(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 November 2014 (13.11.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/183071 A2

(51) International Patent Classification:
C12N 15/85 (2006.01) *C12N 5/078* (2010.01)

(21) International Application Number:
PCT/US2014/037554

(22) International Filing Date:
9 May 2014 (09.05.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/822,071 10 May 2013 (10.05.2013) US

(71) Applicant: **WHITEHEAD INSTITUTE FOR BIOMED¬ICAL RESEARCH** [US/US]; Nine Cambridge Center, Cambridge, MA 02142 (US).

(72) Inventors: **LODISH, Harvey;** 195 Fisher Avenue, Brookline, MA 02446 (US). **PLOEGH, Hidde, L.;** 1 Toxteth Street, Brookline, MA 02445 (US). **LEE, Hsiang-Ying;** 26 John F. Kennedy Street, Apt. 6, Cambridge, MA 02138 (US). **SHI, Jiahai;** 11 Peabody Terrace, Apt. 2202, Cambridge, MA 02138 (US). **KUNDRAT, Lenka;** 10 Roberts Street, Apt. 34, Somerville, MA 02145 (US). **PISHESHA, Novalia;** 70 Pacific Street, Apt. 285A, Cambridge, MA 02139 (US).

(74) Agent: **BAKER, C , Hunter;** Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (US).

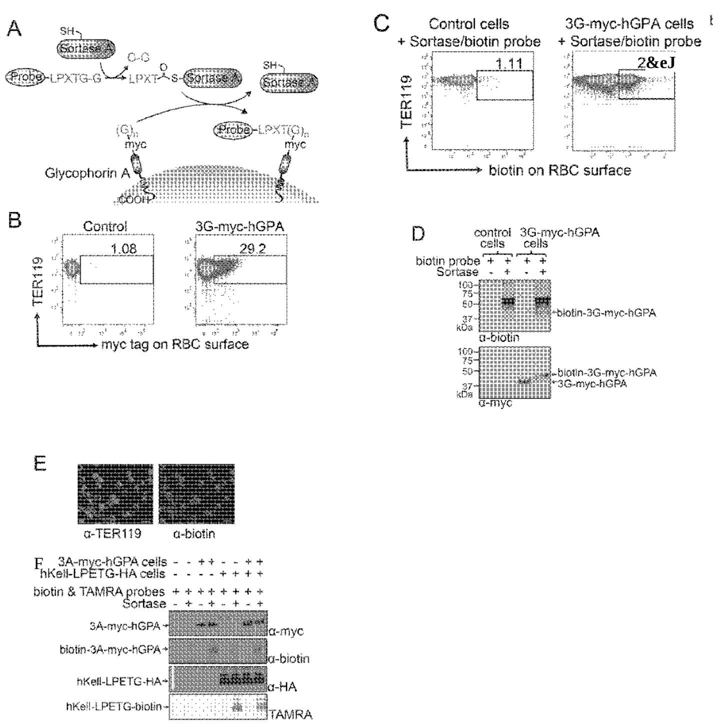
(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV,

[Continued on next page]

(54) **Title:** IN VITRO PRODUCTION OF RED BLOOD CELLS WITH SORTAGGABLE PROTEINS

Figure 16



(57) Abstract: Methods for the in vitro production of enucleated red blood cells and the enucleated red blood cells thus prepared are provided. Such enucleated red blood cells may express a sortaggable surface protein, which allows for surface modification in the presence of a sortase. Also described herein are surface modified enucleated red blood cells, e.g., conjugated with an agent of interest such as a peptide, a detectable label, or a chemotherapeutic agent, and uses thereof in delivering the agent to a subject.

WO 2014/183071 A2

MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

IN VITRO PRODUCTION OF RED BLOOD CELLS WITH SORTAGGABLE PROTEINS

RELATED APPLICATION

This application claims priority under 35 U.S.C. § 119(e) to U.S. Provisional Application, U.S.S.N. 61/822,071, filed on May 10, 2013, the content of which is hereby incorporated by reference in its entirety.

GOVERNMENT SUPPORT

This invention was made with government support under Grant No. HR001 1-12-2-0015, awarded by the Defense Advanced Research Projects Agency (DARPA). The Government has certain rights in this invention.

BACKGROUND OF THE INVENTION

Cell surfaces can be modified in order to modulate surface function or to confer new functions to such surfaces. Surface functionalization may, for example, include an addition of a detectable label, a binding moiety, or a therapeutic agent to a surface protein, allowing for the detection or isolation of the surface-modified cells, for the generation of new cell-cell interactions that do not naturally occur, or for the conjugation of a therapeutic or diagnostic agent to the cell surface.

Cell surface modification can be achieved by genetic engineering or by chemical modifications of cell surface proteins. Both approaches are, however, limited in their capabilities, for example, in that many surface proteins do not tolerate insertions above a certain size without suffering impairments in their function or expression, and in that many chemical modifications require non-physiological reaction conditions.

SUMMARY OF THE INVENTION

The present disclosure is based on the development of an ***in vitro*** multi-phase culturing process for differentiating human $CD34^+$ peripheral blood cells into enucleated red blood cells. This ***in vitro*** culturing process unexpectedly synchronized erythroid expansion, terminal differentiation, and enucleation of mobilized human $CD34^+$ peripheral blood cells. The present disclosure is also based on the development of a sortagging system for surface modification of red blood cells, in which an agent of interest is conjugated to the surface of the enucleated red blood cells. The enucleated red blood cells can be produced by the *in vitro* multi-phase culturing system described herein. Such surface modified enucleated red blood cells can be used for diagnostic and therapeutic purposes.

Accordingly, one aspect of the present disclosure features a method for producing human enucleated red blood cells, the method comprising: (i) providing a population of human $CD34^+$ progenitor cells (*e.g.*, human $CD34^+$ peripheral blood cells); (ii) expanding the population of human $CD34^+$ progenitor cells in a first medium for 0-6 days (*e.g.,* 4 days), wherein the first medium comprises Flt-3 ligand, stem cell factor (SCF), interleukin 3 (IL-3), and interleukin 6 (IL-6); (iii) differentiating the expanded human $CD34^+$ progenitor cells obtained from step (ii) in a second medium for 4-7 days (*e.g.,* 5 days), wherein the second medium comprises dexamethasone, β-estradiol, IL-3, SCF, and erythropoietin (EPO); (iv) differentiating the cells obtained from step (iii) in a third medium for 3-5 days (*e.g.,* 4 days), wherein the third medium comprises SCF and EPO; (v) differentiating the cells obtained from step (iv) in a fourth medium for 4- 12 days (*e.g.,* 9 days) to produce human enucleated red blood cells, wherein the fourth medium comprises EPO; and (vi) collecting the human enucleated red blood cells obtained from step (v). In some embodiments, the total time period for steps (ii)-(v) of the method described above ranges from 11-25 days (*e.g.,* 21 days). Step (v) may be performed in a culturing container coated with a component from the extracellular matrix, such as fibronectin.

In some embodiments, one or more of the second, third, and fourth media can further comprise holo human transferrin and insulin. When necessary, all of the second, third, and fourth media further comprise these two cytokines. In other embodiments, the second, third, and/or fourth media may be free of certain cytokines, for example, thrombopoietin (TPO), granulocyte macrophage colony-stimulating factor (GM-CSF), or both.

In one example, the second medium used in any of the *in vitro* production methods described herein can comprise 250-1500 μg/ml (*e.g.,* 500 μg/ml) holo human transferrin, 5-20 μg/ml insulin (*e.g.,* 10 μg/ml), 100nM~5 μM (*e.g., 2* μM) dexamethasone, 0.5-5 μM (*e.g.,* 1 μM) β-estradiol, 1-10 ng/ml (*e.g.,* 5 ng) IL-3, 10-500 ng/ml (*e.g.,* 100 ng/ml) SCF, and/or 2-10 U (*e.g.,* 6 U) EPO. The second medium can be free of certain cytokines, for example, Flt-3, IL-6, or both.

In another example, the third medium used in any of the *in vitro* production methods described herein can comprise 250-1500 μg/ml *(e.g.,* 500 μg/ml) holo human transferrin, 5-20 μg/ml *(e.g.,* 10 μg/ml) insulin, 10-100 ng/ml *(e.g.,* 50 ng/ml) SCF, and/or 2-10 U *(e.g.,* 6 U) EPO. The third medium may be may be free of certain cytokines, for example, Flt-3, IL-6, dexamethasone, β-estradiol, or any combination thereof..

In yet another example, the fourth medium used in any of the *in vitro* production methods described herein can comprise 250-1500 μg/ml *(e.g.,* 500 μg/ml) holo human transferrin, 5-20 μg/ml *(e.g.,* 10 μg/ml) insulin, and/or 0.5-3 U *(e.g., 2* U) EPO. The fourth medium may be free of certain cytokines, for example, Flt-3, IL-6, dexamethasone, β-estradiol, SCF, or any combination thereof.

In any of the *in vitro* production methods described herein, the human $CD34^+$ progenitor cells can be any of the genetically engineered enucleated blood cells (which are not naturally occurring) as described herein, for example, a human $CD34^+$ expressing a fusion protein comprising a red blood cell membrane protein and a peptide of interest. In some embodiments, the fusion protein comprises a type I red blood cell transmembrane protein *(e.g.*, glycophorin A) fused to an acceptor peptide at the N-terminus of the type I red blood cell transmembrane protein. The acceptor peptide may include an oligoglycine moiety, *e.g.*, a 1-5 glycine fragment, or an oligoalanine *(e.g.,* a 1-5 alanine fragment) moiety. In other embodiments, the fusion protein comprises a type II red blood cell transmembrane protein *(e.g.*, Kell or CD71) fused to a peptide comprising a sequence recognized by a sortase *(e.g.*, *a* sortase A) at the C-terminus of the type II red blood cell transmembrane protein. The sequence recognizable by the sortase can be LPXTG (SEQ ID NO: 1), in which X is any amino acid residue. In still other embodiments, the membrane protein is a type III red blood cell transmembrane protein, such as glucose transporter 1 (GLUT1).

In another aspect, the present disclosure provides a genetically engineered enucleated blood cell which expresses on the surface a first fusion protein comprising a first peptide of interest and a first red blood cell membrane protein. In some embodiments, the first red blood cell membrane protein is a type I red blood cell transmembrane protein *(e.g.,* glycophorin A such as human glycophorin A), and the first peptide of interest is fused to the N-terminus of the type I red blood cell transmembrane protein. In other embodiments, the first red blood cell membrane protein is a type II transmembrane protein *(e.g.,* Kell or CD71), and the first peptide of interest is fused to the C-terminus of the type II transmembrane protein. The first peptide of interest may comprise a sequence recognizable by a sortase, such as sortase A. In one example, the sequence recognizable by the sortase is LPXTG (SEQ ID NO: 1), in which X is any amino acid residue. In still other embodiments, the first red blood cell membrane protein is a type III red blood cell transmembrane protein, such as GLUT1.

In some embodiments, the first fusion protein further comprises a protein of interest *(e.g.,* a cytoplasmic protein, which can be a diagnostic agent or a therapeutic agent). The protein of interest is fused to the terminus of the first red blood cell membrane protein that is exposed to a cytoplasmic space and the first peptide of interest is fused to the terminus of the first red blood cell membrane protein that is exposed to an extracellular or luminal space.

In some examples, the first red blood cell membrane protein can be a type I membrane protein *(e.g.,* a GPA). The protein of interest is fused to the C-terminus of the type I membrane protein and the first peptide of interest is fused to the N-terminus of the type I membrane protein.

In other examples, the first red blood cell membrane protein is a type II membrane protein. The protein of interest is fused to the N-terminus of the type II membrane protein, and the first peptide of interest is fused to the C-terminus of the type I membrane protein.

In yet other embodiments, the genetically engineered enucleated blood cell as described herein can further express on the surface a second fusion protein comprising a second peptide of interest and a second red blood cell membrane protein.

In some examples, the first peptide of interest in the first fusion protein comprises a recognizable site or an acceptable peptide of a first sortase and the second peptide of interest in the second fusion protein comprises a recognizable site or an acceptable peptide of a second sortase. The first sortase and the second sortase use different recognizable sites and different acceptable peptides. In one example, the first peptide of interest comprises the motif of LPXTA (SEQ ID NO:2), in which X is any amino acid residue, or an oligoalanine; and the second peptide of interest comprises the motif LPXTG (SEQ ID NO: 1), or an oligoglycine *(e.g.,* consisting of 1-5 glycine residues). In another example, the first fusion protein comprises Kell, the C-terminus of which is fused to a first peptide of interest comprising the motif LPXTG (SEQ ID NO:l), and the second fusion protein comprises GPA, the C-terminus of which is fused to a second peptide of interest comprising an oligoalanine *(e.g.,* consisting of 1-5 alanine residues).

In any of the genetically engineered enucleated blood cells described herein that express two fusion proteins on the surface, either one or both of the fusion proteins can be conjugated to two different functional moieties.

The above described enucleated blood cell may be prepared by any of the *in vitro* culturing methods described herein.

In some instances, the first peptide of interest, the second peptide of interest, or both, that are fused to the red blood cell membrane protein(s) described herein may comprise a protein drug *(e.g.,* an antibody or an antigen-binding fragment thereof, which can be a single domain antibody), a vaccine antigen, a fluorescent protein, streptavidin, biotin, an enzyme, or a peptide capable of targeting a cell *(e.g.,* a disease cell). In other instances, the peptide of interest is conjugated to a detectable label or a chemotherapeutic agent. For example, the peptide of interest may be conjugated to a lipid, a carbohydrate, a nucleic acid, a binding agent, a click-chemistry handle, a polymer, a peptide, a protein, a metal, a chelator, a radiolabel, or a small molecule.

