSITY

Fig. 4I

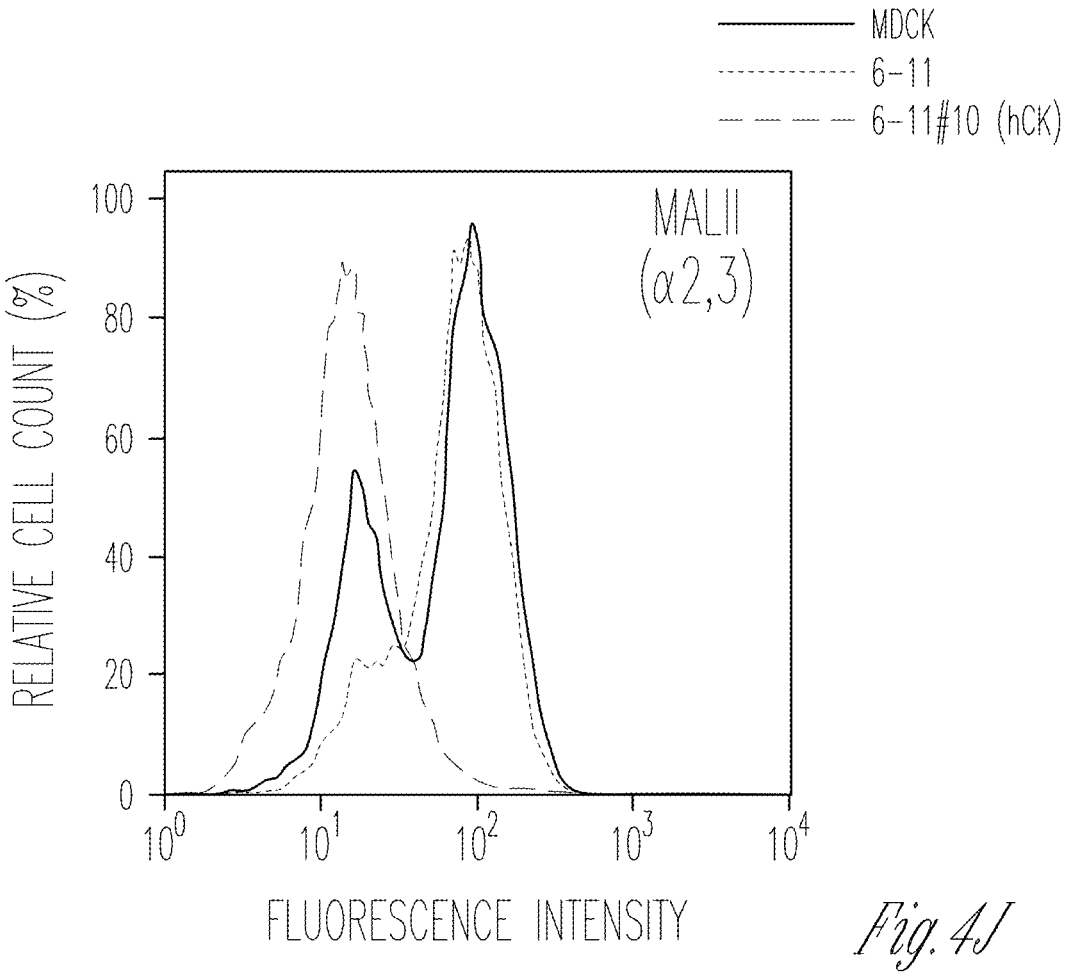
MDCK
6-11
6-11#10 (hCK)
MALII
(α2,3)
RELATIVE CELL COUNT (%)
100
80
60
40
20
0
10^0
10^1
10^2
10^3
10^4
FLUORESCENCE INTENSITY

Fig. 4J

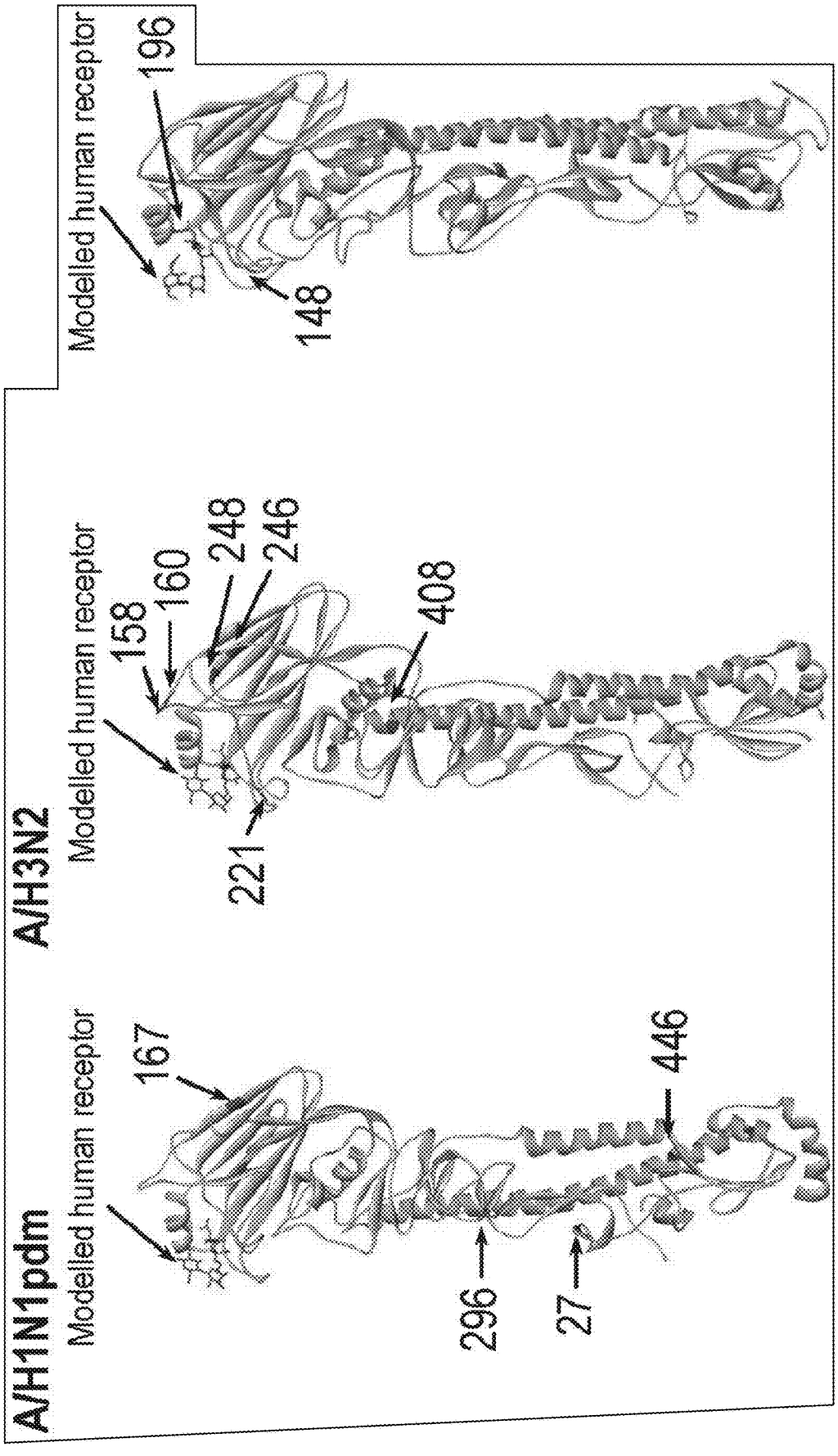

Fig. 5A Fig. 5B Fig. 5C

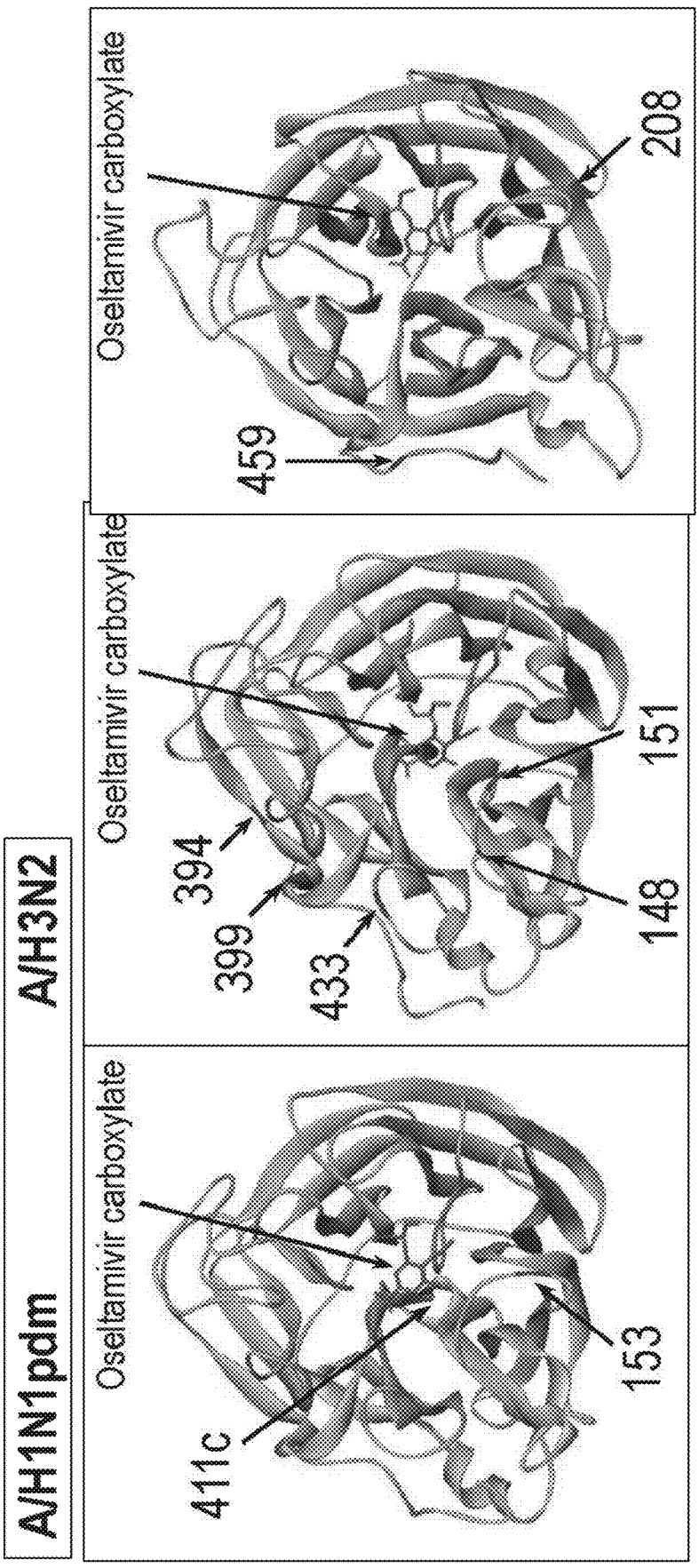

Fig. 5D

Fig. 5E

Fig. 5F

HUMANIZED CELL LINE

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of the filing date of U.S. application No. 62/803,266, filed on Feb. 8, 2019, the disclosure of which is incorporated by reference herein.

STATEMENT OF GOVERNMENT RIGHTS

[0002] This invention was made with government support under HHSN272201400008C awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND

[0003] The influenza A and B viruses possess two major surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). HA recognizes sialic acid-containing receptors on the cell surface, while NA cleaves sialic acids from receptors on cellular surfaces to facilitate the release of progeny virions from the surface of infected cells (Gamblin and Skehel, 2010). HA is also the major antigen stimulating the host's protective immunity, specifically the production of neutralizing antibodies.

[0004] Virus isolation from clinical specimens is an essential tool for the identification and characterization of circulating viruses. Currently, two subtypes of influenza A viruses (A/H1N1 and A/H3N2) and two lineages of influenza B viruses (B/Yamagata- and B/Victoria-lineage) are cocirculating in the human population and cause epidemics of seasonal influenza. Madin-Darby canine kidney (MDCK) cells are the most widely used cell line for isolation and propagation of human influenza viruses. This cell line shows high susceptibility to influenza viruses; however, it supports the growth of recent A/H3N2 viruses poorly. Furthermore, passaging of influenza viruses in MDCK cells often leads to the selection of variants with mutations in their HA and/or NA genes (Chambers et al., 2014; Lee et al., 2013; Tamura et al., 2013; Lin et al., 2017; Li et al., 2009; Oh et al., 2008). The emergence of such variants carrying mutations relevant to adaptation of influenza viruses to cell culture could distort the evaluation of the antigenic, genetic, and antiviral properties of circulating influenza viruses. For example, the emergence of mutations that confer receptor-binding activity to the NA of A/H3N2 viruses, such as the aspartic acid-to-glycine substitution at position 151 (D151G) (Mob et al., 2015; Lin et al., 2010; Zhu et al., 2012), is problematic for characterization of HA antigenicity by means of hemagglutination-inhibition, virus-neutralization, and focus reduction assays because the receptor-binding activity of NA contributes to the results of these assays. Nevertheless, many laboratories use MDCK cells to isolate A/H3N2 viruses. A GISAID EpiFlu database analysis by Lee et al. (2013) showed that approximately 30% of MDCK-cultured A/H3N2 isolates possess an amino acid change at position 151. Therefore, currently circulating A/H3N2 strains should be isolated and propagated in cell lines that can faithfully maintain their characteristics.

[0005] The HAs of human influenza viruses prefer to bind to glycans that end with sialic acid linked to galactose by α2,6-linkages, whereas avian virus HAs preferentially bind to glycans that terminate with sialic acid linked to galactose by α2,3-linkages (Connor et al., 1994; Rogers and Paulson, 1983; Stevens et al., 2006). Correspondingly, epithelial cells in the human upper respiratory tract express predominantly α2,6-sialoglycans (van Riel et al., 2006; Shinya et al., 2006). Although MDCK cells expressing both α2,6- and α2,3-sialoglycans are suitable for the isolation of influenza viruses from multiple animal species, this cell line has been shown to express relatively low levels of α2,6-sialoglycans (Lin et al., 2017; Hatakeyama et al., 2005; Matrosovich et al., 2003). Previously, our group and others engineered MDCK cells to overexpress α2,6-sialoglycans (Hatakeyama et al., 2005; Matrosovich et al., 2003). These modified MDCK cells (designated AX4 or MDCK-SIAT1) displayed a higher sensitivity for human influenza virus isolation than a conventional MDCK cell line (Oh et al., 2008; Hatakeyama et al., 2005), yet they still expressed α2,3-sialoglycans. Importantly, as with conventional MDCK cells, variants with mutations in either HA or NA have been detected when seasonal influenza viruses were passaged through MDCK-SIAT1 cells (Tamura et al., 2013; Li et al., 2009). Therefore, an alternative cell line that supports efficient isolation and propagation of human influenza viruses without any cell culture-adaptive mutations is necessary for accurate characterization of circulating viruses and possibly for efficient vaccine production in cells.

SUMMARY

[0006] The present disclosure relates to a mammalian or avian cell line that is genetically modified to support, for example, more efficient isolation and/or amplification (propagation) of human influenza viruses, and in particular human H3 influenza viruses. The disclosed cell lines may be genetically modified to decrease expression of alpha-2,3-linked sialic acids on the cell surface and to increase expression of alpha-2,6-linked sialic acids relative to a parental cell lines that are not modified to alter expression of alpha-2,3-linked sialic acids, alpha-2,6-linked sialic acids, or both. In one embodiment, the modified mammalian or avian cell lines are modified to express high levels of human influenza virus receptors and low levels of avian influenza virus receptors. In one embodiment, the cell line is a mammalian cell line, e.g., a non-human cell line such as a primate cell line, or a canine cell line. In one embodiment, the modified cell line is a modified MDCK cell line that has decreased expression of alpha-2,3-linked sialic acid relative to AX-4, or increased expression of alpha-2,6-linked sialic acid relative to unmodified MDCK cells. In one embodiment, the modified cell line is hCK, which supports more efficient isolation and amplification of human influenza viruses compared to MDCK and AX-4 cells. In one embodiment, the decrease in expression of alpha-2,3-linked sialic acids is due to a genetic modification that decreases or eliminates expression of one or more sialyltransferases that produce alpha-2,3-linked sialic acids, a genetic modification including but not limited to an insertion of one or more nucleotides, a deletion of one or more nucleotides, a substitution of one or more nucleotides, or any combination thereof, in one or more sialyltransferase genes. In one embodiment, the genetic modification includes an insertion of one or more nucleotides in one or more sialyltransferase genes. In one embodiment, the genetic modification includes a deletion of one or more nucleotides in one or more sialyltransferase genes. In one embodiment, the genetic modification includes a substitution of one or more nucleotides in one or more sialyltransferase genes. In one embodi-

ment, the genetic modification includes an insertion of one or more nucleotides in at least one sialyltransferase gene. In one embodiment, the genetic modification includes a deletion of one or more nucleotides in at least one sialyltransferase gene. In one embodiment, the genetic modification includes a substitution of one or more nucleotides in at least one sialyltransferase gene. The genetic modifications that decrease expression of alpha-2,3-linked sialic acids may be the result of any method that "knocks down" or "knocks out" expression, methods including the uses of recombinase systems such as CRISPR/Cas, TALEN or zinc finger binding proteins. In one embodiment, the increase in expression of alpha-2,6-linked sialic acids is due to a genetic modification that increase expression of one or more sialyltransferases that produce alpha-2,6-linked sialic acids, a genetic modification including but not limited to an expression cassette comprising a nucleotide sequence encoding a sialyltransferase that that produces alpha-2,6-linked sialic acids, e.g., a human β-galactoside α2,6-sialyltransferase I (ST6Gal I) gene.

[0007] In one embodiment, an isolated recombinant mammalian or avian cell comprising a reduced amount of cell surface β-galactoside α2,3 sialyl residues and an increased amount of human β-galactoside α2,6 sialyl residues relative to a corresponding non-recombinant mammalian or avian cell is provided. In one embodiment, the isolated recombinant cell is a non-human mammalian cell. In one embodiment, the isolated recombinant cell is a canine or non-human primate cell. In one embodiment, the reduced amount of surface β-galactoside α2,3 sialyl residues is the result of a reduced amount or activity of one or more α2,3 sialyltransferases, e.g., a reduction in the amount or activity of one or more α2,3 sialyltransferases of at least 5%, 10%, 20%, 50%, 70%, 80%, 90%, 95% or more, which may result in a reduction of at least 5%, 10%, 20%, 50%, 70%, 80%, 90%, 95% or more in α2,3 sialyl residues, in the recombinant cell. In one embodiment, the amount or activity of α2,3 sialyltransferases, or the amount of α2,3 sialyl residues, in the recombinant cell is undetectable. In one embodiment, the α2,3 sialyltransferase gene that is modified encodes an α2,3 sialyltransferase that has at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to any one of SEQ ID Nos. 6, 8, 10, 12, 14, 16 or 18, or a nucleotide sequence having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleotide acid sequence identity to any one of SEQ ID Nos. 5, 7, 9, 11, 13, 15, or 17. In one embodiment, the isolated recombinant cell comprises an expression cassette encoding human β-galactoside α2,6 sialyltransferase I (ST6Gal-I) or ST6Gal-II. In one embodiment, the ST6Gal-I or ST6Gal-II comprises a protein having at least 80% amino acid sequence identity to any one of SEQ ID Nos. 1-4, 101 or 150. In one embodiment, the α2,6 sialyltransferase gene encodes an α2,6 sialyltransferase that has at least 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to any one of SEQ ID Nos. 1-4, 101 or 150, or a nucleotide sequence having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleotide acid sequence identity to SEQ ID Nos. 101 or 151. In one embodiment, the human β-galactoside α2,6 sialyltransferase amount or activity in the recombinant cell in creased by at least 1%, 5%, 10%, 20%, 50%, 70%, 80%, 90%, 95% or more. In one embodiment, one or more β-galactoside α2,3 sialyltransferase genes are mutated so as to reduce the amount of cell surface β-galactoside α2,3 sialyl residues. In one embodiment, two or more of ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV, ST3Gal-V ST3Gal-VI, or ST3Gal-II-like genes are mutated. In one embodiment, three, four, five, six or seven of ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV, ST3Gal-V, ST3Gal-VI, or ST3Gal-II-like genes are mutated. In one embodiment, the ST3 genes have at least 80% nucleic acid sequence identity to any one of SEQ ID Nos. 5, 7, 9, 11, 13, 15, or 17. In one embodiment, the reduction in cell surface β-galactoside α2,3 sialyl residues is the result of reduced expression of one or more ST3 sialyltransferases. In one embodiment, the one or more ST3 sialyltransferases have at least 80% amino acid sequence identity to any one of SEQ ID Nos. 6, 8, 10, 12, 14, 16, or 18. In one embodiment, influenza H3 viruses replicate more efficiently in the recombinant cell relative to the non-recombinant cell.

[0008] Further provided is an isolated recombinant mammalian or avian cell, comprising a reduced amount of cell surface β-galactoside α2,3 sialyl residues relative to a corresponding non-recombinant mammalian or avian cell. In one embodiment, one or more of ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV; ST3Gal-V, ST3Gal-VI, or ST3Gal-II-like genes in the recombinant cell are mutated. In one embodiment, a combination of ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV, ST3Gal-V; ST3Gal-VI, or ST3Gal-II-like genes in ther recombinant cell are mutated. In one embodiment, ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV ST3Gal-V, ST3Gal-VI, and ST3Gal-II-like genes in the recombinant cell are mutated.

[0009] The recombinant cells described herein are useful, for example, in virus isolation, vaccine production and in diagnostics. For example, the recombinant cells allow for isolation and/or amplification of progeny viruses. Moreover, HA assays generally are not used to detect human H3N2 viruses. The recombinant cells may be advantageous in that regard, e.g., to amplify virus.

[0010] Further, recombinant cells that have increased β-galactoside α2,6 sialyl residues can be used as a source of isolated β-galactoside α2,6 sialyl, which in turn may be used to coat surfaces such as beads, to inhibit galectin(s), to isolate or detect *Sambucus nigra* agglutinin (SNA), *Sambucus sieboldiana* (SSA) or *Trichosanthes japonica* agglutinin I (TJA-I).

[0011] In one embodiment, a method of modifying the amount of cell surface β-galactoside α2,3 sialyl residues and human β-galactoside α2,6 sialyl residues on a mammalian or an avian cell is provided. In one embodiment, the method includes mutating one or more β-galactoside α2,3 sialyltransferase (ST3Gal) genes, and overexpressing a human β-galactoside α2,6 sialyltransferase (ST6Gal) gene, in a parental mammalian or avian cell so as to result in a modified mammalian or avian cell having a reduced amount of cell surface β-galactoside α2,3 sialyl residues and an increased amount of human β-galactoside α2,6 sialyl residues on the surface of the modified cell relative to the corresponding parental cell. In one embodiment, the one or more ST3Gal genes are mutated using a genome editing system, e.g., a CRISPR/Cas9, Zinc Finger Nuclease (ZFN) or transcription activator-like effector nuclease (TALEN) system. In one embodiment, the mutations include one or more nucleotide insertions or one or more nucleotide deletions, or both, in one or more ST3 genes. In one embodiment, the modified cell comprises an expression cassette comprising a ST6Gal open reading frame. In one embodi-

ment, the modified cell is a kidney cell. In one embodiment, the modified cell is a canine cell.

[0012] Methods of using the recombinant cell include a method of propagating an influenza virus, e.g., a human influenza virus, for vaccine production. In one embodiment, the influenza virus is an influenza A virus. In one embodiment, the influenza virus is an influenza B virus. In one embodiment, the influenza virus is a H3 virus. In one embodiment, the virus is A/H1N1, A/H3N2, a B/Yamagata-lineage influenza B virus or a B/Victoria-lineage influenza B virus.

[0013] A further method which employs the recombinant cell is a method of isolating an influenza virus. The method includes providing a sample from an avian or a mammal suspected of being infected with an influenza virus; and contacting the recombinant cell with the sample. In one embodiment, the method further includes determining whether the sample is infected with an influenza virus. In one embodiment, the method further includes identifying the HA and/or NA subtype of the virus.

[0014] In one embodiment, the cell line is a modified MDCK cell line, 'hCK' for 'humanized MDCK' cells, which was prepared using CRISPR/Cas-mediated gene knock-out methods to down-regulate sialyltransferases that catalyze the synthesis of alpha-2,3-linked sialic acids, and overexpression of a sialyltransferase that catalyzes the synthesis of alpha-2,6-linked sialic acids. hCK cells express low levels of alpha-2,3-linked sialic acids and high levels of alpha-2,6-linked sialic acids (similar to human epithelial cells in the upper respiratory tract). As disclosed herein, hCK cells allow for the isolation of H3N2 human influenza viruses 10-100 better than the AX-4 cell line. Efficient isolation and amplification of influenza viruses including human influenza viruses is advantageous for vaccine production (possibly supporting better replication), e.g., for seasonal influenza virus vaccine production, as seasonal human influenza viruses often replicate inefficiently in unmodified MDCK cells and even in MDCK (AX-4) cells overexpressing alpha-2,6-linked sialic acids on their surface, to which human influenza viruses bind efficiently. In one embodiment, the titer of human influenza viruses on the modified cell line disclosed herein is at least one log, at least two logs, at least three logs or greater than in unmodified MDCK cells, MDCK (AX-4) cells or MDCK-SIAT1 cells (Li et al., 2009).

BRIEF DESCRIPTION OF THE FIGURES

[0015] FIGS. **1**A-IP. Sensitivity of hCK cells to human influenza virus growth and isolation. A-L) Growth kinetics of seasonal influenza viruses in MDCK, AX-4, and hCK cells. MDCK, AX4, and hCK cells were infected with viruses at a multiplicity of infection (MOI) of 0.002. The supernatants of the infected cells were harvested at the indicated times, and virus titers were determined by means of plaque assays in hCK cells. Error bars indicate standard deviations from three independent experiments. P values were calculated by using the linear mixed model described in the Methods section (*P<0.05; **P<0.01). Red and blue asterisks indicate the comparison of hCK and AX4 cells with MDCK cells; gray asterisks indicate the comparison between the cell lines depicted in red and blue. M-P) Comparative sensitivity of hCK and AX4 cells to seasonal influenza viruses. Serial 2-fold dilutions (2^1 to 2^{20}) of clinical samples were prepared and inoculated into AX4 and hCK cells. Cells were observed for the development of cytopathic effect (CPE) for 7 days. Three wells were inoculated with each virus dilution. The highest dilution showing CPE in all three wells is shown by the horizontal bar. The number at the end of each horizontal bar indicates the ratio of the hCK highest dilution to the AX4 highest dilution.

[0016] FIGS. **2**A-**2**B. Schematic overview of the generation of MDCK cells expressing markedly low levels of α2,3-linked sialic acid and high levels of α2,6-linked sialic acid.

[0017] FIGS. **3**A-**3**C. Flow cytometric analysis of the cell surface expression of α2,6- and α2,3-linked Sias. A-B) Clone 6-11 (orange line open profiles) and parental MDCK cells (black line open profiles) were incubated with biotinylated *Maackia amurensis* II agglutinin (MAL II) lectin (specific for α2,3-linked sialic acid) or *Sambucus Nigra* agglutinin (SNA) lectin (specific for α2,6-linked Sias), followed by Alexa 488-conjugated streptavidin, and then analyzed by flow cytometry. C) Unstained cells served as negative controls (no lectin).

[0018] FIGS. **4**A-**4**J. Characterization of MDCK cells expressing markedly low levels of α2,3-linked sialic acid and high levels of α2,6-linked sialic acid. MDCK cells carrying mutations in seven different β-galactoside α2,3 sialyltransferase (ST3Gal) genes were generated by using the CRISPR/Cas9 genome-editing system, as described in the Methods section. The edited MDCK cells were further modified to overexpress the human β-galactoside α2,6 sialyltransferase I (ST6Gal-I) by transfection of plasmids containing the ST6Gal-I gene. The modified cell clones were selected with puromycin and blasticidin, and characterized. A-D) Flow cytometric analysis of the cell surface expression of α2,6- and α2,3-linked Sias. Modified MDCK cells (green and red line open profiles) and parental MDCK cells (black line open profiles) were incubated with biotinylated *Maackia Amurensis* II agglutinin (MAL II) lectin (specific for α2,3-linked sialic acid) or *Sambucus Nigra* agglutinin (SNA) lectin (specific for α2,6-linked Sias), followed by Alexa 488-conjugated streptavidin, and then analyzed by flow cytometry. E-G) Immunofluorescence analysis of the expression of α2,3-linked sialic acid. Modified MDCK and parental MDCK cells were fixed and stained with a monoclonal antibody (green) that recognizes Siaα2,3Galβ1,4GlcNAc. Nuclei were stained with Hoechst dye (blue). H-J) Flow cytometric analysis of the cell surface expression of α2,6-linked Sias. Modified MDCK cells (red line open profiles), parental MDCK cells (black line open profiles), and AX4 cells (blue line open profiles) were incubated with SNA lectin or *Sambucus sieboldiana* (SSA) lectin (specific for α2,6-linked Sias), followed by Alexa 488-conjugated streptavidin, and then analyzed by flow cytometry. D) Flow cytometric analysis of the cell surface expression of α2,3-linked Sias. Modified MDCK cells (red line open profiles), parental MDCK cells (black line open profiles), and AX4 cells (blue line open profiles) were incubated with MAL II lectin, followed by Alexa 488-conjugated streptavidin, and then analyzed by flow cytometry.

[0019] FIGS. **5**A-**5**F. Localization of amino acid changes in HA and NA proteins. A-C) Shown are the three-dimensional structures of A/California/04/2009 (H1N1pdm) HA (PDB ID: 3UBN), A/Wyoming/3/2003 (H3N2) HA (PDB ID: 6BKR), and B/Hong Kong/8/1973 HA (PDB ID: 2RFU) in complex with human receptor analogues. Mutations identified in this study are shown in red. Mutations in influenza A virus HA are shown with H3 numbering. D-F) Shown are

the three-dimensional structures of A/California/04/2009 (H1N1 pdm) NA (PDB ID: 3TI6), A/Tanzania/205/2010 (H3N2) NA (PDB ID: 4GZP), and B/Brisbane/60/2008 NA (PDB ID: 4CPM) in complex with oseltamivir carboxylate. Mutations identified in this study are shown in red. All mutations are shown with N2 numbering. Images were created with the DS Visualizer v 17.2.

DETAILED DESCRIPTION

Definitions

[0020] A "vector" or "construct" (sometimes referred to as gene delivery or gene transfer "vehicle") refers to a macromolecule or complex of molecules comprising a polynucleotide to be delivered to a host cell, either in vitro or in vivo. The polynucleotide to be delivered may comprise a coding sequence of interest, may comprise sequences for introducing mutations into a host cell genome, or both. Vectors include, for example, plasmids, viral vectors (such as adenoviruses, adeno-associated viruses (AAV), lentiviruses, herpesvirus and retroviruses), liposomes and other lipid-containing complexes, and other macromolecular complexes capable of mediating delivery of a polynucleotide to a host cell. Vectors can also comprise other components or functionalities that further modulate gene delivery and/or gene expression, or that otherwise provide beneficial properties to the targeted cells. Such other components include, for example, components that influence binding or targeting to cells (including components that mediate cell-type or tissue-specific binding); components that influence uptake of the vector nucleic acid by the cell; components that influence localization of the polynucleotide within the cell after uptake (such as agents mediating nuclear localization); and components that influence expression of the polynucleotide. Such components also might include markers, such as detectable and/or selectable markers that can be used to detect or select for cells that have taken up and are expressing the nucleic acid delivered by the vector. Such components can be provided as a natural feature of the vector (such as the use of certain viral vectors which have components or functionalities mediating binding and uptake), or vectors can be modified to provide such functionalities. A large variety of such vectors are known in the art and are generally available. When a vector is maintained in a host cell, the vector can either be stably replicated by the cells during mitosis as an autonomous structure, incorporated within the genome of the host cell, or maintained in the host cell's nucleus or cytoplasm.

[0021] A "recombinant viral vector" refers to a viral vector comprising one or more heterologous genes or sequences. Since many viral vectors exhibit size constraints associated with packaging, the heterologous genes or sequences are typically introduced by replacing one or more portions of the viral genome. Such viruses may become replication-defective (biologically contained), requiring the deleted function (s) to be provided in trans during viral replication and encapsidation (by using, e.g., a helper virus or a packaging cell line carrying genes necessary for replication and/or encapsidation). Modified viral vectors in which a polynucleotide to be delivered is carried on the outside of the viral particle have also been described.

[0022] "Gene delivery," "gene transfer," and the like as used herein, are terms referring to the introduction of an exogenous polynucleotide (sometimes referred to as a "transgene") into a host cell, irrespective of the method used for the introduction. Such methods include a variety of well-known techniques such as vector-mediated gene transfer (by, e.g., viral infection/transfection, or various other protein-based or lipid-based gene delivery complexes) as well as techniques facilitating the delivery of "naked" polynucleotides (such as electroporation, "gene gun" delivery and various other techniques used for the introduction of polynucleotides). The introduced polynucleotide may be stably or transiently maintained in the host cell. Stable maintenance typically requires that the introduced polynucleotide either contains an origin of replication compatible with the host cell or integrates into a replicon of the host cell such as an extrachromosomal replicon (e.g., a plasmid) or a nuclear or mitochondrial chromosome. A number of vectors are known to be capable of mediating transfer of genes to mammalian cells, as is known in the art.

[0023] By "transgene" is meant any piece of a nucleic acid molecule (for example, DNA) which is inserted by artifice into a cell either transiently or permanently, and becomes part of the organism if integrated into the genome or maintained extrachromosomally. Such a transgene may include at least a portion of an open reading frame of a gene which is partly or entirely heterologous (i.e., foreign) to the transgenic organism, or may represent at least a portion of an open reading frame of a gene homologous to an endogenous gene of the organism, which portion optionally encodes a polypeptide with substantially the same activity as the corresponding full-length polypeptide or at least one activity of the corresponding full-length polypeptide.

[0024] By "transgenic cell" is meant a cell containing a transgene. For example, a cell stably or transiently transformed with a vector containing an expression cassette is a transgenic cell that can be used to produce a population of cells having altered phenotypic characteristics. A "recombinant cell" is one which has been genetically modified, e.g., by insertion, deletion or replacement of sequences in a nonrecombinant cell by genetic engineering.

[0025] The term "wild-type" or "native" refers to a gene or gene product that has the characteristics of that gene or gene product when isolated from a naturally occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designated the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product that displays modifications in sequence and or functional properties (i.e., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally-occurring mutants can be isolated; these are identified by the fact that they have altered characteristics when compared to the wild-type gene or gene product.

[0026] The term "transduction" denotes the delivery of a polynucleotide to a recipient cell either in vivo or in vitro, via a viral vector such as a replication-defective viral vector.