In yet another aspect, the present disclosure provides methods for delivering an agent to a subject, the method comprising administering any of the enucleated blood cells described herein to the subject. In some examples, the enucleated blood cell being delivered is derived from the same subject the cell is being delivered to.

Also within the scope of the present disclosure are methods for conjugating any of the peptides of interest described herein to the surface of red blood cells. This method comprises: (i) providing a red blood cell expressing a fusion protein comprising a membrane protein and a first peptide, and (ii) contacting the red blood cell with a peptide of interest in the presence of a sortase *(e.g.,* sortase A) under conditions suitable for the sortase to conjugate the peptide of interest to the first peptide in the fusion protein. Either the first peptide in the fusion protein or the peptide of interest comprises a sequence recognized by the sortase. In one example, the sequence recognizable by the sortase is LPXTG (SEQ ID NO:l), in which X is any amino acid residue.

In some embodiments, the membrane protein is a type I red blood cell transmembrane protein *(e.g.,* glycophorin A) and the first peptide is an acceptor peptide *(e.g.,* including an oligoglycine fragment such as a 1-5 glycine fragment) fused to the N-terminus of the type I red blood cell transmembrane protein, wherein the peptide of interest comprises the sequence recognized by the sortase. In other embodiments, the membrane protein is a type II red blood cell transmembrane protein *(e.g.,* Kell or CD71) and the first peptide is fused to the C-terminus of the type II red blood cell transmembrane protein, wherein the first peptide comprises the sequence recognizable by the sortase. In still other embodiments, the membrane protein is a type III red blood cell transmembrane protein such as GLUT1.

In some embodiments, the first fusion protein in any of the methods described herein may further comprise a protein of interest *(e.g.*, a cytoplasmic protein, which can be a diagnostic agent or a therapeutic agent), which is fused to the terminus of the first red blood cell membrane protein that is exposed to a cytoplasmic space. The first peptide of interest in the fusion protein is fused to the terminus of the first red blood cell membrane protein that is exposed to an extracellular or luminal space.

In some examples, the first red blood cell membrane protein is a type I membrane protein. The first protein of interest is fused to the C-terminus of the type I membrane protein, and the first peptide is rased to the N-terminus of the type I membrane protein. In other examples, the first red blood cell membrane protein is a type II membrane protein. The protein of interest is fused to the N-terminus of the type II membrane protein, and the first peptide is fused to the C-terminus of the type I membrane protein.

In yet other embodiments, any of the methods described herein can further comprise contacting the red blood cell in the presence of a second sortase under conditions suitable for the second sortase to conjugate a second peptide of interest to a second fusion protein expressed on the surface of the red blood cell. The second fusion protein comprises a second red blood cell membrane protein and a second peptide to which the second peptide of interest conjugates. In some examples, either the second peptide in the fusion protein or the second peptide of interest comprises a sequence recognizable by the second sortase. The sequence recognizable by the first sortase differ from the sequence recognizable by the second sortase.

In some examples, the first peptide in the first fusion protein comprises a sequence recognizable by the first sortase or an acceptable peptide of the first sortase; and/or the second peptide in the second fusion protein comprises a sequence recognizable by the second sortase or an acceptable peptide of the second sortase. The acceptable peptide of the first sortase is different from the acceptable peptide of the second sortase. The first sortase may be Sortase A from *Staphylococcus arreus.* One of the first peptide in the first fusion protein and the first peptide of interest comprises the motif LPXTG (SEQ ID NO:l), in which X is any amino acid residue, and the other comprises an acceptable peptide, which is an oligoglycin. The first fusion protein may comprise Kell, the C-terminus of which is fused to the first peptide which comprises the motif LPXTG (SEQ ID NO:l), and the first peptide of interest comprises an oligoglycine *(e.g.,* consisting of 1-5 glycine residues).

In some examples, the second sortase is Sortase A from *Streptococcus pyogenes.* One of the second peptide in the second fusion protein and the second peptide of interest may comprise the motif LPXTA (SEQ ID NO:2), in which X is any amino acid residue, and the other may comprise an acceptable peptide, which can be an oligoalanine *(e.g.,* consisting of 1-5 alanine residues). The second fusion protein may comprise GPA, the N-terminus of which is fused to the second peptide which comprises an oligoglycine. The second peptide of interest may comprise the motif LPXTG (SEQ ID NO:l), in which X is any amino acid residue.

In some embodiments, the first and second peptides of interest comprises or are **conjugated** to two different functional moieties. **Alternatively** or in addition, the first peptide of interest, the second peptide of interest, or both, may be a protein drug, a vaccine antigen, a fluorescent protein, streptavidin, biotin, an enzyme, or a peptide capable of targeting a cell. In one example, one of the first and second peptide of interest is a peptide capable of targeting a disease cell.

Further, the present disclosure provides (a) pharmaceutical compositions for diagnostic or therapeutic uses, the pharmaceutical composition comprising any of the enucleated red blood cells described herein, which carry any of the peptide of interest as described herein, and a pharmaceutically acceptable carrier, and (b) uses of the pharmaceutical compositions for manufacturing medicaments for diagnostic or therapeutic purposes.

The details of one or more embodiments of the invention are set forth in the description below. Other features or advantages of the present invention will be apparent from the following Drawings and Detailed Description of Certain Embodiments, and also from the Claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates an exemplary *in vitro* culturing process for producing enucleated red blood cells from human CD34$^+$ progenitor cells.

Figure 2 includes photographs showing morphology of cells at expansion and different differentiation stages in the *in vitro* culturing process described herein.

Figure 5 is a diagram showing the expression of cell surface markers glycophorin A (CD235) and c-kit in cells at expansion and different differentiation stages in the *in vitro* culturing process described herein, as determined by FACS analysis.

Figure 4 is a diagram showing the expression of cell surface markers glycophorin A (CD235) and transferrin receptor (CD71) in cells at expansion and different differentiation stages in the *in vitro* culturing process described herein, as determined by FACS analysis.

Figure 5 is a diagram showing the enucleation of cells at expansion and different differentiation stages in the *in vitro* culturing process described herein.

Figure 6 includes charts showing cell proliferation at different differentiation stages.

Figure 7 is a chart showing cell proliferation in a window of around 20 days.

Figure 8 include charts showing the expression of globin genes as indicated during hCD34$^+$ cell differentiation.

Figure 9 includes charts showing the expression of various genes as indicated during hCD34$^+$ cell differentiation.

Figure 10 is a diagram showing construction of EFl and MSCV expression vectors for producing 5Gly (SEQ ID NO: 3)-myc-human glycophorin A (hGYPA or hGPA) fusion protein.

Figure 11 is a diagram showing surface expression of sortaggable hGYPA on hCD34$^+$ erythroid progenitors (at different differentiation stages) using EFl expression vectors; 5Gly: (SEQ ID NO:3).

Figure 12 is a diagram showing surface expression of sortaggable hGYPA on hCD34$^+$ erythroid progenitors (at different differentiation stages) using MSCV expression vectors; 5Gly: (SEQ ID NO:3).

Figure 13 is a photograph showing conjugation of biotin to hGYPA on hCD34$^+$ erythroid progenitor cells by sortagging as determined by Western blotting; 5G: (SEQ ID NO:3).

Figure 14 is a diagram showing conjugation of biotin to the surface of human CD34$^+$ cells at the terminal differentiation stage via sortagging; 5Gly: (SEQ ID NO:3).

Figure 15 is a diagram showing conjugation of biotin to the surface of human CD34$^+$

cells at the terminal differentiation stage via sortagging; 5Gly: (SEQ ID NO:3).

Figure 16 is a diagram showing the expression and sortase-mediated N-terminal labeling of glycophorin A on the surface of mature mouse red blood cells. *(A)* Schematic for glycophorin A N-terminal sortase labeling. GPA was extended at the N-terminus to include 3 glycine residues and a myc tag. Pre-incubation of sortase with a probe, which contains a C-terminal sortase recognition motif, LPETG (SEQ ID NO:44). leads to cleavage between T and G and formation of an acyl enzyme intermediate between Cys on sortase and Thr on the probe. Nucleophilic attack of the N-terminal glycine on GPA resolves the intermediate, thus ligating the probe to GPA. The probe can be a peptide, protein, lipid, carbohydrate, small molecule, etc. For conjugation of biotin to GPA, the probe is K(biotin)LPRTGG (SEQ ID NO:45); LPXTGG: (SEQ ID NO:46). (5) Evaluation of mature RBCs for the presence of 3G-myc-hGPA on the cell surface by staining either control blood or blood from mice that have undergone bone marrow transplantation to express 3G-myc-hGPA, with a-TERl 19 and a-myc tag antibodies and analyzing via flow cytometry. (*C*) Evaluation of mature RBCs for sortase-iabeling by incubating control blood or 3G-myc-hGPA blood with sortase and the biotin containing probe, staining with a-TERl 19 and a-biotin antibodies, and analyzing via flow cytometry. (*D)* Evaluation of RBCs for sortase-iabeling by immunoblotting. Control or 3G-myc-hGPA blood was incubated with biotin probe with or without sortase, and total cell protein was resolved by SDS-PAGE and immunoblotted for a-myc tag and a-biotin. The shift in molecular weight of GPA upon biotin conjugation in the a-myc tag immunoblot indicates almost complete modification. Biotin conjugated GPA is indicated in the a-biotin immunoblot with an arrow. *(E)* RBCs conjugated with biotin were sorted by flow cytometry. Immunofluorescence images show biotin (labeled with red) at the N-terminus of hGPA on mature RBCs (labeled with Terl 19 antibody, purple). (*F*) HEK293T cells were transfected with 3A-myc-hGPA, hKell-LPETG (SEQ ID NO:44)-HA, or both. Cells were incubated with *S. pyogenes* sortase and a biotin probe followed by *S. aureus* sortase and TAMRA-containing probe incubation. Specific conjugation of biotin to GPA and of TAMRA to Kell is demonstrated by immunoblotting and fluorescence imaging.

Figure 17 is a diagram showing the overexpression of modified human GPA and mouse GPA containing myctag at N-termini do not affect *in vitro* differentiation of mouse erythroid progenitor. (*A)* Flow cytometry analysis of the differentiation capacity of *in vitro* differentiated murine fetal-liver derived progenitor cells infected with retroviral constructs containing modified human (h) or mouse (m) GPA constructs extended at their N-terminus with myc-tag: myc-hGPA and myc-mGPA. Differentiation capacity in these cells is assessed based on the expressions of Terl 19 and enucleation, i.e. nuclei expulsion, resulting in Terl 19$^+$ reticulocytes. The percentages of reticulocytes produced by erythroid progenitor cells infected with myc-hGPA or myc-mGPA (~20%) are comparable to cells infected with empty (control) vector, indicating that these constructs do not disturb erythroid terminal differentiation. Quantification of enucleation rate represents 3 independent experiments and graphed as mean values +/- standard deviation. (5) Evaluation of murine terminally differentiated erythroblasts, i.e. nucleated erythroblasts and reticulocytes, for the surface expression of myc-hGPA or myc-mGPA. More than 60% of these terminally differentiated cells expressed the desired modified GPA proteins as measured by flow cytometry using a-myc tag antibodies. Percentage of erythroblasts and reticulocytes with myc-tag on cell surface was determined from 3 independent experiments, graphed as mean value $^{+/-}$ standard deviation; ** indicates $p < 0.01$. (*C*) Immunofluorescence images further confirm the surface expression of myc-tag (labelled with red) and enucleation capacity (blue nucleus staining) of the *in vitro* terminally differentiated erythroblasts.

Figure 18 is a diagram showing the overexpression of engineered human GPA and mouse GPA containing myc-tag with multiple (3 or 5) glycines (SEQ ID NO:3) at their N-termini do not affect *in vitro* differentiation of mouse erythroid progenitors. Only cells expressing engineered human GPA with myc-tag and 3 glycines at the N-terminus can be efficiently biotin-labeled by sortagging. *(A)* Flow cytometry analysis on the differentiation capacity of *in vitro* differentiated murine fetal-liver derived progenitor cells infected with retroviral constructs containing modified human (h) or mouse (m) GPA constructs extended at their N-terminus with myc-tag and multiple (3 or 5) glycines (SEQ ID NO:3) (G): 5G (SEQ ID NO:3)-myc-hGPA, 5G (SEQ ID NO:3)-mycmGPA, and 3G-myc-hGPA. Differentiation capacity in these cells is assessed based on the expressions of Terl 19 and enucleation, *i.e.* nuclei expulsion, resulting in Terl 19$^+$ reticulocytes. See upper panel. The percentages of reticulocytes produced by erythroid progenitor cells infected with 5G (SEQ ID NO:3)-myc-hGPA, 5G-myc-mGPA, and 3G-myc-hGPA (-17.5%) are comparable to cells infected with empty (control) vector, indicating that these constructs do not disturb erythroid terminal differentiation. Quantification of enucleation rate represents 3 independent experiments and graphed as mean values $^{+/-}$ standard deviation. See lower panel. *(B)*

Evaluation of murine terminally differentiated erythroblasts, *i.e.*, nucleated erythroblasts and reticulocytes, for the surface expression of 5G (SEQ ID NO:3)-myc-hGPA, 5G (SEQ ID NO:3)-mycmGPA, and 3G-myc-hGPA. More than 50% of these terminally differentiated cells expressed the desired modified GPA proteins as measured by flow cytometry using a-myc tag antibodies. See upper panel. Percentage of erythroblasts and reticulocytes with myc-tag on cell surface was determined from 3 independent experiments, graphed as mean value $^{+/-}$ standard deviation; ** indicates $p < 0.01$. See lower panel. (*C*) HEK 293T cells were transfected to express 5G (SEQ ID NO:3)-myc-hGPA, 5G (SEQ ID NO:3)-myc-mGPA, 3G-myc-mGPA, and 3G-myc-hGPA. These cells were then incubated with a biotin probe with or without sortase, and total cell protein was immunoblotted for the myc-tag and biotin. Immunoblot stained for biotin shows successful biotin conjugation only to human GPA constructs 3G-myc-hGPA (strong band) and 5G (SEQ ID NO:3)-myc-hGPA (weak band) with a much higher efficiency for 3Gmyc (strong band) and 5G (SEQ ID NO:3)-myc-hGPA (weak band) with a much higher efficiency for 3Gmyc-hGPA. a-myc-tag immunoblotting further confirms biotin probe conjugation as indicated by a shift in hGPA molecular weight upon biotin conjugation.