[0027] The term "heterologous" as it relates to nucleic acid sequences such as gene sequences encoding a protein and control sequences, denotes sequences that are not normally joined together, and/or are not normally associated with a particular cell, e.g., are from different sources (for instance, sequences from a virus are heterologous to sequences in the genome of an uninfected cell). Thus, a "heterologous" region of a nucleic acid construct or a vector is a segment of nucleic acid within or attached to another nucleic acid molecule that is not found in association with

the other molecule in nature. For example, a heterologous region of a nucleic acid construct could include a coding sequence flanked by sequences not found in association with the coding sequence in nature, i.e., a heterologous promoter. Another example of a heterologous coding sequence is a construct where the coding sequence itself is not found in nature (e.g., synthetic sequences having codons different from the native gene). Similarly, a cell transformed with a construct which is not normally present in the cell would be considered heterologous.

[0028] By "DNA" is meant a polymeric form of deoxyribonucleotides (adenine, guanine, thymine, or cytosine) in double-stranded or single-stranded form found, inter alia, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having the sequence complementary to the mRNA). The term captures molecules that include the four bases adenine, guanine, thymine, or cytosine, as well as molecules that include base analogues which are known in the art.

[0029] As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (i.e., a sequence of nucleotides) related by the base-pairing rules. For example, the sequence "A-G-T," is complementary to the sequence "T-C-A." Complementarity may be "partial," in which only some of the nucleic acids' bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon binding between nucleic acids.

[0030] DNA molecules are said to have "5' ends" and "3' ends" because mononucleotides are reacted to make oligonucleotides or polynucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage. Therefore, an end of an oligonucleotide or polynucleotide is referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5' phosphate of a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide or polynucleotide, also may be said to have 5' and 3' ends. In either a linear or circular DNA molecule, discrete elements are referred to as being "upstream" or 5' of the "downstream" or 3' elements. This terminology reflects the fact that transcription proceeds in a 5' to 3' fashion along the DNA strand. The promoter and enhancer elements that direct transcription of a linked gene are generally located 5' or upstream of the coding region. However, enhancer elements can exert their effect even when located 3' of the promoter element and the coding region. Transcription termination and polyadenylation signals are located 3' or downstream of the coding region.

[0031] A "gene," "polynucleotide," "coding region," "sequence," "segment," "fragment" or "transgene" which, in one embodiment, "encodes" a particular protein, is a nucleic acid molecule which is transcribed and optionally also translated into a gene product, e.g., a polypeptide, in vitro or in vivo when placed under the control of appropriate regulatory sequences. The coding region may be present in either a cDNA, genomic DNA, or RNA form. When present in a DNA form, the nucleic acid molecule may be single-stranded (i.e., the sense strand) or double-stranded. The boundaries of a coding region are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxy) terminus. A gene can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic DNA sequences. A transcription termination sequence will usually be located 3' to the gene sequence.

[0032] The term "control elements" refers collectively to promoter regions, polyadenylation signals, transcription termination sequences, upstream regulatory domains, origins of replication, internal ribosome entry sites ("IRES"), enhancers, splice junctions, and the like, which collectively provide for the replication, transcription, post-transcriptional processing and translation of a coding sequence in a recipient cell. Not all of these control elements need always be present so long as the selected coding sequence is capable of being replicated, transcribed and translated in an appropriate host cell.

[0033] The term "promoter" is used herein in its ordinary sense to refer to a nucleotide region comprising a DNA regulatory sequence, wherein the regulatory sequence is derived from a gene which is capable of binding RNA polymerase and initiating transcription of a downstream (3' direction) coding sequence.

[0034] By "enhancer" is meant a nucleic acid sequence that, when positioned proximate to a promoter, confers increased transcription activity relative to the transcription activity resulting from the promoter in the absence of the enhancer domain.

[0035] By "operably linked" with reference to nucleic acid molecules is meant that two or more nucleic acid molecules (e.g., a nucleic acid molecule to be transcribed, a promoter, and an enhancer element) are connected in such a way as to permit transcription of the nucleic acid molecule. "Operably linked" with reference to peptide and/or polypeptide molecules is meant that two or more peptide and/or polypeptide molecules are connected in such a way as to yield a single polypeptide chain, i.e., a fusion polypeptide, having at least one property of each peptide and/or polypeptide component of the fusion. The fusion polypeptide may be chimeric, i.e., composed of heterologous molecules.

[0036] "Homology" refers to the percent of identity between two polynucleotides or two polypeptides. The correspondence between one sequence and to another can be determined by techniques known in the art. For example, homology can be determined by a direct comparison of the sequence information between two polypeptide molecules by aligning the sequence information and using readily available computer programs. Alternatively, homology can be determined by hybridization of polynucleotides under conditions which form stable duplexes between homologous regions, followed by digestion with single strand-specific nuclease(s), and size determination of the digested fragments. Two DNA, or two polypeptide, sequences are "substantially homologous" to each other when at least about 80%, at least about 90%, or at least about 95% of the

nucleotides, or amino acids, respectively match over a defined length of the molecules, as determined using the methods above.

[0037] By "mammal" is meant any member of the class Mammalia including, without limitation, humans and non-human primates such as chimpanzees and other apes and monkey species; farm animals such as cattle, sheep, pigs, goats and horses; domestic mammals such as dogs and cats; laboratory animals including rodents such as mice, rats, rabbits and guinea pigs, and the like.

[0038] By "derived from" is meant that a nucleic acid molecule was either made or designed from a parent nucleic acid molecule, the derivative retaining substantially the same functional features of the parent nucleic acid molecule, e.g., encoding a gene product with substantially the same activity as the gene product encoded by the parent nucleic acid molecule from which it was made or designed.

[0039] By "expression construct" or "expression cassette" is meant a nucleic acid molecule that is capable of directing transcription. An expression construct includes, at the least, a promoter. Additional elements, such as an enhancer, and/or a transcription termination signal, may also be included.

[0040] The term "exogenous," when used in relation to a protein, gene, nucleic acid, or polynucleotide in a cell or organism refers to a protein, gene, nucleic acid, or polynucleotide which has been introduced into the cell or organism by artificial or natural means. An exogenous nucleic acid may be from a different organism or cell, or it may be one or more additional copies of a nucleic acid which occurs naturally within the organism or cell. By way of a non-limiting example, an exogenous nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature.

[0041] The term "isolated" when used in relation to a nucleic acid, peptide, polypeptide, virus or cell refers to a nucleic acid sequence, peptide, polypeptide, virus or cell that is identified and separated from at least one contaminant nucleic acid, polypeptide or other biological component with which it is ordinarily associated in its natural source, e.g., so that it is not associated with in vivo substances, or is substantially purified from in vitro substances. Isolated nucleic acid, peptide, polypeptide or virus is present in a form or setting that is different from that in which it is found in nature. For example, a given DNA sequence (e.g., a gene) is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. The isolated nucleic acid molecule may be present in single-stranded or double-stranded form. When an isolated nucleic acid molecule is to be utilized to express a protein, the molecule will contain at a minimum the sense or coding strand (i.e., the molecule may single-stranded), but may contain both the sense and anti-sense strands (i.e., the molecule may be double-stranded).

[0042] As used herein, the term "recombinant nucleic acid" or "recombinant DNA sequence, molecule or segment" refers to a nucleic acid, e.g., to DNA, that has been derived or isolated from a source, that may be subsequently chemically altered in vitro, and includes, but is not limited to, a sequence that is naturally occurring, is not naturally occurring, or corresponds to naturally occurring sequences that are not positioned as they would be positioned in the native genome. An example of DNA "derived" from a source, would be a DNA sequence that is identified as a useful fragment, and which is then chemically synthesized in essentially pure form. An example of such DNA "isolated" from a source would be a useful DNA sequence that is excised or removed from said source by chemical means, e.g., by the use of restriction endonucleases, so that it can be further manipulated, e.g., amplified, for use, by the methodology of genetic engineering.

[0043] The term "recombinant protein" or "recombinant polypeptide" as used herein refers to a protein molecule that is expressed from a recombinant DNA molecule.

[0044] The term "peptide", "polypeptide" and protein" are used interchangeably herein unless otherwise distinguished.

[0045] The term "sequence homology" means the proportion of base matches between two nucleic acid sequences or the proportion amino acid matches between two amino acid sequences. When sequence homology is expressed as a percentage, e.g., 50%, the percentage denotes the proportion of matches over the length of a selected sequence that is compared to some other sequence. Gaps (in either of the two sequences) are permitted to maximize matching; gap lengths of 15 bases or less are usually used, 6 bases or less such as 2 bases or less. When using oligonucleotides as probes or treatments, the sequence homology between the target nucleic acid and the oligonucleotide sequence is generally not less than 17 target base matches out of 20 possible oligonucleotide base pair matches (85%); not less than 9 matches out of 10 possible base pair matches (90%), or not less than 19 matches out of 20 possible base pair matches (95%).

[0046] Generally, the nucleic acid sequence homology between the polynucleotides, oligonucleotides, and fragments, and a nucleic acid sequence of interest is at least 65%, and more typically with y increasing homologies of at least about 70%, about 90%, about 95%, about 98%, and 100%.

[0047] Two amino acid sequences are homologous if there is a partial or complete identity between their sequences. For example, 85% homology means that 85% of the amino acids are identical when the two sequences are aligned for maximum matching. Gaps (in either of the two sequences being matched) are allowed in maximizing matching; gap lengths of 5 or less such as 2 or less. Alternatively, two protein sequences (or polypeptide sequences derived from them of at least 30 amino acids in length) are homologous, as this term is used herein, if they have an alignment score of at more than 5 (in standard deviation units) using the program ALIGN with the mutation data matrix and a gap penalty of 6 or greater. The two sequences or parts thereof may be considered homologous if their amino acids are greater than or equal to 50% identical when optimally aligned using the ALIGN program.

[0048] The term "corresponds to" is used herein to mean that a polynucleotide sequence is homologous (e.g., is identical, not strictly evolutionarily related) to all or a portion of a reference polynucleotide sequence that encodes a polypeptide or its complement, or that a polypeptide sequence is identical in sequence or function to a reference polypeptide sequence. For illustration, the nucleotide sequence "TATAC" corresponds to a reference sequence "TATAC" and is complementary to a reference sequence "GTATA".

[0049] The following terms are used to describe the sequence relationships between two or more polynucleotides: "reference sequence", "comparison window", "sequence identity", "percentage of sequence identity", and "substantial identity". A "reference sequence" is a defined sequence used as a basis for a sequence comparison; a reference sequence may be a subset of a larger sequence, for example, as a segment of a full-length cDNA or gene sequence given in a sequence listing, or may comprise a complete cDNA or gene sequence. Generally, a reference sequence is at least 20 nucleotides in length, frequently at least 25 nucleotides in length, and often at least 50 nucleotides in length. Since two polynucleotides may each (1) comprise a sequence (i.e., a portion of the complete polynucleotide sequence) that is similar between the two polynucleotides, and (2) may further comprise a sequence that is divergent between the two polynucleotides, sequence comparisons between two (or more) polynucleotides are typically performed by comparing sequences of the two polynucleotides over a "comparison window" to identify and compare local regions of sequence similarity.

[0050] A "comparison window", as used herein, refers to a conceptual segment of at least 20 contiguous nucleotides and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Optimal alignment of sequences for aligning a comparison window may be conducted by using local homology algorithms or by a search for similarity method, by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA Genetics Software Package or by inspection, and the best alignment (i.e., resulting in the highest percentage of homology over the comparison window) generated by the various methods is selected.

[0051] The term "sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" means that two polynucleotide sequences are identical (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The terms "substantial identity" as used herein denote a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 85 percent sequence identity, at least 90 to 95 percent sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 nucleotide positions, frequently over a window of at least 20-50 nucleotides, wherein the percentage of sequence identity is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison.

[0052] As applied to polypeptides, the term "substantial identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least about 80% sequence identity, at least about 90% sequence identity, at least about 95% percent sequence identity, and or at least about 99% sequence identity.

[0053] As used herein, "substantially pure" means an object species is the predominant species present (i.e., on a molar basis it is more abundant than any other individual species in the composition), and a substantially purified fraction is a composition wherein the object species comprises at least about 50 percent (on a molar basis) of all macromolecular species present. Generally, a substantially pure composition will comprise more than about 80 percent of all macromolecular species present in the composition, more than about 85%, about 90%, about 95%, and about 99%. The object species may be purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species.

[0054] "Transfected," "transformed" or "transgenic" is used herein to include any host cell or cell line, which has been altered or augmented by the presence of at least one recombinant DNA sequence. The host cells of the present invention are typically produced by transfection with a DNA sequence in a plasmid expression vector, as an isolated linear DNA sequence, or infection with a recombinant viral vector.

Exemplary Cells and Modifications Thereof

[0055] Most influenza vaccines are produced in embryonated chicken eggs, but increasingly influenza vaccines are produced in other systems. MDCK (Madin-Darby Canine Kidney) cells are one of two mammalian cell lines that have been approved for influenza vaccine production. Virus production in cells may be enhanced by altering the host cell or the virus. However, for vaccine production, virus modification during passage is not advantageous.

[0056] As disclosed herein, the genome of cells, e.g., avian cells or mammalian cells including but not limited to canine, feline, equine, bovine, caprine, swine, human or non-human primate, cells may be modified to enhance influenza virus isolation, propagation, or both. For example, certain HA subtypes (HA subtypes H1-H18) may not bind well to certain species or types of cells due to the number of or composition of cell surface receptors for HA. Those cells may be modified by increasing the number of cell surface receptors or modifying the type of molecules found on cell surface receptors, or both. For example, in mammals there are about 20 sialyltransferases that transfer sialic acid residues to oligosaccharide side chains of glycoconjugates. The genes encoding one or more of those enzymes may be modified to decrease, e.g., decrease by 1%, 5%, 10%, 50%, 70%, 80%, 90% or more, or eliminate expression of the encoded enzyme, or the open reading frame for one or more of those enzymes may be expressed in the cell from an exogenously introduced expression cassette, e.g., a plasmid having that expression cassette. For example, α2,6-sialyltransferases transfer sialic acid with an α2,6-linkage to terminal Gal (ST6GalI and II) or GalNAc (ST6GalNAcI-VI); α2,8-sialyltransferases transfer sialic acid with an α2,8-linkage (STSiaI-IV); and α2,3-sialyltransferases transfer sialic acid with an α2,3-linkage to terminal Gal residues

(ST3Gal). ST3GalI-II and IV transfer to the Gal residue located on terminal Galβ1-3GlcNAc, ST3GalIV and VI transfer to the Gal residue located on terminal Galβ1-4GlcNAc, ST3GalV transfers to the Gal residue located on terminal Galβ1-4Glc-Cer, and ST3GalIII transfers to the Gal residue located on terminal Galβ1-3GlcNAc or Galβ1-3GlcNAc. Thus, each of the genes for these sialyltransferases may be employed to prepare a cell disclosed herein. In one embodiment, one or more α2,3-sialyltransferase genes in the genome of a host cell are modified to decrease, e.g., eliminate, expression of the encoded enzyme, and one or more α2,6-sialyltransferase genes are expressed from a recombinant expression vector introduced to the host cell. To decrease expression of a sialyltransferse, one or more vectors, or a combination of vectors and isolated protein, may be introduced to a cell. The vectors and/or protein may be part of a recombinase system that can be targeted to a specific gene in the cell, systems including CRISPR/Cas, TALEN and zinc finger nucleases.

[0057] To prepare expression cassettes (to express RNA such as gRNA or a protein including a recombinase or a sialyltransferase) for transformation herein, the recombinant DNA sequence or segment may be circular or linear, double-stranded or single-stranded. A DNA sequence which encodes an RNA sequence that is substantially complementary to a mRNA sequence encoding a gene product of interest is typically a "sense" DNA sequence cloned into a cassette in the opposite orientation (i.e., 3' to 5' rather than 5' to 3'). Generally, the DNA sequence or segment is in the form of chimeric DNA, such as plasmid DNA, that can also contain coding regions flanked by control sequences which promote the expression of the DNA in a cell. As used herein, "chimeric" means that a vector comprises DNA from at least two different species, or comprises DNA from the same species, which is linked or associated in a manner which does not occur in the "native" or wild-type of the species.

[0058] Aside from DNA sequences that serve as transcription units, or portions thereof, a portion of the DNA may be untranscribed, serving a regulatory or a structural function. For example, the DNA may itself comprise a promoter that is active in eukaryotic cells, e.g., mammalian cells, or in certain cell types, or may utilize a promoter already present in the genome that is the transformation target of the lymphotropic virus. Such promoters include the CMV promoter, as well as the SV40 late promoter and retroviral LTRs (long terminal repeat elements), e.g., the MMTV, RSV, MLV or HIV LTR, although many other promoter elements well known to the art may be employed.

[0059] Other elements functional in the host cells, such as introns, enhancers, polyadenylation sequences and the like, may also be a part of the recombinant DNA. Such elements may or may not be necessary for the function of the DNA, but may provide improved expression of the DNA by affecting transcription, stability of the mRNA, or the like. Such elements may be included in the DNA as desired to obtain the optimal performance of the transforming DNA in the cell.

[0060] The recombinant DNA to be introduced into the cells may contain either a selectable marker gene or a reporter gene or both to facilitate identification and selection of transformed cells from the population of cells sought to be transformed. Alternatively, the selectable marker may be carried on a separate piece of DNA and used in a co-transformation procedure. Both selectable markers and reporter genes may be flanked with appropriate regulatory sequences to enable expression in the host cells. Useful selectable markers are well known in the art and include, for example, antibiotic and herbicide-resistance genes, such as neo, hpt, dhfr, bar, aroA, puro, h g, dapA and the like. See also, the genes listed on Table 1 of Lundquist et al. (U.S. Pat. No. 5,848,956).

[0061] Reporter genes are used for identifying potentially transformed cells and for evaluating the functionality of regulatory sequences. Reporter genes which encode for easily assayable proteins are well known in the art. In general, a reporter gene is a gene which is not present in or expressed by the recipient organism or tissue and which encodes a protein whose expression is manifested by some easily detectable property, e.g., enzymatic activity. Exemplary reporter genes include the chloramphenicol acetyl transferase gene (cat) from Tn9 of *E. coli*, the beta-glucuronidase gene (gus) of the uidA locus of *E. coli*, the green, red, or blue fluorescent protein gene, and the luciferase gene. Expression of the reporter gene is assayed at a suitable time after the DNA has been introduced into the recipient cells.

[0062] The general methods for constructing recombinant DNA which can transform target cells are well known to those skilled in the art, and the same compositions and methods of construction may be utilized to produce the DNA useful herein. For example, Sambrook et al., *Molecular Cloning: A Laboratory Manual* (2002) provides suitable methods of construction.

[0063] The recombinant DNA can be readily introduced into the host cells, e.g., mammalian, yeast or insect cells, by transfection with an expression vector comprising the recombinant DNA by any procedure useful for the introduction into a particular cell, e.g., physical or biological methods, to yield a transformed (transgenic) cell having the recombinant DNA so that the DNA sequence of interest is expressed by the host cell. In one embodiment, at least one of the recombinant DNA which is introduced to a cell is maintained extrachromosomally. In one embodiment, at least one recombinant DNA is stably integrated into the host cell genome.

[0064] Physical methods to introduce a recombinant DNA into a host cell include calcium-mediated methods, lipofection, particle bombardment, microinjection, electroporation, and the like. Biological methods to introduce the DNA of interest into a host cell include the use of DNA and RNA vectors. Viral vectors, e.g., retroviral or lentiviral vectors, have become a widely used method for inserting genes into eukaryotic, such as mammalian, e.g., human, cells. Other viral vectors useful to introduce genes into cells can be derived from poxviruses, e.g., vaccinia viruses, herpes viruses, adenoviruses, adeno-associated viruses, baculoviruses, and the like.

[0065] To confirm the presence of the recombinant DNA sequence in the host cell, a variety of assays may be performed. Such assays include, for example, molecular biological assays well known to those of skill in the art, such as Southern and Northern blotting, RT-PCR and PCR; biochemical assays, such as detecting the presence or absence of a particular gene product, e.g., by immunological means (ELISAs and Western blots) or by other molecular assays.

[0066] To detect and quantitate RNA produced from introduced recombinant DNA segments, RT-PCR may be employed. In this application of PCR, it is first necessary to

reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR techniques amplify the DNA. In most instances PCR techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA product may be obtained by Northern blotting. This technique demonstrates the presence of an RNA species and gives information about the integrity of that RNA. The presence or absence of an RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and only demonstrate the presence or absence of an RNA species.

[0067] While Southern blotting and PCR may be used to detect the recombinant DNA segment in question, they do not provide information as to whether the recombinant DNA segment is being expressed. Expression may be evaluated by specifically identifying the peptide products of the introduced DNA sequences or evaluating the phenotypic changes brought about by the expression of the introduced DNA segment in the host cell.

[0068] For vectors that are used to knock down or knock out expression of one or more sialyltransferases, the vectors harbor sequences that result in one or more mutations in the genome of the cell. The mutation is effective to inhibit or prevent production of at least one functional sialyltransferase. In one embodiment, the mutation is a deletion from 1, 10, 20, 50, 100, 500 and up to thousands of nucleotides, e.g., 1%, 10%, 50%, 90% or more of sequences corresponding to a sialyltransferase gene are deleted, e.g., a deletion in the coding region for a sialyltransferase, e.g., a 2,3-sialyltransferase (ST3). In one embodiment, the deleted sequences correspond to sequences with a substantial identity, e.g., at least 80% or more, e.g., 85%, 90% or 95% and up to 100% or any integer in between, nucleic acid sequence identity, to SEQ ID Nos. 5, 7, 9, 11, 13, 15 or 17, or any combination thereof. In one embodiment, the mutation is an insertion from 1, 2, 3, 5, 10, 20, 50, 100, 500 and up to thousands of nucleotides or more into sequences corresponding to a sialyltransferase gene such as an insertion into the coding region for a sialyltransferase, e.g., a 2,3-sialyltransferase (ST3). In one embodiment, the insertion is in sequences corresponding to sequences with a substantial identity, e.g., at least 80% or more, e.g., 85%, 90% or 95% and up to 100% or any integer in between, nucleic acid sequence identity, to SEQ ID Nos. 5, 7, 9, 11, 13, 15 or 17, or any combination thereof. In one embodiment, the mutation include one or more nucleotide substitutions, e.g., 1, 2, 3, 4, 5, 6, 10 or up to hundreds of nucleotide substitutions in sequences corresponding to the coding region for a sialyltransferase, e.g., a 2,3-sialyltransferase (ST3) such as substitutions in SEQ ID Nos. 5, 7, 9, 11, 13, 15 or 17, or any combination thereof. In one embodiment, a combination of insertions, nucleotide substitutions, and/or deletions in sequences with a substantial identity, e.g., at least 80% or more, e.g., 85%, 90% or 95% and up to 100% or any integer in between, nucleic acid sequence identity, SEQ ID Nos. 5, 7, 9, 11, 13, 15 or 17, or any combination thereof, are in a host cell. In one embodiment, the mutation(s) result in the host cell having reduced expression of one or more ST3 genes, e.g., encoding a protein having at least 80% or more, e.g., 85%, 90% or 95% and up to 100% or any integer in between, amino acid sequence identity to any of SEQ ID Nos. 6, 8, 10, 12, 14, 16, or 18.

[0069] In one embodiment, the host cell expresses one or more ST6 genes, e.g., encoding a protein having at least 80% or more, e.g., 85%, 90% or 95% and up to 100% or any integer in between, amino acid sequence identity to any of SEQ ID Nos. 1-4.

The CRISPR/Cas System

[0070] The Type II CRISPR is a well characterized system that carries out targeted DNA double-strand break in four sequential steps. First, two non-coding RNA, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the repeat regions of the pre-crRNA and mediates the processing of pre-crRNA into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the target DNA via Watson-Crick base-pairing between the spacer on the crRNA and the protospacer on the target DNA next to the protospacer adjacent motif (PAM), an additional requirement for target recognition. Finally, Cas9 mediates cleavage of target DNA to create a double-stranded break within the protospacer. Activity of the CRISPR/Cas system comprises of three steps: (i) insertion of alien DNA sequences into the CRISPR array to prevent future attacks, in a process called 'adaptation,' (ii) expression of the relevant proteins, as well as expression and processing of the array, followed by (iii) RNA-mediated interference with the alien nucleic acid. Thus, in the bacterial cell, several of the so-called 'Cas' proteins are involved with the natural function of the CRISPR/Cas system. The primary products of the CRISPR loci appear to be short RNAs that contain the invader targeting sequences, and are termed guide RNAs

[0071] "Cas1" polypeptide refers to CRISPR associated (Cas) protein. Cas1 (COG 1518 in the Clusters of Orthologous Group of proteins classification system) is the best marker of the CRISPR-associated systems (CASS). Based on phylogenetic comparisons, seven distinct versions of the CRISPR-associated immune system have been identified (CASS1-7). Cas1 polypeptide used in the methods described herein can be any Cas polypeptide present in any prokaryote. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide of an archaeal microorganism. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide of a Euryarchaeota microorganism. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide of a Crenarchaeota microorganism. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide of a bacterium. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide of a gram negative or gram positive bacteria. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide of *Pseudomonas aeruginosa*. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide of *Aquifex aeolicus*. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide that is a member of one of CASs1-7. In certain embodiments, Cas1 polypeptide is a Cas1 polypeptide that is a member of CASS3. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide that is a member of CASS7. In certain embodiments, a Cas1 polypeptide is a Cas1 polypeptide that is a member of CASS3 or CASS7.

[0072] In some embodiments, a Cas polypeptide is encoded by a nucleotide sequence provided in GenBank at, e.g., GeneID number: 2781520, 1006874, 9001811, 947228, 3169280, 2650014, 1175302, 3993120, 4380485, 906625, 3165126, 905808, 1454460, 1445886, 1485099, 4274010, 888506, 3169526, 997745, 897836, or 1193018 and/or an

amino acid sequence exhibiting homology (e.g., greater than 80%, 90 to 99% including 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99%) to the amino acids encoded by these polynucleotides and which polypeptides function as Cas polypeptides.

[0073] There are three types of CRISPR/Cas systems which all incorporate RNAs and Cas proteins. Types I and III both have Cas endonucleases that process the pre-crRNAs, that, when fully processed into crRNAs, assemble a multi-Cas protein complex that is capable of cleaving nucleic acids that are complementary to the crRNA.

[0074] In type II CRISPR/Cas systems, crRNAs are produced using a different mechanism where a trans-activating RNA (tracrRNA) complementary to repeat sequences in the pre-crRNA, triggers processing by a double strand-specific RNase III in the presence of the Cas9 protein. Cas9 is then able to cleave a target DNA that is complementary to the mature crRNA however cleavage by Cas9 is dependent both upon base-pairing between the crRNA and the target DNA, and on the presence of a short motif in the crRNA referred to as the PAM sequence (protospacer adjacent motif)). In addition, the tracrRNA must also be present as it base pairs with the crRNA at its 3' end, and this association triggers Cas9 activity.

[0075] The Cas9 protein has at least two nuclease domains: one nuclease domain is similar to a HNH endonuclease, while the other resembles a Ruv endonuclease domain. The HNH-type domain appears to be responsible for cleaving the DNA strand that is complementary to the crRNA while the Ruv domain cleaves the non-complementary strand.

[0076] The requirement of the crRNA-tracrRNA complex can be avoided by use of an engineered "single-guide RNA" (sgRNA) that comprises the hairpin normally formed by the annealing of the crRNA and the tracrRNA (see Jinek, et al. (2012) Science 337:816 and Cong et al. (2013) Sciencexpress/10.1126/science.1231143). In *S. pyrogenes*, the engineered tracrRNA:crRNA fusion, or the sgRNA, guides Cas9 to cleave the target DNA when a double strand RNA:DNA heterodimer forms between the Cas associated RNAs and the target DNA. This system comprising the Cas9 protein and an engineered sgRNA

[0077] "Cas polypeptide" encompasses a full-length Cas polypeptide, an enzymatically active fragment of a Cas polypeptide, and enzymatically active derivatives of a Cas polypeptide or fragment thereof. Suitable derivatives of a Cas polypeptide or a fragment thereof include but are not limited to mutants, fusions, covalent modifications of Cas protein or a fragment thereof.

RNA Components of CRISPR/Cas

[0078] The Cas9 related CRISPR/Cas system comprises two RNA non-coding components: tracrRNA and a pre-crRNA array containing nuclease guide sequences (spacers) interspaced by identical direct repeats (DRs). To use a CRISPR/Cas system to accomplish genome engineering, both functions of these RNAs must be present (see Cong, et al. (2013) Sciencexpress 1/10.1126/science 1231143). In some embodiments, the tracrRNA and pre-crRNAs are supplied via separate expression constructs or as separate RNAs. In other embodiments, a chimeric RNA is constructed where an engineered mature crRNA (conferring target specificity) is fused to a tracrRNA (supplying interaction with the Cas9) to create a chimeric cr-RNA-tracrRNA hybrid (also termed a single guide RNA). (see Jinek, ibid and Cong, ibid).

[0079] Chimeric or sgRNAs can be engineered to comprise a sequence complementary to any desired target. The RNAs comprise 22 bases of complementarity to a target and of the form G[n 19], followed by a protospacer-adjacent motif (PAM) of the form NGG. Thus, in one method, sgRNAs can be designed by utilization of a known ZFN target in a gene of interest by (i) aligning the recognition sequence of the ZFN heterodimer with the reference sequence of the relevant genome (human, mouse, or of a particular plant species); (ii) identifying the spacer region between the ZFN half-sites; (iii) identifying the location of the motif G[N20]GG that is closest to the spacer region (when more than one such motif overlaps the spacer, the motif that is centered relative to the spacer is chosen); (iv) using that motif as the core of the sgRNA. This method advantageously relies on proven nuclease targets. Alternatively, sgRNAs can be designed to target any region of interest simply by identifying a suitable target sequence that conforms to the G[n20]GG formula.

Donors

[0080] As noted above, insertion of an exogenous sequence (also called a "donor sequence" or "donor" or "transgene" or "gene of interest"), for example for correction of a mutant gene or for increased expression of a wild-type gene. It will be readily apparent that the donor sequence is typically not identical to the genomic sequence where it is placed. A donor sequence can contain a non-homologous sequence flanked by two regions of homology to allow for efficient HDR at the location of interest. Alternatively, a donor may have no regions of homology to the targeted location in the DNA and may be integrated by NHEJ-dependent end joining following cleavage at the target site. Additionally, donor sequences can comprise a vector molecule containing sequences that are not homologous to the region of interest in cellular chromatin. A donor molecule can contain several, discontinuous regions of homology to cellular chromatin. For example, for targeted insertion of sequences not normally present in a region of interest, said sequences can be present in a donor nucleic acid molecule and flanked by regions of homology to sequence in the region of interest.

[0081] The donor polynucleotide can be DNA or RNA, single-stranded and/or double-stranded and can be introduced into a cell in linear or circular form. If introduced in linear form, the ends of the donor sequence can be protected (e.g., from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends. See, for example, Chang, et al. (1987) Proc. Natl. Acad. Sci. USA 84:4959-4963; Nehls, et al. (1996) Science 272:886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues.

[0082] A polynucleotide can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters and genes

encoding antibiotic resistance. Moreover, donor polynucleotides can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses (e.g., adenovirus, AAV, herpesvirus, retrovirus, lentivirus and integrase defective lentivirus (IDLV)).

[0083] The donor is generally inserted so that its expression is driven by the endogenous promoter at the integration site, namely the promoter that drives expression of the endogenous gene into which the donor is inserted (e.g., highly expressed, albumin, AAVS1, HPRT, etc.). However, it will be apparent that the donor may comprise a promoter and/or enhancer, for example a constitutive promoter or an inducible or tissue specific promoter.

[0084] The donor molecule may be inserted into an endogenous gene such that all, some or none of the endogenous gene is expressed. For example, a transgene as described herein may be inserted into an albumin or other locus such that some (N-terminal and/or C-terminal to the transgene encoding the lysosomal enzyme) or none of the endogenous albumin sequences are expressed, for example as a fusion with the transgene encoding the lysosomal sequences. In other embodiments, the transgene (e.g., with or without additional coding sequences such as for albumin) is integrated into any endogenous locus, for example a safe-harbor locus. See, e.g., U.S. Patent Publication Nos. 2008/0299580; 2008/0159996; and 2010/0218264.

[0085] When endogenous sequences (endogenous or part of the transgene) are expressed with the transgene, the endogenous sequences (e.g., albumin, etc.) may be full-length sequences (wild-type or mutant) or partial sequences. In one embodiment, the endogenous sequences are functional. Non-limiting examples of the function of these full length or partial sequences (e.g., albumin) include increasing the serum half-life of the polypeptide expressed by the transgene (e.g., therapeutic gene) and/or acting as a carrier.

[0086] Furthermore, although not required for expression, exogenous sequences may also include transcriptional or translational regulatory sequences, for example, promoters, enhancers, insulators, internal ribosome entry sites, sequences encoding 2A peptides and/or polyadenylation signals.

[0087] Other nucleases, such as zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), or others that are specific for targeted genes and can be utilized such that the transgene construct is inserted by either homology directed repair (HDR) or by end capture during non-homologous end joining (NHEJ) driven processes.

Exemplary Embodiments

[0088] In one embodiment, an isolated recombinant mammalian or avian cell is provided comprising a reduced amount of cell surface β-galactoside α2,3 sialyl residues and an increased amount of human β-galactoside α2,6 sialyl residues relative to a corresponding non-recombinant mammalian or avian cell. In one embodiment, the isolated recombinant cell is a non-human cell. In one embodiment, the isolated recombinant cell is a canine or primate cell. In one embodiment, the isolated recombinant cell comprises an expression cassette encoding human β-galactoside α2,6 sialyltransferase I (ST6Gal-I) or ST6Gal-II. In one embodiment, the ST6Gal-I or ST6Gal-II comprises a protein having at least 80% amino acid sequence identity to any one of SEQ ID Nos. 1-4 or 101. In one embodiment, the one or more β-galactoside α2,3 sialyltransferase genes are mutated in the recombinant cell so as to reduce the amount of cell surface β-galactoside α2,3 sialyl residues. In one embodiment, two or more of ST3Gal-I, ST3Gal-II, ST3Gal-II, ST3Gal-IV, ST3Gal-V, ST3Gal-VI, or ST3Gal-II-like genes are mutated in the recombinant cell. In one embodiment, the ST3 genes have at least 80% nucleic acid sequence identity to any one of SEQ ID Nos. 5, 7, 9, 11, 13, 15, or 17. In one embodiment, the reduction in cell surface β-galactoside α2,3 sialyl residues is the result of reduced expression of one or more ST3 sialyltransferases. In one embodiment, the one or more ST3 sialyltransferases have at least 80% amino acid sequence identity to any one of SEQ ID Nos. 6, 8, 10, 12, 14, 16, or 18. In one embodiment, influenza H3 viruses replicate more efficiently in the recombinant cell relative to the non-recombinant cell.

[0089] In one embodiment, an isolated recombinant mammalian or avian cell is provided comprising a reduced amount of cell surface β-galactoside α2,3 sialyl residues relative to a corresponding non-recombinant mammalian or avian cell. In one embodiment, one or more of ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV, ST3Gal-1; ST3Gal-VI, or ST3Gal-II-like genes are mutated in the recombinant cell. In one embodiment, a combination of ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV, ST3Gal-V, ST3Gal-VI, or ST3Gal-II-like genes are mutated in the recombinant cell. In one embodiment, ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV ST3Gal-V, ST3Gal-VI, and ST3Gal-II-like genes are mutated.

[0090] In one embodiment, a method of modifying the amount of cell surface β-galactoside α2,3 sialyl residues and human β-galactoside α2,6 sialyl residues on a mammalian or an avian cell is provided. In one embodiment, the method includes mutating one or more β-galactoside α2,3 sialyltransferase (ST3Gal) genes, and overexpressing a human β-galactoside α2,6 sialyltransferase (ST6Gal) gene, in a parental mammalian or avian cell so as to result in a modified mammalian or avian cell having a reduced amount of cell surface β-galactoside α2,3 sialyl residues and an increased amount of human β-galactoside α2,6 sialyl residues on the surface of the modified cell relative to the corresponding parental cell. In one embodiment, the one or more ST3Gal genes are mutated using a genome editing system. In one embodiment, the genome editing system comprises a CRISPR/Cas9, Zinc Finger Nuclease (ZFN) or transcription activator-like effector nuclease (TALEN). In one embodiment, the mutations include one or more nucleotide insertions or one or more nucleotide deletions, or both, in one or more ST3 genes. In one embodiment, the modified cell comprises an expression cassette comprising a ST6Gal open reading frame. In one embodiment, the modified cell is a kidney cell. In one embodiment, the modified cell is a canine cell. In one embodiment, the modified cell is a Madin-Darby canine kidney (MDCK) cell.

[0091] In one embodiment, a method of propagating an influenza virus is provided. The method includes infecting the recombinant cell with an influenza virus; and collecting progeny virus. In one embodiment, the influenza virus is a human influenza virus. In one embodiment, the influenza virus is an influenza A virus. In one embodiment, the influenza virus is an influenza B virus. In one embodiment, the influenza virus is a H3 virus. In one embodiment, the

virus is A/H1N1, A/H3N2, a B/Yamagata-lineage influenza B virus or a B/Victoria-lineage influenza B virus.

[0092] In one embodiment, a method of isolating an influenza virus is provided which includes providing a sample from an avian or a mammal suspected of being infected with an influenza virus; and contacting the recombinant cell with the sample. In one embodiment, the method includes determining whether the sample is infected with an influenza virus. In one embodiment, the method includes identifying the HA and/or NA subtype of the virus.

[0093] In one embodiment, a method of diagnosing an influenza virus infection is provided. The method includes contacting the recombinant cell with a sample from an avian or a mammal suspected of being infected with an influenza virus; and determining if the cell is infected with virus. In one embodiment, a plaque assay is employed to determine the presence of amount of virus. In one embodiment, a nucleic acid amplification assay is employed to determine the presence of amount of virus, e.g., in the supernatant of the infected cell.

Exemplary Sialyltransferase Sequences

[0094] Sialyltransferases in higher vertebrates are glycosyltransferases that mediate the transfer of sialic acid residues from activated sugar donors (CMP-β-Neu5Ac, CMP-β-Neu5Gc, and CMP-β-KDN) to terminal non-reducing positions of oligosaccharide chains of glycoproteins and glycolipids. The vertebrate sialyltransferase superfamily is divided into four families, ST6Gal, ST3Gal, ST6GalNAc, and ST8Sia, depending on the glycosidic linkage formed and the monosaccharide acceptor used. Members of the mammalian and avian ST6Gal family catalyze the transfer of sialic acid residues to the terminal galactose residues of the type 2 disaccharide (Gal(NAc)β1,4GlcNAc), resulting in the formation of an α2-6 glycosidic linkage. Unlike the other sialyltransferase families, this family comprises only two paralogs in the human genome named ST6GAL1 and ST6GAL2, respectively. The human ST6GAL1 gene is ubiquitously expressed in a broad variety of tissues, whereas the ST6GAL2 gene is expressed in a tissue-specific (adult brain) and stage-specific (embryonic) manner. Mammalian st6gal1 gene expression is regulated by multiple promoters governing the expression of several transcripts encoding an identical polypeptide enzyme.

[0095] In one embodiment, one or more ST3 genes in an avian or a mammlian cell, e.g., a canine or non-human primate cell, are modified so as to result in decreased expression of α-2,3-linked sialic acids on the cell surface. In one embodiment, one or more human ST6 genes are introduced to a canine or a non-human primate cell. In one embodiment, one or more ST3 genes are modified before one or more ST6 genes are introduced to the cell. In one embodiment, one or more ST6 genes are introduced before ST3 genes are modified in the cell. In one embodiment, concurrently or sequentially ST3 genes are modified and ST6 genes are introduced to the cell.

[0096] In one embodiment, the ST6Gal that is expressed comprises human ST6 (Accession No. KJ897554) comprising

(SEQ ID NO: 1)

MIHTNLKKKFSCCVLVFLLFAVICVWKEKKKGSYYDSFKLOTKE

FQVLKSLGKLAMGSDSQSYSSSSTQDPHRGRQTLGSLRGLAKAKPEASFQVIVNKDSSS

KNLIPRLOKIWKNYLSNINKYKVSYKGPGPGIKFSAEALRCHLRDIIVNVSMVEVTDFPF

NTSENVEGYLPKESIRTKAGPWGRCAVVSSAGSLKSSQLGREIDDHDAVLUNGAPTAN

FQQDVGTKTTIRLMNSQLVTTEKRFLKDSLYNEGILIVWDPSVYHSDIPKWYQNPDYN

FFNNYKTYRKLHPNQPFYILKPQMPWELWDILQEISPEEIQPNPPSSGMLGIIIMIVITL

CDQVDIYEFLPSKRKTDVCYYYQKFFDSACTMGAYHPLLYEKNLVKHLNQGTDEDIYL

LGKATLPURTIHC

or

(SEQ ID NO: 150)

MIHTNLKKKFSCCVLVFLLFAVICVWKEKKKGSYYDSFKLQTKEFQVLKSL

GKLAMGSDSQSVSSSSTQDPHRGRQTLGSLRGLAKAKPEASFQVWNKDSSS

KNLIPRLQKIWKNYLSMNKYKVSYKGPGPGIKFSAEALRCHLRDHVNVSM

VEVTDFPFNTSEWEGYLPKESIRTKAGPWGRCAVVSSAGSLKSSQLGREIDD

HDAVLRFNGAPTANFQQDVGTKTTIRLMNSQLVTTEKRFLKDSLYNEGILIV

WDPSVYHPDIPKWYQNPDYNFFNNYKTYRKLHPNQPFYILKPQMPWELWD

ILQEISPEEIQPNPPSSGMLGIIIMMTLCDQVDIYEFLPSKRRTDVCYYYQKFF

DSACTMGAYHPLITEKNLVKHLNQGTDEDIYLLGKATLPGFRTIHC,

which is encoded by

(SEQ ID NO: 151)

ATGATTCACACCAACCTGAAGAAAAAGTTCAGCTGCTGCGTCCTGGTCT

TTCTTCTGTTTGCAGTCATCTGTGTGTGGAAGGAGAAGAAGAAAGGGAG

-continued

TTACTATGATTCCTTTAAATTGCAAACCAAGGAATTCCAGGTGTTAAAGA

GTCTGGGGAAATTGGCCATGGGGTCTGATTCCCAGTCTGTATCCTCAAGC

AGCACCCAGGACCCCCACAGGGGCCGCCAGACCCTCGGCAGTCTCAGAG

GCCTAGCCAAGGCCAAACCAGAGGCCTCCTTCCAGGTGTGGAACAAGGA

CAGCTCTTCCAAAAACCTTATCCCTAGGCTGCAAAAGATCTGGAAGAAT

TACCTAAGCATGAACAAGTACAAAGTGTCCTACAAGGGGCCAGGACCA

GGCATCAAGTTCAGTGCAGAGGCCCTGCGCTGCCACCTCCGGGACCATG

TGAATGTATCCATGGTAGAGGTCACAGATTTTCCCTTCAATACCTCTGAA

TGGGAGGGTTATCTGCCCAAGGAGAAGCATTAGGACCAAGGCTGGGCCTT

GGGGCAGGTGTGCTGTTGTGTCGTCAGCGGGATCTCTGAAGTCCTCCCA

ACTAGGCAGAGAAATCGATGATCATGACGCAGTCCTGAGGTTTAATGGG

GCACCCACAGCCAACTTCCAACAAGATGTGGGCACAAAAACTACCATTC

GCCTGATGAACTCTCAGTTGGTTACCACAGAGAAGCGCTTCCTCAAAGA

CAGTTTGTACAATGAAGGAATCCTAATTGTATGGGACCCATCTGTATACC

ACCCAGATATCCCAAAGTGGTACCAGAATCCGGATTATAATTTCTTTAAC

AACTACAAGACTTATCGTAAGCTGCACCCCAATCAGCCCTTITACATCCT

CAAGCCCCAGATGCCTTGGGAGCTATGGGACATTCTTCAAGAAATCTCC

CCAGAAGAGATTCAGCCAAACCCCCCATCCTCTGGGATGCTTGGTATCA

TCATCATGATGACGCTGTGTGACCAGGTGGATATTTATGAGTTCCTCCCA

TCCAAGCGCAGGACTGACGTGTGCTACTACTACCAGAAGTTCTTCGATA

GTGCCTGCACGATGGGTGCCTACCACCCGCTGCTCTTTGAGAAGAATTTG

GTGAAGCATCTCAACCAGGGCACAGATGAGGACATCTACCTGCTTGGAA

AAGCCACACTGCCTGGCTTCCGGACCATTCACTGCTAA;

a human ST6 (Accession No. BAC24793) comprising

(SEQ ID NO: 2)

mkphlkqwrq rmlfgifawg llfllifiyf tdsnpaepvp sslsfletrr llpvqgkqra

imgaahepsp pggldargal prahpagsth agpgdlqkwa qsqdgfehke ffssqvgrks

qsafypeddd yffaagqpgw hshtqgtlgf pspgepgpre gafpaaqvqr rrvklahrrq

rrshvleegd dgdrlyssms raflyrlwkg nvsskmlnpr lqkamkdylt ankhgvrfrg

kreaglsraq llcqlrsrar vrtldgteap fsalgwrrlv pavplsqlhp rglrscavvm

sagailnssl geeidshdav lrfnsaptrg yekdvgnktt iriinsqilt npshhfidss

lykdvilvaw dpapysanln lwykkpdynl ftpyiqhrqr npnqpfyilh pkfiwqlwdi

iqentkekiq pnppssgfig ilimmsmcre vhvyeyipsv rqtelchyhe lyydaactlg

ayhpllyekl lvqrinmgtq gdlhrkgkvv lpgfqavhcp apspviphs;

a human ST6 (Accession No. SJL87798) comprising

(SEQ ID NO: 3)

aamgsdsqsv sssstqdphr grqtlgslrg lakakpeasf qvwnkdsssk nliprlqkiw

knylsmnkyk vsykgpgpgi kfsaealrch lrdhynysmv evtdfpfnts ewegylpkes

irtkagpwgr cavvssagsl kssqlgreid dhdavIrfng aptanfqqdv gtkttirlmn

sqlvttekrf lkdslynegi livwdpsyyh sdipkwyqnp dynffnnykt yrklhpnqpf

-continued

yilkpqmpwe lwdilqeisp eeiqpnppss gmlgiiimmt lcdqvdiyef lpskrktdvc

yyyqkffdsa ctmgayhpll yeknlvkhln qgtdediyll gkatlpgfit ihc; or

a human ST6 Gal-II

(SEQ ID NO : 4)

MKPHLKQWRQRMLFGIFAWGLLFLLIFINTTDSNPAEPVPSSLS

FLETRRLLPVQGKQRAIMGAAHEPSPPGGLDARQALPRAHPAGSFHAGPGDLQKWAQS

QDGFEHKEFFSSQVGRKSQSAFYPEDDDYFFAAGQPGWHSHTQGTLGFPSPGEPGPRE

GAFPAAQVQRRRVKKRHRRQRRSHNTEEGDDGDRLYSSIVISRAFLYRLWKGNVSSKMLN

PRLQKAMKDYLTANKHGVRFRGKREAGLSRAQLLCQLRSRARVRTLDGTEAPFSALGW

RRINPAVPLSQLHPRGLRSCAVVMSAGAILNSSLGEE1DSHDAVLRFNSAPTRGYEKD

VGNKTTIRTINSQILTNPSHHFIDSSLYKDVILVAWDPAPYSANLNLWYKKPDYNLFT

PYIQHRQRNPNQPFYILHPKFIWQLWDITQENTKEKIQPNPPSSGFIGILIMMSIVICRE

VIWYEYIPSVRQTELCHYHELYYDAACTLGAYHPLLYEKLLVQRLNIVIGTQGDLHRKGK

VVLPGFQAVHCPAPSPVIPHS

human ST6Gal1 encoded by

(SEQ ID NO: 100)

ATGATTCACACCAACCTGAAGAAAAAGTTCAGCTGCTGCGTCCTGGTCTTTCTTCTGTTTGCAGTCA

TCTGTGTGTGGAAGGAGAAGAAGAAAGGGAGTTACTATGATTCCTTTAAATTGCAAACCAAGGAA

TTCCAGGTGTTAAAGAGTCTGGGGAAATTGGCCATGGGGTCTGATTCCCAGTCTGTATCCTCAAGC

AGCACCCAGGACCCCCACAGGGGCCGCCAGACCCTCGGCAGTCTCAGAGGCCTAGCCAAGGCCAA

ACCAGAGGCCTCCTTCCAGGTGTGGAACAAGGACAGCTCTTCCAAAAACCTTATCCCTAGGCTGCA

AAAGATCTGGAAGAATTACCTAAGCATGAACAAGTACAAAGTGTCCTACAAGGGGCCAGGACCAG

GCATCAAGTTCAGTGCAGAGGCCCTGCGCTGCCACCTCCGGGACCATGTGAATGTATCCATGGTA

GAGGTCACAGATTTTCCCTTCAATACCTCTGAATGGGAGGGTTATCTGCCCAAGGAGAGCATTAGG

ACCAAGGCTGGGCCTTGGGGCAGGTGTGCTGTTGTGTCGTCAGCGGGATCTCTGAAGTCCTCCCA

ACTAGGCAGAGAAATCGATGATCATGACGCAGTCCTGAGGTTTAATGGGGCACCCACAGCCAACT

TCCAACAAGATGTGGGCACAAAAACTACCATTCGCCTGATGAACTCTCAGTTGGTTACCACAGAGA

AGCGCTTCCTCAAAGACAGTTTGTACAATGAAGGAATCCTAATTGTATGGGACCCATCTGTATACC

ACCCAGATATCCCAAAGTGGTACCAGAATCCGGATTATAATTTCTTTAACAACTACAAGACTTATCG

TAAGCTGCACCCCAATCAGCCCTTTTACATCCTCAAGCCCCAGATGCCTTGGGAGCTATGGGACAT

TCTTCAAGAAATCTCCCCAGAAGAGATTCAGCCAAACCCCCCATCCTCTGGGATGCTTGGTATCATC

ATCATGATGACGCTGTGTGACCAGGTGGATATTTATGAGTTCCTCCCATCCAAGCGCAGGACTGAC

GTGTGCTACTACTACCAGAAGTTCTTCGATAGTGCCTGCACGATGGGTGCCTACCACCCGCTGCTC

TTTGAGAAGAATTTGGTGAAGCATCTCAACCAGGGCACAGATGAGGACATCTACCTGCTTGGAAA

AGCCACACTGCCTGGCTTCCGGACCATTCACTGCTAA;

human ST6Ga1 I comprising

(SEQ ID NO: 101)

MIHTNLKKKFSCCVLVFLLFAVICVWKEKKKGSYYDSFKLQTKEFQVLKSLGKLAMGSDSQSVSSSSTQD

PHRGRQTLGSLRGLAKAKPEASFQVWNKDSSSKNLIPRLQKIWKNYLSMNKYKVSYKGPGPGIKFSAE

ALRCHLRDHVNVSMVEVTDFPFNTSEWEGYLPKESIRTKAGPWGRCAVVSSAGSLKSSQLGREIDDHD

AVLRFNGAPTANFQQDVGTKTTIRLMNSQLVTTEKRFLKDSLYNEGILIVWDPSVYHPDIPKWYQNPD

-continued

YNFFNNYKTYRKLHPNQPFYILKPQMPWELWDILQEISPEEIQPNPPSSGMLGIIIMMTLCDQVDIYEFL

PSKRRTDVCYYYQKFFDSACTMGAYHPLLFEKNLVKHLNQGTDEDIYLLGKATLPGFRTIHC;

or a protein having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to any one of SEQ ID Nos. 1-4, 101 or 150 or a nucleotide sequence having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleotide acid sequence identity to any one of SEQ ID Nos. 100 or 151.

[0097] In one embodiment, the ST3 gene that is mutated has at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleic acid sequence identity to a canine ST3GalI comprising

(Accession No. XM_022426722)

(SEQ ID NO: 5)

gtcaggtctc agaaaagtct agaataaggc ttacaggcac tttgttcagt tgtggaacac

atggaaactt catacaccgc ccccctcctt gcagaccgga gagctctctg ctctaatttg

ggcacaggcg cccaaccttg ggccccagag gaagcctgtt ctactcccag ggccaacgct

caattgcttg ataacttgac acccacactg ttccaggctc tggtctcaat ttcactctct

gcaaaatgag aggctcagag ttccgctgac atggcctcgg ggccgttaaa acctttcttc

atgaagacag cctgcaccct tcctgcttct gctccaggtc catcctcaga accttccaga

aagtcctggc acgtcagata acagggccaa cccggcagtg atgccccacc ccccacccct

tcccttactc atagagcctg ctccagactc tcagagccca cacccacttg gtgaagtcat

ttgccagtaa ttcttcgcac tggacattga gaggtttcag accccagaag tctcaggcgc

tgggtctgaa agtgggcaga gcccaggtga catttgtgga gactctcagt ggtgcgtata

gccgccggga ccattttcag actcaacctt tctgacctgg aaatgccaat agaagtaatc

atcatcgcca agggctgtag tgagcaaatg cttagagctc tgtggccaag gcagtttcat

ttgaggacca gagatggaca atccctcagc ctactgagat gaagaaactg agtctcagag

aggttaagga actcccccga ggttgcacca ctgaggaaga ttgacctgac ttcccaagac

catacatctg gtaaaccgga acctgcacct gcgccatctc aagcctactc tggaggcccg

aggctaattg gcagagtttg aaggctgaga tgacagacaa ccctcagtgc cttcatcgga

cgggctgctt acctgcacat ccctggtgac agcatgggaa agaccgcttc taattaagcg

tcatcacaca tcacccttttt ccggaggaag agaggcaaag acagctcccc tctatcctgc

actgtgaacc ttcctccgag gtcctccct caccccgag agtccttgcc ctgtcaccaa

gattaattac ccctcaaccc ctttgaatgg caaaggcagt cattttatta agtttaatta

aagcttcaag agacattgcc ggatgtttca ggactgctga caaagcagcc tgcttgtttc

ctgaaaagac caattatatg caagaagcgt cagccccact ccccgagggg tccacttagc

ctccgaccac cacagggtga tgtccagcgc accggtgtgg ccatcacctg gcgggagtga

ggtcggccgt agcaagtctg gaggcccccc tgcaagtccc tgcctcaccc tatgactgaa

cctctcgtct ctgggaattt tgttccaaat tccccgtcta ccaggtgtga tttcctccag

ccccaccagc cctgggaggc gcccatccag agagcagaga tagtgaccat gagaaaaagg

actctcaaag tactcaccct cctcgtcctc ttcatcttcc tcacctcctt cttcctaaat

tactcccaca ccatggtgac caccacctgg tttcccaagc agatggttgt cgagctctca

gagaacttta agaagttcat gaaatacact cacaggcctt gcacctgtgc ccgctgcatc

gggcagcaga aggtctcgac ctggttcgat gagaggttca accggtccat gcagccgctg

-continued

ctgaccgccc agaacgccct cttggaggag gacacctaca gctggtggct gaggctccag

cgggagaagc aacccaacaa cttgaacgat accatcaggg agctgttcca ggtggtgccc

gggaacgtgg acccccctgct gaagaagagg tcggtggact gccgacgctg cacagtcgta

ggcaactctg gcaacctccg aaagtcctgg tacgggcctc agatcgacag ccacgacttc

gtgctcagga taaacaaggc ccccacagcg ggcttcgaga tagatgtcga gagcaaaacc

acccaccacc tggtgtaccc cgagagcttc agggagctgg cggagaatgt cagcatggtc

ctggtgccct tcaagaccac cgacctggag tgggtggtca gtgccaccac cacaggcacc

atctctcaca cctatgttcc tgttcctgca aagatcaaag tgaaaaagaa taagatcctc

atctaccacc cggccttcat caagtacgtc ttcgacagct ggctgcaggg ccacgggcgg

tacccgtcca ccggcatcct ctccgtcatc ttctcgctgc acatctgcga cgaggtggac

ttgtacggct tcggggcaga cagtaagggg aactggcatc actactggga gaacaatcca

tcggcggagg ctttccgcaa gaccggggtg cacgacgaag actttgagtc caacgtgaca

gccaccttgg catccatcaa taagatccgg atatttaagg gaagatgaca ctgccgagga

gcaccggagc ccgcctcttt ggccagcccc agcctctgct ggagccgatc tgtgctgggg

gctttgaggg ccagcctcgg gggcgtgttc aggtgcccct gftgcccccct cgcaccccga

catttggcag catcgactca gcaagacccc agaaccccgc tgggtctgca gagcgagtgt

cagaactgat cttggatggg gacacccccc ctccacctcc ctgcaccgcc actgctgtcc

tggagcaggg ggcaggtggg gagacgcggg aggaggtccg gtctccaaac gctcaagtca

tttcggcttt ggggggcaga gggggagata tcgaagactc tgtgttctaa agtgaaggac

actggccctg ggcaaagggg tcccgggctg actctctaac tctgatgctt gctgggtgaa

gacgacctcg gaacagagcc acatcgagcc acgtagacgc taggggtgaa aaggcacctt

cctctgcacg atgcccggcc cctccctcac cgcctctgcg gtcttcccga agctcctccg

tggcggccag gagaggcgcc ctgcgccgag ggtcctcaca gatgcttggc caagtgtttc

agactccagc aggtgtcttc ttcgcaagct gagactccct tgagtgatcg atctttgtgg

ccataaataa tggctaagag caaatctgac tacttccacg tgcctttgtg tctgggggag

agctgtgcgc attggctgaa ataaggcaaa agccttaatt cgggagtggg gagctccccc

ctctccccgc ccccagcaat gccacccccc tgctctggag ctgggtaaca tctttactag

tttcctgagg cggtaccgga gctggaatga agctaggatg atgctcaacg gcgtccagaa

gtgcatcagc cacggcacgt cattcttagg ttcacagact gatggattta gggcccatga

ccttccaagg ccacacacag aacctccacc agccagcact ttgccccagg agcccaccct

ccaggccaag tccctccggg ccacctgcac agtgtcctgg ggtttgtggg gccatcccag

gaacatgacc ctcggagcag gagaaaacat acctcatcct ctcacttctg ggctgctctg

attcctcctc atgattattt attgattttt ttttaattca gctgacattg cgtgcacttt

gtggctccgt ggctggtcct gatgttttaa ttaagcttgc tctcgcttca cctggcagcc

ggggcatggg gggcttaagt caaggtttgc aggagtcctc agacttggga gggggatgca

tatctagggc ttgggggctc gtcggtgggc aaatgccagg gtttcaggtt ggtaggttct

ccgagtgctc tgtacctcct tcccctcagc tctgcctccc ttctccatca ttgcggattg

ggatcattct caggctaaga atctgcaaaa cccagtgaga cggccttggg gccaggtggg

gagtggagtc acacaggcag ggcagcgatg gttggctccg gttgctgaca ctgaatcaga

-continued

aaatccacgt ttcctattga gagcatttcc taacaggcca ctgctacttc gaggaggtgt

gacagtgtcc tggctgtcac gggggcccgt ctectgtgct ccttgcagca gcactccggc

aagttctgct ctcctggctt cagtctctgc ctctgcaaac ggagagggag gcatttgggg

agctcagggg gccacaggtg gtttcagggg gttcacaagc gtccccagag agagcacacc

acgcccgtgt tttaccccaa gcctgtgtgt atgagcatat attcctgggg aacggtgcca

tagcgttcat tcaattgtca gaggttcaga acccagacaa tggtttacaa aaacacgaaa

cgacagcaag caaacaaaaa cccgttcgag ctttaaaatc ctccaacact tagactctct

tgacccctaa aggctcataa atcactacga ggtcagcaag ctgatcataa atcccaaaga

ctatatagca ggcatctgtg ggttggctga tgaatccgct gagagaagtg ggtgcggagc

ggcctccgag ctcctccccg agtcctggaa gagccgagtc cagtcagaac tgatgagtcc

agagcgctca gcctgcgtcg ggactgtgct gagtcaggac aggtccccag agtaaggggg

ggggtctctc tccctcccca ccccactgag ccgtccaggc tgttgagagg acttcccacc

tgccttccca tgtaacactt gagaaaagcg aggcccggag aagggaagtg gcttgcaaag

tgaacccatg gtgctggaca acgcaagtca gacccctttga gcaaagatag aacacgatta

ctacatgttg aacacacctc gaggtgggct ttcccagccg gcttccatcc agcctcccca

aggactgtga ccgggcacag gtgggctcct caggctgcac gtggtggccc ccgcaagctc

cggtgtctgc cttttccccc catccctcca gtgagggaga cacccttgct cttccctcag

ccccagcaaa cttgtctttg tgttcattgg cttagcaaag aggtctgacc cgaatacacc

aacgggcttc agccccctgg catctgtacc tggggtaagg gtggctgaga atccaaccag

agtatttccc ccagagctcg tggctgtggg caggttaacg agaaccaccc gccccccccа

ctcggaggag gctcctgaat gcccacccca gtgttgggct cccagcctag ggccctgctt

gctcctcaag cacaaagctc ccttggaggc cagtcctaga cccctgggca gagagcagaa

gcatccctgc ctctctcagt gctgagcccg aagccaggaa gaccagggcc gaggagagag

agagagattt taacaccctc cccggccctg gggagccagg agctggagtc acctgccgca

aatgcaaggc ctggttccac gtgcccccct ccgcccctgt cctcaccccc ttgttagggt

gggtttgtag cgtctggctc ccagaaccat cagacaaaaa agaaagttct taaagagatg

cttttagggg cctgtgctta gtgggagaag ccagattcaa gaaaatggga cttaaaacgg

ggaagccaga cgagaaggtg gggggggggt tgtccgggtg tgattcttac aagtgctgag

gccagaagga tggaccctct ccaggtccag agtagagtgg agtcaacggg gaatatttag

gactgtgatg tttgcagtaa tagtcctctt caaggcaacc tggacacaaa gcaaaatcga

aaacagaaaa gaactcacca tcccttgcgt gggtaggacc aaaacagctt tctccaggga

ggcgcaagcc ctgagtggcc gccgtctggc caaagccggc ctcgtgccat cgggaaccgg

ccaccgccca tgcccctccg gagagctcgg ggtcgtcatc gtgtccacca cgggcatctg

gtcagaagca aacgttccca caggccactg gctagtcatc atcccctgtg gtctgagctg

accggtttga gaccaagtcc gctgccttgc tccggcctgt cctgaagaca ccagcgcccc

ggcctgtggg gtgccccgtg atgcctccat gcagggaaga gcaggggacc agggaggaag

agcagagaca ggtgaagcga cagtccccgc gtcccagcct cagcattcgc atcctcttgg

cccctacttt tcctctccgc ccagcagaca tctgccctgc ccttgccctt gaccccattg

ctgcgcttcc ctcaaggacg ggcctggcct tggtggccac ctgcggacag ccctgcgccc

-continued

gacgcccgct tcaccccggg gcccgggtct ggaggggccg cccccaggac gaacgcggct

gccccacggg gccggcccct caccggcttc gcgtccaagc caaagtttct cgagcacttt

tttgttcttt gcaatcatgt tgggttcatt gttggtgttt taaaattttg cttccctctc

cctctggcct cgctcctgtg tgtgttttgt agccgagcgc taacctggat gctttttga

atgacctttg caagagcctg ccttcctcgg cctctgctct gttttattta ttgttgaata

tttccaatga tccaaatcaa agtgaattaa aacaaagcta ttttatcgtt;

a ST3 gene encoding a protein having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to

(SEQ ID NO: 6; Accession No. XP_022282430)

mvtmrkrtlk vitilvlfif ltsfflnysh tmvtttwfpk qmvvelsenf kkfmkythrp

ctcarcigqq rvsawfderf nrsmqpilta qnalleedty swwirlqrek qpnnindtir

elfqvvpanv dpliekrsvg crrcavvgns gnireswygp qidshdfvlr mnkaptagfe

mdvgsktthh lvypesfrel aenvsmvlvp fkttdlewvv satttgtish tyvpvpakik

vkkdkiliyh pafikyvfds wlqghgryps tgilsvifsl hicdevdivg fgadskgnwh

hywennpsag afrktgvhdg dfesnvtati asinkirifk gr;

a ST3 gene having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleic acid sequence identity to a canine ST3GalI gene comprising

(Accession No. XM_014114023)

(SEQ ID NO: 7)

cggctcgctc cgggagggca gagccggcaa gggcgggact ggcctgctgc gggcggacgg

gggagccgcg gagctaacgg gtccggacgt cgccagcggc ggagctttct gcacgggcgg

aggcggcagc aggaacggca aacatggccc gcgcggcacc cggcactgac taagggtcat

gctgoggcag cctggcgcct gacgagtgag ggggaacctg ccaggggatt actgctagca

cgagctaaca gcaggacagc aggcctcagg ggagtgggga tgggcatagc agcctgcctg

caacgctggg ccagtggcac caaggaggcc caggccctgg cagggagctc ggcacacgct

cagctgaccc agacagcatt caaagcaggt ccctgctgag cagcagctat toccaagocc

tggaggaccg caaacgcctc agcaaccaca tttggggctg tgaccatcag gagaagggac

caacacctcg aatgggaagc agaagggcag aacctgctgc caggaggatc ttggtctctg

cagcaaagtc ttccgacacc cgtggagctt cttgggtttc tcctctcagg actggggcct

tgtttcctcc tctgatggag gaggtaggcc aggagggcca gctgaggtgg gaacgactct

cggatggaat ccaggccact tgaagcccat gggctctctg ctgtgattgg ggttcgaggg

cttocccact gcaggggacg aggggccgcc tccgtgtctg ctcatgaacc acaaggaccc

cgaatgctcc agactggacc atttcgagcc accaaagagt ggggccccca gagctggcag

ccaacagctc caagggacta gagagctggg atgaactgac ctcccoctca ccaggacatt

aggagagtca gagcctctgt ggcccagcta gtgacagaga gacccgatga ggccataagc

tggcacccga gtgactcagg gacaggagag ccactcctgc ctactgtgct ttctotttac

attgagggag ggtgtggtaa gggacccctg cctctgttcc ctcctcagca cgtgccccta

tgccotttgc acgtagtgcc aagagacgca ggctacactg tggctggccc atcagtgggc

-continued

tgggaaggga gtggccacgg tgacacccac ccgcccacca ggctggtggt ctagoccccc

ggggcgccaa accatgaagt gctocctgcg ggtgtggttc ctotctgtgg cottcctgct

ggtgttcatc atgtcactgc tctttaccta ctcccaccac agcatggcca cottgoccta

cctggactca ggggccctgg gcggtaccca ccgggtgaag ctggtgcctg gctatgctgg

tctgcagcgc ctcagcaagg aggggctcac cggtaagagc tgtgcctgcc gccgctgcat

gggtgacact agcgcctctg actggtttga cagccacttc aacagcaaca tttoccctgt

gtggacccga gagaatatgg atctgcctcc agatgtccag aggtggtgga tgatgctgca

gocccagttc aagtcacaca acaccaacga ggtactggag aagctgttcc agatagtacc

aggcgagaac coctaccgtt tcogggaccc ccaccagtgc cggcactgtg cagtagtgga

gaactcaagc aacctgcggg gctctggcta cggcccaaat gtggatgggc ataacttcat

catgaggatg aatcaggcgc caaccgtggg ctttgagcag gatgttggca gccgaactac

ccaccatttc atgtacccog agagtgccaa gaaccttcct gccaacgtca gctttgtgtt

ggtgccottc aaagctctgg acctactatg gattgccagc gctttgtcca cagggcaaat

cagattcacc tatgcgccag tgaagtcctt ccttcgagtg aacaaagaaa aggttcagat

ctacaacccg gcattcttca agtacatcca cgaccggtgg acagagcatc acgggcggta

ccottccaca gggatgctgg tgctottctt tgccctgcat gtttgtgatg aggtgaacgt

gtacgggttc ggggccgaca gccggggcaa ctggcaccac tactgggaga ataaccggta

tgcgggcaag ttccagaaga caggagtgca cgacgccaac ttcgaggccc acatcatcga

catgctggcc aaggccagca agatcgaggt ctaccgaggc aactgagccc ggcctcgccg

cgacccttcc ggcccagccg cgaggctgcg acgctcgctc cgcagccggg actcccggcc

agcccgaggg cggcggcctt ggcgagggcg cggcgccogg ttgggcgtct ccagccctgg

gagcgacgcc aagtcgcgat ctggaccaat catactgcaa atccagcgag cgccgactgt

ccoccgccaa tcaggagact ttgggggctg gcccaggcct ggcacccaat cagcgctgca

gtgggagcgg aggctctttc tcccagccaa tcatgcgact caaggagaac ttccggcgct

gggcccggtc toctccaatc aatggccttc ggaggcgggc cggccgccgc tgaatcocca

ctoccotatg ctttaggtag gattttattt tatgcttttt aaggagtagt gattggttcc

ggcctcagtg gagtactttc ctcaggctct gcgggaggag tgttggtggc ctgtoggcgg

tactcggcca ggggcaccga ggaggaagcg ggggggaagg tgcggggcag cagcggctgg

gcctccottg gccgggggcc cctcgcgacc tcggggcggg cgggggggcg ggggcgttgg

cctcccgctt ctggaggtcc gggggaaatc aggtggtttc cggggagcac gctttcatcc

ccgggaagag ctagatccct ctcaaacctt ttcaggcctc agagcactct agaccgcgta

tttcctttat ctgtogggcc cagatgggtg agtgtaaacc caccaaagaa aggcagtgga

gcgtgggtcc cctcgtoctc cottgtoccc atccccacct ttggccaccc tatgggatgg

cottcctaac caggacattg aacatcccac ctggaaacta gaactgtatt cacccgttgc

tagggctcgg agttcgccga tggcctaact cgagcgcagg ggtaaggaac agggcagggg

acccggaagc cccgtcactt cagatgtaag gtgcttctca ctctcgtgta ctctcgcggc

cccttactgt tcgcccacaa ctttttttagt gtcccttctt aagocctggg cctcctcacc

agcctgctcg tcctgggatt gggggtgggg ggtagggcac tgctggcttc tccaaccocc

taccoctoct cgctcgtott tagccggctc tagggagagg aaggcaggct ggagatgggg

-continued

agcccagctg cctggtgcat gcaccgtttt cctccgccca tcaccccaaa gaggagtagg

aaacctcttg cttgggggtg gaatttgctt tggtctccta atttagttaa cttgaggtta

ccagggatqg ctgaccaaca aaqattottt taaaattcca ggctgqccat gcaaattgct

gggatcctag ctggggagga gtcgactgac ttgcccgcct tacatgtotc ctctoctgcc

cctgcgtocc ctocctotgc cacgctcact toctacctca totctccaac ccattttcca

ttttcagctc tagaagggca gggacgctta caaacaggag ttacatctgq aagttacttc

caagactgaa cccagcttaa gtccctagag gaagctgctg atgatattct cacccttcaa

ggttggggaa atttcggaag gggaaagtgc ttctgtgaag cttccaaacc actaatagga

tccoccttcc caacaatgag gaacacaaac accacccttt atcttagttg ataccaccaa

gcagcctcct ggccattggg gtaattcctg cagctggctg gggtaaccag caqgggagta

tattagaaga ggattggggc aaggcagtgg gcaccoctaa agttaatata ttgagaactt

agcttaaacc taagtcttag ttcottocca attccaaaag taggaggagc aacgagtgga

ggtgaatttg gaggggccta tcctggaatg cctotctcag gacttocccc accattttag

agagtcaagg caccagccat tcatgccagt ctcctctcag tqcttcctga agaggctgtt

tggaqtgttc gqaaaatgaa aaaaacaatg caattatgcc aaacagtatt gagcagaata

atttatttct tttttgttcc tttttcttct ttttgttttg tttaaaacat taataaatcc

cotttctgaa agaggtaggt cccagcatcc agcccagatc tocttttctg caatagttat

ttaaacaaat gtttgtttgt ttttttattt tcttcccttt ctctctottt ctqaattaaa

aaaaaagaaa actccta;

a ST3 gene encoding a protein having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to