Figure 19 is a diagram showing the expression and sortase-mediated N-terminal labeling of glycophorin A on the surface of *in vitro* differentiated erythroblasts. *(A)* Flow cytometry analysis of *in vitro* differentiated erythroblasts for the presence of 3G-myc-hGPA and for sortase-labelled biotin-3G-myc-hGPA on the cell surface by staining with a-myc and a-biotin tag antibodies. See upper panel. Percentage of erythroblasts and reticulocytes with myc-tag on cell surface and percentage of myc-tag positive cells sortagged with a biotin probe was determined from 3 independent experiments, graphed as mean value $^{+/-}$ standard deviation; ** indicates $p < 0.01$. See lower panel. *(B)* Evaluation of *in vitro* differentiated erythroblasts for sortase-labeling by incubating control or 3G-myc-hGPA erythroblasts with sortase and the biotin containing probe by means of immunoblotting with a-myc tag and a-biotin antibodies. The shift in molecular weight of hGPA upon biotin conjugation in the a-myc blot indicates complete modification of hGPA. Biotin-conjugated GPA is denoted by an arrow in the a-biotin immunoblot. (*C*) Immunofluorescence shows biotin (labelled with red) at the N-terminus of hGPA on the surface of differentiated erythroblasts. Nucleus was stained by Hoechst (blue). The arrow indicates reticulocytes, while the arrowheads indicate enucleating erythroblasts. The scale bar is 10 μπι.

Figure 20 is a schematic illustration of engineered RBC production by mouse bone marrow/fetal liver cell transplantation. Isolated murine bone marrow cells or fetal liver cells were transduced with a retrovirus carrying hKell-LPETG (SEQ ID NO:44)-HA or 3G-myc-hGPA. Lethally irradiated mice were reconstituted with transduced bone marrow or fetal liver cells. After 1 month and onward, a significant percentage of mature RBCs are derived from transduced donor progenitors.

Figure 21 is a diagram showing expression and sortase-mediated C-terminal labeling of Kell on the surface of mature mouse red blood cells. *(A)* Schematic for Kell C-terminal sortase labeling. Kell was extended at the C-terminus to include the sortase recognition motif LPETG (SEQ ID NO: 44) followed by an HA epitope tag. Incubation of sortase with cells containing Kell-LPETG (SEQ ID NO:44)-HA on the surface leads to peptide cleavage between T and G in the recognition motif and the formation of an acyl-enzyme intermediate between Cys on sortase and the sortase motif on Kell. Addition of a nucleophilic probe with N-terminal glycine residues (GGG-K(biotin); SEQ ID NO: 47; for biotin conjugation) resolves the intermediate, thus iigaiing the probe to the C-terminus of Kell. The probe can be a peptide, protein, lipid, carbohydrate, small molecule, etc; LPXTG (SEQ ID NO:l). *(B)* Evaluation of mature RBCs for the presence of Kell-LPETG (SEQ ID NO:44)-HA on the cell surface by staining either control blood or blood from mice that have undergone bone marrow transplantation to express Kell-LPETG (SEQ ID NO:44)-HA with a-HA tag antibody and analyzing via flow cytometry (gated on mature RBCs). (*C*) Evaluation of mature RBCs for sortase-labeling by incubating control blood or Kell-LPETG (SEQ ID NO:44)-HA blood with sortase and the biotin containing probe, staining with a-biotin antibody, and analyzing via flow cytometry (gated on mature RBCs). (*D)* Evaluation of RBCs for sortase-labeling by immune-blotting. Control or Kell-LPETG (SEQ ID NO:44)-HA blood was incubated with biotin probe with or without sortase. Total cell protein was resolved by SDS-PAGE and immunoblotted for a-HA tag and a-biotin. Loss of HA tag upon sortase labeling indicates complete modification. *(E)* Immunofluorescence images show biotin conjugated to the C-terminus of hKell on mature RBCs (labeled with Terl 19 antibody).

Figure 22 is a diagram showing expression and sortase-mediated C-terminal labeling of Kell on the surface of *in vitro* differentiated erythroblasts. *(A)* Flow cytometry of *in vitro* differentiated erythroblasts for the presence of hKell-LPETG (SEQ ID NO: 44)-HA and for sortase-labelled hKell-LPETG (SEQ ID NO: 44)-biotin on the cell surface by staining with α-HA and α-biotin tag antibodies. See upper panel. Percentage of erythroblasts and reticulocytes with HA-tag on cell surface and percentage of HA-tag positive cells sortagged with a biotin probe was determined from 3 independent experiments, graphed as mean value +/- standard deviation; ** indicates $p < 0.01$. See lower panel. *(B)* Evaluation of *in vitro* differentiated erythroblasts for sortase-labeling by incubating control or hKell-LPETG (SEQ ID NO: 44)-HA erythroblasts with sortase and the biotin containing probe followed by immunoblotting for biotin and Kell. (*C*) Immunofluorescence shows biotin (labelled with red) at the C-terminus of hKell on the surface of differentiated erythroblasts. Nucleus was stained by Hoechst (blue). The arrow indicates reticulocytes, while the arrowhead indicates enucleating erythroblasts. The scale bar is 10 μηι.

Figure 23 is a diagram showing sortase-mediated dual labeling at the N-terminus of hGPA and C-terminus of hKell on the surface of mature RBCs. Mature RBCs expressing hKell-LPETG (SEQ ID NO: 44)-HA and 3A-myc-hGPA were incubated with sortase A from *S. pyogenes* and K(biotin)LPETAA (SEQ ID NO: 45) probe for N-terminal GPA labeling (second panel), or with sortase A from S. aureus and GGGC(Aiexa-647; SEQ ID NO: 48) probe for C-terminal Kell labeling (third panel), or both sequential label with both sortase enzymes and their corresponding probes (fourth panel). Flow cytometry analysis of mature RBCs for the presence of both hKell-LPET (SEQ ID NO: 49)-Alexa-647 and biotin-myc-hGPA indicate successful dual labeling.

Figure 24 is a number of charts showing survival of engineered red blood cells in vivo. (*A*) Wild-type RBCs subjected to mock C-terminal sortagging reacting with or without sortase (n=3, each group) or left untreated (n=3) were reacted with CFSE and transfused into recipient mice via intravenous injection. Their *in vivo* survival was tracked via CFSE fluorescence and flow cytometry. *(B)* mKell-LPETG(SEQ ID NO: 44) RBCs were obtained from mice genetically engineered to have endogenous Kell extended at is C- terminus with an LPETG (SEQ ID NO: 44) sequence. Wild-type RBCs (n=6), mKell-LPETG (SEQ ID NO: 44) RBCs (n=6), and mKell-LPETG (SEQ ID NO: 44)- RBCs sortagged with biotin (n=6) were stained with CFSE, transfused into recipients, and their survival monitored via CFSE fluorescence or anti-biotin staining using flow cytometry. (C) CFSE-stained wild-type RBCs (n=6) and hKell-LPETG (SEQ ID NO: 44) RBCs isolated from transplanted mice (See Example 5) and sortase- labeled with biotin (n=2) were transfused into recipient mice and their survival in circulation was tracked via CFSE fluorescence (wild-type RBCs) or anti-biotin staining (hKell-LPETG (SEQ ID NO: 44)-biotin RBCs) using flow-cytometry. *(D)* Wild-type RBCs (n=6) and sortase-labeled biotin-3G-myc-hGPA RBCs (n=6) were CFSE stained, transfused into mice and tracked via CFSE fluorescence or anti-biotin staining using flow-cytometry. Internal control RBCs are cells that originate from 3G-myc-hGPA transplanted mice, were subjected to the sortagging reaction and CFSE-staining, but are wild-type. Error bars represent standard deviation; * denotes $p < 0.01$.

Figure 25 is a diagram showing conjugation of single domain antibody to engineered red blood cells and cell type-specific targeting. *(A)* Mouse mature red blood engineered to contain 3G-myc-GPA on the cell surface were incubated with Sortase with or without VHH7-LPETG (SEQ ID NO: 44), a single domain antibody with binding specificity for mouse MHC class II molecules and sortase motif complementary to 3G. An immunoblot against myc epitope tag on glycophorin A reveals fusion of VHH7 to GPA, as an increase in GPA molecular weight. *(B)* Either wild type mouse B cells (expressing MHC class II) or B cells derived from MHC class II knock-out mice were immobilized on magnetic streptavidin beads via biotinyiated a-CDI9 antibody, incubated with 3G-rnyc-GPA red blood cells or red ceils that contained sortase-labeled VHH7-LPETG (SEQ ID NO: 44)-myc GPA on their surface, and washed. Immobilization of B cells in all experimental set-ups was tested by presence of CD19 in a-CD19 immunoblot and binding between B cells and red blood cells was evaluated by the presence of hemoglobin in the a-hemoglobin immunoblot. (*C*) Schematic for experiment shown in *(B)*; LPETG: (SEQ ID NO: 44).

Figure 26 is a diagram showing expression and sortase-mediated labeling of glycophorin A on the surface of human *in vitro* differentiated reticulocytes. *(A)* Human granulocyte - colony stimulating factor- mobilized peripheral blood CD34+ stem cells were virus transduced either with empty vector or a vector containing 3G-myc-GPA-GFP and differentiated for 18 days. Flow cytometry analysis of reticulocytes reveals the percentage of transduced cells(GFP^+) and percentage of enucleated cells (lower panel with Hoechst staining). Reticulocytes, either control or 3G-myc-GPA-GFP (3G-myc-hGPA-GFP-IRES-GFP) transduced, were incubated with sortase and a biotin-containing probe, stained for biotin with a-biotin-PE antibody, analyzed via flow cytometry, and gated on enucleated cells. *(B)* 3G-myc-GPA-GFP transduced and *in vitro* differentiated reticulocytes were incubated with a biotin probe with or without sortase, and total cell protein immunoblotted for the myc tag and biotin. Both monomers as well as dimers (higher molecular weight) of 3G-myc-

GPA-GFP are visible in both immunoblots. A shift in GPA-GFP molecular weight upon biotin conjugation in a-myc tag immunoblot indicates complete modification.

DETAILED DESCRIPTION CERTAIN EMBODIMENTS OF THE INVENTION

Red blood cells are the most numerous cell type in blood and account for a quarter of the total number of cells in the human body. RBCs possesses many unique characteristics that make them an attractive tool in therapeutics and diagnostics for various purposes, *e.g.,* for *in vivo* delivery of natural or synthetic payloads. Yoo et al., 2011, *Nature Reviews. Drug Delivery* 10(7):521-535. For example, mature red blood cells do not have nuclei (i.e., enucleated) and thus there will be no risk of delivering remnants of foreign genes into a host. Thus, the possibility of tumorigenicity a key risk of stem cell-based therapies (Gruen et al., 2006 *Stem cells* 24(10):2162-2169) is thereby eliminated. Further, red blood cells have a long lifespan *in vivo, e.g.,* about 120 days in the human blood stream, and about 50 days in mice, and presence throughout the macro- and micro-circulation. Modification of red cells with bioavailable therapeutics might thus lead to prolonged efficacy and coverage of all areas perfused by the circulation *in vivo.* Moreover, RBCs have large cell surface areas of about 140 μm^2 with a favorable surface to volume ratio. Also, red blood cells have good biocompability when used as carriers for delivery of therapeutic or diagnostic agents. Finally, old or damaged RBCs can be removed by cells of the reticuloendothelial system. Thus, any modification made to the DNA of RBC precursors is eliminated upon their enucleation and cannot lead to abnormal growth or tumorigenicity after their transfusion into a recipient.

Accordingly, it is of great interest to develop methods for producing enucleated red blood cells that carry an agent of interest, such as diagnostic or therapeutic agents. Such enucleated red blood cells can be used for, *e.g.,* delivering the agent of interest into a subject.

Engineered RBCs have been generated using encapsulation (Biagiotti et al., 2011, *IUBMB life* 63(8):621-631; Godfrin et al., 2012, *Expert Opinion on Biological Therapy* 12(1):127-133; and Muzykantov, 2010, *Expert Opinion on Drug Delivery* 7(4):403-427), by non-covalent attachment of foreign peptides, or through installation of proteins by fusion to a monoclonal antibody specific for a RBC surface protein(Murciano, 2003, *Nature Biotechnology* 21(8):891-896; and Zaitsev et al., 2010, *Blood* 115(25):5241-5248). Modified RBCs face limitations if intended for application *in vivo.* Encapsulation allows entrapment of sizable quantities of material, but at the expense of disrupting plasma membrane integrity, with a concomitant reduction in circulatory half-life of the modified red blood cells. Osmosisdriven entrapment limits the chemical nature of materials that can be successfully encapsulated, the site of release is difficult to control, and encapsulated enzymes are functional only at the final destination, compromising reusability at other sites. Murciano et al., 2003 and Zaitsev et al., 2010. Targeting of cargo to RBCs by fusion to an RBC-specific antibody, *(e.g.,* antiglycophorin antibody), has its limitations because this mode of attachment to the RBC is non-covalent and readily dissociates, thus reducing circulatory half life and mass of cargo available for delivery. Murciano et al., 2003 and Zaitsev et al., 2010. Other developments that exploit RBCs for targeted delivery include nanoparticles enveloped by an RBC-mimicking membrane as well as RBC-shaped polymers. Yoo et al., 201 1 *Nature Reviews. Drug Discovery* 10(7):521-535. The short *in vivo* survival rate of these RBC-inspired carriers (~ 7 days maximum) may limit their therapeutic utility.