(Accession No. XP_013969498)

(SEQ ID NO: 8)

mkcslrvwfl svafllvfim sllftyshhs matlpyldsg alggthrvkl vpgyaglgrl

skegltgksc acrrcmgdtg asdwfdshfn snispvwtre nmdippdvqr wwmmlqpqfk

shntnevlek lfqivpgenp vrfrdphqcr rcavvgnsgn lrgsgygpdv dghnfimrmn

gaptvgfegd vgsrtthhfm ypesaknlpa nvsfvlvpfk aldilwiasa istgqirfty

apvksfirvd kekvqiynpa ffkyibdrwt ehhgrypstg mlviffaihv cdevnvygfg

adsrgnwhhy wennryagef rktgvhdadf eahiidmlak askievyrgn;

a ST3 gene having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleic acid sequence identity to a canine ST3GalIII gene comprising

(Accession No. XM_025420404)

(SEQ ID NO: 9)

ttgatggtcg cgctccgccc gccgctgcgt coccaccatg acggcgcccg tgcagcccac

cgcgtcgtag gcgcccgccg ggctocccog cggctgtgac ggccgcccgc gcctoggcct

ccgcctcccc gcccgcgccg gccogggcgc cgcctccccg ctgcctccgt ctccgctgcg

qtcatgtaqg aaatcqtaaa tcatgtgaag atgggactct tggtatttgt acqcaatctg

ctgctagccc tctgcctttt tctggtactg ggatttttgt attattctgc gtggaagcta

-continued

catttactcc aatgggagga ctccaattca gtggttcttt cotttgactc cgctggacaa

acactaggct cagagtatga tcggttgggt ttcctcctga agctggactc taaactgcct

gctgagttag ccaccaagta tgcaaacttt tcagagggag cttgcaagcc tggctatgct

tcgaccttga tgactgccat ctttoccogg ttctccaagc cagcacccat gttcctggat

aactctttcc gcaagtgggc taagattcgg aagtttgtac cgccttttgg gatcaaaggt

caagacaatc tgatcaaagc catcttgtca gtcaccaaag agtaccgcct gacccctgcc

ttggacagcc tcagctgccg ccqctgcatc atcgtgggca acggaggtgt cctagccaac

aagtctctqg ggtcacgaat tgatgactat qacattgtqg tcagactgaa ctccgcacca

gtgaaaggct ttgagaagga cgtgggcagc aaaactacac tqcgcatcac ctaccctgag

ggcgccatgc aqcggcctga gcaatatgaa cgcgattctc tatttgtcct cgctggcttc

aagtqgcagg acttcaagtq gttgaaqtac atcgtctaca aggagagagt gctctgqgcc

cgcagggata cctgccaatc tgtctgqgcc catccocctc toccotccac cagctgtcac

caqccacccc aggggaqqag tcctqcagag ttcaggccat tottottcca atacccgagc

ctcctactqg aggagaatga tgacagacag cctctggcga caagtqcatc agatggcttc

tggaaatccg tggccacacg agtgcccaag gagccccctg agattcgcat cctcaacccg

tacttcatcc agqaagccgc cttcaccotc atcggactgc ccttcaacaa cagcctcata

ggccgcqaga acatcccgac ccttggcaqt gtggcaqtga ccatagcgct acacggctqt

gatgaggtgg cagtcgcagq ctttggctac gacatgagca cacccaacgc gccoctqcac

tactatgaga ccgtgcgcat ggcagccatc aaagaggtca ccagcgactc agctcaaggc

tgccaaatcc aqtggacaca tggaagcctc atctttcctg acctcccaga aatgcttttt

ctqttgacca ctccttcctc tttqaaactt ttectgctca aactgtcctg gacacacaat

atccagcgag agaaagagtt tctgcgcaag ctggtgaaqg cgcgcqtcat caccgaccta

accagcggca tctgaggtgg gcccagcaca tggccacgga ggtcctggca ccqccaagag

gaagccgcag ccactqccac ctqtccactt cattggcctc ggtctqgctc tgcctgaaag

gcgcaggaqt cttcaqaccc agagaaggac agtgccaaqg gg;

a ST3 gene encoding a protein having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to

(Accession No. XE 025276189)

(SEQ ID NO: 10)

mgilvfvrnl lla1clfivl gf1yvsawkl hllqwedsns vvlsfdsagq tlgsevdrlg

flikldsklp aelatkyanf segackpgya salmtaifpr fskpapmfld dsfrkwarir

efvppfgikg qdnlikails vtkeyrltpa ldslscrrci ivgnggvian kslgsriddv

divvrinsap vkgfekdvgs kttlritype gamqrpeqve rdslfvlagf kwqdfkwlkv

ivykervlwa rrdtcqsvwa hpplpstsch qppqgrgpae frpfffqyps illeenddrq

platsasdgf wksvatrvpk eppeirilnp yfigeaaftl iglpfnnglm grgniptlgs

vavtmalhgc devavagfgv dmstpnaplh yyetvrmaai kevtsdsaqg cqiqwthgsl

ifpdlpemlf llttpsslkl fllrlswthn iqrekeflrk lvkarvitdl tsgi;

a ST3 gene which has at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleic acid sequence identity to a canine ST3GalIV gene comprising

(Accession No. XM_014113293)

(SEQ ID NO: 11)

gccctacagg cccgagctgc cagggtcggg cctccccagg ttcccgctcc caggtcctcc

tggacacacc gacctggcct ggctcccggg gaactctcgt ctgctagcga ggagcctccc

tccgcctcgc ccacgggcac ccctcccacc cagtatcctt gqcctcttgc aggtggcccg

aggcagccgg gatgacagct ctccccagga accctgctac cctctgagaa acatgatcag

caaatcccgc tggaagctcc tggccatgtt ggctctggtc ctggtcgtca tggtgtggta

ttccatctcc cgagaagaca ggtacattga acttttttat tttcccatcc caaagaagaa

ggaaccgtqc ttccaqggtg agqcagagag aaaggcctct aagctctttg gcaactactc

ccgagatcag cccatcttcc tgcagatgaa qgattatttc tgggtcaaga caccgtctgc

ctacgagctg ccctatggga ccaaggggag cgaagacctg ctcctccggg ttctagccat

caccagctac tccattccag agagcatcca gagtctcaag tgtcgccgct gcgtggtggt

gggcaatggg catcggctgc gcaacaqctc gctgqgagat gccatcaaca agtacgacgt

ggtcatcaga ctgaacaacg cccccgtggc tggctacgag ggtgacgtgg gctcgaagac

caccatgcgt ctcttctacc cggagtcagc ccacttcaac cccaaagtgq agaacaaccc

aqacacactt ctcgtcctag tggccttcaa gqcaatggac ttccactqaa ttgaqaccat

cctgagtgat aagaagaggg tacgaaaggg cttctggaag cagcctcccc tcatctggga

cgtcaacccc aggcaqgttc ggattctcaa ccctttcttt atggaqattg caqctgacaa

actgctqaac ctqccaatga aacagccacq caagatttcc cagaagccca ccacgggcct

gctggccatc acqctggctc tccacctctq cgacctgatg cacatcgccg qcttcggcta

cccggacgcc cacaacagga agcagaccat tcactactat gaacagatca cgctcaagtc

catgqcgggg tcaggccaca acgtctccca ggagqccctg gccatcaagc ggatgctgga

gatcqgagca gtcaagaacc tcacgttctt ctgacgggga caggagctct agccgtcagt

ctqcccgccc tgccgcctaa gcgaccaacc acqactgtgq aggcgccqac gtgacctgct

tggattcccc ctccccgtgt ggagaggggg cctggtacag gcgggccctg agatggggcc

qcgcccctqg ctgctcttgg ggcggccgga tccagtcaqg gtggaqgccc cgqgtggcgg

qaggccttcc gaggcqcggg gtqtgtggct qaggcacccc ttctcaccag ccccgggagc

ttatttaatg ggctatttaa ttaaaagqqt aggaatqtgc ctcqagctgg tcccatgqca

tccggaaacg gqggcatagc acagtggtct gcccactgtg gataaaaaca cacaagtgct

tggcccacta gagcctagaq ccagagcagg cctcccagga gqgcaggggc gtctggagcg

ggtgqgtgcc ctccagagaq gggctgctac ctcccagcgg gcatgggaaq agcattqgga

tqaagtccca cggagaatag gacctcatgt aqaaaagagg tttgaaacct aacattaaac

tatttttcc taaaacggaa;

a ST3 gene encoding a protein having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to

(Accession No. NP_013968768)

(SEQ ID NO: 12)

misksrwkll amlalvlvvm vwysisredr yielfyfpip ekkepcfqge aerkasklfg

nysrdqpifl qmkdyfwvkt psayelpygt kgsedlllry laitsysipe siqsikorrc

vvvgnghrlr nssigdaink ydvvirinna pvagyegdvg skttmrlfvp esahfnpkve

-continued

nnpdtlivlv afkamdfhwi etilsdkkry rkgfwkqppl iwdvnprqvr ilnpffmeia

adkllnipmk qprkisqkpt tgllaitlai hladlvhiag fgypdahnrk atihyyeqit

lksmagsahn vsqealaikr mleigavkni tff;

a ST3 gene having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleic acid sequence identity to a canine GalV comprising

(Accession No. XM_022404744)

(SEQ ID NO: 13)

cgctctggaa ccacttacag ccacctggtg catcctcctt tggggtgcgt ttqgagggcc

tggttcctqc tcagccacat cttctgccac tttcaccaqc aatgcccagt gaqtataact

atgtaaaact gagaaqcgat cgctcaagac cctctctgca atggtacacc cgagctcaaa

acaagataag aaqacccaac ttgttgttaa aagacatcct taaqtgtaca ttgcttgtqt

ttggagtgtg gatcctttat attctcaagt taaattatac tactgaagaa tgtgacatga

aaaaaatgca ttatgtggac ccagaccgtg taaagagagc tcagaaatat gctcagcaag

tcttgcaaaa ggagtqccga cccaagtttg cgaagaagtc gatggcgcag ttqttcgagc

acagatacag cacggactta ccacctttcg tgaaagagac coccaaaata aatgaaaccg

agtacaagta taatcctcct tttggattcc gaaaattctc cagtgaagtc cagaccctgt

tggaaatact gcccgagcat gacatgcccg aacacttgag aqcaaagagc tgtaggcgtt

gtgtqgtcat cqgaagcggt ggcatactcc acggactagc actgggccaq gccctcaacc

aattcgatgt agtgataaag ttaaacagtg caccagttga aggatattct gagcatgttg

gtaataaaac tactataagg atgacttatc cagagggcgc gccactgtct gaccttgaat

attattccaa tgacttgttt gttgctgttt tattcaagag tgttgacttc aactggcttc

aagcaatagt aaaaaatgaa accctgccat tttggatacg gctottcttt tagaagcaga

tggcgaaaaa aatcccacta cagccaaaac atttcagaat tttgaatcca gttattatca

aagaaactgc ctttgacatc cttcaatact cagaacccca gtcaaggttc tggggccgag

ataaqaacgt gcccaccatt ggtgtcattg ccgttgtott aqccacacat ctgtgtqatg

aagtcagctt gqcaggcttt ggatatqacc tcaatcaacc caaaacacct ttgcactact

ttgacaatct ctgcatagct gccataaact ttcaaaccat acataatgtg acaacagaga

ccaggttcct cctcaagctg gtcaaagagg acgtggtgaa ggatctcagc ggaggcatcc

attgtgaatt ttgaacacag ggaaacctca tgtgacaatg caactctgac tctgaaggct

qtttttcgta gccttctcga tgcagcgcat cctgcaaaat acttagaggt gcagctgggg

ttttt;

a ST3 gene encoding a protein having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to

(Accession No. XP_022260452)

(SEQ ID NO: 14)

mpseynvvkl rsdrsrpslq wvtraqnkmr rpnlllkdil kctllvfgvw ilyilkinvt

teecdmkkmh yvdpdrvkra akyaqqvlqk ecrpkfakks maqlfehrys tdlppfvket

pkmneaeyky dppfafrkfs sevqtlleil pehdmpehlr akscrrovvi gsggilhgla

-continued

lggalnqfdv virinsapve gysehvgnkt tirmtypega plsdleyysn dlfvavlfks

vdfnwlqamv knetlpfwvr lffwkqvaek iplqpkhfri lnpviiketa fdilaysepq

srfwgrdknv ptigviavvl athlcdevsl agfgydinqp ktpihyfdnl cmaamnfqtm

hnvttetrfl lklvkegyvk dlsggihcef;

a ST3 gene having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleic acid sequence identity to a canine GalVI comprising

(Accession No. XM_005639375)

(SEQ ID NO: 15)

ggtcgattgc cccttggctg ctgtggaggc tgtgatgacc tccaaggccg cagccctcca

ggcgatgctt ctccaggggc tgaggccaac gcagaactcc catggcaccc actcggactc

gcggcgtgtt cacatgtggq gttttattaa atcctcccac caaccgtgtq agacaggaac

agttagcccc gqtgtgtccq ccaagattgc cccgcacaag tqgctccgga tggatcacac

gaaacacttg caagtgaaaa agcagcacag ccotttatct tgggctattt cctgtagaga

aactccaaca atttaacagc caagctcctg agcctctgag accctcacca catcacatcc

ttcaccttca ggagcagagc gcctttggga aacagacttc taaaaqtgca ggtgggccag

ccatgagaqg gtacctagtg gccatattcc tgagtgctqt ctttctctat tatgtgctgc

attgtatatt gtggagaaca aacatctatt gggtgccacc tgtgaaaatg aagcggagaa

ataaaatcca gccttgttta gcgaagccag cttttgcctc tctoctgaga tttcatcagt

ttcacccttt tctgtgtgca gctgatttta aaaaqattgc ttccttgtat ggtagcqata

agtttgatct gccctatggg ataagaacat cagcggaata ttttcqactc gctctttcaa

aactacagag ttgtgatctc tttgataagt ttgacaatgt gccgtgtaaa aagtgcatgg

tggttggtaa tagaggagtt ctgaagaata agacattagg aaaaaaaatt gactcctatg

atgtcataat aagaatgaat aatggtcctg ttttaggaca tqaagaggaa gttgggagaa

ggacaacctt ccgacttttt tatccaqaat ctgtttttc aqatcccaat cacaatqatc

ctaatactac agcgattctc actgctttta agccgcttga cttaaagtag ctgtgagaag

tgttgacggg tggcaaaata aacactaatg gtttttggaa gaaaccagct ttaaacttga

tctacaaacc ttatcaaatc agaatattag atcctttcat tatcagaatg gcagcttatg

aactgcttca cttoccaaaa gtatttccca aaaaccaaaa acccaaacac ccaacaacaa

gaattattgc catcacgctg gcctttcaca tatgtcacga agttcacctt gctggtttta

aatacaattt ttctgacctc aagagccott tacactatta taggaacgca accatgtctt

tgatqaataa gaatgcgtat cacaatqtga cagcqgaaca gctottttg aaggacattc

tagaaaaaaa ctttgtaatc aacttgactg aagattgacc ctacagactc tgcagatgat

gctaagagta ttagttttat ttttatactg caatttttag tttatttta aatatattgg

atgcacttat caaaaaattg tgtatagtca atctattgct gcctgatgat tcataaccac

cagcttaatt tctgtgaata tatttaattt ataaaaacca agaagatatg cttagatatc

cgggaagttt tgattqcgtt ggttttaaaa caaccttaqt tctctqaagt gtttttaaac

atctttttta atagttactt catctttgac ttctgagagc atgtaacgtc caagtaagga

gotttagctt gaccaccaca aactctaaac agagttggtg gcggattcga ctactgtaaa

ttggtgggga atagccatgt gattgtgcaa actggaaccg gtttaggcaa gtatcgagtt

-continued

cctttttact gaacccgagq aaacggattt gaatcttaaa gcaggcccaa ccatagcagt

aggtacggtt atgaaatcta agatcataat ggtttcatta agctttttt cctgtaagta

aaccagatta taaaatgaaa ggtgtttgtt tttaaggtag aggaaacagg ctacatgtga

aattctggat gagtaaacaa cctaggaatg caattactaa agtctggtgg ctgcattatt

ttaaagttca tacaaagaag cagagctagg ccacctcaag gagacagttc ttaaacgtca

tcttttgcct gccttaatat gttaaaattt ggaagtttac tatttgaaat aagaaagata

aatacggcac aataggtaaa tocttcagac tcctcaggct gtttttggat ttaaatagtc

ctttcgtgaa aaatctcact tgtccacggt gaaatcccat cttcaaaggg aaggcttacc

cggctaccta gqgtgcatca gagaagagtc ctgctggatg cagacaagtc aaaaccagcc

tgtccaacaa acgtgcgccc gtotctottc tcaaagaggg atggaatgaa cagctctcag

aagaggtaag agttgaagga cttgttatcc tctgagcgat aatcgtcatg gagagacact

gctggtattc ctgaaaacca gcctgcctct gagtctcaga gacaaaatat gagagcagcc

actgggataa atcgtqaagc acqgcataag qggggagaag cctcgtagtt gattgaaccc

atgtctacgt ggcttcagct gattcccctg taacggaagt ggaaagttcc cacacgtaca

cagctgcacg ctgcagccta gcggctagga ttccatgggt gaactcattc agggtacaaa

gacagtcctg gctgcaaagt gaaaaacccc aggtggcatt ttcaagtgtt tatggactga

aataatggct gtacggtatc tggcggatgc tcaacttgag gaatcggcat ttttgtacag

tggaagctga agctataaac ctcagcgtgg cttcacataa accagaagaa actctcagcc

cgatacatat gtacaattta ttaaaaacac atgaacacat taaaatctca ctatttatac

aatctacatt ctagcaacat atacaaatac cgagtgacta cagtacatgc cgaggtaaga

aaagtacatt cggggagact atcactgaca ctcaagccat ttttatttcc aatatgtttt

gotttcacct ttcccagtgc caaaaaaaaa aaaaaaaaa;

a ST3 gene4 encoding a protein having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to

(Accession No. XP_005639432)

(SEQ ID NO: 16)

mrgylvaifl savflyyvin cilwgtniyw vppvemkrrn kiqpclakpa fasllrfhaf

hpflcaadfk kiaslygsdk fdlpygirts aeyfrlalsk lqscdlfdef dnvpckkcvv

vgnggviknk tlgekidsvd viirmnngpv lgheeevarr ttfrlfypes vfsdpnhndp

nttailtafk pldikwlwev itggkintng fwkkpainli ykpyqirild pfiirmaaye

llhfpkvfpk nqkpknpttg iiaitlafhi chevhlagfk vnfsdlkspl hyygnatmsl

mnknayhnvt aeqlflkdil eknfvinite d;

a ST3 gene having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% nucleic acid sequence identity to a canine ST3GalII-like comprising

(Accession No. XM_025469036)

(SEQ ID NO: 17)

aaagacttca ctgggtatca gtotcctttg ggagaccaca gqacacgtgt cacctctccc

atcctctcag cctccagccc agaccttggc agagttcctt ttaggagtta gcaagtggct

-continued

gaggaggcaa gaggtgccaq agccaatcta ctatctgctg gqgggatgatt gccagggcca

gagatgaggq ctcaatactt gaaqtagggt ctgatagctq cctgtataat tacqttatgg

ctgatgatqa tgaacttcct ggaccaggag ttcaaacaqa atgacttccc taaaaagaca

agaatacaat tatgccactg coccaggaac tctttcagaa agtgtaggtg ttcgtttgag

atccgcaagt gctctqcctg cctccgcgta cgtggaacqt ctgtctggtt tgatgaacgc

ttcgaaacgg ctattgagcc tatgcagaqa ccagaagatc ccatatcctc taatgctcta

atattgtggt taggtatcca atcaaagagg gagtttgaga ctcagaagcc aatagaagag

cctcctgggc aaccactggg ctacgtggag tccagttgtc ggacctgtgc agtggttgga

aactcaaggt gcctacgagg ctctggccat ggattcagga ttaaccaaaa tgacatggtc

ctcaggatga accaggcccc cgtccaagga tttaagatgq atgtgggqgaa cacaaccacc

atgcgcataa tgtaccccga tatggctagc acgcagaatc ctggcaccaa attgctgctg

cttcctctga attcatctgg tctaaagtgg tttatggaag tactacagga acagagcttc

agaaagccca taaaccctgg atttcagata gtccagtttc ctgatggaag taacacgagc

aaagacgagg tcttagtgat cagcctcacc tttcttcagt acatccagga tcattggcta

cgaaaacgtc atcgttttcc atcottaggg tttgtgggtc tattatatgc cctgcacact

tgtgaccagg tatccttatt tggttttggg acagatcagc tcatgaggtg gtcccattac

tgggatgata aatatcggtt cgagagtaac atgcacagtt tcaaagaaga gcagaagctc

atcctccagc tgcaatgtaa gggqaagatt gttatctaca actgacatat ttctgtoctg

ttcagcccac tggaggcccc aggaggctga caggtagtca aggggaccac agagtgtcag

agagggactg gggcttcaag tggaccctgg atatagatca gtctgctgct aaataaaact

acagcttatt tctccca;

or

a ST3 gene that encodes a protein having at least 80%, 85%, 87%, 90%, 92%, 95%, 96%, 97%, 98%, or 99% amino acid sequence identity to

(Accession No. XP_025324321)

(SEQ ID NO: 18)

mragylkwql vaacivtiwl mmmnfldqef kqndfpkktr iqlchcprns frkcrcsfei

rkcsacirvr gtsvwfderf etaiepvqrp edpissdali lwlgvqskre fetqkpieep

pgqplgyves scrtcavvgn srclrgsghg frinqndmvl rmnqapvqgf emdvgntttm

rimvpdmast qnpgtkllll pinssglkwf mevlqeqsfr kpinpgfqiv qfpggsntsk

devlvisltf lqyiqdhwlr krhrfpslgf vgilyaihtc dqvslfgfgt dqlmrwshyw

ddkyrfesnm hsfkeeqkli lqlqcegkiv iys.

[0098] The invention will be further described by the following non-limiting example.

Example

Methods

[0099] Cells.

[0100] MDCK and AX4 cells were maintained in Eagle's minimal essential media (MEM) containing 5% newborn calf serum (NCS) or 10% fetal calf serum (FCS). All cells were incubated at 37° C. with 5% CO_2, and regularly tested for *mycoplasma* contamination by using PCR and were confirmed to be *mycoplasma*-free.

[0101] Clinical Specimens.

[0102] Respiratory specimens were obtained from patients with influenza-like symptoms who visited clinics in Yokohama city, Japan during the 2017-2018 influenza season, and were submitted to the Yokohama City Institute of Public Health for virus isolation. These clinical specimens were collected under the National Epidemiological Surveillance of Infectious Diseases program in Japan. Respiratory specimens were also obtained from patients with influenza-like

symptoms who visited clinics in Tokyo, Japan during the 2013-2014, 2015-2016, 2016-2017, and 2017-2018 seasons, and were submitted to the Division of Virology, Department of Microbiology and Immunology, Institute of Medical Science, the University of Tokyo for virus isolation. These specimens were collected by attending physicians after informed consent was obtained. Our research protocol was approved by the Research Ethics Review Committee of the Institute of Medical Science of the University of Tokyo (approval no. 26-42-0822). Samples that were positive by real-time RT-PCR (see below) or rapid diagnostic kits were used in this study.

[0103] Viruses.

[0104] Human influenza viruses were propagated in hCK cells in MEM containing 1 ug of L-1-Tosylamide-2-phenylethyl chloromethyl ketone (TPCK)-trypsin/ml.

[0105] Real-Time RT-PCR.

[0106] RNA was extracted from clinical specimens by using the Simply RNA Tissue Kit (Promega) or RNeasy Mini Kit (Qiagen). Amplification and detection by real-time PCR were performed with the Applied Biosystems 7900HT Fast Real-Time PCR System (Applied Biosystems) or StepOnePlus Real-Time PCR System (Applied Biosystems). RT-PCR was carried out using the QuantiTect multiplex RT-PCR kit (Qiagen) or QuantiTect Probe RT-PCR Kit (Qiagen). The probes contained oligonucleotides with the 6-carboxyfluorescein (FAM) or the hexacholoro-6-carboxyfluorescein (HEX) reporter dye at the 5' end, and the Black Hole Quencher-1 (BHQ-1) or 6-carboxytetramethylrhodamine (TAMRA) quencher dye at the 3' end. The list of primers and probes used is provided in Table 5.

[0107] Virus Isolation.

[0108] MDCK, AX4, and hCK cells grown in 12-well plates were inoculated with 0.2 mL per well of the clinical samples and incubated at 34° C. for at least 30 minutes. One microliter of MEM containing 2.5 μg/mL acetylated trypsin was then added to cells. The cultures were then incubated for up to 7 days, until CPE was evident. Cell culture supernatants were harvested and subjected to hemagglutination assays using guinea pig red blood cells (see below).

[0109] Hemagglutination Assay.

[0110] Viruses (50 μL) were serially diluted with 50 μL of PBS in a microtiter plate. An equal volume (i.e., 50 μL) of a 0.75% (vol/vol) guinea pig red blood cell suspension was added to each well. The plates were kept at 4° C. and hemagglutination was assessed after a 90-minute incubation.

[0111] RT-PCR and Sequencing of Viral Genes.

[0112] Viral RNA was extracted from 140 ul of culture supernatants using the QIAamp Viral RNA Mini kit (Qiagen). Samples were amplified using the SuperScript III One-step RT-PCR System with Platinum Taq High Fidelity DNA Polymerase (Invitrogen) and specific primers of HA or NA genes. PCR products were then analyzed by means of 1.5% agarose gel electrophoresis in tris-buffer, and target bands were visualized by staining with GelRed (Biotium). The PCR products were purified and subjected to direct sequencing. The level of mutation frequencies were examined based on the height of the waves at each position on the sequencing chromatogram. The detection limit for a minor population was 10%-20%. The list of primers used is provided in Table 5.

[0113] Serial Passages of Human Influenza Viruses.

[0114] Ten-fold serial dilutions (10' to 10^6) of viruses were prepared in MEM. Each dilution was inoculated into MDCK, AX4, and hCK cell monolayers in 24-well culture plates using one well per dilution. The plates were incubated at 33° C. for 3 days. The end point was taken as the highest dilution of the sample showing CPE. Culture supernatants were harvested from wells inoculated with the 10-fold higher concentration of dilution than the end point dilution, and were used for the next round of infection. Viruses sampled after the first and sixth passages in the supernatants of each cell were subjected to sequence analysis.

[0115] Statistical Analysis.

[0116] Data are expressed as the mean±SD. For the analysis of the growth curve data, we performed a linear mixed effects analysis. As fixed effects, the different cell lines, and the time of the measurement (with an interaction term between those fixed effects), were used. As random effects, intercepts for the individual animals were used. The virus titer values were transformed to the log 10 scale, and the R statistical package (www.r-project.org), lme4 (Bates et al., 2015), and the lsmeans package (Lenth, 2016) for the group comparisons, were used. The p-values were adjusted using Holm's method and considered significant if less than 0.05.

Generation of MDCK Cells Expressing Markedly Low Levels of α2,3-Linked Sialic Acid and High Levels of α2,6-Linked Sialic Acid

[0117] To mimic the expression pattern of sialic acid (Sia) molecules on the surface of human upper airway epithelial cells, we first attempted to knockout the β-galactoside α2,3 sialyltransferase (ST3Gal) genes, whose products catalyze the transfer of Sia with an α2,3-linkage to terminal galactose (Gal) residues, by using the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) gene editing system (Cong et al., 2013; Jinek et al., 2012; Han et al., 2018; Shalem et al., 2014). Dogs have seven different ST3Gal proteins (ST3Gal-I, —II, —III, —IV, —V, —VI, and ST3Gal-II-like protein) each of which is encoded by a distinct gene. ST3Gal-I, —II, —III, —IV, and -VI use oligosaccharides on glycoproteins, or on glycoproteins and glycolipids, as acceptor substrates, whereas ST3Gal-V utilizes oligosaccharides on glycolipids only (Takashima and Tsuji, 2011). A previous study reported that N-linked glycoprotein is required for productive entry of influenza viruses into host cells (Chu and Whittaker, 2004). Therefore, to inhibit the transfer of α2,3-linked Sias to glycoproteins, MDCK cells were transfected with a mixture of six plasmids, each containing a Cas9 gene expression cassette and an expression cassette for the individual guide RNA (gRNA) targeting the ST3Gal-I, —II, —III, —IV, —VI, or ST3Gal-II-like protein gene (FIG. **2**). After transfection, puromycin was added to the cells, and 33 drug-resistant clones were randomly picked up. Genomic DNA analysis revealed that only one clone (6-11) contained mutations in the gRNA target regions for the six ST3Gal genes (data not shown). Cell surface Sias was measured by flow cytometry using the *Maackia Amurensis* II agglutinin (MAL II) lectin specific for α2,3-linked Sias and the *Sambucus Nigra* agglutinin (SNA) lectin specific for α2,6-linked Sias. Unexpectedly, the reactivity with MALII was very similar between the parental MDCK cells and clone 6-11, indicating that the clone still expressed high levels of α2,3-linked Sias (FIG. **3**). This may have been due to the compensatory activity of ST3Gal-V.

[0118] To inhibit the transfer of α2,3-linked Sias more efficiently and to express high levels of α2,6-linked Sias on

the cell surface, clone 6-11 was co-transfected with a plasmid encoding human β-galactoside α2,6 sialyltransferase I (ST6Gal-I), which catalyzes the addition of α2,6-linked Sia to Gal-containing glycans, and a plasmid containing expression cassettes for Cas9 and a gRNA targeting ST3Gal-V. Eighteen cell clones were selected with blasticidin and subjected to genomic DNA analysis. Among the drug-resistant clones, 9 possessed a mutation in the gRNA target region for the ST3Gal-V gene (data not shown). Flow cytometric analysis using the MAL II and SNA lectins revealed that two (clones 6-11#2 and 6-11#10) of the nine clones had markedly decreased expression of α2,3-linked Sias compared with the parental MDCK cells and higher expression levels of α2,6-linked Sias than those of the parental cells (Sy FIG. **4***a*; data for only clones 6-11#2 and 6-11#10 are shown). Terminal Sia is attached to several types of oligosaccharide structures on glycoproteins or glycolipids, such as Galβ1,4GlcNAc (GlcNAc; N-acetylglucosamine), Galβ1,3GalNAc (GalNAc; N-acetylgalactosamine), and Galβ1,4Glc (Glc; glucose) (Takashima and Tsuji, 2011). The MAL II lectin preferentially recognizes the Siaα2,3Galβ1,3GalNAc structure (Hidari et al., 2013). To assess whether the two clones express different types of α2,3-linked oligosaccharide structures on the cell surface, an indirect immunofluorescence assay (WA) analysis was performed using a monoclonal antibody against Siaα2, 3Galβ1,4GlcNAc7. IFA showed that levels of Siaα2,3Gal1, 4GlcNAc were undetectable or markedly low in one (6-11#10) of these two clones (FIG. **4***b*), suggesting that in clone 6-11#10, multiple types of oligosaccharides containing terminal α2,3-linked Sias are expressed at lower levels than in the parental cells. Next, the cell surface expression levels of α2,6-linked Sias on AX4 cells and clone 6-11#10 were compared by using the SNA and *Sambucus sieboldiana* (SSA) lectins, both of which recognize the Siaα2,6Gal or Siaα2,6GalNAc structure (Shibuya et al., 1989). Flow cytometric analysis indicated that there were no differences in the expression level of α2,6-linked Sias between AX4 cells and clone 6-11#10 (FIG. **4***c*). However, the expression level of α2,3-linked Sias, as measured by using the MAL II lectin, was markedly lower in clone 6-11#10 compared to AX4 cells (FIG. **4***d*). It was confirmed that clone 6-11#10 contained the desired mutations in the gRNA target regions for the seven ST3Gal genes (Table 1). These results show that clone 6-11#10 expresses mainly human virus receptors and limited amounts of avian virus receptors. The resulting clone, 6-11#10, was designated hCK, and subsequently expanded for further analysis.

TABLE 1

Amino acid changes in the HA and NA of viruses analyzed after passages in MDCK, AX4, or hCK cells[a].

			HA[b]			NA[c]		
Virus type	Sample ID	Cell	P1	P6	P10	P1	P6	P10
H1N1pdm	BB139	MDC	—[d]	—	T167T/I[e]	—	—	—
		AX4	—	—	—	—	—	—
		hCK	N296N/S[f]	N296N/S[f]	N296S	S153G	S153G	S153G
	BB131	MDC	—	—	N446N/S	—	—	H411cH/
		AX4	—	—	—	—	—	C53C/Y[i]
		hCK	—	—	—	—	—	—
	HP79	MDC	—	—	—	—	—	—
		AX4	—	—	—	—	—	—
		hCK	—	—	D27N	—	—	—
H3N2	DA30[j]	AX4	—	N158N/K	N158K	—	—	—
		hCK	—	—	—	—	—	—
	DA29-1[j]	AX4	—	—	—	—	T148K	T148K
		hCK	—	—	D408D/N	—	—	—
	DA23-1[j]	AX4	—	—	—	—	—	—
		hCK	—	—	—	—	—	—
B/Yamagata	HP70-2	MDC	—	—	—	—	—	—
		AX4	—	—	—	—	—	—
		hCK	—	S148S/N	S148N	—	—	—
	BB005	MDC	—	—	—	—	—	—
		AX4	—	—	—	—	—	—
		hCK	—	—	—	—	—	—
	DA09-2	MDC	—	—	—	—	—	—
		AX4	—	—	—	—	—	—
		hCK	—	—	—	—	—	—
B/Victoria	HP015	MDC	—	—	—	—	—	D459D/N
		AX4	—	—	—	G208G/R	G208G/R	G208G/R
		hCK	—	—	—	—	—	—
	WD28	MDC	—	—	N196N/S	—	—	—
		AX4	—	—	—	—	—	—
		hCK	—	—	—	—	L72L/F[q]	L72F

TABLE 1-continued

Amino acid changes in the HA and NA of viruses analyzed after passages in MDCK, AX4, or hCK cells[a].

			HA[b]			NA[c]		
Virus type	Sample ID	Cell	P1	P6	P10	P1	P6	P10
	DA25-2	MDC	—	—	—	—	—	—
		AX4	—	—	—	—	—	—
		hCK	—	—	—	—	—	—

[a]Influenza viruses isolated form the clinical specimens were passaged ten times in MDCK, AX4, or hCK cells. The sequences of the HA and NA genes of the viruses were determined after a single passage (P1), the sixth passage (P6), and tenth passage (P10).

[b]Mutations of influenza A viruses are shown with H3 numbering.

[c]All mutations are shown with N2 numbering.

d—, No mutation was detected compared to the sequences from the original clinical specimens.

[e]T/I, mixture of threonine and isoleucine at position 167

[f]N/S, mixture of asparagine and serine at position 296.

[g]N/S, mixture of asparagine and serine at position 446.

[h]H/Y, mixture of histidine and tyrosine at position 411c.

[i]C/Y, mixture of cysteine and tyrosine at position 53.

[j]Influenza viruses were not isolated from the clinical specimens in MDCK cells.

[k]N/K, mixture of asparagine and lysine at position 158.

[l]D/N, mixture of aspartic acid and asparagine at position 408.

[m]S/N, mixture of serine and asparagine at position 148.

[n]D/N, mixture of aspartic acid and asparagine at position 459.

[o]G/R, mixture of glycine and arginine at position 208.

[p]N/S, mixture of asparagine and serine at position 196.

[q]L/F, mixture of leucine and phenylalanine at position 72.

[0119] Establishment of a Stable Cell Line Possessing Mutations in its ST3Gal Genes and Expressing the ST6Gal-I and HAT Genes.

[0120] gRNA sequences each targeting the ST3Gal-I, —II, —III, —IV, V, —VI, and ST3Gal-II-like protein genetic loci were designed using the sgRNA Design Tool from the Michael Boutros lab (http://www.e-crisp.org/E-CRISP/). The oligo DNA for the gRNA was cloned into the Cas9/gRNA dual expression vector pSpCas9(BB)-2APuro (PX459), encoding puromycin resistance (addgene). The resulting constructs were designated PX459-ST3Gal-I, PX459-ST3Gal-II, PX459-ST3Gal-III, PX459-ST3Gal-IV, PX459-ST3Gal-V, PX459-ST3Gal-VI, and PX459-ST3Gal-II-like, which express gRNA targeting ST3Gal-I, —II, —III, —IV, V, —VI, and ST3Gal-II-like protein genes, respectively. Human ST6Gal-I genes were amplified by PCR from the pCAGGS-FLAG-PUR-ST6Gal-I plasmid (Hatakeyama et al., 2005) and were then digested with NotI and XhoI. The digested fragment was cloned between the NotI and XhoI sites of the eukaryotic expression vector pCAG-Bsd, which encodes blasticidin resistance (Wako). The resulting construct was designated pCAG-Bsd-ST6Gal-I, which expresses ST6Gal-I. All constructs were sequence verified by Sanger sequencing. Cycle sequencing was performed using BigDye Terminator version 3.1 Cycle Sequencing Kits (Thermo Fisher Scientific), and sequences were analyzed on an ABI Prism 3130×1 Genetic Analyzer (Thermo Fisher Scientific).

[0121] Electroporation was performed using the AMAXA Nucleofector II machine (Lonza) according to the manufacturer's instructions. Briefly, 5×10^5 MDCK cells were resuspended in 100 μL of the desired electroporation buffer and mixed with either 5 μg of Cas9/gRNA dual expression vectors (1 μg PX459-ST3Gal-I, 1 μg PX459-ST3Gal-II, 1 μg PX459-ST3Gal-III, 1 μg PX459-ST3Gal-IV, 1 μg PX459-ST3Gal-VI, and 1 μg PX459-ST3Gal-II-like) or 1.7 μg of PX459-ST3Gal-V and 1.7 μg of pCAG-Bsd-ST6Gal-I. The resuspended cells were transferred to cuvettes and immediately electroporated using the program A-024. The cells were cultured in the presence of 2 μg/mL puromycin or 10 μg/ml blasticidin in MEM supplemented with 5% NCS to select for transfected cells. Clones were isolated using cloning rings, dissociated using trypsin and EDTA, and expanded. Genomic DNA was isolated using a genome isolation kit (Promega) according to the manufacturer's instructions. The target region was amplified by PCR using primers surrounding each target site, and amplification products were cloned by using a Zero Blunt TOPO PCR Cloning Kit (Invitrogen). At least eight clones were randomly selected for each gene and the isolated plasmids were sequenced. The list of primers used is provided in Table 5.

[0122] Flow Cytometric Analysis.

[0123] Cells were detached by incubation for 10 min in PBS containing 0.125% Trypsin-20 mM EDTA (Dojindo). After being washed with PBS, the cells were blocked with Carbo-Free Blocking Solution (Vector) at 4° C. for 15 minutes. The cells were incubated with either biotinylated MAL II, SNA, or SSA at 4° C. for 30 minutes. The cells were then rinsed with PBS before being incubated with Alexa 488-conjugated streptavidin for 30 minutes at 4° C. (Invitrogen). Fluorescence was measured using a FACS Calibur or a FACS Verse (Becton Dickinson) and analyzed using FlowJo software (Becton Dickinson).

[0124] To confirm sialic acid-specific lectin binding, cells were treated, before incubation with lectin, with *Clostridium perflingense* (Roche) for 1 h at 37° C. Lectins bound to cells were detected as described above.

[0125] Immunofluorescence Staining.

[0126] Cells grown in 24-well plates were incubated with a mouse monoclonal antibody, which recognizes Siaα2,

3Galβ1,4GlcNAc (HYB4: Wako) at 4° C. After incubation, the cells were fixed with 10% trichloroacetic acid for 10 minutes at −20° C. Cells were then washed with PBS and incubated for 30 minutes with Alexa 488-conjugated goat anti-mouse immunoglobulin G (IgG) (Invitrogen). Cell nuclei were counterstained with Hoechst 33342, trihydrochloride, trihydrate (Molecular Probes). The samples were examined by using Zeiss fluorescence microscopy (model Imager Z1; Carl Zeiss).

TABLE A

Sequence analysis of the CRISPR/Cas9 target sites in hCK cells[a].

α2,3-sialyltransferase gene	Mutation type
ST3Gal-I	2 nucleotide deletion
ST3Gal-II	1 nucleotide insertion
ST3Gal-III	1 nucleotide deletion, 1 nucleotide insertion
ST3Gal-IV	1 nucleotide deletion, 236 nucleotide insertion
ST3Gal-V	8 nucleotide deletion, 1 nucleotide deletion, 2 nucleotide deletion, 1 nucleotide insertion
ST3Gal-VI	1 nucleotide deletion
ST3Gal-II like	1 nucleotide deletion

[a]PCR products of each gene were cloned into blunt-end vectors and subjected to sequencing analysis.

TABLE 2

Isolation of human influenza viruses from clinical specimens[a].

Virus type	Total number of specimens	Number of virus isolates recovered (isolation efficiency)[b] MDCK cells	AX4 cells	hCK cells
A/H1N1pdm	30	30 (100%)	30 (100%)	30 (100%)
A/H3N2	30	25 (83%)	28 (93%)	30 (100%)
B	30	30 (100%)	30 (100%)	30 (100%)

[a]Clinical specimens shown to be influenza virus-positive by real-time RT-PCR or rapid diagnostic kits were used for virus isolation.
[b]Clinical specimens were inoculated into MDCK, AX4, and hCK cells. Cells were observed for the development of cytopathic effect (CPE) for 7 days. Supernatants from CPE-negative cell culture samples were tested by using rapid diagnostic kits and hemagglutination assays with guinea pig red blood cells at 7 days after inoculation.

TABLE 3

Sample ID	Virus	Cell	Amino acid substitutions HA	NA
I-1202	A/Yokohama/146/2017	MDCK	—b	D399G
		AX4	—	—
		hCK	—	—
I-1205	A/Yokohama/147/2017	MDCK	—	D151N
		AX4	—	—
		hCK	—	—

TABLE B

Mutations in α2,3-sialyltransferase genes caused by each gRNA

Target gene	Sequence
ST3Gal-I	CCTCCT**TCTTCCTGAATTACTCCCAC**ACC
	CCTCCTT--TTCTGAATTACTCCCACACC
	CCTCCT---TTCCTGAATTACTCCCACACC(SEQ ID NO: 20)
ST3Gal-II	CTT**TACCTACTCCCACCACAGCA**TGGCCA
	CTTTACCTACTCCCACCACAAGCATGGCCA(SEQ ID NO: 21)
ST3Gal-III	CTCCCC**CGCGGCTGTGGCGGCCGCCC**GCG
	CTCCCCCGCGGGCTGTGGCGGCCGCCCGCG
	CTCCCCCGC-GCTGTGGCGGCCGCCCGCG(SEQ ID NO: 22)
ST3Gal-IV	TCCCCA**GGAACCCTGCTACCCTCTGA**GAA
	TCCCCAGGA-CCCTGCTACCCTCTGAGAA
	TCCCCAGGACGATGGTGGTGTCAAAGTACTTGAAGGCGGCAGGGGCT
	CCCAGATTGGTCAGGGTAAACAGGTGGATGATATTCTCGGCCTGCTCT
	CTGATGGGCTTATCCCGGTGCTTGTTGTAGGCGGACAGCACTTTGTCC
	AGATTAGCGTCGGCCAGGATCACTCTCTTGGAGAACTCGCTGATCTGC
	TCGATGATCTCGTCCAGGTAGTGCTTGTGCTGTTCCACAAACAGCTGTTTC
	ACCCTGCTACCCTCTGAGAA (SEQ ID NO: 23)
ST3Gal-V	ATC**GCTCAAGACCCTCTCTGCAA**TGGTAC
	ATCGCTCAAGACCCTCTCTG--ATGGTAC
	ATCGCTCAAGAC----CAATGGTAC
	ATCGCTCAAGACCCTCTCTGGCAATGGTAC
	ATCGCTCAAGACCCTCTCT-CAATGGTAC (SEQ ID NO: 24)
ST3Gal-VI	AGC**GATAAGTTTGATCTGCCCTA**TGGGATA
	AGCGATAAGTTTGATCTGCC-TATGGGATA(SEQ ID NO: 25)
ST3Gal-II-like	CGCCCC**GTCCAAGGATTTGAGATGGA**TGT
	GCCCCCGTC-AAGGATTTGAGATGGATGT(SEQ ID NO: 26)

The sequence of sgRNA is shown by bold letters. The underlined sequence shows the PAM sequence.
[1]The sequence matched some sequence ot the PX459 vector.

TABLE 3-continued

Sample ID	Virus	Cell	Amino acid substitutions HA	NA
I-1218	A/Yokohama/160/2017	MDCK	—	—
		AX4	—	—
		hCK	—	—
I-1221	A/Yokohama/181/2017	MDCK	—	D151D/G[c]
		AX4	—	—
		hCK	—	—
P-9211	A/Yokohama/199/2017	MDCK	N158K	R394K/R[d]
		AX4	—	—
		hCK	—	—
I-1250	A/Yokohama/240/2017	MDCK	—	—
		AX4	—	—
		hCK	—	—
I-1244	A/Yokohama/1/2018	MDCK	—	—
		AX4	—	—
		hCK	—	—
I-1248	A/Yokohama/8/2018	MDCK	—	D151D/N[e]
		AX4	—	—
		hCK	—	—
P-9256	A/Yokohama/10/2018	MDCK	P221S/P[f], N246D/N[g]	—
		AX4	—	—
		hCK	—	—
P-9265	A/Yokohama/14/2018	MDCK	T248I	—
		AX4	—	D151D/N[e]
		hCK	—	—
P-9279	A/Yokohama/15/2018	MDCK	T160K/I[h]	—
		AX4	—	—
		hCK	—	—
P-9281	A/Yokohama/16/2018	MDCK	—	—
		AX4	—	—
		hCK	—	—
P-9288	A/Yokohama/20/2018	MDCK	T160K/T[i]	D151D/N[e]
		AX4	—	—
		hCK	—	—
P-9291	A/Yokohama/21/2018	MDCK	—	D151D/N[e]
		AX4	—	—
		hCK	—	—
P-9301	A/Yokohama/29/2018	MDCK	—	—
		AX4	—	T148K/T[j], D151D/N[e]
		hCK	—	—
I-1271	A/Yokohama/33/2018[k]	hCK	—	—
I-1275	A/Yokohama/32/2018	MDCK	—	—
		AX4	—	—
		hCK	—	—
P-9307	A/Yokohama/34/2018	MDCK	—	—
		AX4	—	—
		hCK	—	—
P-9315	A/Yokohama/36/2018[l]	AX4	—	—
		hCK	—	—
I-1279	A/Yokohama/37/2018[l]	AX4	—	—
		hCK	—	—
I-1280	A/Yokohama/38/2018[l]	AX4	—	—
		hCK	—	—
P-9328	A/Yokohama/40/2018	MDCK	—	—
		AX4	—	T148T/I[m]
		hCK	—	—
I-1288	A/Yokohama/41/2018	MDCK	—	—
		AX4	—	—
		hCK	—	—
P-9330	A/Yokohama/43/2018	MDCK	—	D151D/G[c]
		AX4	—	E433E/K[n]
		hCK	—	—
P-9333	A/Yokohama/44/2018	MDCK	—	D151D/N[e]
		AX4	—	D151D/N[e]
		hCK	—	—
P-9334	A/Yokohama/45/2018	MDCK	—	—
		AX4	—	D151D/G[c]
		hCK	—	—
P-9352	A/Yokohama/48/2018	MDCK	—	—
		AX4	—	—
		hCK	—	—
P-9356	A/Yokohama/49/2018[k]	hCK	—	S44S/P[o]

TABLE 3-continued

Sample ID	Virus	Cell	Amino acid substitutions HA	NA
I-1295	A/Yokohama/50/2018	MDCK	T160K	—
		AX4	—	D151D/N[e]
		hCK	—	—
I-1296	A/Yokohama/51/2018	MDCK	N158K	—
		AX4	—	—
		hCK	—	—

[a]Influenza viruses isolated from the clinical specimens in MDCK, AX4, or hCK cells. The sequences of the HA and NA genes of the viruses were determined after isolation in MDCK, AX4, or hCK cells.

b—, No mutation was detected compared to the sequences from the original clinical specimens.

[c]D/G, mixture of aspartic acid and glycine at position 151.

[d]K/R, mixture of Lysine and arginine at position 394.

[e]D/N, mixture of aspartic acid and asparagine at position 151.

[f]S/P, mixture of serine and proline at position 221.

[g]D/N, mixture of aspartic acid and asparagine at position 246.

[h]K/I, mixture of lysine and isoleucine at position 160.

[i]K/T, mixture of lysine and threonine at position 160.

[j]K/T, mixture of lysine and threonine at position 148.

[k]Influenza viruses were not isolated from the clinical specimens in MDCK and AX4 cells.

[l]Influenza viruses were not isolated from the clinical specimens in MDCK cells.

[m]T/I, mixture of threonine and isoleucine at position 148.

[n]E/K, mixture of glutamic acid and lysine at position 433.

[o]S/P, mixture of serine and proline at position 44.

TABLE 4

Comparison of the sensitivity of hCK and AX4 cells to human influenza viruses[a].

Virus type	Influenza season	Sample ID	Highest dilution of clinical sample showing CPE observed in[b] AX4 cells	hCK cells	Ration (hCK highest dilution/AX4 highest dilution
A/H1N1pdm	2017-18	BB139	8192	32768	4
		UT001-1	16	512	32
		HP79	65536	16384	0.25
	2016-17	P-8848	8192	8192	1
		BB131	524288	524288	1
	2013-14	IMS1	8192	16384	2
A/H3N2	2017-18	DA29-1	<2	64	>64
		DA30	<2	2048	>2048
		HP62	32	4096	128
	2016-17	DA23-1	16	2048	128
		DA16-2	<2	16	>16
		DA19-2	256	2048	8
B/Yamagata	2017-18	BB140	16384	16384	1
		HP70-2	512	4096	8
		BB152	262144	262144	1
	2016-17	BB005	128	256	2
	2015-16	DA09-2	65536	524288	8
		DA07-2	64	64	1
B/Victoria	2017-18	WD28	16384	131072	8
	2016-17	DA25-2	8192	32768	4
		BB078	524288	524288	1
		BB130	4096	4096	1
	2015-16	HP015	2048	16384	8
		HP009	262144	262144	1

[a]Clinical specimens shown to be influenza virus-positive by real-time RT-PCR were used for virus isolation.

[b]Serial 2-fold dilutions (2^1 to 2^{20}) of clinical samples were prepared and inoculated into AX4 and hCK cells. Cells were observed for the development of CPE for 7 days. Three wells were used to infect with the same dilutions of virus, and the highest dilution showing CPE in all three wells is shown.

TABLE 5

List of primers used.

Primer or probe	Target gene	Sequence (5'-3')[a]	Orientation
ST3Gal-I-F	Canis lupus familiaris ST3Gal-I	CCCTCCTCGTCCTCTTCATC (SEQ ID NO: 27)	Forward
ST3Gal-I-R	Canis lupus familiaris ST3Gal-I	AGGCAGAGAGAGACCAGAGA (SEQ ID NO: 28)	Reverse
ST3Gal-II-F	Canis lupus familiaris ST3Gal-II	CCAAACCATGAAGTGCTCCC (SEQ ID NO: 29)	Forward
ST3Gal-II-R	Canis lupus familiaris ST3Gal-II	AGGGGCTTGAAGAGTGACTC (SEQ ID NO: 30)	Reverse
ST3Gal-III-F	Canis lupus familiaris ST3Gal-III	ATGAGACTTGCTTGCATCCC (SEQ ID NO: 31)	Forward
ST3Gal-III-R	Canis lupus familiaris ST3Gal-III	CTTTGGTTGGCCTCTCTGTCTC (SEQ ID NO: 32)	Reverse
ST3Gal-III-seq-R	Canis lupus familiaris ST3Gal-III	CGTTAGCCGCGCGCACAG (SEQ ID NO: 33)	Reverse
ST3Gal-IV-F	Canis lupus familiaris ST3Gal-IV	CCGGGATGACAGCTCTC (SEQ ID NO: 34)	Forward
ST3Gal-IV-R	Canis lupus familiaris ST3Gal-IV	ACATGGAAGCTGGACTCAC(SEQ ID NO: 35)	Reverse
ST3Gal-V-F	Canis lupus familiaris ST3Gal-V	CATCATCACAAGGATCCTGC (SEQ ID NO: 36)	Forward
ST3Gal-V-R	Canis lupus familiaris ST3Gal-V	CTCTCCCATGAAAACCTGG (SEQ ID NO: 37)	Reverse
ST3Gal-VI-F	Canis lupus familiaris ST3Gal-VI	GTTTTAAATTTGGGAGCGGCC (SEQ ID NO: 38)	Forward
ST3Gal-VI-R	Canis lupus familiaris ST3Gal-VI	TGGCTCACATCAAACACCAC (SEQ ID NO: 39)	Reverse
ST3Gal-II-like-F	Canis lupus familiaris ST3Gal-II-like	GGTTGGAAACTCAAGGTGCC (SEQ ID NO: 40)	Forward
ST3Gal-II-like-R	Canis lupus familiaris ST3Gal-II-like	TGACTCCTTCCCCTTTTCCC (SEQ ID NO: 41)	Reverse
RT/PCR-A/H1N1 pdm-HA-F	A/H1N1pdm virus HA	GTTACGCGCCAGCAAAAGCAGGG GAAAACAAAAGCAA (SEQ ID NO:42)	Forward
RT/PCR-A/H1N1 pdm-HA-R	A/H1N1pdm virus HA	GTTACGCGCCAGTAGAAACAAGGG TGTTTTTCTCATGC (SEQ ID NO: 43)	Reverse

TABLE 5-continued

List of primers used.

Primer or probe	Target gene	Sequence (5'-3')[a]	Orientation
RT/PCR-A/H1N1 pdm-NA-F	A/H1N1pdm virus NA	GTTACGCGCCAGCAAAAGCAGGA GTTTAAAAT (SEQ ID NO: 44)	Forward
RT/PCR-A/H1N1 pdm-NA-R	A/H1N1pdm virus NA	GTTACGCGCCAGTAGAAACAAGGA GTTTTTTGAACAAC (SEQ ID NO: 45)	Reverse
RT/PCR-A/H3N2-HA-F	A/H3N2 virus HA	GTTACGCGCCAGCAAAAGCAGGG GATAATTCTATTAA (SEQ ID NO: 46)	Forward
RT/PCR-A/H3N2-HA-R	A/H3N2 virus HA	GTTACGCGCCAGTAGAAACAAGGG TGTTTTTTAATTAATG (SEQ ID NO: 47)	Reverse
RT/PCR-A/H3N2-NA-F	A/H3N2 virus NA	GTTACGCGCCAGCAAAAGCAGGA GTAAAGATG (SEQ ID NO: 48)	Forward
RT/PCR-A/H3N2-NA-R	A/H3N2 virus NA	GTTACGCGCCAGTAGAAACAAGGA GTTTTTTCTAAAATTGC(SEQ ID NO: 49)	Reverse
RT/PCR-IBV-HA-F	Influenza B virus HA	GTTACGCGCCAGCAGAAGCAGAGC ATTTTCTAATATCC(SEQ ID NO: 50)	Forward
RT/PCR-IBV-HA-R	Influenza B virus HA	GTTACGCGCCAGTAGTAACAAGAG CATTTTTCAATAACGTTTC (SEQ ID NO: 51)	Reverse
RT/PCR-IBV-NA-F	Influenza B virus NA	GTTACGCGCCAGCAGAAGCAGAGC ATCTTCTCAAAACTG (SEQ ID NO: 52)	Forward
RT/PCR-IBV-NA-R	Influenza B virus NA	GTTACGCGCCAGTAGTAACAAGAG CATTTTTCAGAAAC (SEQ ID NO: 53)	Reverse
qPCR-A/H1N1 pdm-F	A/H1N1pdm virus HA	AGAAAAGAATGTAACAGTAACAC ACTCTGT (SEQ ID NO: 54)	Forward
qPCR-A/H1N1 pdm-R	A/H1 N1pdm virus HA	TGTTTC CAC AATGTARGAC CAT (SEQ ID NO: 55)	Reverse
qPCR-A/H3N2-F	A/H3N2 virus HA	CTATTGGACAATAGTAAAACCGGG RGA (SEQ ID NO: 56)	Forward
qPCR-A/H3N2-R	A/H3N2 virus HA	GTCATTGGGRATGCTTCCATTTGG (SEQ ID NO: 57)	Reverse
qPCR-B/Victoria-HA-F	B/Victoria virus HA	CCTGTTACATCTGGGTGCTTTCCTA TAATG (SEQ ID NO: 59)	Forward
qPCR-B/Victoria-HA-R	B/Victoria virus HA	GTTGATARCCTGATATGTTCGTATC CTCKG (SEQ ID NO: 60)	Reverse
qPCR-B/Yamagata-HA-F	B/Yamagata virus HA	CCTGTTACATCCGGGTGCTTYCCTA TAATG (SEQ ID NO: 61)	Forward

TABLE 5-continued

List of primers used.

Primer or probe	Target gene	Sequence (5'-3')[a]	Orientation
qPCR-B/Yamagata-HA-R	B/Yamagata virus HA	GTTGATAACCTKATIVITTTTCATAT CCTCTG (SEQ ID NO: 62)	Reverse
MP-39-67 For	Type A virus M	CCMAGGTCGAAACGTAYGTTCTCT CTATC(SEQ ID NO: 63)	Forward
MP-183-153 Rev	Type A virus M	TGACAGRATYGGTCTTGTCTTTAG CCAYTCCA (SEQ ID NO: 64)	Reverse
NIID-TypeB TM Primer-F1	Type B virus NS	GGAGCAACCAATGCCAC (SEQ ID NO: 65)	Forward
NIID-TypeB TM Primer-R1	Type B virus NS	GTKTAGGCGGTCTTGACCAG (SEQ ID NO: 66)	Reverse
FAM-A/H1N1pdm-HA-Probe	A/H1N1pdm virus HA	(FAM)CAGCCAGCAATRTTRCATTT ACC(BHQ-1) (SEQ ID NO: 67)	
NIID-swH1 Probe2	A/H1N1pdm virus HA	(FAM)CAGCCAGCAATRTTRCATTT ACC(MGB/TAMRA) (SEQ ID NO: 68)	
HEX-A/H3N2-HA-Probe	A/H3N2 virus HA	(HEX)AAGTAACCCCKAGGAGCAAT TAG(BHQ-1) (SEQ ID NO: 69)	
NIID-H3 Probe1	A/H3N2 virus HA	(FAM)AAGTAACCCCKAGGAGCAA TTAG(NIGB/TAMRA) (SEQ ID NO: 70)	
FAM-B/Victoria-HA-Probe	B/Victoria virus HA	(FAM)TTAGACAGCTGCCTAACC(B HQ-1) (SEQ ID NO: 71)	
FAM-Type B HA	Victoria B/Victoria virus HA	(FAM)TTAGACAGCTGCCTAACC(M GB/TAMRA) (SEQ ID NO: 72)	
HEX-B/Yamagata-HA-Probe	B/Yamagata virus HA	(HEX)TCAGGCAACTASCCAATC(BH Q-1) (SEQ ID NO: 73)	
FAM-Type B HA	Yamagata B/Yamagata virus HA	(FAM)TCAGGCAACTASCCAATC(M GB/TAMRA) (SEQ ID NO: 74)	
MP-96-75 Probe	As Type A virus M	(FAM)ATYTCGGCTTTGAGGGGGCC TG(MGB/TAMRA) (SEQ ID NO: 75)	
NIID-TypeB Probe1	Type B virus NS	(FAM)ATAAACTTTGAAGCAGGAAT (MGB/TAMRA) (SEQ ID NO: 76)	

[a]FAM, 6-carboxyfluorescein;
HEX, hexacholoro-6-carboxyfluorescein;
BHQ-1, black hole quencher;
MGB, minor groove binder;
TAMRA, 6-carboxytetramethylrhodamine.

Results

[0127] A new MDCK cell line (designated hCK) was prepared that overexpresses α2,6-sialoglycans and expresses extremely low levels of α2,3-sialoglycans to mimic the sialic acid expression pattern of human upper respiratory epithelial cells (see FIGS. **2-4** and Table A).

[0128] To determine whether hCK cells could support efficient replication of human influenza viruses, the growth kinetics of viruses [3 A/H1N1 2009 pandemic (A/H1N1pdm), 3 A/H3N2, 3 B/Yamagatalineage, and 3 B/Victoria-lineage] were examined in hCK cells. The three A/H1N1pdm isolates grew efficiently in MDCK, AX4, and hCK cells, and no substantial differences in titers were observed (FIG. **1**A). The six influenza B isolates also replicated with similar efficiency in all three cell lines. By contrast, for A/H3N2 viruses, all three isolates grew much faster and to higher titers (2.03 to 2.91 log units higher at 48 h post-infection) in hCK cells than in AX4 cells. As reported elsewhere (Chambers et al., 2014), in MDCK cells, these recent A/H3N2 isolates replicate poorly. These findings demonstrate that hCK cells, which express very low levels of α2,3-sialoglycans and high levels of α2,6-sialoglycans, more efficiently support the replication of recent A/H3N2 viruses than do either MDCK or AX4 cells.

[0129] To evaluate the susceptibility of hCK cells for isolation of human influenza viruses, aliquots of 90 respiratory specimens (30 A/H1N1pdm, 30 A/H3N2, and 30 B/Yamagata-lineage) were inoculated into MDCK, AX4, and hCK cells. The cells were observed for the development of cytopathic effect (CPE) for 7 days. For MDCK, AX4, and hCK cells, A/H1 N1pdm viruses were successfully recovered from all of the RT-PCR-positive samples without the need for blind passages (100% isolation efficiency) (Table 2). Similarly, these three cell lines showed 100% efficiency for the isolation of influenza B viruses. For the A/H3N2-positive samples, 5 and 2 viruses were not recovered from MDCK and AX4 cells, respectively. These results are consistent with previous reports (Oh et al., 2008; Hatakeyama et al., 2005) that conventional MDCK cells have relatively low sensitivity for the detection of recent A/H3N2 viruses.

[0130] The agglutination of red blood cells by influenza viruses is thought to be due to the virus binding to sialic acids on the surface of the cell. Since 2005, A/H3N2 isolates have lost their ability to agglutinate turkey red blood cells (Lin et al., 2013). In addition, current A/H3N2 isolates show reduced or no agglutination of guinea pig red blood cells (Lin et al., Influenza Other Respir Viruses, 2017), indicating a change in their avidity for sialic acid receptors. Indeed, Lin et al. (2013) measured the avidity of recent A/H3N2 viruses for α2,6-linked sialic acid receptors and showed that it has decreased drastically. Glycan array analysis has revealed that recent A/H3N2 isolates prefer binding to branched sialylated N-linked glycans with extended poly-N acetyllactosamine chains (Peng et al., Cell Host Microbe., 2017).

[0131] By contrast, virus isolation from hCK cells was successful with all samples without any subsequent blind passage, suggesting that this cell line is more effective than AX4 or MDCK cells for the isolation of human A/H3N2 viruses from clinical specimens.

[0132] During replication of recent A/H3N2 human isolates in MDCK cells, the viruses rapidly acquired amino acid changes at positions 148 and 151 of the NA protein (e.g., T148I and D151G), which affect the biological properties of NA. To examine whether the A/H3N2 viruses isolated from the three cell lines possessed mutations in their HA and NA proteins, the nucleotide sequences of the HA and NA segments of the isolates were determined by means of Sanger sequencing (Table 3). Sequence analysis revealed that 7 out of 25 MDCK-grown isolates contained an amino acid change at position 151 of NA compared with the sequence from the original specimens: NA-151N, NA-151D/G, and NA-151D/N (mixed populations of amino acids at position 151). Amino acid changes leading to the loss of the glycosylation site at position 158 of HA were found among virus populations of some other MDCK-grown isolates: HA-158K, HA-160K, HA-160K/I, and HA-160K/T. These changes are known to alter the antigenic properties of HA (Lin et al., 2017; Chambers et al., 2015; Skowronski et al., 2016). Importantly, cell culture-adaptive mutations were also found in the NA protein of several isolates propagated in AX4 cells: NA-148K/T, NA-148T/I, NA-151D/N, and NA-151D/G. Strikingly, no mutations were detected in hCK-grown isolates, except for only one isolate that possessed an S44P mutation in its NA stalk. These findings strongly suggest that hCK cells support the efficient growth of A/H3N2 viruses without accompanying cell culture-adaptive mutations.

[0133] Seasonal influenza viruses from clinical specimens grow better in AX4 cells than in MDCK cells (Hatakeyama et al., 2005). To determine whether hCK cells are superior to AX4 cells for virus isolation, the sensitivity of hCK and AX4 cells were compared by testing serial 2-fold dilutions of specimens. Aliquots of 24 specimens (6 A/H1N1pdm, 6 A/H3N2, 6 B/Yamagata-lineage, and 6 B/Victoria-lineage) were inoculated into AX4 and hCK cells in triplicate. All culture wells were examined for CPE on day 7 post-inoculation, and the ratios of the highest dilutions showing CPE observed in hCK cells to those in AX4 cells were determined. For one of the six A/H1N1pdm-positive samples (sample ID, HP79), hCK cells were slightly less sensitive than AX4 cells (FIG. **1**B and Table 4). For the remaining samples, however, the sensitivity of hCK cells was similar to or greater than that of AX4 cells. For the B/Yamagata- and B/Victoria-lineage-positive samples, hCK cells showed sensitivities equal to or somewhat greater than that of AX4 cells. For all of the A/H3N2-positive samples, hCK cells showed greater sensitivity than AX4 cells; for some samples, hCK cells were approximately 100- to 2,000-fold more sensitive than AX4 cells. Taken together, these results indicate that hCK cells are more suitable than AX4 or MDCK cells for the primary isolation of recent seasonal A/H3N2 viruses.

[0134] To evaluate the genetic stability of the HA and NA genes of viruses isolated in hCK cells, aliquots of 12 clinical specimens (3 A/H1N1pdm, 3 A/H3N2, and 3 B/Yamagata-lineage, and 3 B/Victoria-lineage) were inoculated into MDCK, AX4, and hCK cells, and the isolates were sequentially passaged ten times. After the first, sixth, and tenth passages, the HA and NA sequences of the viruses were determined by Sanger sequencing, and the sequences were compared to those in the clinical specimens. For A/H1N1pdm-positive specimens, a mixed viral population encoding either N or S at position 296 of HA was detected in one out of the three hCK-grown viruses (BB139) after the first passage (Table 1). The hCK-grown virus also possessed an S153G substitution mutation in its NA. Another hCK-grown virus (HP79) encoded a D27N substitution in its HA after the tenth passage. A mixed population encoding either

T or I at position 167 of HA was found in one MDCK-grown virus after the tenth passage (BB139). Another MDCK-grown virus (BB131) had a mixed population encoding HA-446N and HA-446S at passage ten. The MDCK-grown virus also contained a mixed population encoding either H or Y at position 411c of NA. A mixture of C53Y/C in NA was observed in one AX4-grown virus after the tenth passage (BB131).