Another technical difficulty associated with engineering RBCs is lack of a suitable *in vitro* system for cuituring and differentiation of RBCs. Human $CD34^{*}$ cell culture systems have been disclosed in Miharada et al., *Nat. BiotechnoL,* 24(10):1255-1256, 2006; Sankaran et al., *Science,* 322(5909):1839-1842, 2008, and Giarratana et al., *Blood,* 118(19):5071-5079, 2011. However, the mature enucleated red blood cells obtained from these cell culture systems did not show similar hemoglobin contents and/or cell sizes as compared to normal human reticulocytes or red blood cells, did not show synchronized expression of cell surface differentiation markers during the culture process. Also, the cell culture system taught in Sankaran *et al.* did not generate terminally-differentiated erythroid cells or enucleated red cells.

In light of the disadvantages associated with the existing technology, there is a need to develop new methodology for engineering RBCs, such that they can carry a wide variety of useful cargoes to specific locations in the body.

In the present studies, modified red blood cells were developed to serve as carriers for systemic delivery of a wide array of payloads. These RBCs contain modified proteins on their plasma membrane, which can be labeled in a sortase-catalyzed reaction under native conditions without inflicting damage on the target membrane or cell. Sortase accommodates a wide range of natural and synthetic payloads that allow modification of RBCs with substituents that cannot be encoded genetically. The present studies demonstrate the successful site-specific conjugation of biotin to *in vitro* differentiated mouse erythroblasts as well as to mature mouse RBCs. Unexpectedly, these modified red cells remain in the bloodstream for up to 28 days, which is far in excess of red blood cells engineered by methods known in the art. A single domain antibody attached enzymatically to RBCs enables them to bind specifically to target cells that express the antibody target. This study was extended to human red cells and demonstrate unexpected efficient sortase-mediated labeling of *in vitro* differentiated human reticulocytes.

The engineering and labeling processes described herein do not damage the cells or affect their survival *in vivo.* Most importantly, the engineered red blood cells can be labeled with a wide array of functional probes, including small molecules, peptides, and proteins and thus have the potential to be carriers of a variety of therapeutic substances into the bloodstream. Thus, the methods described herein offer advantages over the osmosis-driven encapsulation methods known in the art.

Sortase-mediated cell surface labeling has been exploited previously, for example to explore trafficking of flu glycoproteins. Popp et al, 2012 *PLos Pathogens* 8(3):ei002604; and Sanyal et al., 2013 *Cell Host & Microbe* 14(5):510-521. This methodology was modified and applied to primary cells as a platform for diagnostic or therapeutic purposes. A sortase A variant from S. *aureus* active at 0 °C (Chen et al, 201 1 *PNAS* 108(28):1 1399-11404) to reduce the risk of inflicting cellular damage in the course of labeling.

There are many other possible applications for the methods presented here, where the wide variety of possible payloads -ranging from proteins and peptides to synthetic compounds and fluorescent probes may serve as a guide. For example, such methods would enable the targeting of the modified red cells to a specific cell type. Further, the methods described herein can be combined with established protocols of small molecule encapsulation(Godfrin et al., 2012). In this scenario, engineered red cells loaded in the cytosol with a therapeutic agent and modified on the surface with a cell-type specific recognition module could be used to deliver payloads to a precise tissue/location in the body. It has been demonstrated herein attachment of two different functional probes to the surface of red blood cells, exploiting the subtly different recognition specificities of two distinct sortases. It should therefore be possible to attach both a therapeutic moiety as well as a targeting module to the red cell surface, to direct the engineered red cells to tumors or other diseased cells. Conjugation of an imaging probe, i.e., a radioisotope, together with such a targeting moiety could also be used for diagnostic purposes.

The *in vitro* generation of human RBCs and genetic engineering of their precursors as described herein may provide a robust platform for application of this surface engineering method, e.g., in conjunction with cytosolic modification, to clinical applications(Liu et al., 2010 *Blood* 115(10):2021-2027; Cong et al., 2013 *Science* 339(6121):819-826; Mali et al, *Science* 339(6121):823-826; Giarratana et al., 201 1 *Blood* 118(19):5071-5079; Douay et al, 2009 *Blood* 105(l):85-94; and Griffiths et al., 2012 *Blood* 119(26):6296-6306.) Moreover, the established safety of blood transfusions inspires confidence that these engineered red blood cells will indeed find use in humans.

Accordingly, described herein is an *in vitro* multi-phase culturing process for producing enucleated red blood cells from mobilized $CD34^+$ progenitor cells *(e.g.,* human mobilized $CD34^+$ peripheral blood cells). Such enucleated red blood cells can be genetically engineered such that they express proteins of interest, e.g., sortaggable surface proteins. Also described herein are methods for conjugating one or more agents of interest to the surface of the genetically engineered enucleated red blood cells *via* a sortase-mediated transpeptide reaction, as well as methods for producing cytoplasmically disposed protein of interest in mature RBCs such as human RBCs while retaining the ability to selectively target the RBCs to an intended target via sortagging of a modified membrane protein with a target moiety of interest.

I. *In vitro* Culturing Systems for Producing Enucleated Red Blood Cells

Described herein is an *in vitro* culturing process for producing mature enucleated red blood cells from $CD34^+$ progenitor cells *(e.g.,* from a human subject). This culturing process involves multiple differentiation stages *(e.g., 2,* 3, or more) and optionally an expansion stage prior to the differentiation phases. The total time period for the *in vitro* culturing process described herein can range from 11-25 days *(e.g.,* 15-25 days, 15-20 days, or 18-21 days). In one example, the total time period is 2 1 days.

a. $CD34^+$ progenitor cells

CD34 is a cell surface glycoprotein and functions as a cell-cell adhesion factor. Many human progenitor cells express this cell surface marker. Novershtern et al., *Cell* 144:296-309, 201 1. A progenitor cell, like a stem cell, has a tendency to differentiate into a specific type of cell. Progenitor cells are usually more specific than stem cells and are often pushed to differentiate into the target cells. Any type of CD34+ progenitor cells that possess the tendency of differentiating into red blood cells can be used in the *in vitro* culturing process described herein. Such progenitor cells are well known in the art. See, *e.g.*, Novershtern et al., Cell 144:296-309, 201 1. In some examples, the *in vitro* culturing process described herein utilizes mobilized CD34+ peripheral blood cells as the progenitor cells for differentiation into enucleated red blood cells. CD34+ progenitor cells can also be derived from other sources *(e.g.,* bone marrow).

Various techniques can be used to separate or isolate the $CD34^+$ cell population from a suitable source such as peripheral blood cells. For example, antibodies such as monoclonal antibodies binding to CD34 can be used to enrich or isolate $CD34^+$ cells. The anti-CD34 antibodies can be attached to a solid support such that cells expressing these surface markers are immobilized, thereby allowing for the separation of $CD34^+$ cells from cells that do not express this surface marker. The separation techniques used should maximize the retention of viable cells to be collected. Such separation techniques can result in sub-populations of cells where up to 10%, usually not more than about 5%, preferably not more than about 1%, of the selected cells do not express CD34. The particular technique employed will depend upon the efficiency of separation, associated cytotoxicity, ease and speed of performance, and necessity for sophisticated equipment and/or technical skill.

An "isolated" or "purified" population of $CD34^+$ cells for use in the *in vitro* culturing process described herein is substantially free of cells and materials with which it is associated in nature, in particular, free of cells that lack the desired phenotype, e.g., expressing CD34. Substantially free or substantially purified includes at least 50% $CD34^+$ cells, preferably at least 70%, more preferably at least 80%, and even more preferably at least 90% $CD34^+$ cells.

Procedures for separating the $CD34^+$ population of cells can include, but are not limited to, physical separation, magnetic separation, antibody-coated magnetic beads, affinity chromatography, cytotoxic agents joined to a monoclonal antibody or used in conjunction with a monoclonal antibody, including, but not limited to, complement and cytotoxins, and "panning" with antibody attached to a solid matrix, *e.g.,* plate, elutriation or any other convenient technique.

The use of physical separation techniques also include those based on differences in physical (density gradient centrifugation and counter-flow centrifugal elutriation), cell surface (lectin and antibody affinity), and vital staining properties (mitochondria-binding dye rhol23 and DNA-binding dye Hoechst 33342). These procedures are well known to those of skill in this art.

Techniques providing accurate separation of $CD34^+$ cells further include flow cytometry, which can have varying degrees of sophistication, e.g., a plurality of color channels, low angle and obtuse light scattering detecting channels, impedance channels. $CD34^+$ cells also can be selected by flow cytometry based on light scatter characteristics, where the target cells are selected based on low side scatter and low to medium forward scatter profiles.

b. Expansion Stage

Optionally, the *in vitro* culturing process described herein includes an expansion stage, in which the $CD34^+$ progenitor cells are allowed to proliferate. As used herein, expansion or proliferation includes any increase in cell number. Expansion includes, for example, an increase in the number of CD34+ cells over the number of CD34+ cells present in the cell population used to initiate the culture. Expansion can also include increased survival of existing CD34+ cells. The term survival refers to the ability of a cell to continue to remain alive or function.

A population of $CD34^+$ progenitor cells (e.g., mobilized $CD34^+$ peripheral blood cells) can be placed in a suitable container for expanding the CD34+ cells. For example, suitable containers for culturing the population of cells include flasks, tubes, or plates. In one embodiment, the flask can be T-flask such as a 12.5 cm^2, or a 75 cm^2 T-flask. The plate can be a 10 cm plate, a 3.5 cm plate, or a multi-welled plate such as a 12, 24, or 96 well plate. The wells can be flat, v-bottom, or u-bottom wells. The containers can be treated with any suitable treatment for tissue culture to promote cell adhesion or to inhibit cell adhesion to the surface of the container. Such containers are commercially available from Falcon, Corning and Costar. As used herein, "expansion container" also is intended to include any chamber or container for expanding cells whether or not free standing or incorporated into an expansion apparatus.

The cell density of the cultured population of $CD34^+$ cells can be at least from about $1x10^2$ cells to about $1x10^7$ cells/mL. Preferably, the cell density can be from about $1x10^5$ to about $lxlO^6$ cells/mL. The cells can be cultured at an oxygen concentration of from about 2 to 20%.

Various media can be used to expand the population of $CD34^+$ progenitor cells, including, but not limited to, Dulbecco's MEM, IMDM, X-Vivo 15 (serum-depleted) and RPMI-1640. Such culture media can be serum free. In one embodiment, the medium is serum free StemSpan (Stem Cell Technologies), which can be supplemented with 10 μg/ml heparin.

The culture medium for use in the expansion stage can contain one or more cytokines. As used herein, cytokines are factors that exert a variety of effects on cells, for example, growth or proliferation. Non-limiting examples of the cytokines that may be used in one or more stages of the *in vitro* culturing process *(e.g.*, in the expansion stage or any of the differentiation stages described below) include interleukin-2 (IL-2), interleukin 3 (IL-3), interleukin 6 (IL-6) including soluble IL-6 receptor, interleukin 12 (IL12), G-CSF, granulocyte-macrophage colony stimulating factor (GM-CSF), interleukin 1 alpha (IL-1 .alpha.), interleukin 11 (IL-1 1), MIP-loc, leukemia inhibitory factor (LIF), c-kit ligand, and fit3 ligand. In some examples, the *in vitro* culturing process described herein, or any stages thereof, can include culture conditions, in which one or more cytokine is specifically excluded from the culture medium. Cytokines are commercially available from several vendors such as, for example, Amgen (Thousand Oaks, Calif), R & D Systems and Immunex (Seattle, Wash.). Cytokine can also include fibroblast growth factor (FGF) *(e.g.,* FGF-1 or FGF-2), insulin-like growth factor *(e.g.,* IGF-2, or IGF-1), thrombopoietin (TPO), and stem cell factor (SCF), or analogs and equivalents thereof. Equivalents thereof include molecules having similar biological activity to these factors *(e.g.,* FGF, TPO, IGF, and SCF) in wild-type or purified form (e.g., recombinantly produced). Analogs include fragments retaining the desired activity and related molecules. For example, TPO is a ligand of the mpl receptor, thus molecules capable of binding the mpl receptor and initiating one or more biological actions associated with TPO binding to mpl are also within the scope of the invention. An example of a TPO mimetic is found in Cwirla et. al. (1997) Science 276:1696.

In one example, the expansion stage of the *in vitro* culturing process described herein can be performed as follows. A population of human mobilized $CD34^+$ peripheral blood cells is placed in an expansion container at a cell density of $10x10^4$-$10x10^6$ *(e.g.,* $lxlO^5$) cells/mL. The $CD34^+$ cells are cultured in an expansion medium (e.g., StemSpan serum-free medium) supplemented with a cytokine mixture of Flt-3 ligand, SCF, IL-3, and IL-6 and 2% penicillin and streptomycin under suitable conditions *(e.g.,* 37 °C) for 1-6 days *(e.g.,* 2-5 days, 3-4 days, or 4 days). The expanded CD34+ cells can be collected and subjected to further *in vitro* culturing under conditions allowing for differentiation toward mature enucleated red blood cells.