[0135] For A/H3N2-positive samples, viruses that were recovered from AX4 and hCK, but not MDCK, cells were serially passaged. After the sixth passage, a mixed population encoding HA-158N and HA-158K (leading to the loss of the glycosylation site at position 158 of HA) was detected in one of the three AX4-grown viruses (DA30). In addition, another AX4-grown virus (DA29-1) encoded a T148K substitution in its NA after the sixth passage. A mixed population encoding HA-408D and HA-408N was detected in one hCK-grown virus after the tenth passage (DA29-1).

[0136] For B/Yamagata-lineage viruses, no changes were detected in any isolates after the first, sixth, or tenth passages, with the exception of a mixed population encoding HA-148S and HA-148N detected in one hCK-grown virus at passage six (HP70-2). For B/Victoria lineage viruses, a mixed population encoding NA-208G and NA-208R was found in one of the three AX4-grown viruses (BB139) after the first passage. After the sixth passage, one hCK-grown viruses encoded a mixture of L72L/F in its NA (WD28). At passage ten, one MDCK-grown virus (WD28) contained an N196S mutation known to lead to the loss of the glycosylation site at position 196 of HA (B/Victoria-lineage), which can significantly alter the antigenicity of influenza B viruses. Another MDCK-grown virus (HP015) had a mixture of D459D/N in its NA.

[0137] Overall, A/H1N1pdm and B viruses were slightly more variable when passaged in MDCK or hCK cells than in AX4 cells. In contrast, A/H3N2 viruses propagated in hCK cells maintained higher genetic stability than those in AX4 cells.

[0138] In conclusion, a cell line derived from MDCK cells, hCK, expresses large amounts of α2,6-sialoglycans and small amounts of α2,3-sialoglycans that will be useful for influenza virus research, particularly studies involving human A/H3N2 influenza viruses and possibly for vaccine production.

REFERENCES

[0139] Bates et al., *J. Stat. Softw.,* 61:1 (2015).

[0140] Chambers et al., *Cell Rep.,* 12:1 (2015).

[0141] Chambers et al., *J. Viral,* 88:10986 (2014).

[0142] Chu and Whittaker, *Proc. Natl. Acad. Sci. U.S.A.,* 101:18153 (2004).

[0143] Cong et al., *Science,* 339:819 (2013).

[0144] Connor et al., *Virology,* 205:17 (1994).

[0145] Gamblin and Skehel, *J. Biol. Chem.,* 285:28403 (2010).

[0146] Han et al., *Cell Rep.,* 23:596 (2018).

[0147] Hatakeyama et al., *J. Clin. Microbiol.,* 43:4139 (2005).

[0148] Hegde, *Hum. Vaccine Immunother.,* 11:1223 (2015).

[0149] Hidari et al., *Biochem. Biophys. Res. Commun.,* 436:394 (2013).

[0150] Jinek et al., *Science,* 337:816 (2012).

[0151] Lee et al., *PLoS One* 8 (2013).

[0152] Lenth, *J. Stat. Softw.,* 69 (2016) Li et al., *J. Clin. Microbial.,* 47:466 (2009).

[0153] Lin et al., *Influenza Other Respir Viruses,* 11:263 (2017).

[0154] Lin et al., *J. Viral.,* 84:6769 (2010).

[0155] Lin et al., *PLoS One,* 12, 72299 (2017).

[0156] Matrosovich et al., *Journal of Virology,* 77:8418 (2003).

[0157] Mohr et al., *Viral J.,* 12:67 (2015).

[0158] Oh et al., *J. Clin. Microbial.,* 46:2189 (2008).

[0159] Rogers and Paulson, *Virology,* 127:361 (1983).

[0160] Shalem et al., *Science.* 343:84 (2014).

[0161] Shibuya et al., *J Biochem.,* 106:1098 (1989).

[0162] Shinya et al., *Nature,* 440:435 (2006).

[0163] Skowronski et al., *Euro. Surveill.,* 21:30112 (2016).

[0164] Stevens et al., *J. Mal. Biol.,* 355:1143 (2006).

[0165] Takashima and Tsuji, *Trends in Glycoscience and Glycotechnology,* 23:178 (2011).

[0166] Tamura et al., *Antimicrob. Agents Chemother.,* 57:6141 (2013).

[0167] van Riel et al., *Science,* 312:399 (2006).

[0168] Zhu et al., *J. Viral,* 86:13371 (2012).

[0169] All publications, patents and patent applications are incorporated herein by reference. While in the foregoing specification this invention has been described in relation to certain preferred embodiments thereof, and many details have been set forth for purposes of illustration, it will be apparent to those skilled in the art that the invention is susceptible to additional embodiments and that certain of the details described herein may be varied considerably without departing from the basic principles of the invention.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 151

<210> SEQ ID NO 1
<211> LENGTH: 406
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 1

```
Met Ile His Thr Asn Leu Lys Lys Lys Phe Ser Cys Cys Val Leu Val
1               5                   10                  15

Phe Leu Leu Phe Ala Val Ile Cys Val Trp Lys Glu Lys Lys Lys Gly
            20                  25                  30
```

-continued

```
Ser Tyr Tyr Asp Ser Phe Lys Leu Gln Thr Lys Glu Phe Gln Val Leu
        35                  40                  45

Lys Ser Leu Gly Lys Leu Ala Met Gly Ser Asp Ser Gln Ser Val Ser
    50                  55                  60

Ser Ser Ser Thr Gln Asp Pro His Arg Gly Arg Gln Thr Leu Gly Ser
65                  70                  75                  80

Leu Arg Gly Leu Ala Lys Ala Lys Pro Glu Ala Ser Phe Gln Val Trp
                85                  90                  95

Asn Lys Asp Ser Ser Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile
            100                 105                 110

Trp Lys Asn Tyr Leu Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly
        115                 120                 125

Pro Gly Pro Gly Ile Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu
    130                 135                 140

Arg Asp His Val Asn Val Ser Met Val Glu Val Thr Asp Phe Pro Phe
145                 150                 155                 160

Asn Thr Ser Glu Trp Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr
                165                 170                 175

Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser
            180                 185                 190

Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val
        195                 200                 205

Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly
    210                 215                 220

Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu
225                 230                 235                 240

Lys Arg Phe Leu Lys Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val
                245                 250                 255

Trp Asp Pro Ser Val Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn
            260                 265                 270

Pro Asp Tyr Asn Phe Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His
        275                 280                 285

Pro Asn Gln Pro Phe Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu
    290                 295                 300

Trp Asp Ile Leu Gln Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro
305                 310                 315                 320

Pro Ser Ser Gly Met Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp
                325                 330                 335

Gln Val Asp Ile Tyr Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val
            340                 345                 350

Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala
        355                 360                 365

Tyr His Pro Leu Leu Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln
    370                 375                 380

Gly Thr Asp Glu Asp Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly
385                 390                 395                 400

Phe Arg Thr Ile His Cys
                405
```

<210> SEQ ID NO 2
<211> LENGTH: 529

-continued

<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 2

Met Lys Pro His Leu Lys Gln Trp Arg Gln Arg Met Leu Phe Gly Ile
1 5 10 15

Phe Ala Trp Gly Leu Leu Phe Leu Leu Ile Phe Ile Tyr Phe Thr Asp
            20 25 30

Ser Asn Pro Ala Glu Pro Val Pro Ser Ser Leu Ser Phe Leu Glu Thr
        35 40 45

Arg Arg Leu Leu Pro Val Gln Gly Lys Gln Arg Ala Ile Met Gly Ala
    50 55 60

Ala His Glu Pro Ser Pro Pro Gly Gly Leu Asp Ala Arg Gln Ala Leu
65 70 75 80

Pro Arg Ala His Pro Ala Gly Ser Phe His Ala Gly Pro Gly Asp Leu
                85 90 95

Gln Lys Trp Ala Gln Ser Gln Asp Gly Phe Glu His Lys Glu Phe Phe
            100 105 110

Ser Ser Gln Val Gly Arg Lys Ser Gln Ser Ala Phe Tyr Pro Glu Asp
        115 120 125

Asp Asp Tyr Phe Phe Ala Ala Gly Gln Pro Gly Trp His Ser His Thr
    130 135 140

Gln Gly Thr Leu Gly Phe Pro Ser Pro Gly Glu Pro Gly Pro Arg Glu
145 150 155 160

Gly Ala Phe Pro Ala Ala Gln Val Gln Arg Arg Arg Val Lys Lys Arg
                165 170 175

His Arg Arg Gln Arg Arg Ser His Val Leu Glu Glu Gly Asp Asp Gly
            180 185 190

Asp Arg Leu Tyr Ser Ser Met Ser Arg Ala Phe Leu Tyr Arg Leu Trp
        195 200 205

Lys Gly Asn Val Ser Ser Lys Met Leu Asn Pro Arg Leu Gln Lys Ala
    210 215 220

Met Lys Asp Tyr Leu Thr Ala Asn Lys His Gly Val Arg Phe Arg Gly
225 230 235 240

Lys Arg Glu Ala Gly Leu Ser Arg Ala Gln Leu Leu Cys Gln Leu Arg
                245 250 255

Ser Arg Ala Arg Val Arg Thr Leu Asp Gly Thr Glu Ala Pro Phe Ser
            260 265 270

Ala Leu Gly Trp Arg Arg Leu Val Pro Ala Val Pro Leu Ser Gln Leu
        275 280 285

His Pro Arg Gly Leu Arg Ser Cys Ala Val Val Met Ser Ala Gly Ala
    290 295 300

Ile Leu Asn Ser Ser Leu Gly Glu Glu Ile Asp Ser His Asp Ala Val
305 310 315 320

Leu Arg Phe Asn Ser Ala Pro Thr Arg Gly Tyr Glu Lys Asp Val Gly
                325 330 335

Asn Lys Thr Thr Ile Arg Ile Ile Asn Ser Gln Ile Leu Thr Asn Pro
            340 345 350

Ser His His Phe Ile Asp Ser Ser Leu Tyr Lys Asp Val Ile Leu Val
        355 360 365

Ala Trp Asp Pro Ala Pro Tyr Ser Ala Asn Leu Asn Leu Trp Tyr Lys
    370 375 380

-continued

```
Lys Pro Asp Tyr Asn Leu Phe Thr Pro Tyr Ile Gln His Arg Gln Arg
385                 390                 395                 400

Asn Pro Asn Gln Pro Phe Tyr Ile Leu His Pro Lys Phe Ile Trp Gln
                405                 410                 415

Leu Trp Asp Ile Ile Gln Glu Asn Thr Lys Glu Lys Ile Gln Pro Asn
            420                 425                 430

Pro Pro Ser Ser Gly Phe Ile Gly Ile Leu Ile Met Met Ser Met Cys
        435                 440                 445

Arg Glu Val His Val Tyr Glu Tyr Ile Pro Ser Val Arg Gln Thr Glu
    450                 455                 460

Leu Cys His Tyr His Glu Leu Tyr Tyr Asp Ala Ala Cys Thr Leu Gly
465                 470                 475                 480

Ala Tyr His Pro Leu Leu Tyr Glu Lys Leu Leu Val Gln Arg Leu Asn
                485                 490                 495

Met Gly Thr Gln Gly Asp Leu His Arg Lys Gly Lys Val Val Leu Pro
            500                 505                 510

Gly Phe Gln Ala Val His Cys Pro Ala Pro Ser Pro Val Ile Pro His
        515                 520                 525

Ser


<210> SEQ ID NO 3
<211> LENGTH: 353
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 3

Ala Ala Met Gly Ser Asp Ser Gln Ser Val Ser Ser Ser Ser Thr Gln
1               5                   10                  15

Asp Pro His Arg Gly Arg Gln Thr Leu Gly Ser Leu Arg Gly Leu Ala
            20                  25                  30

Lys Ala Lys Pro Glu Ala Ser Phe Gln Val Trp Asn Lys Asp Ser Ser
        35                  40                  45

Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile Trp Lys Asn Tyr Leu
    50                  55                  60

Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly Pro Gly Pro Gly Ile
65                  70                  75                  80

Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu Arg Asp His Val Asn
                85                  90                  95

Val Ser Met Val Glu Val Thr Asp Phe Pro Phe Asn Thr Ser Glu Trp
            100                 105                 110

Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr Lys Ala Gly Pro Trp
        115                 120                 125

Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser Leu Lys Ser Ser Gln
    130                 135                 140

Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val Leu Arg Phe Asn Gly
145                 150                 155                 160

Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly Thr Lys Thr Thr Ile
                165                 170                 175

Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu Lys Arg Phe Leu Lys
            180                 185                 190

Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val Trp Asp Pro Ser Val
        195                 200                 205
```

-continued

```
Tyr His Ser Asp Ile Pro Lys Trp Tyr Gln Asn Pro Asp Tyr Asn Phe
    210                 215                 220

Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His Pro Asn Gln Pro Phe
225                 230                 235                 240

Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu Trp Asp Ile Leu Gln
                245                 250                 255

Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro Pro Ser Ser Gly Met
            260                 265                 270

Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp Gln Val Asp Ile Tyr
        275                 280                 285

Glu Phe Leu Pro Ser Lys Arg Lys Thr Asp Val Cys Tyr Tyr Tyr Gln
    290                 295                 300

Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala Tyr His Pro Leu Leu
305                 310                 315                 320

Tyr Glu Lys Asn Leu Val Lys His Leu Asn Gln Gly Thr Asp Glu Asp
                325                 330                 335

Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly Phe Arg Thr Ile His
            340                 345                 350

Cys


<210> SEQ ID NO 4
<211> LENGTH: 529
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 4

Met Lys Pro His Leu Lys Gln Trp Arg Gln Arg Met Leu Phe Gly Ile
1               5                   10                  15

Phe Ala Trp Gly Leu Leu Phe Leu Leu Ile Phe Ile Tyr Phe Thr Asp
            20                  25                  30

Ser Asn Pro Ala Glu Pro Val Pro Ser Ser Leu Ser Phe Leu Glu Thr
        35                  40                  45

Arg Arg Leu Leu Pro Val Gln Gly Lys Gln Arg Ala Ile Met Gly Ala
    50                  55                  60

Ala His Glu Pro Ser Pro Pro Gly Gly Leu Asp Ala Arg Gln Ala Leu
65                  70                  75                  80

Pro Arg Ala His Pro Ala Gly Ser Phe His Ala Gly Pro Gly Asp Leu
                85                  90                  95

Gln Lys Trp Ala Gln Ser Gln Asp Gly Phe Glu His Lys Glu Phe Phe
            100                 105                 110

Ser Ser Gln Val Gly Arg Lys Ser Gln Ser Ala Phe Tyr Pro Glu Asp
        115                 120                 125

Asp Asp Tyr Phe Phe Ala Ala Gly Gln Pro Gly Trp His Ser His Thr
    130                 135                 140

Gln Gly Thr Leu Gly Phe Pro Ser Pro Gly Glu Pro Gly Pro Arg Glu
145                 150                 155                 160

Gly Ala Phe Pro Ala Ala Gln Val Gln Arg Arg Arg Val Lys Lys Arg
                165                 170                 175

His Arg Arg Gln Arg Arg Ser His Val Leu Glu Glu Gly Asp Asp Gly
            180                 185                 190

Asp Arg Leu Tyr Ser Ser Met Ser Arg Ala Phe Leu Tyr Arg Leu Trp
        195                 200                 205
```

-continued

```
Lys Gly Asn Val Ser Ser Lys Met Leu Asn Pro Arg Leu Gln Lys Ala
    210                 215                 220

Met Lys Asp Tyr Leu Thr Ala Asn Lys His Gly Val Arg Phe Arg Gly
225                 230                 235                 240

Lys Arg Glu Ala Gly Leu Ser Arg Ala Gln Leu Leu Cys Gln Leu Arg
                245                 250                 255

Ser Arg Ala Arg Val Arg Thr Leu Asp Gly Thr Glu Ala Pro Phe Ser
            260                 265                 270

Ala Leu Gly Trp Arg Arg Leu Val Pro Ala Val Pro Leu Ser Gln Leu
        275                 280                 285

His Pro Arg Gly Leu Arg Ser Cys Ala Val Val Met Ser Ala Gly Ala
    290                 295                 300

Ile Leu Asn Ser Ser Leu Gly Glu Glu Ile Asp Ser His Asp Ala Val
305                 310                 315                 320

Leu Arg Phe Asn Ser Ala Pro Thr Arg Gly Tyr Glu Lys Asp Val Gly
                325                 330                 335

Asn Lys Thr Thr Ile Arg Ile Ile Asn Ser Gln Ile Leu Thr Asn Pro
            340                 345                 350

Ser His His Phe Ile Asp Ser Ser Leu Tyr Lys Asp Val Ile Leu Val
        355                 360                 365

Ala Trp Asp Pro Ala Pro Tyr Ser Ala Asn Leu Asn Leu Trp Tyr Lys
    370                 375                 380

Lys Pro Asp Tyr Asn Leu Phe Thr Pro Tyr Ile Gln His Arg Gln Arg
385                 390                 395                 400

Asn Pro Asn Gln Pro Phe Tyr Ile Leu His Pro Lys Phe Ile Trp Gln
                405                 410                 415

Leu Trp Asp Ile Ile Gln Glu Asn Thr Lys Glu Lys Ile Gln Pro Asn
            420                 425                 430

Pro Pro Ser Ser Gly Phe Ile Gly Ile Leu Ile Met Met Ser Met Cys
        435                 440                 445

Arg Glu Val His Val Tyr Glu Tyr Ile Pro Ser Val Arg Gln Thr Glu
    450                 455                 460

Leu Cys His Tyr His Glu Leu Tyr Tyr Asp Ala Ala Cys Thr Leu Gly
465                 470                 475                 480

Ala Tyr His Pro Leu Leu Tyr Glu Lys Leu Leu Val Gln Arg Leu Asn
                485                 490                 495

Met Gly Thr Gln Gly Asp Leu His Arg Lys Gly Lys Val Val Leu Pro
            500                 505                 510

Gly Phe Gln Ala Val His Cys Pro Ala Pro Ser Pro Val Ile Pro His
        515                 520                 525

Ser
```

<210> SEQ ID NO 5
<211> LENGTH: 6830
<212> TYPE: DNA
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 5

```
gtcaggtctc agaaaagtct agaataaggc ttacaggcac tttgttcagt tgtggaacac     60

atggaaactt catacaccgc ccccctcctt gcagaccgga gagctctctg ctctaatttg    120

ggcacaggcg cccaaccttg ggccccagag gaagcctgtt ctactcccag ggccaacgct    180
```

-continued

caattgcttg gtaacttggc acccacactg ttccaggctc tggtctcagt ttcactctct 240

gcaaaatgag aggctcagag ttccgctgac atggcctcgg ggccgttaaa acctttcttc 300

atgaagacag cctgcaccct tcctgcttct gctccaggtc catcctcaga accttccaga 360

aagtcctggc acgtcagata acagggccaa cccggcagtg atgccccacc ccccaccct 420

tcccttactc atagagcctg ctccagactc tcagagccca cacccacttg gtgaagtcat 480

ttgccagtaa ttcttcgcac tggacattga gaggtttcag accccagaag tctcaggcgc 540

tgggtctgaa agtgggcaga gcccaggtga catttgtgga gactctcagt ggtgcgtata 600

gccgccggga ccattttcag actcaacctt tctgacctgg aaatgccaat agaagtaatc 660

atcatcgcca ggggctgtgg tgagcaaatg cttggagctc tgtggccagg gcagtttcat 720

ttgaggacca gagatggaca atccctcagc ctactgagat gaagaaactg agtctcagag 780

aggttaagga actcccccga ggttgcacca ctgaggaaga ttgacctgac ttcccaagac 840

catacatctg gtaaaccgga acctgcacct gcgccatctc aagcctactc tggaggcccg 900

aggctaattg gcagagtttg agggctgaga tgacagacaa ccctcagtgc cttcatcggg 960

cgggctgctt acctgcacat ccctggtgac agcatgggaa agaccgcttc taattaagcg 1020

tcatcacaca tcacccttttt ccggaggaag agaggcaaag acagctcccc tctatcctgc 1080

actgtgaacc ttcctccgag gtcctcccct caccccccgag agtccttgcc ctgtcaccaa 1140

gattaattac ccctcaaccc ctttggatgg caaaggcagt cattttatta agtttgatta 1200

aagcttcaag agacattgcc ggatgtttca ggactgctga caaagcagcc tgcttgtttc 1260

ctgaaaagac caattatatg caagaagcgt cagccccact ccccgagggg tccacttagc 1320

ctccgaccac cacagggtga tgtccagcgc accggtgtgg ccatcacctg gcgggagtga 1380

ggtcggccgt agcaggtctg gaggcccccc tgcaagtccc tgcctcaccc tgtgactgag 1440

cctctcgtct ctgggaattt tgttccaaat tccccgtcta ccaggtgtga tttcctccag 1500

ccccaccagc cctgggaggc gcccatccag agagcagaga tggtgaccat gagaaagagg 1560

actctcaaag tgctcaccct cctcgtcctc ttcatcttcc tcacctcctt cttcctgaat 1620

tactcccaca ccatggtgac caccacctgg tttcccaagc agatggttgt cgagctctca 1680

gagaacttta agaagttcat gaaatacact cacaggcctt gcacctgtgc ccgctgcatc 1740

gggcagcaga gggtctcggc ctggttcgat gagaggttca accggtccat gcagccgctg 1800

ctgaccgccc agaacgccct cttggaggag gacacctaca gctggtggct gaggctccag 1860

cgggagaagc aacccaacaa cttgaacgat accatcaggg agctgttcca ggtggtgccc 1920

gggaacgtgg acccccctgct ggagaagagg tcggtgggct gccggcgctg cgcagtcgtg 1980

ggcaactctg gcaacctccg agagtcctgg tacgggcctc agatcgacag ccacgacttc 2040

gtgctcagga tgaacaaggc ccccacggcg ggcttcgaga tggatgtcgg gagcaagacc 2100

acccaccacc tggtgtaccc cgagagcttc agggagctgg cggagaatgt cagcatggtc 2160

ctggtgccct tcaagaccac cgacctggag tgggtggtca gtgccaccac cacaggcacc 2220

atctctcaca cctatgttcc tgttcctgca aagatcaaag tgaaaaagga taagatcctc 2280

atctaccacc cggccttcat caagtacgtc ttcgacagct ggctgcaggg ccacgggcgg 2340

tacccgtcca ccggcatcct ctccgtcatc ttctcgctgc acatctgcga cgaggtggac 2400

ttgtacggct tcggggcaga cagtaagggg aactggcatc actactggga gaacaatcca 2460

-continued

tcggcggggg ctttccgcaa gaccggggtg cacgacggag actttgagtc caacgtgacg 2520

gccaccttgg cgtccatcaa taagatccgg atatttaagg ggagatgacg ctgccgagga 2580

gcaccggagc ccgcctcttt ggccagcccc agcctctgct ggagccgatc tgtgctgggg 2640

gctttgaggg ccagcctcgg gggcgtgttc aggtgcccct cgtgcccccc cgcaccccga 2700

catttggcag catcgactca gcaaggcccc agagccccgc tgggtctgca gagcgagtgt 2760

cagaactggt cttggatggg gacacccccc ctccacctcc ctgcaccgcc actgctgtcc 2820

tggagcaggg ggcaggtggg gagacgcggg aggaggtccg gtctccaaac gctcaagtca 2880

tttcggcttt ggggggcaga gggggagata tcgaagactc tgtgttctaa agtgaaggac 2940

actggccctg ggcagagggg tcccgggctg actctctgac tctgatgctt gctgggtgag 3000

gacgacctcg ggacagagcc acatcgggcc acgtggacgc tgggggtgag aaggcacctt 3060

cctctgcacg atgcccggcc cctccctcac cgcctctgcg gtcttcccga agctcctccg 3120

tggcggccag gagaggcgcc ctgcgccgag ggtcctcaca gatgcttggc caagtgtttc 3180

agactccagc aggtgtcttc ttcgcaagct gaggctccct tgagtgatcg atctttgtgg 3240

ccataaatga tggctgagag cagatctgac tacttccacg tgcctttgtg tctgggggag 3300

agctgtgcgc attggctgaa ataaggcaaa agccttaatt cgggagtggg gagctccccc 3360

ctctccccgc ccccagcaat gccacccccc tgctctggag ctgggtaaca tctttactag 3420

tttcctgagg cggtgccgga gctggaatga agctaggatg atgctcaacg gcgtccagag 3480

gtgcgtcagc cgcggcacgt cattcttagg ttcacagact gatggatttg gggcccgtga 3540

ccttccaagg ccacacacag aacctccacc agccagcact ttgccccagg agcccaccct 3600

ccaggccaag tccctccggg ccacctgcac agtgtcctgg ggtttgtggg gccatcccag 3660

gaacatgacc ctcggagcgg gagaagacat acctcatcct ctcacttctg ggctgctctg 3720

attcctcctc atgattattt attgattttt ttttaattca gctgacattg cgtgcacttt 3780

gtggctccgt ggctggtcct gatgttttaa ttaagcttgc tctcgcttca cctggcagcc 3840

ggggcatggg gggcttaagt caaggtttgc aggagtcctc agacttggga gggggatgca 3900

tatctagggc ttgggggctc gtcggtgggc aaatgccagg gtttcaggtt ggtaggttct 3960

ccgagtgctc tgtacctcct tcccctcagc tctgcctccc ttctccatca ttgcgggttg 4020

ggatcattct caggctaaga atctgcaaaa cccagtgaga cggccttggg gccaggtggg 4080

gagtggagtc acacaggcag ggcagcgatg gttggctccg gttgctgaca ctgaatcaga 4140

aaatccacgt ttcctattga gagcatttcc taacaggcca ctgctacttc gaggaggtgt 4200

gacagtgtcc tggctgtcac gggggcccgt ctcctgtgct ccttgcagca gcactccggc 4260

aagttctgct ctcctggctt cagtctctgc ctctgcaaac ggagagggag gcatttgggg 4320

agctcagggg gccacaggtg gtttcagggg gttcacaagc gtccccagag agagcacacc 4380

acgcccgtgt tttaccccaa gcctgtgtgt atgagcatat attcctgggg agcggtgccg 4440

tagcgttcat tcaattgtca gaggttcaga acccagacaa tggtttacaa aaacacgaaa 4500

cgacagcaag caaacaaaaa cccgttcgag ctttaaaatc ctccaacact tagactctct 4560

tgacccctaa aggctcataa atcactacga ggtcagcaag ctgatcataa atcccaaaga 4620

ctatatagca ggcatctgtg ggttggctga tgaatccgct gagagaagtg ggtgcggagc 4680

ggcctccgag ctcctccccg agtcctggaa gagccgagtc cagtcagagc tgatgggtcc 4740

-continued

agagcgctca gcctgcgtcg gggctgtgct gagtcaggac aggtccccag agtaaggggg 4800

ggggtctctc tccctcccca ccccactgag ccgtccaggc tgttgagagg acttcccacc 4860

tgccttccca tgtaacactt gagaaaagcg aggcccggag aagggaagtg gcttgcaaag 4920

tgaacccgtg gtgctggaca acgcaagtca gacccctttga gcaaagatag agcacgattg 4980

ctacatgttg aacacacctc gaggtgggct ttcccagccg gcttccatcc agcctcccca 5040

aggactgtga ccgggcacag gtgggctcct caggctgcac gtggtggccc ccgcaagctc 5100

cggtgtctgc cttttccccc catccctcca gtgagggaga cacccttgct cttccctcag 5160

ccccagcaaa cttgtctttg tgttcgttgg cttagcaaag gggtctgacc cgaatgcacc 5220

aacgggcttc agccccctgg catctgtacc tggggtaagg gtggctgaga atccaaccag 5280

agtatttccc ccagagctcg tggctgtggg caggttaacg agaaccaccc gccccccccа 5340

ctcggaggag gctcctgaat gcccacccca gtgttgggct cccagcctag ggccctgctt 5400

gctcctcaag cacagagctc ccttggaggc cagtcctgga cccctgggca gggagcagag 5460

gcatccctgc ctctctcagt gctgagcccg aagccaggaa gaccagggcc gaggagggag 5520

agagagattt taacaccctc cccggccctg gggagccagg agctggagtc acctgccgca 5580

aatgcaaggc ctggttccac gtgcccccct ccgcccctgt cctcaccccc ttgttagggt 5640

gggtttgtag cgtctggctc ccagagccat caggcaaaaa agaaagttct taaagagatg 5700

cttttagggg cctgtgctta gtgggagaag ccagattcaa gaaaatggga cttaaaacgg 5760

ggaagccaga cgagaaggtg gggggggggt tgtccgggtg tgattcttac aagtgctgag 5820

gccagaagga tggaccctct ccaggtccag agtagagtgg agtcaacggg gaatatttag 5880

gactgtgatg tttgcagtaa tggtcctctt caaggcagcc tggacacaaa gcaaaatcgg 5940

aaacagaaaa gaactcacca tcccttgcgt gggtaggacc aaaacagctt tctccaggga 6000

ggcgcaagcc ctgagtggcc gccgtctggc caaagccggc ctcgtgccat cgggaaccgg 6060

ccaccgccca tgcccctccg gagagctcgg ggtcgtcatc gtgtccacca cgggcatctg 6120

gtcggaagca aacgttccca caggccactg gctggtcatc gtcccctgtg gtctgggctg 6180

accggtttga gaccaagtcc gctgccttgc tccggcctgt cctgaagaca ccagcgcccc 6240

ggcctgtggg gtgccccgtg atgcctccat gcagggaaga gcaggggacc agggaggaag 6300

agcagagaca ggtgaagcga cagtccccgc gtcccagcct cagcattcgc atcctcttgg 6360

cccctacttt tcctctccgc ccagcagaca tctgccctgc ccttgccctt gaccccattg 6420

ctgcgcttcc ctcaaggacg ggcctggcct tggtggccac ctgcggacag ccctgcgccc 6480

gacgcccgct tcaccccggg gcccgggtct ggaggggccg cccccaggac gaacgcggct 6540

gccccacggg gccggccccc caccggcttc gcgtccaagc caaagtttct cgagcacttt 6600

tttgttcttt gcaatcatgt tgggttcatt gttggtgttt taaaattttg cttccctctc 6660

cctctggcct cgctcctgtg tgtgttttgt agccgagcgc taacctggat gctttttga 6720

atgacctttg caagagcctg ccttcctcgg cctctgctct gttttattta ttgttgaata 6780

tttccaatga tccaaatcaa agtgaattaa aacaaagcta ttttatcgtt 6830

<210> SEQ ID NO 6
<211> LENGTH: 342
<212> TYPE: PRT
<213> ORGANISM: Canis familiaris

-continued

<400> SEQUENCE: 6

Met Val Thr Met Arg Lys Arg Thr Leu Lys Val Leu Thr Leu Leu Val
1 5 10 15

Leu Phe Ile Phe Leu Thr Ser Phe Phe Leu Asn Tyr Ser His Thr Met
20 25 30

Val Thr Thr Thr Trp Phe Pro Lys Gln Met Val Val Glu Leu Ser Glu
35 40 45

Asn Phe Lys Lys Phe Met Lys Tyr Thr His Arg Pro Cys Thr Cys Ala
50 55 60

Arg Cys Ile Gly Gln Gln Arg Val Ser Ala Trp Phe Asp Glu Arg Phe
65 70 75 80

Asn Arg Ser Met Gln Pro Leu Leu Thr Ala Gln Asn Ala Leu Leu Glu
85 90 95

Glu Asp Thr Tyr Ser Trp Trp Leu Arg Leu Gln Arg Glu Lys Gln Pro
100 105 110

Asn Asn Leu Asn Asp Thr Ile Arg Glu Leu Phe Gln Val Val Pro Gly
115 120 125

Asn Val Asp Pro Leu Leu Glu Lys Arg Ser Val Gly Cys Arg Arg Cys
130 135 140

Ala Val Val Gly Asn Ser Gly Asn Leu Arg Glu Ser Trp Tyr Gly Pro
145 150 155 160

Gln Ile Asp Ser His Asp Phe Val Leu Arg Met Asn Lys Ala Pro Thr
165 170 175

Ala Gly Phe Glu Met Asp Val Gly Ser Lys Thr Thr His His Leu Val
180 185 190

Tyr Pro Glu Ser Phe Arg Glu Leu Ala Glu Asn Val Ser Met Val Leu
195 200 205

Val Pro Phe Lys Thr Thr Asp Leu Glu Trp Val Val Ser Ala Thr Thr
210 215 220

Thr Gly Thr Ile Ser His Thr Tyr Val Pro Val Pro Ala Lys Ile Lys
225 230 235 240

Val Lys Lys Asp Lys Ile Leu Ile Tyr His Pro Ala Phe Ile Lys Tyr
245 250 255

Val Phe Asp Ser Trp Leu Gln Gly His Gly Arg Tyr Pro Ser Thr Gly
260 265 270

Ile Leu Ser Val Ile Phe Ser Leu His Ile Cys Asp Glu Val Asp Leu
275 280 285

Tyr Gly Phe Gly Ala Asp Ser Lys Gly Asn Trp His His Tyr Trp Glu
290 295 300

Asn Asn Pro Ser Ala Gly Ala Phe Arg Lys Thr Gly Val His Asp Gly
305 310 315 320

Asp Phe Glu Ser Asn Val Thr Ala Thr Leu Ala Ser Ile Asn Lys Ile
325 330 335

Arg Ile Phe Lys Gly Arg
340

<210> SEQ ID NO 7
<211> LENGTH: 4577
<212> TYPE: DNA
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 7