C. Multiple Differentiation Stages

The *in vitro* culturing process described herein involves multiple differentiation stages (2, 3, 4, or more), in which CD34+ progenitor cells differentiate into mature enucleated red blood cells. In each differentiation stage, CD34+ progenitor cells (either obtained from the expansion stage or collected from the original source) or cells obtained from the preceding differentiation stage can be cultured in a medium comprising one or more suitable cytokines (e.g., those described herein) under suitable conditions for a suitable period of time. Biological properties of the cells, such as cell size and expression of surface markers, may be monitored during the course or at the end of each differentiation stage to evaluate the status of erythropoiesis. Whenever necessary, cytokines can be timely supplied and/or withdrawn at each differentiation stage to achieve optimal erythroid differentiation and/or synchronizing the cell population in culture.

At each of the differentiation stages, $CD34^+$ progenitor cells, either obtained from the expansion stage described herein or isolated from an original source (e.g., human peripheral blood), or cells obtained from the preceding differentiation stage can be cultured in a suitable medium, such as those described above, supplemented with one or more cytokines under suitable culturing conditions for a suitable period of time. In one example, the culture medium *(e.g.,* IMDM) is supplemented with holo human transferrin and insulin at suitable concentrations. For example, the concentration of holo human transferrin ranges from 250-1,500 μg/ml *(e.g.,* 250-1,000 μg/ml; 300-800 μg/ml, or 400-600 μg/ml); and the concentration of insulin can range from 5-20 μg/ml *(e.g.,* 5-15 μg/ml, 5-10 μg/ml, or 10-20 μg/ml). The culture medium may also be supplemented with other components commonly used in cell culture, *e.g.,* fetal bovine serum, glutamine, bovine serum albumin, one or more antibiotics *(e.g.,* penicillin and streptomycin), or any combination thereof.

In some embodiments, the *in vitro* culturing process described herein includes three differentiation stages, Differentiation stage I ('Dif. I), Differentiation stage II ("Dif. II"), and

Differentiation stage III ("Dif. III").

In Dif. I, the cells may be cultured in the presence of a mixture of cytokines including a glucocorticoid *(e.g.,* dexamethasone), β-estradiol, IL-3, SCF, and EPO (including human EPO, EPO from other species, or EPO analogs such as Epoetin alfa, Epoetin beta, or Darbepoetin alfa) at suitable concentrations for a suitable period of time *(e.g.,* 4-7 days, 4-6 days, 5-7 days, or 5-6 days). In some examples, the concentration of dexamethasone may range from 100 nM to 5 μM *(e.g.,* 100 nM to 2 μM, 500 nM to 5 μM, 1 μM to 3 μM, 1 μM to 2 μM, or 2 μM to 5 μM). The concentration of β-estradiol may range from 0.5-5 μM *(e.g.,* 0.5-4 μM, 1-5 μM, 0.5-3μM, 0.5-2 μM, 0.5-1 μM, 1-2 μM, 1-3 μM, or 3-5 μM). The concentration of IL-3 may range from 1-10 ng/ml *(e.g.,* 1-8 ng/ml, 1-5 ng/ml, 3-6 ng/ml, 4-8 ng/ml, or 5-10 ng/ml). The concentration of SCF may range from 10-500 ng/ml *(e.g.,* 10-300 ng/ml, 50-500 ng/ml, 50-200 ng/ml, 50-100 ng/ml, 100-200 ng/ml, or 100-400 ng/ml). Alternatively or in addition, the amount of EPO may range from 2-10 U *(e.g.,* 2-8 U, 2-6 U, 5-10 U, or 6-10 U). As well known in the art, one EPO unit elicits the same erythropoiesis stimulating response in rodents (historically: fasted rats) as five micromoles of cobaltous chloride. See, *e.g.,* Jelkmann, Nephrol Dial. Transplant, 2009. In some examples, the medium used in Dif. I contains holo human transferrin, insulin, dexamethasone, β-estradiol, IL-3, SCF, and EPO. This medium may be substantially free of certain other cytokines, such as Flt-3 ligand or IL-6, or substantially free of other cytokines.

In Dif. II, the cells obtained from Dif. I may be cultured in the presence of a mixture of cytokines including SCF and EPO at suitable concentrations for a suitable period of time *(e.g.,* 3-5 days, 3-4 days, or 4-5 days). In some examples, the concentration of SCF can range from 10-100 ng/ml *(e.g.,* 10-80 ng/ml, 20-80 ng/ml, 20-50 ng/ml, 30-50 ng/ml, 40-50 ng/ml, 50-80 ng/ml, or 50-60 ng/ml). Alternatively or in addition, the amount of EPO can range from 2-10 U *(e.g.,* 2-8 U, 2-6 U, 5-10 U, or 6-10 U). In some examples, the medium used in Dif. II contains holo human transferrin, insulin, SCF, and EPO. This medium may be substantially free of certain cytokines, such as Flt-3, IL-6, dexamethasone, β-estradiol, IL-3, or any combination thereof, or may be substantially free of other cytokines.

In Dif. Ill, the cells obtained from Dif. II may be cultured in the presence of EPO at a suitable concentration for a suitable period of time *(e.g.,* 4-12 days, 5-10 days, 8-12 days, or 8-10 days). In some examples, the amount of EPO may range from 0.5-3 U *(e.g.,* 1-3 U, 0.5-

2 U, 1-2 U, or 2-3 U). In some examples, the medium used in Dif. Ill contains holo human transferrin, insulin, and EPO. This medium may be substantially free of certain cytokines, for example, Flt-3, IL-6, dexamethasone, β-estradiol, IL-3, SCF, or any combination thereof, or substantially free of other cytokines.

In some embodiments, the in vitro culturing process described herein may include one or any combination of the differentiation stages described herein, for example, Dif. I and Dif. Ill, Dif. II and Dif. Ill, or Dif. I and Dif. II.

Prior to the differentiation stages, the $CD34^+$ progenitor cells may be genetically modified such that they express surface proteins of interest, for example, sortaggable surface proteins, which are discussed in detail below.

II. <u>Preparation of Red Blood Cells Expressing Surface Proteins of Interest</u>

Also described herein are methods of preparing genetically engineered red blood cells capable of expressing surface proteins of interest, such as sortaggable surface proteins, fluorescent proteins such as green fluorescent protein (GFP), or protein drugs *(e.g.,* antibodies or antigen-binding fragments thereof).

(a) Genetic Modification of Progenitor Cells

Expression vectors for producing the surface protein of interest may be introduced into $CD34^+$ progenitor cells, which can be isolated from an original source or obtained from the expansion stage described above via routine recombinant technology. In some instances, the expression vectors can be designed such that they can incorporate into the genome of cells by homologous or non-homologous recombination by methods known in the art. Methods for transferring expression vectors into $CD34^+$ progenitor cells include, but are not limited to, viral mediated gene transfer, liposome mediated transfer, transformation, transfection and transduction, *e.g.,* viral mediated gene transfer such as the use of vectors based on DNA viruses such as adenovirus, adeno-associated virus and herpes virus, as well as retroviral based vectors. Examples of modes of gene transfer include e.g., naked DNA, $CaP0_4$ precipitation, DEAE dextran, electroporation, protoplast fusion, lipofection, cell microinjection, and viral vectors, adjuvant-assisted DNA, gene gun, catheters. In one example, a viral vector is used. To enhance delivery of non-viral vectors to a cell, the nucleic acid or protein can be conjugated to antibodies or binding fragments thereof which bind cell surface antigens, *e.g.,* CD34. Liposomes that also include a targeting antibody or fragment thereof can be used in the methods described herein.

A "viral vector" as described herein refers to a recombinantly produced virus or viral particle that comprises a polynucleotide to be delivered into a host cell, either *in vivo*, *ex vivo* or *in vitro*. Examples of viral vectors include retroviral vectors such as lentiviral vectors, adenovirus vectors, adeno-associated virus vectors and the like. In aspects where gene transfer is mediated by a retroviral vector, a vector construct refers to the polynucleotide comprising the retroviral genome or part thereof, and a therapeutic gene.

A gene encoding the surface protein of interest can be inserted into a suitable vector *(e.g.,* a retroviral vector) using methods well known in the art. Sambrook et al., Molecular Cloning, A Laboratory Mannual, 3rd Ed., Cold Spring Harbor Laboratory Press. For example, the gene and vector can be contacted, under suitable conditions, with a restriction enzyme to create complementary ends on each molecule that can pair with each other and be joined together with a ligase. Alternatively, synthetic nucleic acid linkers can be ligated to the termini of a gene. These synthetic linkers contain nucleic acid sequences that correspond to a particular restriction site in the vector. Additionally, the vector can contain, for example, some or all of the following: a selectable marker gene, such as the neomycin gene for selection of stable or transient transfectants in mammalian cells; enhancer/promoter sequences from the immediate early gene of human CMV for high levels of transcription; transcription termination and RNA processing signals from SV40 for mRNA stability; SV40 polyoma origins of replication and ColEl for proper episomal replication; versatile multiple cloning sites; and T7 and SP6 RNA promoters for in vitro transcription of sense and antisense RNA. Suitable vectors and methods for producing vectors containing transgenes are well known and available in the art. Sambrook et al., Molecular Cloning, A Laboratory Mannual, 3rd Ed., Cold Spring Harbor Laboratory Press.

Modification of $CD34^+$ progenitor cells can comprise the use of an expression cassette created for either constitutive or inducible expression of the introduced gene. Such an expression cassette can include regulatory elements such as a promoter, an initiation codon, a stop codon, and a polyadenylation signal. The elements are preferably operable in the progenitor cells or in cells that arise from the progenitor cells *(e.g.,* enucleated red blood cells) after administration *(e.g.,* infusion) into an individual. Moreover, the elements can be operably linked to the gene encoding the surface protein of interest such that the gene is operational *(e.g.,* is expressed) in the progenitor cells or red blood cells derived therefrom.

A variety of promoters can be used for expression of the surface protein of interest. Promoters that can be used to express the protein are well known in the art. Promoters include cytomegalovirus (CMV) intermediate early promoter, a viral LTR such as the Rous sarcoma virus LTR, HIV-LTR, HTLV-1 LTR, the simian virus 40 (SV40) early promoter, E. coli lac UV5 promoter and the herpes simplex tk virus promoter.

Regulatable promoters can also be used. Such regulatable promoters include those using the lac repressor from *E. coli* as a transcription modulator to regulate transcription from lac operator-bearing mammalian cell promoters [Brown, M. et al., Cell, 49:603-612 (1987)], those using the tetracycline repressor (tetR) [Gossen, M., and Bujard, H., Proc. Natl. Acad. Sci. USA 89:5547-5551 (1992); Yao, F. et al., Human Gene Therapy, 9:1939-1950 (1998); Shockelt, P., et al., Proc. Natl. Acad. Sci. USA, 92:6522-6526 (1995)]. Other systems include FK506 dimer, VP16 or p65 using astradiol, RU486, diphenol murislerone or rapamycin. Inducible systems are available from Invitrogen, Clontech and Ariad.

Regulatable promoters that include a repressor with the operon can be used. In one embodiment, the lac repressor from *E. coli* can function as a transcriptional modulator to regulate transcription from lac operator-bearing mammalian cell promoters [M. Brown et al., Cell, 49:603-612 (1987)]; Gossen and Bujard (1992); [M. Gossen et al., Natl. Acad. Sci. USA, 89:5547-5551 (1992)] combined the tetracycline repressor (tetR) with the transcription activator (VP 16) to create a tetR-mammalian cell transcription activator fusion protein, tTa (tetR-VP 16), with the tetO-bearing minimal promoter derived from the human cytomegalovirus (hCMV) major immediate-early promoter to create a tetR-tet operator system to control gene expression in mammalian cells. In one embodiment, a tetracycline inducible switch is used. The tetracycline repressor (tetR) alone, rather than the tetR-mammalian cell transcription factor fusion derivatives can function as potent trans-modulator to regulate gene expression in mammalian cells when the tetracycline operator is properly positioned downstream for the TATA element of the CMVIE promoter [F. Yao et al., Human Gene Therapy, supra]. One particular advantage of this tetracycline inducible switch is that it does not require the use of a tetracycline repressor-mammalian cells transactivator or repressor fusion protein, which in some instances can be toxic to cells [M. Gossen et al., Natl. Acad. Sci. USA, 89:5547-5551 (1992); P. Shockett et al., Proc. Natl. Acad. Sci. USA, 92:6522-6526 (1995)], to achieve its regulatable effects.

The effectiveness of some inducible promoters can be increased over time. In such cases one can enhance the effectiveness of such systems by inserting multiple repressors in tandem, *e.g.,* TetR linked to a TetR by an internal ribosome entry site (IRES). Alternatively, one can wait at least 3 days before screening for the desired function. While some silencing may occur, it can be minimized by using a suitable number of cells, preferably at least lxlO^{4}, more preferably at least lxlO^{5}, still more preferably at least lxlO^{6}, and even more preferably at least lxl 0^{7}. One can enhance expression of desired proteins by known means to enhance the effectiveness of this system. For example, using the Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element (WPRE). See Loeb, V. E., et al., Human Gene Therapy 10:2295-2305 (1999); Zufferey, R., et al., J. of Virol. 73:2886-2892 (1999); Donello, J. E., et al., J. of Virol. 72:5085-5092 (1998).

Examples of polyadenylation signals useful to practice the methods described herein include, but are not limited to, human collagen I polyadenylation signal, human collagen II polyadenylation signal, and SV40 polyadenylation signal.

The exogenous genetic material that includes the surface protein-encoding gene operably linked to the regulatory elements may remain present in the cell as a functioning cytoplasmic molecule, a functioning episomal molecule or it may integrate into the cell's chromosomal DNA. Exogenous genetic material may be introduced into cells where it remains as separate genetic material in the form of a plasmid. Alternatively, linear DNA, which can integrate into the chromosome, may be introduced into the cell. When introducing DNA into the cell, reagents, which promote DNA integration into chromosomes, may be added. DNA sequences, which are useful to promote integration, may also be included in the DNA molecule. Alternatively, RNA may be introduced into the cell.

Selectable markers can be used to monitor uptake of the desired transgene into the progenitor ceils described herein. These marker genes can be under the control of any promoter or an inducible promoter. These are known in the art and include genes that change the sensitivity of a cell to a stimulus such as a nutrient, an antibiotic, etc. Genes include those for neo, puro, tk, multiple drug resistance (MDR), etc. Other genes express proteins that can readily be screened for such as green fluorescent protein (GFP), blue fluorescent protein (BFP), luciferase, and LacZ.

(b) Genetic Modificationfor Expressing Sortaggable Surface Proteins

In some embodiments, the CD34^{+} progenitor cells can be genetically modified using any of the methods described herein or known in the art such that they are capable of expressing a sortaggable surface protein, such as a fusion protein comprising a red blood cell membrane protein and a peptide *(e.g.,* a peptide heterologous to the membrane protein). A sortaggable surface protein can be conjugated to another peptide via a sortase-mediated transpeptidation reaction. Strijbis et al., *Traffic* 13(6):780-789, 2012. Preferably the transduction of the $CD34^{+}$ progenitor cells with a gene encoding a sortaggable surface protein is via a viral vector such as a retroviral vector (as described in for example, in WO 94/29438, WO 97/21824 and WO 97/21825).

A sortaggable surface protein can be a fusion protein comprising a membrane protein and at least one heterologous protein. In some embodiments, a membrane protein for use in the methods described herein, which may present on mature RBCs, inhibit neither erythroid differentiation nor be targeted for degradation during the extensive membrane remodeling that occurs during enucleation and at the later reticulocyte stage. Such membrane proteins are known in the art. See, *e.g.,* Liu et al., 2010, *Blood,* 115(10):2021-2027. For example, red blood cell precursors express high levels of the transferrin receptor (Tfr), a type II membrane protein, but since it is no longer present on mature RBCs, Tfr may not be a suitable target for use in preparing a sortaggable surface protein. Any membrane proteins present in mature RBCs can be used for constructing the sortaggable surface protein as described herein. The membrane protein can be fused to a heterologous peptide at the terminus that is exposed to the extracellular or luminal space. In some examples, the terminus of the membrane protein that is exposed to cytoplasm may also be fused to a second heterologous protein of interest, which can be a cytoplasmic protein.

When the N-terminus of the membrane protein is exposed to the extracellular or luminal space (e.g., a Type I membrane protein), the heterologous protein may comprise an acceptor peptide to which another peptide can conjugate via a sortase-catalyzed transpeptidation reaction. Typically, the acceptor peptide is an oligoglycine or oligoalanine, such as a 1-5 glycine fragment or a 1-5 alanine fragment. In some examples, the oligoglycine consists of 3 or 5 glycine (SEQ ID NO: 3) residues. In other examples, the oligoalanine consists of 3 or 5 alanine residues (SEQ ID NO: 79).

In one example, the sortaggable surface protein is a fusion protein comprising a type I red blood cell transmembrane protein, such as glycophorin A, intercellular adhesion molecule 4 (ICAM-4), Lutheran glycoprotein (CD329), Basigin (CD 147), and a peptide that comprises an acceptor peptide *(e.g.,* a peptide that includes an oligoglycine moiety, such as a 1-5 glycine fragment). These membrane proteins can be human proteins. The acceptor peptide is fused to the N-terminus of the type I transmembrane protein. Type I transmembrane proteins are single-pass transmembrane proteins which have their N-termini exposed to the extracellular or luminal space.

In another example, the sortaggable surface protein is a fusion protein comprising a type II red blood cell transmembrane protein, ***e.g.,*** Kell, or CD71, and a peptide that comprises a sequence recognizable by a sortase *(e.g.,* sortase A). Type II transmembrane proteins are single-pass transmembrane proteins which have their C-termini exposed to the extracellular or luminal space.

Motifs recognizable by a sortase are well known in the art. One exemplary motif recognizable by a sortase, particularly sortase A, is LPXTG (SEQ ID NO: 1), in which X can be any amino acid residue (naturally-occurring or non-naturally occurring), ***e.g.,*** any of the 20 standard amino acids found most commonly in proteins found in living organisms. In some examples, the recognition motif is LPXTG (SEQ ID NO: 1) or LPXT (SEQ ID NO: 50), in which X is D, E, A, N, Q, K, or R. In other examples, X is selected from K, E, N, Q, A in an LPXTG (SEQ ID NO: 1) or LPXT (SEQ ID NO: 50)motif, which are recognizable by a sortase A. In yet other examples, X is selected from K, S, E, L, A, N in an LPXTG (SEQ ID NO: 1) or LPXT (SEQ ID NO: 50) motif, which are recognizable by a class C sortase. Exemplary sortase recognition motifs include, but are not limited to, LPKTG (SEQ ID NO: 51), LPITG (SEQ ID NO: 52), LPDTA (SEQ ID NO: 53), SPKTG (SEQ ID NO: 54) , LAETG (SEQ ID NO: 55), LAATG (SEQ ID NO: 56), LAHTG (SEQ ID NO: 57), LASTG (SEQ ID NO: 58), LPLTG (SEQ ID NO: 59), LSRTG (SEQ ID NO: 60), LPETG (SEQ ID NO: 44), VPDTG (SEQ ID NO: 61), IPQTG (SEQ ID NO: 62), YPRRG (SEQ ID NO: 63), LPMTG (SEQ ID NO: 64), LAFTG (SEQ ID NO: 65), LPQTS (SEQ ID NO: 66), LPXT (SEQ ID NO: 50), LAXT (SEQ ID NO: 67), LPXA (SEQ ID NO: 68), LGXT (SEQ ID NO: 69), IPXT (SEQ ID NO: 70), NPXT (SEQ ID NO: 71), NPQS (SEQ ID NO: 72), LPST (SEQ ID NO: 73), NSKT (SEQ ID NO: 74), NPQT (SEQ ID NO: 75), NAKT (SEQ ID NO: 76), LPIT (SEQ ID NO: 77), or LAET (SEQ ID NO: 78)

The sortaggable surface protein described herein can also be a fusion protein comprising a type III red cell transmembrane protein and a heterologous peptide. Type III membrane proteins are multi-pass structures, which usually have their N-termini exposed to the extracellular or luminal space. Examples of type III red cell transmembrane proteins include GLUT1, Aquaporin 1, and Band 3. In one example, a type III transmembrane protein can be fused with an acceptor peptide at the N-terminus of the transmembrane protein.

In some embodiments, a sortase recognition sequence as described above can further comprises one or more additional amino acids, ***e.g.,*** at the N or C terminus. For example, one or more amino acids ***(e.g.,*** up to 5 amino acids) having the identity of amino acids found immediately N-terminal to, or C-terminal to, a five (5) amino acid recognition sequence in a naturally occurring sortase substrate may be incorporated. Such additional amino acids may provide context that improves the recognition of the recognition motif.

In some embodiments, a sortase recognition motif can be masked. In contrast to an unmasked sortase recognition motif, which can be can be recognized by a sortase, a masked sortase recognition motif is a motif that is not recognized by a sortase but that can be readily modified ("unmasked") such that the resulting motif is recognized by the sortase. For example, in some embodiments at least one amino acid of a masked sortase recognition motif comprises a side chain comprising a moiety that inhibits, ***e.g.,*** prevents, recognition of the sequence by a sortase of interest, e.g., SrtAaureus. Removal of the inhibiting moiety, in turn, allows recognition of the motif by the sortase. Masking may, for example, reduce recognition by at least 80%, 90%, 95%, or more (e.g., to undetectable levels) in certain embodiments. By way of example, in certain embodiments a threonine residue in a sortase recognition motif such as LPXTG (SEQ ID NO: 1) may be phosphorylated, thereby rendering it refractory to recognition and cleavage by SrtA. The masked recognition sequence can be unmasked by treatment with a phosphatase, thus allowing it to be used in a SrtA-catalyzed transamidation reaction.

When necessary, the genetically modified membrane protein of a CD34+ progenitor cell can further include additional suitable tags, which include, but are not limited to, amino acids, nucleic acids, polynucleotides, sugars, carbohydrates, polymers, lipids, fatty acids, and small molecules. Other suitable tags will be apparent to those of skill in the art and the invention is not limited in this aspect.

In some embodiments, such a tag comprises a sequence useful for purifying, expressing, solubilizing, and/or detecting a polypeptide. In some embodiments, a tag can serve multiple functions. In some embodiments, the tag is relatively small, ***e.g.,*** ranging from a few amino acids up to about 100 amino acids long. In some embodiments, a tag is more than 100 amino acids long, ***e.g.,*** up to about 500 amino acids long, or more. In some embodiments, a tag comprises an HA, TAP, Myc, 6xHis, Flag, streptavidin, biotin, or GST tag, to name a few examples. In some embodiments, a tag comprises a solubility-enhancing tag *(**e.g.,*** a SUMO tag, NUS A tag, SNUT tag, or a monomeric mutant of the Ocr protein of bacteriophage T7). See, ***e.g.,*** Esposito D and Chatterjee DK. Curr Opin Biotechnol.; 17(4):353-8 (2006). In some embodiments, a tag is cleavable, so that it can be removed, ***e.g.,*** by a protease. In some embodiments, this is achieved by including a protease cleavage site in the tag, ***e.g.,*** adjacent or linked to a functional portion of the tag. Exemplary proteases include, e.g., thrombin, TEV protease, Factor Xa, PreScission protease, etc. In some embodiments, a "self-cleaving" tag is used. See, ***e.g.,*** Wood et al, International PCT Application PCT/US2005/05763, filed on February 24, 2005, and published as WO/2005/086654 on September 22, 2005.

In some embodiments, the genetically modified red blood cells may express a fusion protein on the surface, the fusion protein comprising a red blood cell membrane protein as described herein, a first heterologous peptide fused to the terminus of the membrane protein that is exposed to the extracellular or luminal space, and a second heterologous protein fused to the terminus of the membrane protein that is exposed to the cytoplasmic space. The first heterologous peptide may comprise an acceptor peptide of a sortase (any acceptor peptides as described herein), if it is fused to the N-terminus of the membrane protein (e.g., a Type I membrane protein). Alternatively, the first heterologous peptide may comprise a sequence recognizable by a sortase (e.g., LPXTG (SEQ ID NO: 1) or LPXT (SEQ ID NO: 50)), if it is fused to the C-terminus of the membrane protein *(**e.g.,*** a Type II membrane protein).

When the membrane protein is a Type I *(**e.g.,*** GPA) or Type III membrane protein, the second heterologous protein is fused to the C-terminus of the membrane protein. The expression cassette for producing such a fusion protein may contain two copies of sequences encoding the second heterologous protein, one being fused in-frame with the membrane protein while the other being located 3' downstream to an IRES site located between the two copies. This design would allow for production of the second heterologous protein in free form and in fusion form with the membrane protein. See, ***e.g.,*** Examples below. Alternatively, when the membrane protein is a Type II membrane protein *(**e.g.,*** Kell), the second heterologous protein is fused to the N-terminus of the membrane protein. The expression cassette for producing such a fusion protein may contain two copies of the sequence encoding the second heterologous protein, which are separated by an IRES site, so as to produce the second heterologous protein in both free and fusion form.

The second heterologous protein can be a cytoplasmic protein. In some examples, the heterologous protein is a diagnostic protein, ***e.g.,*** a protein such as a fluorescent protein ***(e.g.,*** GFP) which can release a detectable signal under suitable conditions. In other examples, the heterologous protein is a therapeutic protein, ***e.g.,*** a protein drug.

When necessary, a targeting agent may be conjugated to the first heterologous peptide as described above via a sortase reaction. Such a target agent ***(e.g.,*** an antibody such as a single domain antibody) may specifically bind to a specific type of cells ***(e.g.,*** disease cells such as cancer cells). The resultant genetically modified red blood cells are useful in delivering the cytoplasmically deposed protein of interest (the second heterologous protein) to intended target cells recognizable by the target agent.

Alternatively or in addition, the present disclosure provides genetically engineered red blood cells that express two sortaggable surface proteins, which can be conjugated to different functional moieties via reactions catalyzed by sortase enzymes having different substrate specificity. See below discussions. In one example, both of the two sortaggable surface proteins comprise a Type I membrane protein (e.g., GPA). In one sortaggable surface protein, the N-terminus of the Type I membrane protein is fused to an oligoglycine (e.g., G_3 or G_5 (SEQ ID NO: 3)); in the other sortaggable surface protein, the N-terminus of the Type I membrane protein is fused to an oligoalanine (e.g., A_3 or A_5 (SEQ ID NO: 79)). In another example, both of the two sortaggable surface proteins comprise a Type II membrane protein (e.g., Keil or CD71). In one sortaggable surface protein, the C-terminus of the Type II membrane protein is fused to the motif LPXTG (SEQ ID NO: 1); in the other sortaggable surface protein, the C-terminus of the Type II membrane protein is fused to the motif LPXTA (SEQ ID NO: 2). In yet another example, one of the sortaggable surface protein comprises a Type I membrane protein (e.g., GPA), the N-terminus of which is fused to an acceptor peptide, and the other sortaggable surface protein comprises a Type II membrane protein (e.g., Kell or CD71), the C-terminus of which is fused to a sequence recognizable by a sortase. When the acceptor peptide is an oligoglycine ***(e.g.,*** G_3 or G_5 (SEQ ID NO: 3)), the sequence recognizable by a sortase may be the motif of LPXTA (SEQ ID NO: 2). When the acceptor peptide is an oligoalanine *(e.g.,* A_3 or A_5 (SEQ ID NO: 79)), the sequence recognizable by a sortase may be the motif of LPXTG (SEQ ID NO: 1).