-continued

cggctcgctc cgggagggca gagccggcaa gggcgggact ggcctgctgc gggcggacgg 60

gggagccgcg gagctaacgg gtccggacgt cgccagcggc ggagctttct gcacgggcgg 120

aggcggcggc aggagcggca agcatggccc gcgcggcgcc cggcgctgac tgagggtcat 180

gctgcggcag cctggcgcct gacgagtgag ggggaacctg ccaggggatt actgctagca 240

cgagctaaca gcaggacagc aggcctcagg ggagtgggga tgggcatagc agcctgcctg 300

caacgctggg ccagtggcac caaggaggcc caggccctgg cagggagctc ggcacacgct 360

cagctgaccc agacagcatt caaagcaggt ccctgctgag cagcagctgt tcccaggccc 420

tggaggaccg caaacgcctc agcaaccaca tttggggctg tgaccatcag gagaagggac 480

caacacctcg aatgggaagc agaagggcag aacctgctgc caggaggatc ttggtctctg 540

cagcaaagtc ttccgacacc cgtggagctt ctttgggttc tcctctcagg actggggcct 600

tgtttcctcc tctggtggag ggggtaggcc aggagggcca gctgaggtgg ggacgactct 660

cggatggaat ccaggccact tgaagcccat gggctctctg ctgtgattgg ggttcgaggg 720

cttccccact gcaggggacg aggggccgcc tccgtgtctg ctcatgaacc acaaggaccc 780

cgaatgctcc agactggacc atttcgagcc accaaagagt ggggccccca gagctggcag 840

ccagcagctc caagggacta gagagctggg atggactgac ctccccctca ccagggcatt 900

aggagagtca gagcctctgt ggcccagcta gtgacagaga gacccgatga ggccataagc 960

tggcacccga gtgactcagg gacaggagag ccactcctgc ctactgtgct ttctctttac 1020

attgagggag ggtgtggtaa gggacccctg cctctgttcc ctcctcagca cgtgccccta 1080

tgccctttgc acgtggtgcc aggagacgca ggctacactg tggctggccc atcagtgggc 1140

tgggaaggga gtggccacgg tgacacccac ccgcccacca ggctggtggt ctagcccccc 1200

ggggcgccaa accatgaagt gctccctgcg ggtgtggttc ctctctgtgg ccttcctgct 1260

ggtgttcatc atgtcactgc tctttaccta ctcccaccac agcatggcca ccttgcccta 1320

cctggactca ggggccctgg gcggtaccca ccgggtgaag ctggtgcctg gctatgctgg 1380

tctgcagcgc ctcagcaagg aggggctcac cggtaagagc tgtgcctgcc gccgctgcat 1440

gggtgacact ggcgcctctg actggtttga cagccacttc aacagcaaca tttcccctgt 1500

gtggacccga gagaatatgg atctgcctcc agatgtccag aggtggtgga tgatgctgca 1560

gccccagttc aagtcacaca acaccaacga ggtactggag aagctgttcc agatagtacc 1620

aggcgagaac ccctaccgtt tccgggaccc ccaccagtgc cggcgctgtg cggtagtggg 1680

gaactcaggc aacctgcggg gctctggcta cggcccagat gtggatgggc ataacttcat 1740

catgaggatg aatcaggcgc caaccgtggg ctttgagcag gatgttggca gccgaactac 1800

ccaccatttc atgtaccccg agagtgccaa gaaccttcct gccaacgtca gctttgtgtt 1860

ggtgcccttc aaagctctgg acctactatg gattgccagc gctttgtcca cagggcaaat 1920

cagattcacc tatgcgccag tgaagtcctt ccttcgagtg gacaaagaaa aggttcagat 1980

ctacaacccg gcattcttca agtacatcca cgaccggtgg acagagcatc acgggcggta 2040

cccttccaca gggatgctgg tgctcttctt tgccctgcat gtttgtgatg aggtgaacgt 2100

gtacgggttc ggggccgaca gccggggcaa ctggcaccac tactgggaga ataaccggta 2160

tgcgggcgag ttccggaaga cgggagtgca cgacgccgac ttcgaggccc acatcatcga 2220

catgctggcc aaggccagca agatcgaggt ctaccgaggc aactgagccc ggcctcgccg 2280

-continued

cgacccttcc ggcccagccg cgaggctgcg acgctcgctc cgcagccggg actcccggcc 2340

agcccgaggg cggcggcctt ggcgagggcg cggcgcccgg ttgggcgtct ccagccctgg 2400

gagcgacgcc aagtcgcggt ctggaccaat catgctgcaa gtccagcgag cgccggctgt 2460

cccccgccaa tcaggagact ttgggggctg gcccaggcct ggcacccaat cagcgctgca 2520

gtgggagcgg aggctctttc tcccagccaa tcatgcgact caaggagaac ttccggcgct 2580

gggcccggtc tcctccaatc aatggccttc ggaggcgggc cggccgccgc tgaatcccca 2640

ctcccctatg ctttgggtag gattttattt tatgcttttt aaggagtagt ggttggttcc 2700

ggcctcagtg gagtactttc ctcaggctct gcgggaggag tgttggtggc ctgtcggcgg 2760

tactcggcca ggggcaccga ggaggaagcg ggggggaagg tgcggggcag cagcggctgg 2820

gcctcccttg gccgggggcc cctcgcgacc tcggggcggg cgggggggcg ggggcgttgg 2880

cctcccgctt ctggaggtcc gggggggaatc aggtggtttc cggggagcgc gctttcatcc 2940

ccgggaagag ctagatccct ctcaaacctt ttcaggcctc agagcactct agaccgcgta 3000

tttcctttat ctgtcgggcc cagatgggtg agtgtaaacc caccaaagaa aggcagtgga 3060

gcgtgggtcc cctcgtcctc ccttgtcccc atccccacct ttggccaccc tatgggatgg 3120

ccttcctgac cagggcattg agcatcccac ctggaaacta gaactgtatt cacccgttgc 3180

tagggctcgg agttcgccga tggcctaact cgagcgcagg ggtaaggaac agggcagggg 3240

acccggaagc cccgtcactt cagatgtaag gtgcttctca ctctcgtgta ctctcgcggc 3300

cccttactgt tcgcccacaa ctttttttagt gtcccttctt aagccctggg cctcctcacc 3360

agcctgctcg tcctgggatt gggggtgggg ggtggggcac tgctggcttc tccaaccccc 3420

taccctcct cgctcgtctt tagccggctc tagggagagg aaggcaggct ggagatgggg 3480

agcccagctg cctggtgcat gcaccgtttt cctccgccca tcaccccaaa gaggagtagg 3540

aaacctcttg cttgggggtg gaatttgctt tggtctccta atttagttaa cttgaggtta 3600

ccagggatgg ctgaccaaca aagattcttt taaaattcca ggctggccat gcaaattgct 3660

gggatcctag ctggggagga gtcgactgac ttgcccgcct tgcatgtctc ctctcctgcc 3720

cctgcgtccc ctccctctgc cacgctcact tcctgcctca tctctccaac ccattttcca 3780

ttttcagctc tagaagggca gggacgctta caaacaggag ttacatctgg aagttacttc 3840

caagactgaa cccagcttaa gtccctagag gaagctgctg atgatattct cacccttcaa 3900

ggttggggaa gtttcggagg gggaaagtgc ttctgtgaag cttccaaacc actaatagga 3960

tccccttcc caacaatgag gaacacaaac accacccttt atcttagttg ataccaccaa 4020

gcagcctcct ggccattggg gtaattcctg cagctggctg gggtaaccag caggggagta 4080

tattagagga ggattggggc agggcagtgg gcacccctaa agttaatata ttgagaactt 4140

agcttaagcc taagtcttag ttccttccca attccaaaag taggaggagc aacgagtggg 4200

ggtggatttg ggggggccta tcctggaatg cctctctcag ggcttccccc accattttag 4260

agagtcaagg caccagccat tcatgccagt ctcctctcag tgcttcctga agaggctgtt 4320

tggagtgttc ggaaaatgaa aaaaacaatg caattatgcc aaacagtatt gagcagaata 4380

atttatttct tttttgttcc tttttcttct ttttgttttg tttaaaacat taataaatcc 4440

cctttctgga agaggtaggt cccagcatcc agcccagatc tccttttctg caatagttat 4500

ttaaacaaat gtttgtttgt ttttttattt tcttcccttt ctctctcttt ctgaattaaa 4560

-continued

aaaaaagaaa actccta 4577

<210> SEQ ID NO 8
<211> LENGTH: 350
<212> TYPE: PRT
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 8

Met Lys Cys Ser Leu Arg Val Trp Phe Leu Ser Val Ala Phe Leu Leu
1 5 10 15

Val Phe Ile Met Ser Leu Leu Phe Thr Tyr Ser His His Ser Met Ala
20 25 30

Thr Leu Pro Tyr Leu Asp Ser Gly Ala Leu Gly Gly Thr His Arg Val
35 40 45

Lys Leu Val Pro Gly Tyr Ala Gly Leu Gln Arg Leu Ser Lys Glu Gly
50 55 60

Leu Thr Gly Lys Ser Cys Ala Cys Arg Arg Cys Met Gly Asp Thr Gly
65 70 75 80

Ala Ser Asp Trp Phe Asp Ser His Phe Asn Ser Asn Ile Ser Pro Val
85 90 95

Trp Thr Arg Glu Asn Met Asp Leu Pro Pro Asp Val Gln Arg Trp Trp
100 105 110

Met Met Leu Gln Pro Gln Phe Lys Ser His Asn Thr Asn Glu Val Leu
115 120 125

Glu Lys Leu Phe Gln Ile Val Pro Gly Glu Asn Pro Tyr Arg Phe Arg
130 135 140

Asp Pro His Gln Cys Arg Arg Cys Ala Val Val Gly Asn Ser Gly Asn
145 150 155 160

Leu Arg Gly Ser Gly Tyr Gly Pro Asp Val Asp Gly His Asn Phe Ile
165 170 175

Met Arg Met Asn Gln Ala Pro Thr Val Gly Phe Glu Gln Asp Val Gly
180 185 190

Ser Arg Thr Thr His His Phe Met Tyr Pro Glu Ser Ala Lys Asn Leu
195 200 205

Pro Ala Asn Val Ser Phe Val Leu Val Pro Phe Lys Ala Leu Asp Leu
210 215 220

Leu Trp Ile Ala Ser Ala Leu Ser Thr Gly Gln Ile Arg Phe Thr Tyr
225 230 235 240

Ala Pro Val Lys Ser Phe Leu Arg Val Asp Lys Glu Lys Val Gln Ile
245 250 255

Tyr Asn Pro Ala Phe Phe Lys Tyr Ile His Asp Arg Trp Thr Glu His
260 265 270

His Gly Arg Tyr Pro Ser Thr Gly Met Leu Val Leu Phe Phe Ala Leu
275 280 285

His Val Cys Asp Glu Val Asn Val Tyr Gly Phe Gly Ala Asp Ser Arg
290 295 300

Gly Asn Trp His His Tyr Trp Glu Asn Asn Arg Tyr Ala Gly Glu Phe
305 310 315 320

Arg Lys Thr Gly Val His Asp Ala Asp Phe Glu Ala His Ile Ile Asp
325 330 335

Met Leu Ala Lys Ala Ser Lys Ile Glu Val Tyr Arg Gly Asn
340 345 350

-continued

<210> SEQ ID NO 9
<211> LENGTH: 1782
<212> TYPE: DNA
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 9

ttggtggtcg cgctccgccc gccgctgcgt ccccaccatg gcggcgcccg tgcagcccac 60

cgcgtcgtgg gcgcccgccg ggctcccccg cggctgtggc ggccgcccgc gcctcggcct 120

ccgcctcccc gcccgcgccg gcccgggcgc cgcctccccg ctgcctccgt ctccgctgcg 180

gtcatgtagg aaatcgtaaa tcatgtgaag atgggactct tggtatttgt acgcaatctg 240

ctgctagccc tctgcctttt tctggtactg ggatttttgt attattctgc gtggaagctg 300

catttactcc agtgggagga ctccaattca gtggttcttt cctttgactc cgctggacaa 360

acactaggct cagagtatga tcggttgggt ttcctcctga agctggactc taaactgcct 420

gctgagttag ccaccaagta tgcaaacttt tcagagggag cttgcaagcc tggctatgct 480

tcggccttga tgactgccat ctttccccgg ttctccaagc cagcacccat gttcctggat 540

gactctttcc gcaagtgggc taggattcgg gagtttgtgc cgccttttgg gatcaaaggt 600

caagacaatc tgatcaaagc catcttgtca gtcaccaaag agtaccgcct gacccctgcc 660

ttggacagcc tcagctgccg ccgctgcatc atcgtgggca acggaggtgt cctagccaac 720

aagtctctgg ggtcacgaat tgatgactat gacattgtgg tcagactgaa ctccgcacca 780

gtgaaaggct ttgagaagga cgtgggcagc aaaactacac tgcgcatcac ctaccctgag 840

ggcgccatgc agcggcctga gcaatatgaa cgcgattctc tatttgtcct cgctggcttc 900

aagtggcagg acttcaagtg gttgaagtac atcgtctaca aggagagagt gctctgggcc 960

cgcagggata cctgccaatc tgtctgggcc catccccctc tcccctccac cagctgtcac 1020

cagccacccc aggggagggg tcctgcagag ttcaggccat tcttcttcca atacccgagc 1080

ctcctactgg aggagaatga tgacagacag cctctggcga caagtgcatc agatggcttc 1140

tggaaatccg tggccacacg agtgcccaag gagccccctg agattcgcat cctcaacccg 1200

tacttcatcc aggaggccgc cttcaccctc atcggactgc ccttcaacaa cggcctcatg 1260

ggccgcggga acatcccgac ccttggcagt gtggcagtga ccatggcgct acacggctgt 1320

gatgaggtgg cagtcgcagg ctttggctac gacatgagca cacccaacgc gcccctgcac 1380

tactatgaga ccgtgcgcat ggcagccatc aaagaggtca ccagcgactc agctcaaggc 1440

tgccaaatcc agtggacaca tggaagcctc atctttcctg acctcccaga aatgcttttt 1500

ctgttgacca ctccttcctc tttgaaactt ttcctgctca gactgtcctg gacacacaat 1560

atccagcgag agaaagagtt tctgcgcaag ctggtgaagg cgcgcgtcat caccgaccta 1620

accagcggca tctgaggtgg gcccagcaca tggccacgga ggtcctggca ccgccaagag 1680

gaagccgcag ccactgccac ctgtccactt cattggcctc ggtctggctc tgcctgaaag 1740

gcgcaggagt cttcagaccc agagaaggac agtgccaagg gg 1782

<210> SEQ ID NO 10
<211> LENGTH: 474
<212> TYPE: PRT
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 10

-continued

```
Met Gly Leu Leu Val Phe Val Arg Asn Leu Leu Leu Ala Leu Cys Leu
1               5                  10                  15

Phe Leu Val Leu Gly Phe Leu Tyr Tyr Ser Ala Trp Lys Leu His Leu
            20                  25                  30

Leu Gln Trp Glu Asp Ser Asn Ser Val Val Leu Ser Phe Asp Ser Ala
        35                  40                  45

Gly Gln Thr Leu Gly Ser Glu Tyr Asp Arg Leu Gly Phe Leu Leu Lys
    50                  55                  60

Leu Asp Ser Lys Leu Pro Ala Glu Leu Ala Thr Lys Tyr Ala Asn Phe
65                  70                  75                  80

Ser Glu Gly Ala Cys Lys Pro Gly Tyr Ala Ser Ala Leu Met Thr Ala
                85                  90                  95

Ile Phe Pro Arg Phe Ser Lys Pro Ala Pro Met Phe Leu Asp Asp Ser
            100                 105                 110

Phe Arg Lys Trp Ala Arg Ile Arg Glu Phe Val Pro Pro Phe Gly Ile
        115                 120                 125

Lys Gly Gln Asp Asn Leu Ile Lys Ala Ile Leu Ser Val Thr Lys Glu
    130                 135                 140

Tyr Arg Leu Thr Pro Ala Leu Asp Ser Leu Ser Cys Arg Arg Cys Ile
145                 150                 155                 160

Ile Val Gly Asn Gly Gly Val Leu Ala Asn Lys Ser Leu Gly Ser Arg
                165                 170                 175

Ile Asp Asp Tyr Asp Ile Val Val Arg Leu Asn Ser Ala Pro Val Lys
            180                 185                 190

Gly Phe Glu Lys Asp Val Gly Ser Lys Thr Thr Leu Arg Ile Thr Tyr
        195                 200                 205

Pro Glu Gly Ala Met Gln Arg Pro Glu Gln Tyr Glu Arg Asp Ser Leu
    210                 215                 220

Phe Val Leu Ala Gly Phe Lys Trp Gln Asp Phe Lys Trp Leu Lys Tyr
225                 230                 235                 240

Ile Val Tyr Lys Glu Arg Val Leu Trp Ala Arg Arg Asp Thr Cys Gln
                245                 250                 255

Ser Val Trp Ala His Pro Pro Leu Pro Ser Thr Ser Cys His Gln Pro
            260                 265                 270

Pro Gln Gly Arg Gly Pro Ala Glu Phe Arg Pro Phe Phe Phe Gln Tyr
        275                 280                 285

Pro Ser Leu Leu Leu Glu Glu Asn Asp Asp Arg Gln Pro Leu Ala Thr
    290                 295                 300

Ser Ala Ser Asp Gly Phe Trp Lys Ser Val Ala Thr Arg Val Pro Lys
305                 310                 315                 320

Glu Pro Pro Glu Ile Arg Ile Leu Asn Pro Tyr Phe Ile Gln Glu Ala
                325                 330                 335

Ala Phe Thr Leu Ile Gly Leu Pro Phe Asn Asn Gly Leu Met Gly Arg
            340                 345                 350

Gly Asn Ile Pro Thr Leu Gly Ser Val Ala Val Thr Met Ala Leu His
        355                 360                 365

Gly Cys Asp Glu Val Ala Val Ala Gly Phe Gly Tyr Asp Met Ser Thr
    370                 375                 380

Pro Asn Ala Pro Leu His Tyr Tyr Glu Thr Val Arg Met Ala Ala Ile
385                 390                 395                 400
```

-continued

```
Lys Glu Val Thr Ser Asp Ser Ala Gln Gly Cys Gln Ile Gln Trp Thr
                405                 410                 415

His Gly Ser Leu Ile Phe Pro Asp Leu Pro Glu Met Leu Phe Leu Leu
            420                 425                 430

Thr Thr Pro Ser Ser Leu Lys Leu Phe Leu Leu Arg Leu Ser Trp Thr
        435                 440                 445

His Asn Ile Gln Arg Glu Lys Glu Phe Leu Arg Lys Leu Val Lys Ala
    450                 455                 460

Arg Val Ile Thr Asp Leu Thr Ser Gly Ile
465                 470
```

<210> SEQ ID NO 11
<211> LENGTH: 1819
<212> TYPE: DNA
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 11

```
ccctacaggc ccgagctgcc ggggtcgggc ctccccgggt tcccgctccc gggtcctcct     60

ggacacaccg gcctggcctg gctcccgggg aactctcgtc tgctagcggg gagcctccct    120

ccgcctcgcc cacgggcacc cctcccaccc agtatccttg gcctcttgca ggtggcccga    180

ggcagccggg atgacagctc tccccaggaa ccctgctacc ctctgagaaa catgatcagc    240

aagtcccgct ggaagctcct ggccatgttg gctctggtcc tggtcgtcat ggtgtggtat    300

tccatctccc gagaagacag gtacattgag ctttttttatt ttcccatccc agagaagaag    360

gaaccgtgct tccagggtga ggcagagaga aaggcctcta agctctttgg caactactcc    420

cgagatcagc ccatcttcct gcagatgaag gattatttct gggtcaagac accgtctgcc    480

tacgagctgc cctatgggac caaggggagc gaagacctgc tcctccgggt tctagccatc    540

accagctact ccattccaga gagcatccag agtctcaagt gtcgccgctg cgtggtggtg    600

ggcaatgggc atcggctgcg caacagctcg ctgggagatg ccatcaacaa gtacgacgtg    660

gtcatcagac tgaacaacgc ccccgtggct ggctacgagg gtgacgtggg ctcgaagacc    720

accatgcgtc tcttctaccc ggagtcagcc cacttcaacc ccaaagtgga gaacaaccca    780

gacacacttc tcgtcctggt ggccttcaag gcaatggact tccactggat tgagaccatc    840

ctgagtgata agaagagggt acgaaagggc ttctggaagc agcctcccct catctgggac    900

gtcaacccca ggcaggttcg gattctcaac cctttcttta tggagattgc agctgacaaa    960

ctgctgaacc tgccaatgaa acagccacgc aagatttccc agaagcccac cacgggcctg   1020

ctggccatca cgctggctct ccacctctgc gacctggtgc acatcgccgg cttcggctac   1080

ccggacgccc acaacaggaa gcagaccatt cactactatg aacagatcac gctcaagtcc   1140

atggcggggt caggccacaa cgtctcccag gaggccctgg ccatcaagcg gatgctggag   1200

atcggagcag tcaagaacct cacgttcttc tgacggggac aggagctcta gccgtcagtc   1260

tgcccgccct gccgcctgag cgaccaacca cggctgtggg ggcgccggcg tgacctgctt   1320

ggattccccc tccccgtgtg gagaggggggc ctggtacagg cggggcctga gatggggccg   1380

cgcccctggc tgctcttggg gcggccggat ccagtcaggg tggaggcccc gggtggcggg   1440

aggccttccg aggcgcgggg tgtgtggctg aggcacccct tctcaccagc cccgggagct   1500

tatttaatgg gctatttaat taaaagggta ggaatgtgcc tcgggctggt cccatggcat   1560

ccggaaacgg gggcatagca cagtggtctg cccactgtgg ataaaaacac acaagtgctt   1620
```

-continued

ggcccactag agcctagagc cagagcaggc ctcccaggag ggcaggggcg tctggagcgg 1680

gtgggtgccc tccagagagg ggctgctacc tcccagcggg catgggaaga gcattgggat 1740

gaggtcccac ggagaatggg acctcatgta gaaaagaggt ttgaaaccta acattaaact 1800

atttttcct aaaacggaa 1819

<210> SEQ ID NO 12
<211> LENGTH: 333
<212> TYPE: PRT
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 12

Met Ile Ser Lys Ser Arg Trp Lys Leu Leu Ala Met Leu Ala Leu Val
1 5 10 15

Leu Val Val Met Val Trp Tyr Ser Ile Ser Arg Glu Asp Arg Tyr Ile
20 25 30

Glu Leu Phe Tyr Phe Pro Ile Pro Glu Lys Lys Glu Pro Cys Phe Gln
35 40 45

Gly Glu Ala Glu Arg Lys Ala Ser Lys Leu Phe Gly Asn Tyr Ser Arg
50 55 60

Asp Gln Pro Ile Phe Leu Gln Met Lys Asp Tyr Phe Trp Val Lys Thr
65 70 75 80

Pro Ser Ala Tyr Glu Leu Pro Tyr Gly Thr Lys Gly Ser Glu Asp Leu
85 90 95

Leu Leu Arg Val Leu Ala Ile Thr Ser Tyr Ser Ile Pro Glu Ser Ile
100 105 110

Gln Ser Leu Lys Cys Arg Arg Cys Val Val Val Gly Asn Gly His Arg
115 120 125

Leu Arg Asn Ser Ser Leu Gly Asp Ala Ile Asn Lys Tyr Asp Val Val
130 135 140

Ile Arg Leu Asn Asn Ala Pro Val Ala Gly Tyr Glu Gly Asp Val Gly
145 150 155 160

Ser Lys Thr Thr Met Arg Leu Phe Tyr Pro Glu Ser Ala His Phe Asn
165 170 175

Pro Lys Val Glu Asn Asn Pro Asp Thr Leu Leu Val Leu Val Ala Phe
180 185 190

Lys Ala Met Asp Phe His Trp Ile Glu Thr Ile Leu Ser Asp Lys Lys
195 200 205

Arg Val Arg Lys Gly Phe Trp Lys Gln Pro Pro Leu Ile Trp Asp Val
210 215 220

Asn Pro Arg Gln Val Arg Ile Leu Asn Pro Phe Phe Met Glu Ile Ala
225 230 235 240

Ala Asp Lys Leu Leu Asn Leu Pro Met Lys Gln Pro Arg Lys Ile Ser
245 250 255

Gln Lys Pro Thr Thr Gly Leu Leu Ala Ile Thr Leu Ala Leu His Leu
260 265 270

Cys Asp Leu Val His Ile Ala Gly Phe Gly Tyr Pro Asp Ala His Asn
275 280 285

Arg Lys Gln Thr Ile His Tyr Tyr Glu Gln Ile Thr Leu Lys Ser Met
290 295 300

Ala Gly Ser Gly His Asn Val Ser Gln Glu Ala Leu Ala Ile Lys Arg
305 310 315 320

-continued

Met Leu Glu Ile Gly Ala Val Lys Asn Leu Thr Phe Phe
325 330

<210> SEQ ID NO 13
<211> LENGTH: 1385
<212> TYPE: DNA
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 13

cgctctggaa ccacttacag ccacctggtg catcctcctt tggggtgcgt ttggagggcc 60

tggttcctgc tcagccacat cttctgccac tttcaccagc aatgcccagt gagtataact 120

atgtaaaact gagaagcgat cgctcaagac cctctctgca atggtacacc cgagctcaaa 180

acaagatgag aagacccaac ttgttgttaa aagacatcct taagtgtaca ttgcttgtgt 240

ttggagtgtg gatcctttat attctcaagt taaattatac tactgaagaa tgtgacatga 300

aaaaaatgca ttatgtggac ccagaccgtg taaagagagc tcagaaatat gctcagcaag 360

tcttgcaaaa ggagtgccga cccaagtttg cgaagaagtc gatggcgcag ttgttcgagc 420

acaggtacag cacggacttg ccacctttcg tgaaggagac ccccaaaatg aatgaagccg 480

agtacaagta tgatcctcct tttggattcc gaaagttctc cagtgaagtc cagaccctgt 540

tggaaatact gcccgagcat gacatgcccg aacacttgag agcaaagagc tgtaggcgtt 600

gtgtggtcat cggaagcggt ggcatactcc acggactagc actgggccag gccctcaacc 660

agttcgatgt ggtgataagg ttaaacagtg caccagttga gggatattct gagcatgttg 720

gtaataaaac tactataagg atgacttatc cagagggcgc gccactgtct gaccttgaat 780

attattccaa tgacttgttt gttgctgttt tattcaagag tgttgacttc aactggcttc 840

aagcaatggt aaaaaatgaa accctgccat tttgggtgcg gctcttcttt tggaagcagg 900

tggcggaaaa aatcccacta cagccaaaac atttcaggat tttgaatcca gttattatca 960

aagagactgc ctttgacatc cttcaatact cagagcccca gtcaaggttc tggggccgag 1020

ataagaacgt gcccaccatt ggtgtcattg ccgttgtctt agccacacat ctgtgtgatg 1080

aagtcagctt ggcaggcttt ggatatgacc tcaatcaacc caaaacacct ttgcactact 1140

ttgacaatct ctgcatggct gccatgaact ttcaaaccat gcataatgtg acaacggaga 1200

ccaggttcct cctcaagctg gtcaaagagg gcgtggtgaa ggatctcagc ggaggcatcc 1260

attgtgaatt ttgaacacag ggaaacctca tgtgacaatg caactctgac tctgaaggct 1320

gtttttcgta gccttctcga tgcagcgcat cctgcaaaat acttagaggt gcagctgggg 1380

ttttt 1385

<210> SEQ ID NO 14
<211> LENGTH: 390
<212> TYPE: PRT
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 14

Met Pro Ser Glu Tyr Asn Tyr Val Lys Leu Arg Ser Asp Arg Ser Arg
1 5 10 15

Pro Ser Leu Gln Trp Tyr Thr Arg Ala Gln Asn Lys Met Arg Arg Pro
20 25 30

Asn Leu Leu Leu Lys Asp Ile Leu Lys Cys Thr Leu Leu Val Phe Gly
35 40 45

-continued

```
Val Trp Ile Leu Tyr Ile Leu Lys Leu Asn Tyr Thr Thr Glu Glu Cys
    50                  55                  60

Asp Met Lys Lys Met His Tyr Val Asp Pro Asp Arg Val Lys Arg Ala
65                  70                  75                  80

Gln Lys Tyr Ala Gln Gln Val Leu Gln Lys Glu Cys Arg Pro Lys Phe
                85                  90                  95

Ala Lys Lys Ser Met Ala Gln Leu Phe Glu His Arg Tyr Ser Thr Asp
            100                 105                 110

Leu Pro Pro Phe Val Lys Glu Thr Pro Lys Met Asn Glu Ala Glu Tyr
        115                 120                 125

Lys Tyr Asp Pro Pro Phe Gly Phe Arg Lys Phe Ser Ser Glu Val Gln
    130                 135                 140

Thr Leu Leu Glu Ile Leu Pro Glu His Asp Met Pro Glu His Leu Arg
145                 150                 155                 160

Ala Lys Ser Cys Arg Arg Cys Val Val Ile Gly Ser Gly Gly Ile Leu
                165                 170                 175

His Gly Leu Ala Leu Gly Gln Ala Leu Asn Gln Phe Asp Val Val Ile
            180                 185                 190

Arg Leu Asn Ser Ala Pro Val Glu Gly Tyr Ser Glu His Val Gly Asn
        195                 200                 205

Lys Thr Thr Ile Arg Met Thr Tyr Pro Glu Gly Ala Pro Leu Ser Asp
    210                 215                 220

Leu Glu Tyr Tyr Ser Asn Asp Leu Phe Val Ala Val Leu Phe Lys Ser
225                 230                 235                 240

Val Asp Phe Asn Trp Leu Gln Ala Met Val Lys Asn Glu Thr Leu Pro
                245                 250                 255

Phe Trp Val Arg Leu Phe Phe Trp Lys Gln Val Ala Glu Lys Ile Pro
            260                 265                 270

Leu Gln Pro Lys His Phe Arg Ile Leu Asn Pro Val Ile Ile Lys Glu
        275                 280                 285

Thr Ala Phe Asp Ile Leu Gln Tyr Ser Glu Pro Gln Ser Arg Phe Trp
    290                 295                 300

Gly Arg Asp Lys Asn Val Pro Thr Ile Gly Val Ile Ala Val Val Leu
305                 310                 315                 320

Ala Thr His Leu Cys Asp Glu Val Ser Leu Ala Gly Phe Gly Tyr Asp
                325                 330                 335

Leu Asn Gln Pro Lys Thr Pro Leu His Tyr Phe Asp Asn Leu Cys Met
            340                 345                 350

Ala Ala Met Asn Phe Gln Thr Met His Asn Val Thr Thr Glu Thr Arg
        355                 360                 365

Phe Leu Leu Lys Leu Val Lys Glu Gly Val Val Lys Asp Leu Ser Gly
    370                 375                 380

Gly Ile His Cys Glu Phe
385                 390


<210> SEQ ID NO 15
<211> LENGTH: 3159
<212> TYPE: DNA
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 15

ggtcgattgc cccttggctg ctgtggaggc tgtgatgacc tccagggccg cggccctccg     60
```

-continued

ggcgatgctt ctccaggggc tgaggccaac gcagaactcc cgtggcaccc actcggactc 120

gcggcgtgtt cacatgtggg gttttattaa atcctcccac caaccgtgtg agacaggaac 180

agttagcccc ggtgtgtccg ccaagattgc cccgcacaag tggctccgga tggatcacac 240

gaagcacttg caagtgaaga agcagcacag ccctttatct tgggctattt cctgtggaga 300

gactccaaca atttagcagc caggctcctg ggcctctggg accctcacca catcacatcc 360

ttcaccttca ggagcagagc gcctttggga aacagacttc taaaagtgca ggtgggccag 420

ccatgagagg gtacctagtg gccatattcc tgagtgctgt ctttctctat tatgtgctgc 480

attgtatatt gtggggaaca aacatctatt gggtgccacc tgtggaaatg aagcggagaa 540

ataagatcca gccttgttta gcgaagccag cttttgcctc tctcctgagg tttcatcagt 600

ttcacccttt tctgtgtgca gctgatttta aaaagattgc ttccttgtat ggtagcgata 660

agtttgatct gccctatggg ataagaacat cagcggaata ttttcgactc gctctttcaa 720

aactgcagag ttgtgatctc tttgatgagt ttgacaatgt gccgtgtaaa aagtgcgtgg 780

tggttggtaa tggaggagtt ctgaagaata agacattagg agaaaaaatt gactcctatg 840

atgtcataat aagaatgaat aatggtcctg ttttaggaca tgaagaggaa gttgggagaa 900

ggacaacctt ccgacttttt tatccagaat ctgtttttttc agatcccaat cacaatgatc 960

ctaatactac agcgattctc actgctttta agccgcttga cttaaagtgg ctgtgggaag 1020

tgttgacggg tggcaaaata aacactaatg gtttttggaa gaaaccagct ttaaacttga 1080

tctacaaacc ttatcaaatc agaatattag atcctttcat tatcagaatg gcagcttatg 1140

aactgcttca cttcccaaaa gtgtttccca aaaaccagaa acccaaacac ccaacaacag 1200

gaattattgc catcacgctg gcctttcaca tatgtcacga agttcacctt gctggtttta 1260

aatacaattt ttctgacctc aagagccctt tacactatta tgggaacgcg accatgtctt 1320

tgatgaataa gaatgcgtat cacaatgtga cagcggaaca gctctttttg aaggacattc 1380

tagaaaaaaa ctttgtaatc aacttgactg aagattgacc ctacagactc tgcagatgat 1440

gctaagagta ttagttttat ttttatactg caatttttag tttatttttа aatatgttgg 1500

atgcacttgt caaaaaattg tgtatagtca gtctgttgct gcctggtgat tcataaccac 1560

cagcttaatt tctgtgaata tatttaattt ataaaaacca agaagatatg cttagatatc 1620

cgggaagttt tgattgcgtt ggttttaaaa caaccttagt tctctgaagt gtttttaaac 1680

atctttttta atagttactt catctttgac ttctgagggc atgtgacgtc caagtaaggg 1740

gctttagctt gaccaccaca aactctgaac agagttggtg gcggattcgg ctactgtaaa 1800

ttggtgggga atagccatgt gattgtgcaa actggaaccg gtttaggcaa gtatcgagtt 1860

ccttttтact gaacccgagg aaacggattt gaatcttaaa gcaggcccaa ccatagcagt 1920

aggtacggtt atgaaatcta agatcataat ggtttcatta agctttttтt cctgtaagta 1980

aaccagatta taaaatgaaa ggtgtttgtt tttaaggtgg aggaaacagg ctacatgtga 2040

aattctggat gagtaaacaa cctaggaatg caattactaa agtctggtgg ctgcattatt 2100

ttaaagttca tacaaagaag cagagctagg ccacctcaag gagacagttc ttaaacgtca 2160

tcttttgcct gccttaatat gttaaaattt ggaagtttac tatttgaaat aggaaagatg 2220

aatacggcac agtaggtaaa tccttcagac tcctcaggct gtttttggat ttaaatggtc 2280

ctttcgtgaa aaatctcact tgtccacggt gaaatcccat cttcaaaggg aaggcttacc 2340

-continued

cggctaccta gggtgcatca gagaagagtc ctgctggatg cagacaagtc aaaaccagcc 2400

tgtccaacaa acgtgcgccc gtctctcttc tcaaagaggg atggaatgaa cagctctcag 2460

aagaggtaag agttgaagga cttgttatcc tctgagcgat aatcgtcatg gagagacact 2520

gctggtgttc ctgaaaacca gcctgcctct gagtctcaga gacaaaatat gagagcagcc 2580

actgggataa atcgtgaagc acggcataag gggggagaag cctcgtagtt gattgaaccc 2640

atgtctacgt ggcttcagct gattcccctg taacggaggt ggaaagttcc cgcacgtaca 2700

cagctgcacg ctgcagcctg gcggctggga ttccatgggt ggactcattc agggtacaaa 2760

gacagtcctg gctgcaaagt gaaaaacccc aggtggcatt ttcaagtgtt tatggactga 2820

aataatggct gtacggtatc tggcggatgc tcaacttgag gaatcggcat ttttgtacag 2880

tgggagctga ggctataaac ctcagcgtgg cttcacataa gccagaagaa actctcagcc 2940

cgatacatat gtacaattta ttaaaaacac atgaacacgt taaaatctca ctatttatac 3000

aatctacatt ctagcaacat atacaaatac cgagtgacta cagtacatgc cgaggtaaga 3060

aaagtacatt cggggagact atcactgaca ctcaagccat ttttatttcc aatatgtttt 3120

gctttcacct ttcccagtgc caaaaaaaaa aaaaaaaaa 3159

<210> SEQ ID NO 16
<211> LENGTH: 331
<212> TYPE: PRT
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 16