Any of the genetically modified CD34+ progenitor cells described herein can be cultured under suitable conditions allowing for differentiation into mature enucleated red blood cells, *e.g.,* the *in vitro* culturing process described herein. The resultant enucleated red blood cells are capable of expressing the surface protein of interest, such as a sortaggable surface protein as described herein, which can be evaluated and confirmed by routine methodology *(e.g.*, Western blotting or FACS analysis).

III. Conjugation of Agents to Cells Expression Sortaggable Surface Proteins via Sortagging

Any of the genetically modified cells *(e.g.,* CD34+ progenitor cells or mature enucleated red blood cells) expressing sortagaggle surface protein can be modified in the presence of a sortase to conjugate an agent of interest to the surface of the cells, a process known as sortagging. The term "sortagging," as used herein, refers to the process of adding a tag, *e.g.,* a moiety or molecule *(e.g.,* a protein, polypeptide, detectable label, binding agent, or click chemistry handle, onto a target molecule, for example, a target protein on the surface of a red blood cell via a sortase-mediated transpeptidation reaction.

(a) Sortase

Sortase is a family of enzymes capable of carrying out a transpeptidation reaction conjugating the C-terminus of a protein to the N-terminus of another protein via transamidation. Sortases are also referred to as transamidases, and typically exhibit both a protease and a transpeptidation activity. Various sortases from prokaryotic organisms have been identified. For example, some sortases from Gram-positive bacteria cleave and translocate proteins to proteoglycan moieties in intact cell walls. Among the sortases that have been isolated from *Staphylococcus aureus,* are sortase A (Sit A) and sortase B (Srt B). Thus, in certain embodiments, a transamidase used in accordance with the conjugation methods described herein is sortase A, *e.g.,* that from *S. aureus,* also referred to herein as SrtAaureus- In other embodiments, a transamidase is a sortase B, *e.g.,* from *S. aureus,* also referred to herein as $SrtB_{aure}$us-

Sortases have been classified into four classes, designated A, B, C, and D *(i.e.,* sortase A, sortase B, sortase C, and sortase D, respectively) based on sequence alignment and phylogenetic analysis of 61 sortases from Gram-positive bacterial genomes (Dramsi et al., *Res Microbiol.* 156(3):289-97, 2005; the entire contents of which are incorporated herein by reference). These classes correspond to the following subfamilies, into which sortases have also been classified by Comfort and Clubb (Comfort et al., *Infect Immun.,* 72(5):2710-22, 2004; the entire contents of which are incorporated herein by reference): Class A (Subfamily 1), Class B (Subfamily 2), Class C (Subfamily 3), Class D (Subfamilies 4 and 5). The aforementioned references disclose numerous sortases and their recognition motifs. See also Pallen et al, *TRENDS in Microbiology,* 2001, 9(3), 97-101; the entire contents of which are incorporated herein by reference). Those skilled in the art will readily be able to assign a sortase to the correct class based on its sequence and/or other characteristics such as those described in Drami, *et al.*, supra.

The term "sortase A" is used herein to refer to a class A sortase, usually named SrtA in any particular bacterial species, *e.g.,* SrtA from *S. aureus.* Likewise "sortase B" is used herein to refer to a class B sortase, usually named SrtB in any particular bacterial species, *e.g.,* SrtB from *S. aureus.* The present disclosure encompasses embodiments relating to any of the sortase classes known in the art *(e.g.,* a sortase A from any bacterial species or strain, a sortase B from any bacterial species or strain, a class C sortase from any bacterial species or strain, and a class D sortase from any bacterial species or strain).

Amino acid sequences of Srt A and Srt B and the nucleotide sequences that encode them are known to those of skill in the art and are disclosed in a number of references cited herein, the entire contents of all of which are incorporated herein by reference. The amino acid sequences of *S. aureus* SrtA and SrtB are homologous, sharing, for example, 22% sequence identity and 37% sequence similarity. The amino acid sequence of a sortase-transamidase from *Staphylococcus aureus* also has substantial homology with sequences of enzymes from other Gram-positive bacteria, and such transamidases can be utilized in the ligation processes described herein. For example, for SrtA there is about a 31% sequence identity (and about 44% sequence similarity) with best alignment over the entire sequenced region of the *S. pyogenes* open reading frame. There is about a 28% sequence identity with best alignment over the entire sequenced region of the *A. naeslundii* open reading frame. It will be appreciated that different bacterial strains may exhibit differences in sequence of a particular polypeptide, and the sequences herein are exemplary.

In certain embodiments a transamidase bearing 18% or more sequence identity, 20% or more sequence identity, or 30% or more sequence identity with an *S. pyogenes, A. naeslundii, S. mutans, E. faecalis* or *B. subtilis* open reading frame encoding a sortase can be screened, and enzymes having transamidase activity comparable to Srt A or Srt B from *S. aureas* can be utilized *{e.g.,* comparable activity sometimes is 10% of Srt A or Srt B activity or more).

In some embodiments, the conjugation methods described herein use a sortase A (SrtA). SrtA recognizes the motif LPXTX (SEQ ID NO: 80; wherein each occurrence of X represents independently any amino acid residue), with common recognition motifs being, *e.g.,* LPKTG (SEQ ID NO: 81), LPATG (SEQ ID NO: 82), LPNTG (SEQ ID NO: 83). In some embodiments LPETG (SEQ ID NO: 44) is used as the sortase recognition motif. However, motifs falling outside this consensus may also be recognized. For example, in some embodiments the motif comprises an 'A' rather than a 'T' at position 4, *e.g.,* LPXAG (SEQ ID NO: 84), *e.g.,* LPNAG (SEQ ID NO: 85). In some embodiments the motif comprises an 'A' rather than a 'G5 at position 5, *e.g.,* LPXTA (SEQ ID NO: 2), *e.g.,* LPNTA (SEQ ID NO: 86). In some embodiments the motif comprises a 'G' rather than 'P' at position 2, *e.g.,* LGXTG (SEQ ID NO: 87), *e.g.,* LGATG (SEQ ID NO: 88). In some embodiments the motif comprises an T rather than 'L' at position 1, *e.g.,* IPXTG (SEQ ID NO: 89), *e.g.,* IPNTG (SEQ ID NO: 90) or IPETG (SEQ ID NO: 91). Additional suitable sortase recognition motifs will be apparent to those of skill in the art, and the invention is not limited in this respect. It will be appreciated that the terms "recognition motif and "recognition sequence", with respect to sequences recognized by a transamidase or sortase, are used interchangeably. Such sortase recognition motifs can be used for constructing the sortaggable surface proteins described herein.

In some embodiments of the invention the sortase is a sortase B (SrtB), *e.g., a* sortase B of *S. aureus, B. anthracis, or L. monocytogenes.* Motifs recognized by sortases of the B class (SrtB) often fall within the consensus sequences NPXTX (SEQ ID NO: 92), *e.g.,* NP[Q/K]-[T/s]-[N/G/s] (SEQ ID NO: 93), such as NPQTN (SEQ ID NO: 94) or NPKTG (SEQ ID NO: 95). For example, sortase B of *S. aureus* or B. anthracis cleaves the NPQTN (SEQ ID NO: 94) or NPKTG (SEQ ID NO: 95) motif of IsdC in the respective bacteria (see, *e.g.,* Marraffini et al., *Journal of Bacteriology,* 189(17): 6425-6436, 2007). Other recognition motifs found in putative substrates of class B sortases are NSKTA (SEQ ID NO:

96), NPQTG (SEQ ID NO: 97), NAKTN (SEQ ID NO: 98), and NPQSS (SEQ ID NO: 99). For example, SrtB from *L. monocytogenes* recognizes certain motifs lacking P at position 2 and/or lacking Q or K at position 3, such as NAKTN (SEQ ID NO: 98) and NPQSS (SEQ ID NO: 99) (Mariscotti et al., *J Biol Chem.* 2009 Jan 7). Such sortase recognition motifs can also be used for constructing the sortaggable surface proteins described herein.

Using sortases with distinct substrate specificity, it is possible to combine N-terminal and C-terminal labeling strategies (Antos et al., 2009, *J Am. Chem. Soc,* 131(31):10800-10801) to generate multi-labeled RBCs. For example, unlike Sortase A from *Staphylococcus aureus,* Sortase A derived from *Streptococcus pyogenes* recognizes LPXTA (SEQ ID NO: 2) motifs and accepts oligo-alanine probes as nucleophiles. Therefore, the sortase reactions of both enzymes can be performed as orthogonal reactions. Utilization of such sortase reactions with suitable sortase(s) is also within the scope of the present disclosure.

In some embodiments, the sortase is a sortase C (Srt C). Sortase C may utilize LPXTX (SEQ ID NO: 80) as a recognition motif, with each occurrence of X independently representing any amino acid residue. This recognition motif can be used for constructing the sortaggable surface proteins described herein.

In yet other embodiments, the sortase is a sortase D (Srt D). Sortases in this class are predicted to recognize motifs with a consensus sequence NA-[E/A/S/H]-TG (SEQ ID NO: 100; Comfort D, supra). Sortase D has been found, *e.g.,* in *Streptomyces spp., Corynebacterium spp., Tropheryma whipplei, Thermobifida fusca,* and *Bifidobacterium longhum.* LPXTA (SEQ ID NO: 2) or LAXTG may serve as a recognition sequence for sortase D, *e.g.,* of subfamilies 4 and 5, respectively subfamily-4 and subfamily-5 enzymes process the motifs LPXTA (SEQ ID NO: 2) and LAXTG (SEQ ID NO: 101), respectively. For example, *B. anthracis* Sortase C has been shown to specifically cleave the LPNTA (SEQ ID NO: 102) motif in *B. anthracis* Basl and BasH (see Marrafini, *supra).*

Additional sortases, including, but not limited to, sortases recognizing additional sortase recognition motifs are also suitable for use in some embodiments of this invention. For example, sortases described in Chen et al., *Proc Natl Acad Sci USA.* 201 1 Jul 12;108(28):1 1399, the entire contents of which are incorporated herein; and a sortase that recognizes QVPTGV (SEQ ID NO: 103) motif as described in Barnett et al., *Journal of Bacteriology,* Vol. 184, No. 8, p. 2181-2191, 2002; the entire contents of which are incorporated herein by reference).

The use of sortases found in any gram-positive organism, such as those mentioned herein and/or in the references (including databases) cited herein is contemplated in the context of some embodiments of this invention. Also contemplated is the use of sortases found in gram negative bacteria, *e.g., Colwellia psychrerythraea,Microbulbifer degradans, Bradyrhizobium japonicum, Shewanella oneidensis,* and *Shewanella putrefaciens.* Such sortases recognize sequence motifs outside the LPXTX (SEQ ID NO: 80) consensus, for example, LP[Q/K]T[A/S]T (SEQ ID NO: 104). In keeping with the variation tolerated at position 3 in sortases from gram-positive organisms, a sequence motif LPXT[A/S] (SEQ ID NO: 105*), e.g.*, LPXTA (SEQ ID NO: 2) or LPSTS (SEQ ID NO: 106) may be used.

(b) Agent for Conjugation to Cell Surface

Agents that can be conjugated to cell surfaces via sortagging (directly or indirectly) can be any molecule, entity, or moiety, including, but are not limited to a protein, an amino acid, a peptide, a polynucleotide, a carbohydrate, a detectable label, a binding agent, a tag, a metal atom, a contrast agent, a catalyst, a non-polypeptide polymer, a synthetic polymer, a recognition element, a lipid, a linker, or chemical compound, such as a small molecule. In some embodiments, the agent is a binding agent, for example, a ligand or a ligand-binding molecule such as streptavidin/biotin, and an antibody or an antibody fragment.

When the agent is a polypeptide, it can be conjugated directly to a cell expressing a sortaggable surface protein in the presence of a sortase. The terms "protein," "peptide" and "polypeptide" are used interchangeably herein, and refer to a polymer of amino acid residues linked together by peptide (amide) bonds. The terms refer to a protein, peptide, or polypeptide of any size, structure, or function. Typically, a protein, peptide, or polypeptide will be at least three amino acids long. A protein, peptide, or polypeptide may refer to an individual protein or a collection of proteins. One or more of the amino acids in a protein, peptide, or polypeptide may be modified, for example, by the addition of a chemical entity such as a carbohydrate group, a hydroxyl group, a phosphate group, a farnesyl group, an isofarnesyl group, a fatty acid group, a linker for conjugation, functionalization, or other modification, *etc.* A protein, peptide, or polypeptide may also be a single molecule or may be a multi-molecular complex. A protein, peptide, or polypeptide may be just a fragment of a naturally occurring protein or peptide. A protein, peptide, or polypeptide may be naturally occurring, recombinant, or synthetic, or any combination thereof.