Met Arg Gly Tyr Leu Val Ala Ile Phe Leu Ser Ala Val Phe Leu Tyr
1 5 10 15

Tyr Val Leu His Cys Ile Leu Trp Gly Thr Asn Ile Tyr Trp Val Pro
20 25 30

Pro Val Glu Met Lys Arg Arg Asn Lys Ile Gln Pro Cys Leu Ala Lys
35 40 45

Pro Ala Phe Ala Ser Leu Leu Arg Phe His Gln Phe His Pro Phe Leu
50 55 60

Cys Ala Ala Asp Phe Lys Lys Ile Ala Ser Leu Tyr Gly Ser Asp Lys
65 70 75 80

Phe Asp Leu Pro Tyr Gly Ile Arg Thr Ser Ala Glu Tyr Phe Arg Leu
85 90 95

Ala Leu Ser Lys Leu Gln Ser Cys Asp Leu Phe Asp Glu Phe Asp Asn
100 105 110

Val Pro Cys Lys Lys Cys Val Val Val Gly Asn Gly Gly Val Leu Lys
115 120 125

Asn Lys Thr Leu Gly Glu Lys Ile Asp Ser Tyr Asp Val Ile Ile Arg
130 135 140

Met Asn Asn Gly Pro Val Leu Gly His Glu Glu Glu Val Gly Arg Arg
145 150 155 160

Thr Thr Phe Arg Leu Phe Tyr Pro Glu Ser Val Phe Ser Asp Pro Asn
165 170 175

His Asn Asp Pro Asn Thr Thr Ala Ile Leu Thr Ala Phe Lys Pro Leu
180 185 190

Asp Leu Lys Trp Leu Trp Glu Val Leu Thr Gly Gly Lys Ile Asn Thr
195 200 205

-continued

```
Asn Gly Phe Trp Lys Lys Pro Ala Leu Asn Leu Ile Tyr Lys Pro Tyr
    210                 215                 220

Gln Ile Arg Ile Leu Asp Pro Phe Ile Ile Arg Met Ala Ala Tyr Glu
225                 230                 235                 240

Leu Leu His Phe Pro Lys Val Phe Pro Lys Asn Gln Lys Pro Lys His
                245                 250                 255

Pro Thr Thr Gly Ile Ile Ala Ile Thr Leu Ala Phe His Ile Cys His
            260                 265                 270

Glu Val His Leu Ala Gly Phe Lys Tyr Asn Phe Ser Asp Leu Lys Ser
        275                 280                 285

Pro Leu His Tyr Tyr Gly Asn Ala Thr Met Ser Leu Met Asn Lys Asn
    290                 295                 300

Ala Tyr His Asn Val Thr Ala Glu Gln Leu Phe Leu Lys Asp Ile Leu
305                 310                 315                 320

Glu Lys Asn Phe Val Ile Asn Leu Thr Glu Asp
                325                 330


<210> SEQ ID NO 17
<211> LENGTH: 1337
<212> TYPE: DNA
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 17

aaagacttca ctgggtatca gtctcctttg ggagaccaca ggacacgtgt cacctctccc     60

atcctctcag cctccagccc agaccttggc agagttcctt ttaggagtta gcaagtggct    120

gaggaggcaa gaggtgccag agccaatcta ctatctgctg ggggatgatt gccagggcca    180

gagatgaggg ctcaatactt gaagtggggt ctggtagctg cctgtatagt tacgttatgg    240

ctgatgatga tgaacttcct ggaccaggag ttcaaacaga atgacttccc taaaaagaca    300

agaatacaat tatgccactg ccccaggaac tctttcagaa agtgtaggtg ttcgtttgag    360

atccgcaagt gctctgcctg cctccgcgta cgtggaacgt ctgtctggtt tgatgaacgc    420

ttcgaaacgg ctattgagcc tgtgcagaga ccagaagatc ccatatcctc tgatgctctg    480

atattgtggt tgggtgtcca atcaaagagg gagtttgaga ctcagaagcc aatagaagag    540

cctcctgggc aaccactggg ctacgtggag tccagttgtc ggacctgtgc agtggttgga    600

aactcaaggt gcctacgggg ctctggccat ggattcagga ttaaccaaaa tgacatggtc    660

ctcaggatga accaggcccc cgtccaagga tttgagatgg atgtggggaa cacaaccacc    720

atgcgcataa tgtaccccga tatggctagc acgcagaatc ctggcaccaa attgctgctg    780

cttcctctga attcatctgg tctaaagtgg tttatggaag tactacagga acagagcttc    840

agaaagccca taaaccctgg atttcagata gtccagtttc ctggtggaag taacacgagc    900

aaagacgagg tcttggtgat cagcctcacc tttcttcagt acatccagga tcattggctg    960

cgaaaacgtc atcgttttcc atccttgggg tttgtgggtc tgttatatgc cctgcacact   1020

tgtgaccagg tatccttatt tggttttggg acagatcagc tcatgaggtg gtcccattac   1080

tgggatgata aatatcggtt cgagagtaac atgcacagtt tcaaagaaga gcagaagctc   1140

atcctccagc tgcaatgtga ggggaagatt gttatctaca gctgacatgt ttctgtcctg   1200

ttcagcccac tggaggcccc aggaggctga caggtagtca aggggaccac agagtgtcag   1260

agagggactg gggcttcaag tggaccctgg atatagatca gtctgctgct aaataaaact   1320
```

-continued

acagcttatt tctccca 1337

<210> SEQ ID NO 18
<211> LENGTH: 333
<212> TYPE: PRT
<213> ORGANISM: Canis familiaris

<400> SEQUENCE: 18

Met Arg Ala Gln Tyr Leu Lys Trp Gly Leu Val Ala Ala Cys Ile Val
1 5 10 15

Thr Leu Trp Leu Met Met Met Asn Phe Leu Asp Gln Glu Phe Lys Gln
20 25 30

Asn Asp Phe Pro Lys Lys Thr Arg Ile Gln Leu Cys His Cys Pro Arg
35 40 45

Asn Ser Phe Arg Lys Cys Arg Cys Ser Phe Glu Ile Arg Lys Cys Ser
50 55 60

Ala Cys Leu Arg Val Arg Gly Thr Ser Val Trp Phe Asp Glu Arg Phe
65 70 75 80

Glu Thr Ala Ile Glu Pro Val Gln Arg Pro Glu Asp Pro Ile Ser Ser
85 90 95

Asp Ala Leu Ile Leu Trp Leu Gly Val Gln Ser Lys Arg Glu Phe Glu
100 105 110

Thr Gln Lys Pro Ile Glu Glu Pro Pro Gly Gln Pro Leu Gly Tyr Val
115 120 125

Glu Ser Ser Cys Arg Thr Cys Ala Val Val Gly Asn Ser Arg Cys Leu
130 135 140

Arg Gly Ser Gly His Gly Phe Arg Ile Asn Gln Asn Asp Met Val Leu
145 150 155 160

Arg Met Asn Gln Ala Pro Val Gln Gly Phe Glu Met Asp Val Gly Asn
165 170 175

Thr Thr Thr Met Arg Ile Met Tyr Pro Asp Met Ala Ser Thr Gln Asn
180 185 190

Pro Gly Thr Lys Leu Leu Leu Leu Pro Leu Asn Ser Ser Gly Leu Lys
195 200 205

Trp Phe Met Glu Val Leu Gln Glu Gln Ser Phe Arg Lys Pro Ile Asn
210 215 220

Pro Gly Phe Gln Ile Val Gln Phe Pro Gly Gly Ser Asn Thr Ser Lys
225 230 235 240

Asp Glu Val Leu Val Ile Ser Leu Thr Phe Leu Gln Tyr Ile Gln Asp
245 250 255

His Trp Leu Arg Lys Arg His Arg Phe Pro Ser Leu Gly Phe Val Gly
260 265 270

Leu Leu Tyr Ala Leu His Thr Cys Asp Gln Val Ser Leu Phe Gly Phe
275 280 285

Gly Thr Asp Gln Leu Met Arg Trp Ser His Tyr Trp Asp Asp Lys Tyr
290 295 300

Arg Phe Glu Ser Asn Met His Ser Phe Lys Glu Glu Gln Lys Leu Ile
305 310 315 320

Leu Gln Leu Gln Cys Glu Gly Lys Ile Val Ile Tyr Ser
325 330

<210> SEQ ID NO 19

-continued

<400> SEQUENCE: 19

000

<210> SEQ ID NO 20
<211> LENGTH: 83
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 20

cctccttctt cctgaattac tcccacaccc ctccttttct gaattactcc cacacccctc 60

ctttcctgaa ttactcccac acc 83

<210> SEQ ID NO 21
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 21

ctttacctac tcccaccaca gcatggccac tttacctact cccaccacaa gcatggcca 59

<210> SEQ ID NO 22
<211> LENGTH: 87
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 22

ctcccccgcg gctgtggcgg ccgcccgcgc tcccccgcgg gctgtggcgg ccgcccgcgc 60

tcccccgcgc tgtggcggcc gcccgcg 87

<210> SEQ ID NO 23
<211> LENGTH: 319
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 23

tccccaggaa ccctgctacc ctctgagaat ccccaggacc ctgctaccct ctgagaatcc 60

ccaggacgat ggtggtgtca aagtacttga aggcggcagg ggctcccaga ttggtcaggg 120

taaacaggtg gatgatattc tcggcctgct ctctgatggg cttatcccgg tgcttgttgt 180

aggcggacag cactttgtcc agattagcgt cggccaggat cactctcttg gagaactcgc 240

tgatctgctc gatgatctcg tccaggtagt gcttgtgctg ttccacaaac agctgtttca 300

ccctgctacc ctctgagaa 319

<210> SEQ ID NO 24
<211> LENGTH: 135
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 24

-continued

atcgctcaag accctctctg caatggtaca tcgctcaaga ccctctctga tggtacatcg 60

ctcaagacca atggtacatc gctcaagacc ctctctggca atggtacatc gctcaagacc 120

ctctctcaat ggtac 135

<210> SEQ ID NO 25
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 25

agcgataagt ttgatctgcc ctatgggata agcgataagt ttgatctgcc tatgggata 59

<210> SEQ ID NO 26
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 26

gcccccgtcc aaggatttga gatggatgtg cccccgtcaa ggatttgaga tggatgt 57

<210> SEQ ID NO 27
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 27

ccctcctcgt cctcttcatc 20

<210> SEQ ID NO 28
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 28

aggcagagag agaccagaga 20

<210> SEQ ID NO 29
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 29

ccaaaccatg aagtgctccc 20

<210> SEQ ID NO 30
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 30

-continued

aggggcttga agagtgactc 20

<210> SEQ ID NO 31
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 31

atgagacttg cttgcatccc 20

<210> SEQ ID NO 32
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 32

ctttggttgg cctctctgtc tc 22

<210> SEQ ID NO 33
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 33

cgttagccgc gcgcacag 18

<210> SEQ ID NO 34
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 34

ccgggatgac agctctc 17

<210> SEQ ID NO 35
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 35

acatggaagc tggactcac 19

<210> SEQ ID NO 36
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 36

catcatcaca aggatcctgc 20

<210> SEQ ID NO 37

-continued

<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 37

ctctcccatg aaaacctgg 19

<210> SEQ ID NO 38
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 38

gttttaaatt tgggagcggc c 21

<210> SEQ ID NO 39
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 39

tggctcacat caaacaccac 20

<210> SEQ ID NO 40
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 40

ggttggaaac tcaaggtgcc 20

<210> SEQ ID NO 41
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 41

tgactccttc ccctttccc 20

<210> SEQ ID NO 42
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 42

gttacgcgcc agcaaaagca ggggaaaaca aaagcaa 37

<210> SEQ ID NO 43
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

-continued

<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 43

gttacgcgcc agtagaaaca agggtgtttt tctcatgc 38

<210> SEQ ID NO 44
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 44

gttacgcgcc agcaaaagca ggagtttaaa at 32

<210> SEQ ID NO 45
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 45

gttacgcgcc agtagaaaca aggagttttt tgaacaac 38

<210> SEQ ID NO 46
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 46

gttacgcgcc agcaaaagca ggggataatt ctattaa 37

<210> SEQ ID NO 47
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 47

gttacgcgcc agtagaaaca agggtgtttt taattaatg 39

<210> SEQ ID NO 48
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 48

gttacgcgcc agcaaaagca ggagtaaaga tg 32

<210> SEQ ID NO 49
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 49

-continued

gttacgcgcc agtagaaaca aggagttttt tctaaaattg c 41

<210> SEQ ID NO 50
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 50

gttacgcgcc agcagaagca gagcattttc taatatcc 38

<210> SEQ ID NO 51
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 51

gttacgcgcc agtagtaaca agagcatttt tcaataacgt ttc 43

<210> SEQ ID NO 52
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 52

gttacgcgcc agcagaagca gagcatcttc tcaaaactg 39

<210> SEQ ID NO 53
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 53

gttacgcgcc agtagtaaca agagcatttt tcagaaac 38

<210> SEQ ID NO 54
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 54

agaaaagaat gtaacagtaa cacactctgt 30

<210> SEQ ID NO 55
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 55

tgtttccaca atgtargacc at 22

<210> SEQ ID NO 56

-continued

<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 56

ctattggaca atagtaaaac cgggrga 27

<210> SEQ ID NO 57
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 57

gtcattgggr atgcttccat ttgg 24

<210> SEQ ID NO 58

<400> SEQUENCE: 58

000

<210> SEQ ID NO 59
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 59

cctgttacat ctgggtgctt tcctataatg 30

<210> SEQ ID NO 60
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 60

gttgatarcc tgatatgttc gtatcctckg 30

<210> SEQ ID NO 61
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 61

cctgttacat ccgggtgctt ycctataatg 30

<210> SEQ ID NO 62
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 62

gttgataacc tkatmttttc atatcctctg 30

-continued

<210> SEQ ID NO 63
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 63

ccmaggtcga aacgtaygtt ctctctatc 29

<210> SEQ ID NO 64
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 64

tgacagraty ggtcttgtct ttagccaytc ca 32

<210> SEQ ID NO 65
<211> LENGTH: 17
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 65

ggagcaacca atgccac 17

<210> SEQ ID NO 66
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 66

gtktaggcgg tcttgaccag 20

<210> SEQ ID NO 67
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 67

cagccagcaa trttrcattt acc 23

<210> SEQ ID NO 68
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 68

cagccagcaa trttrcattt acc 23

<210> SEQ ID NO 69
<211> LENGTH: 23

-continued

<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 69

aagtaacccc kaggagcaat tag 23

<210> SEQ ID NO 70
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 70

aagtaacccc kaggagcaat tag 23

<210> SEQ ID NO 71
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 71

ttagacagct gcctaacc 18

<210> SEQ ID NO 72
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 72

ttagacagct gcctaacc 18

<210> SEQ ID NO 73
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 73

tcaggcaact asccaatc 18

<210> SEQ ID NO 74
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 74

tcaggcaact asccaatc 18

<210> SEQ ID NO 75
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

-continued

<400> SEQUENCE: 75

atytcggctt tgaggggcc tg 22

<210> SEQ ID NO 76
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: A synthetic oligonucleotide

<400> SEQUENCE: 76

ataaactttg aagcaggaat 20

<210> SEQ ID NO 77

<400> SEQUENCE: 77

000

<210> SEQ ID NO 78

<400> SEQUENCE: 78

000

<210> SEQ ID NO 79

<400> SEQUENCE: 79

000

<210> SEQ ID NO 80

<400> SEQUENCE: 80

000

<210> SEQ ID NO 81

<400> SEQUENCE: 81

000

<210> SEQ ID NO 82

<400> SEQUENCE: 82

000

<210> SEQ ID NO 83

<400> SEQUENCE: 83

000

<210> SEQ ID NO 84

<400> SEQUENCE: 84

000

<210> SEQ ID NO 85

-continued

<400> SEQUENCE: 85

000

<210> SEQ ID NO 86

<400> SEQUENCE: 86

000

<210> SEQ ID NO 87

<400> SEQUENCE: 87

000

<210> SEQ ID NO 88

<400> SEQUENCE: 88

000

<210> SEQ ID NO 89

<400> SEQUENCE: 89

000

<210> SEQ ID NO 90

<400> SEQUENCE: 90

000

<210> SEQ ID NO 91

<400> SEQUENCE: 91

000

<210> SEQ ID NO 92

<400> SEQUENCE: 92

000

<210> SEQ ID NO 93

<400> SEQUENCE: 93

000

<210> SEQ ID NO 94

<400> SEQUENCE: 94

000

<210> SEQ ID NO 95

<400> SEQUENCE: 95

000

<210> SEQ ID NO 96

-continued

<400> SEQUENCE: 96

000

<210> SEQ ID NO 97

<400> SEQUENCE: 97

000

<210> SEQ ID NO 98

<400> SEQUENCE: 98

000

<210> SEQ ID NO 99

<400> SEQUENCE: 99

000

<210> SEQ ID NO 100
<211> LENGTH: 1221
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 100

atgattcaca ccaacctgaa gaaaaagttc agctgctgcg tcctggtctt tcttctgttt 60

gcagtcatct gtgtgtggaa ggagaagaag aaagggagtt actatgattc ctttaaattg 120

caaaccaagg aattccaggt gttaaagagt ctggggaaat tggccatggg gtctgattcc 180

cagtctgtat cctcaagcag cacccaggac ccccacaggg gccgccagac cctcggcagt 240

ctcagaggcc tagccaaggc caaaccagag gcctccttcc aggtgtggaa caaggacagc 300

tcttccaaaa accttatccc taggctgcaa aagatctgga agaattacct aagcatgaac 360

aagtacaaag tgtcctacaa ggggccagga ccaggcatca agttcagtgc agaggccctg 420

cgctgccacc tccgggacca tgtgaatgta tccatggtag aggtcacaga ttttcccttc 480

aatacctctg aatgggaggg ttatctgccc aaggagagca ttaggaccaa ggctgggcct 540

tggggcaggt gtgctgttgt gtcgtcagcg ggatctctga agtcctccca actaggcaga 600

gaaatcgatg atcatgacgc agtcctgagg tttaatgggg cacccacagc caacttccaa 660

caagatgtgg gcacaaaaac taccattcgc ctgatgaact ctcagttggt taccacagag 720

aagcgcttcc tcaaagacag tttgtacaat gaaggaatcc taattgtatg ggacccatct 780

gtataccacc cagatatccc aaagtggtac cagaatccgg attataattt ctttaacaac 840

tacaagactt atcgtaagct gcaccccaat cagccctttt acatcctcaa gccccagatg 900

ccttgggagc tatgggacat tcttcaagaa atctccccag aagagattca gccaaacccc 960

ccatcctctg ggatgcttgg tatcatcatc atgatgacgc tgtgtgacca ggtggatatt 1020

tatgagttcc tcccatccaa gcgcaggact gacgtgtgct actactacca gaagttcttc 1080

gatagtgcct gcacgatggg tgcctaccac ccgctgctct ttgagaagaa tttggtgaag 1140

catctcaacc agggcacaga tgaggacatc tacctgcttg gaaaagccac actgcctggc 1200

ttccggacca ttcactgcta a 1221

-continued

<210> SEQ ID NO 101
<211> LENGTH: 406
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 101

Met Ile His Thr Asn Leu Lys Lys Lys Phe Ser Cys Cys Val Leu Val
1 5 10 15

Phe Leu Leu Phe Ala Val Ile Cys Val Trp Lys Glu Lys Lys Lys Gly
20 25 30

Ser Tyr Tyr Asp Ser Phe Lys Leu Gln Thr Lys Glu Phe Gln Val Leu
35 40 45

Lys Ser Leu Gly Lys Leu Ala Met Gly Ser Asp Ser Gln Ser Val Ser
50 55 60

Ser Ser Ser Thr Gln Asp Pro His Arg Gly Arg Gln Thr Leu Gly Ser
65 70 75 80

Leu Arg Gly Leu Ala Lys Ala Lys Pro Glu Ala Ser Phe Gln Val Trp
85 90 95

Asn Lys Asp Ser Ser Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile
100 105 110

Trp Lys Asn Tyr Leu Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly
115 120 125

Pro Gly Pro Gly Ile Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu
130 135 140

Arg Asp His Val Asn Val Ser Met Val Glu Val Thr Asp Phe Pro Phe
145 150 155 160

Asn Thr Ser Glu Trp Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr
165 170 175

Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser
180 185 190

Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val
195 200 205

Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly
210 215 220

Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu
225 230 235 240

Lys Arg Phe Leu Lys Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val
245 250 255

Trp Asp Pro Ser Val Tyr His Pro Asp Ile Pro Lys Trp Tyr Gln Asn
260 265 270

Pro Asp Tyr Asn Phe Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His
275 280 285

Pro Asn Gln Pro Phe Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu
290 295 300

Trp Asp Ile Leu Gln Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro
305 310 315 320

Pro Ser Ser Gly Met Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp
325 330 335

Gln Val Asp Ile Tyr Glu Phe Leu Pro Ser Lys Arg Arg Thr Asp Val
340 345 350

Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala
355 360 365

-continued

```
Tyr His Pro Leu Leu Phe Glu Lys Asn Leu Val Lys His Leu Asn Gln
    370                 375                 380

Gly Thr Asp Glu Asp Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly
385                 390                 395                 400

Phe Arg Thr Ile His Cys
                405
```

<210> SEQ ID NO 102

<400> SEQUENCE: 102

000

<210> SEQ ID NO 103

<400> SEQUENCE: 103

000

<210> SEQ ID NO 104

<400> SEQUENCE: 104

000

<210> SEQ ID NO 105

<400> SEQUENCE: 105

000

<210> SEQ ID NO 106

<400> SEQUENCE: 106

000

<210> SEQ ID NO 107

<400> SEQUENCE: 107

000

<210> SEQ ID NO 108

<400> SEQUENCE: 108

000

<210> SEQ ID NO 109

<400> SEQUENCE: 109

000

<210> SEQ ID NO 110

<400> SEQUENCE: 110

000

<210> SEQ ID NO 111

-continued

<400> SEQUENCE: 111

000

<210> SEQ ID NO 112

<400> SEQUENCE: 112

000

<210> SEQ ID NO 113

<400> SEQUENCE: 113

000

<210> SEQ ID NO 114

<400> SEQUENCE: 114

000

<210> SEQ ID NO 115

<400> SEQUENCE: 115

000

<210> SEQ ID NO 116

<400> SEQUENCE: 116

000

<210> SEQ ID NO 117

<400> SEQUENCE: 117

000

<210> SEQ ID NO 118

<400> SEQUENCE: 118

000

<210> SEQ ID NO 119

<400> SEQUENCE: 119

000

<210> SEQ ID NO 120

<400> SEQUENCE: 120

000

<210> SEQ ID NO 121

<400> SEQUENCE: 121

000

<210> SEQ ID NO 122

-continued

<400> SEQUENCE: 122

000

<210> SEQ ID NO 123

<400> SEQUENCE: 123

000

<210> SEQ ID NO 124

<400> SEQUENCE: 124

000

<210> SEQ ID NO 125

<400> SEQUENCE: 125

000

<210> SEQ ID NO 126

<400> SEQUENCE: 126

000

<210> SEQ ID NO 127

<400> SEQUENCE: 127

000

<210> SEQ ID NO 128

<400> SEQUENCE: 128

000

<210> SEQ ID NO 129

<400> SEQUENCE: 129

000

<210> SEQ ID NO 130

<400> SEQUENCE: 130

000

<210> SEQ ID NO 131

<400> SEQUENCE: 131

000

<210> SEQ ID NO 132

<400> SEQUENCE: 132

000

-continued

<210> SEQ ID NO 133

<400> SEQUENCE: 133

000

<210> SEQ ID NO 134

<400> SEQUENCE: 134

000

<210> SEQ ID NO 135

<400> SEQUENCE: 135

000

<210> SEQ ID NO 136

<400> SEQUENCE: 136

000

<210> SEQ ID NO 137

<400> SEQUENCE: 137

000

<210> SEQ ID NO 138

<400> SEQUENCE: 138

000

<210> SEQ ID NO 139

<400> SEQUENCE: 139

000

<210> SEQ ID NO 140

<400> SEQUENCE: 140

000

<210> SEQ ID NO 141

<400> SEQUENCE: 141

000

<210> SEQ ID NO 142

<400> SEQUENCE: 142

000

<210> SEQ ID NO 143

<400> SEQUENCE: 143

000

-continued

<210> SEQ ID NO 144

<400> SEQUENCE: 144

000

<210> SEQ ID NO 145

<400> SEQUENCE: 145

000

<210> SEQ ID NO 146

<400> SEQUENCE: 146

000

<210> SEQ ID NO 147

<400> SEQUENCE: 147

000

<210> SEQ ID NO 148

<400> SEQUENCE: 148

000

<210> SEQ ID NO 149

<400> SEQUENCE: 149

000

<210> SEQ ID NO 150
<211> LENGTH: 406
<212> TYPE: PRT
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 150

Met Ile His Thr Asn Leu Lys Lys Lys Phe Ser Cys Cys Val Leu Val
1 5 10 15

Phe Leu Leu Phe Ala Val Ile Cys Val Trp Lys Glu Lys Lys Lys Gly
            20 25 30

Ser Tyr Tyr Asp Ser Phe Lys Leu Gln Thr Lys Glu Phe Gln Val Leu
        35 40 45

Lys Ser Leu Gly Lys Leu Ala Met Gly Ser Asp Ser Gln Ser Val Ser
    50 55 60

Ser Ser Ser Thr Gln Asp Pro His Arg Gly Arg Gln Thr Leu Gly Ser
65 70 75 80

Leu Arg Gly Leu Ala Lys Ala Lys Pro Glu Ala Ser Phe Gln Val Trp
                85 90 95

Asn Lys Asp Ser Ser Ser Lys Asn Leu Ile Pro Arg Leu Gln Lys Ile
            100 105 110

Trp Lys Asn Tyr Leu Ser Met Asn Lys Tyr Lys Val Ser Tyr Lys Gly
        115 120 125

Pro Gly Pro Gly Ile Lys Phe Ser Ala Glu Ala Leu Arg Cys His Leu
    130 135 140

-continued

```
Arg Asp His Val Asn Val Ser Met Val Glu Val Thr Asp Phe Pro Phe
145                 150                 155                 160

Asn Thr Ser Glu Trp Glu Gly Tyr Leu Pro Lys Glu Ser Ile Arg Thr
                165                 170                 175

Lys Ala Gly Pro Trp Gly Arg Cys Ala Val Val Ser Ser Ala Gly Ser
            180                 185                 190

Leu Lys Ser Ser Gln Leu Gly Arg Glu Ile Asp Asp His Asp Ala Val
        195                 200                 205

Leu Arg Phe Asn Gly Ala Pro Thr Ala Asn Phe Gln Gln Asp Val Gly
    210                 215                 220

Thr Lys Thr Thr Ile Arg Leu Met Asn Ser Gln Leu Val Thr Thr Glu
225                 230                 235                 240

Lys Arg Phe Leu Lys Asp Ser Leu Tyr Asn Glu Gly Ile Leu Ile Val
                245                 250                 255

Trp Asp Pro Ser Val Tyr His Pro Asp Ile Pro Lys Trp Tyr Gln Asn
            260                 265                 270

Pro Asp Tyr Asn Phe Phe Asn Asn Tyr Lys Thr Tyr Arg Lys Leu His
        275                 280                 285

Pro Asn Gln Pro Phe Tyr Ile Leu Lys Pro Gln Met Pro Trp Glu Leu
    290                 295                 300

Trp Asp Ile Leu Gln Glu Ile Ser Pro Glu Glu Ile Gln Pro Asn Pro
305                 310                 315                 320

Pro Ser Ser Gly Met Leu Gly Ile Ile Ile Met Met Thr Leu Cys Asp
                325                 330                 335

Gln Val Asp Ile Tyr Glu Phe Leu Pro Ser Lys Arg Arg Thr Asp Val
            340                 345                 350

Cys Tyr Tyr Tyr Gln Lys Phe Phe Asp Ser Ala Cys Thr Met Gly Ala
        355                 360                 365

Tyr His Pro Leu Leu Phe Glu Lys Asn Leu Val Lys His Leu Asn Gln
    370                 375                 380

Gly Thr Asp Glu Asp Ile Tyr Leu Leu Gly Lys Ala Thr Leu Pro Gly
385                 390                 395                 400

Phe Arg Thr Ile His Cys
                405


<210> SEQ ID NO 151
<211> LENGTH: 1221
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<400> SEQUENCE: 151

atgattcaca ccaacctgaa gaaaaagttc agctgctgcg tcctggtctt tcttctgttt     60

gcagtcatct gtgtgtggaa ggagaagaag aaagggagtt actatgattc ctttaaattg    120

caaaccaagg aattccaggt gttaaagagt ctggggaaat tggccatggg gtctgattcc    180

cagtctgtat cctcaagcag cacccaggac ccccacaggg gccgccagac cctcggcagt    240

ctcagaggcc tagccaaggc caaaccagag gcctccttcc aggtgtggaa caaggacagc    300

tcttccaaaa accttatccc taggctgcaa aagatctgga agaattacct aagcatgaac    360

aagtacaaag tgtcctacaa ggggccagga ccaggcatca agttcagtgc agaggccctg    420

cgctgccacc tccgggacca tgtgaatgta tccatggtag aggtcacaga ttttcccttc    480
```

-continued

aatacctctg aatgggaggg ttatctgccc aaggagagca ttaggaccaa ggctgggcct	540
tggggcaggt gtgctgttgt gtcgtcagcg ggatctctga agtcctccca actaggcaga	600
gaaatcgatg atcatgacgc agtcctgagg tttaatgggg cacccacagc caacttccaa	660
caagatgtgg gcacaaaaac taccattcgc ctgatgaact ctcagttggt taccacagag	720
aagcgcttcc tcaaagacag tttgtacaat gaaggaatcc taattgtatg ggacccatct	780
gtataccacc cagatatccc aaagtggtac cagaatccgg attataattt ctttaacaac	840
tacaagactt atcgtaagct gcaccccaat cagccctttt acatcctcaa gccccagatg	900
ccttgggagc tatgggacat tcttcaagaa atctccccag aagagattca gccaaacccc	960
ccatcctctg ggatgcttgg tatcatcatc atgatgacgc tgtgtgacca ggtggatatt	1020
tatgagttcc tcccatccaa gcgcaggact gacgtgtgct actactacca gaagttcttc	1080
gatagtgcct gcacgatggg tgcctaccac ccgctgctct ttgagaagaa tttggtgaag	1140
catctcaacc agggcacaga tgaggacatc tacctgcttg gaaaagccac actgcctggc	1200
ttccggacca ttcactgcta a	1221

What is claimed is:

1. An isolated recombinant mammalian or avian cell, comprising a reduced amount of cell surface β-galactoside α2,3 sialyl residues relative to a corresponding non-recombinant mammalian or avian cell.

2. The isolated recombinant cell of claim **1** further comprising an increased amount of human β-galactoside α2,6 sialyl residues relative to a corresponding non-recombinant mammalian or avian cell.

3. The isolated recombinant cell of claim **1** which is a canine or primate cell.

4. The isolated recombinant cell of claim **2** which comprises an expression cassette encoding human β-galactoside α2,6 sialyltransferase I (ST6Gal-I) or ST6Gal-II.

5. The isolated recombinant cell of claim **1** wherein one or more β-galactoside α2,3 sialyltransferase genes are mutated so as to reduce the amount of the cell surface β-galactoside α2,3 sialyl residues.

6. The recombinant cell of claim **1** wherein the reduction in cell surface β-galactoside α2,3 sialyl residues is the result of reduced expression of one or more ST3 sialyltransferases.

7. The recombinant cell of claim **1** wherein one or more of ST3Gal-I, ST3Gal-II, ST3Gal-III, ST3Gal-IV, ST3Gal-V, ST3Gal-VI, or ST3Gal-II-like genes are mutated in the recombinant cell.

8. A method of modifying the amount of cell surface β-galactoside α2,3 sialyl residues and human β-galactoside α2,6 sialyl residues on a mammalian or an avian cell, comprising:

mutating one or more β-galactoside α2,3 sialyltransferase (ST3Gal) genes, and overexpressing a human β-galactoside α2,6 sialyltransferase (ST6Gal) gene, in a parental mammalian or avian cell so as to result in a modified mammalian or avian cell having a reduced amount of cell surface β-galactoside α2,3 sialyl residues and an increased amount of human β-galactoside α2,6 sialyl residues on the surface of the modified cell relative to the corresponding parental cell.

9. The method of claim **8** wherein the mutations include one or more nucleotide insertions or one or more nucleotide deletions, or both, in one or more ST3 genes.

10. The method of claim **8** wherein the modified cell comprises an expression cassette comprising a ST6Gal open reading frame.

11. A method of detecting or propagating an influenza virus, comprising:

infecting the recombinant cell of claim **2** with a sample having or suspected of having an influenza virus.

12. The method of claim **11** further comprising collecting progeny virus.

13. The method of claim **11** wherein the sample is from an avian or a mammal suspected of being infected with an influenza virus.

14. The method of claim **11** wherein the influenza virus is a human influenza virus.

15. The method of claim **11** wherein the influenza virus is an influenza A virus.

16. The method of claim **11** wherein the influenza virus is an influenza B virus.

17. The method of claim **11** wherein the influenza virus is a H3 virus.

18. The method of claim **11** wherein the influenza virus is A/H1N1, A/H3N2, a B/Yamagata-lineage influenza B virus or a B/Victoria-lineage influenza B virus.

19. The method of claim **11** further comprising detecting whether the sample is infected with an influenza virus.

20. The method of claim **19** further comprising identifying the HA and/or NA subtype of the virus.

* * * * *