A protein to be conjugated to cell surfaces by the conjugation methods described herein can be any protein having a desired bioactivity, e.g., diagnostic or therapeutic. In some examples, the protein is a protein drug ***{e.g.,*** an antibody or a fragment thereof), a peptide capable of targeting a specific type of cells (e.g., disease cells such as cancer cells), or an immunogenic peptide capable of eliciting desired immune responses ***(e.g.,*** B cell responses or T cell responses). Such a protein can also be a protein-based binding agent, ***e.g.,*** streptavidin.

Without limiting the present disclosure in any way, this section discusses certain target proteins. In general, any protein or polypeptide can be modified to carry a click chemistry handle and/or conjugated to another molecule via click chemistry according to methods provided herein. In some embodiments the target protein comprises or consists of a polypeptide that is at least 80%, or at least 90%, ***e.g.,*** at least 95%, 86%, 97%, 98%, 99%, 99.5%, or 100% identical to a naturally occurring protein or polypeptide. In some embodiments, the target protein has no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 amino acid differences relative to a naturally occurring sequence. In some embodiments the naturally occurring protein is a mammalian protein, ***e.g.,*** of human origin.

The term "antibody," as used herein, refers to a protein belonging to the immunoglobulin superfamily. The terms antibody and immunoglobulin are used interchangeably. With some exceptions, mammalian antibodies are typically made of basic structural units each with two large heavy chains and two small light chains. There are several different types of antibody heavy chains, and several different kinds of antibodies, which are grouped into different isotypes based on which heavy chain they possess. Five different antibody isotypes are known in mammals, IgG, IgA, IgE, IgD, and IgM, which perform different roles, and help direct the appropriate immune response for each different type of foreign object they encounter. In some embodiments, an antibody is an IgG antibody, ***e.g.,*** an antibody of the IgGl, 2, 3, or 4 human subclass. Antibodies from mammalian species ***(e.g.,*** human, mouse, rat, goat, pig, horse, cattle, camel) are within the scope of the term, as are antibodies from non-mammalian species ***(e.g.,*** from birds, reptiles, amphibia) are also within the scope of the term, ***e.g.,*** IgY antibodies.

Only part of an antibody is involved in the binding of the antigen, and antigen-binding antibody fragments, their preparation and use, are well known to those of skill in the art. As is well-known in the art, only a small portion of an antibody molecule, the paratope, is involved in the binding of the antibody to its epitope (see, in general, Clark, W.R. (1986) The Experimental Foundations of Modern Immunology Wiley & Sons, Inc., New York; Roitt, I. (1991) Essential Immunology, 7th Ed., Blackwell Scientific Publications, Oxford). Suitable antibodies and antibody fragments for use in the context of some embodiments of the present invention include, for example, human antibodies, humanized antibodies, domain antibodies, F(ab'), F(ab')2, Fab, Fv, Fc, and Fd fragments, antibodies in which the Fc and/or FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; antibodies in which the FR and/or CDR1 and/or CDR2 and/or light chain CDR3 regions have been replaced by homologous human or non-human sequences; and antibodies in which the FR and/or CDR1 and/or CDR2 regions have been replaced by homologous human or non-human sequences. In some embodiments, so-called single chain antibodies (e.g., ScFv), (single) domain antibodies, and other intracellular antibodies may be used in the context of the present invention. Domain antibodies (single domain antibodies), camelid and camelized antibodies and fragments thereof, for example, VHH domains, or nanobodies, such as those described in patents and published patent applications of Ablynx NV and Domantis are also encompassed in the term antibody. A single domain antibody may consist of a single monomeric variable antibody domain and is capable of specifically binding to an antigen. Further, chimeric antibodies, *e.g.,* antibodies comprising two antigen-binding domains that bind to different antigens, are also suitable for use in the context of some embodiments of the present invention.

The term "antigen-binding antibody fragment," as used herein, refers to a fragment of an antibody that comprises the paratope, or a fragment of the antibody that binds to the antigen the antibody binds to, with similar specificity and affinity as the intact antibody. Antibodies, *e.g.,* fully human monoclonal antibodies, may be identified using phage display (or other display methods such as yeast display, ribosome display, bacterial display). Display libraries, *e.g.,* phage display libraries, are available (and/or can be generated by one of ordinary skill in the art) that can be screened to identify an antibody that binds to an antigen of interest, *e.g.,* using panning. See, *e.g.,* Sidhu, S. (ed.) Phage Display in Biotechnology and Drug Discovery (Drug Discovery Series; CRC Press; 1st ed., 2005; Aitken, R. (ed.) Antibody

Phage Display: Methods and Protocols (Methods in Molecular Biology) Humana Press; 2nd ed., 2009.

Exemplary antibodies include, but are not limited to, Abciximab (glycoprotein Ilb/IIIa; cardiovascular disease), Adalimumab (TNF-α, various auto-immune disorders, e.g., rheumatoid arthritis), Alemtuzumab (CD52; chronic lymphocytic leukemia), Basiliximab (IL-2Ra receptor (CD25); transplant rejection), Bevacizumab (vascular endothelial growth factor A; various cancers, e.g., colorectal cancer, non-small cell lung cancer, glioblastoma, kidney cancer; wet age-related macular degeneration), Catumaxomab, Cetuximab (EGF receptor, various cancers, e.g., colorectal cancer, head and neck cancer), Certolizumab (e.g., Certolizumab pegol) (TNF alpha; Crohn's disease, rheumatoid arthritis), Daclizumab (IL-2Ra receptor (CD25); transplant rejection), Eculizumab (complement protein C5; paroxysmal nocturnal hemoglobinuria), Efalizumab (CD1 la; psoriasis), Gemtuzumab (CD33; acute myelogenous leukemia (e.g., with calicheamicin)), Ibritumomab tiuxetan (CD20; Non-Hodgkin lymphoma (e.g., with yttrium-90 or indium- 111)), Infliximab (TNF alpha; various autoimmune disorders, e.g., rheumatoid arthritis) Muromonab-CD3 (T Cell CD3 receptor; transplant rejection), Natalizumab (alpha-4 (a4) integrin; multiple sclerosis, Crohn's disease), Omalizumab (IgE; allergy-related asthma), Palivizumab (epitope of RSV F protein; Respiratory Syncytial Virus infection), Panitumumab (EGF receptor; cancer, e.g., colorectal cancer), Ranibizumab (vascular endothelial growth factor A; wet age-related macular degeneration) Rituximab (CD20; Non-Hodgkin lymphoma), Tositumomab (CD20; Non-Hodgkin lymphoma), Trastuzumab (ErbB2; breast cancer), and any antigen-binding fragment thereof.

The term "binding agent," as used herein refers to any molecule that binds another molecule with high affinity. In some embodiments, a binding agent binds its binding partner with high specificity. Examples for binding agents include, without limitation, antibodies, antibody fragments, receptors, ligands, aptamers, and adnectins.

In some embodiments, the protein of interest to be conjugated to red blood cells is a cytokine, ***e.g.,*** a type I cytokine. In some embodiments of particular interest, the target protein is a four-helix bundle protein, ***e.g.,*** a four-helix bundle cytokine. Exemplary four-helix bundle cytokines include, ***e.g.,*** certain interferons ***(e.g.,*** a type I interferon, ***e.g.,*** IFN-a), interleukins ***(e.g.,*** IL-2, IL -3, IL-4, IL-5, IL-6, IL-7, IL-12), and colony stimulating factors *(e.g.,* G-CSF, GM-CSF, M-CSF). The IFN can be, *e.g.,* interferon alpha 2a or interferon alpha 2b. See, *e.g.,* Mott HR and Campbell ID. *"Four-helix bundle growth factors and their receptors: protein-protein interactions."* Curr Opin Struct Biol. 1995 Feb;5(l): 114-21; Chaiken IM, Williams WV. *"Identifying structure-function relationships in four-helix bundle cytokines: towards de novo mimetics design."* Trends Biotechnol. 1996 Oct;14(10):369-75; Klaus W, et al., *"The three-dimensional high resolution structure of human interferon alpha-l a determined by heteronuclear NMR spectroscopy in solution". J. Mol Biol.,* 274(4):661-75, 1997, for further discussion of certain of these cytokines.

The protein of interest may also be a cytokine protein that has a similar structure to one or more of the afore-mentioned cytokines. For example, the cytokine can be an IL-6 class cytokine such as leukemia inhibitory factor (LIF) or oncostatin M. In some embodiments, the cytokine is one that in nature binds to a receptor that comprises a GP130 signal transducing subunit. Other four-helix bundle proteins of interest include growth hormone (GH), prolactin (PRL), and placental lactogen. In some embodiments, the target protein is an erythropoiesis stimulating agent, *e.g.,* (EPO), which is also a four-helix bundle cytokine. In some embodiments, an erythropoiesis stimulating agent is an EPO variant, *e.g.,* darbepoetin alfa, also termed novel erythropoiesis stimulating protein (NESP), which is engineered to contain five N-linked carbohydrate chains (two more than recombinant HuEPO). In some embodiments, the protein comprises five helices. For example, the protein can be an interferon beta, *e.g.,* interferon beta-l a or interferon beta-lb, which (as will be appreciated) is often classified as a four-helix bundle cytokine. In some embodiments, a target protein is IL-9, IL-10, IL-1 1, IL-13, or IL-15. See, *e.g.,* Hunter, CA, Nature Reviews Immunology 5, 521-531, 2005, for discussion of certain cytokines. See also Paul, WE (ed.), Fundamental Immunology, Lippincott Williams & Wilkins; 6th ed., 2008. Any protein described in the references cited herein, all of which are incorporated herein by reference, can be used as a target protein.

In addition, the protein of interest may be a protein that is approved by the US Food & Drug Administration (or an equivalent regulatory authority such as the European Medicines Evaluation Agency) for use in treating a disease or disorder in humans. Such proteins may or may not be one for which a PEGylated version has been tested in clinical trials and/or has been approved for marketing.

The protein of interest may also be a neurotrophic factor, *i.e.*, a factor that promotes survival, development and/or function of neural lineage cells (which term as used herein includes neural progenitor cells, neurons, and glial cells, *e.g.*, astrocytes, oligodendrocytes, microglia). For example, in some embodiments, the target protein is a factor that promotes neurite outgrowth. In some embodiments, the protein is ciliary neurotrophic factor (CNTF; a four-helix bundle protein) or an analog thereof such as Axokine, which is a modified version of human Ciliary neurotrophic factor with a 15 amino acid truncation of the C terminus and two amino acid substitutions, which is three to five times more potent than CNTF in in vitro and in vivo assays and has improved stability properties.

In another example, the protein of interest is a protein that forms homodimers or heterodimers, (or homo- or heterooligomers comprising more than two subunits, such as tetramers). In certain embodiments the homodimer, heterodimer, or oligomer structure is such that a terminus of a first subunit is in close proximity to a terminus of a second subunit. For example, an N-terminus of a first subunit is in close proximity to a C-terminus of a second subunit. In certain embodiments the homodimer, heterodimer, or oligomer structure is such that a terminus of a first subunit and a terminus of a second subunit are not involved in interaction with a receptor, so that the termini can be joined via a non-genetically encoded peptide element without significantly affecting biological activity. In some embodiments, termini of two subunits of a homodimer, heterodimer, or oligomer are conjugated via click chemistry using a method described herein, thereby producing a dimer (or oligomer) in which at least two subunits are covalently joined. For example, the neurotrophins nerve growth factor (NGF); brain-derived neurotrophic factor (BDNF); neurotrophin 3 (NT3); and neurotrophin 4 (NT4) are dimeric molecules which share approximately 50% sequence identity and exist in dimeric forms. See, *e.g.*, Robinson RC, et al., *"Structure of the brain-derived neurotrophicfactor/neurotrophin 3 heterodimer."*, Biochemistry. 34(13):4139-46, 1995; Robinson RC, et al., *"The structures of the neurotrophin 4 homodimer and the brain-derived neurotrophicfactor/neurotrophin 4 heterodimer reveal a common Trk-binding site."* Protein Sci. 8(12):2589-97, 1999, and references therein. In some embodiments, the dimeric protein is a cytokine, *e.g.*, an interleukin.

Alternatively, the protein of interest is an enzyme, *e.g.*, an enzyme that is important in metabolism or other physiological processes. As is known in the art, deficiencies of enzymes or other proteins can lead to a variety of disease. Such diseases include diseases associated with defects in carbohydrate metabolism, amino acid metabolism, organic acid metabolism, porphyrin metabolism, purine or pyrimidine metabolism, lysosomal storage disorders, blood clotting, etc. Examples include Fabry disease, Gaucher disease, Pompe disease, adenosine deaminase deficiency, asparaginase deficiency, porphyria, hemophilia, and hereditary angioedema. In some embodiments, a protein is a clotting or coagulation factor,(e.g., factor VII, Vila, VIII or IX). In other embodiments a protein is an enzyme that plays a role in carbohydrate metabolism, amino acid metabolism, organic acid metabolism, porphyrin metabolism, purine or pyrimidine metabolism, and/or lysosomal storage, wherein exogenous administration of the enzyme at least in part alleviates the disease.

Moreover, the protein of interest may be a receptor or receptor fragment ***(e.g.,*** extracellular domain). In some embodiments the receptor is a TNFoc receptor. In certain embodiments, the target protein comprises urate oxidase.

In some embodiments, the protein of interest is an antigenic protein, which may derive from a pathogen ***(e.g.,*** a virus or bacterium). An antigenic protein may be naturally occurring or synthetic in various embodiments. It may be naturally produced by and/or comprises a polypeptide or peptide that is genetically encoded by a pathogen, an infected cell, or a neoplastic cell ***(e.g.,*** a cancer cell). In some examples, the antigenic protein is an autoantigen ("self antigen"), that has the capacity to initiate or enhance an autoimmune response. In other examples, the antigenic protein is produced or genetically encoded by a virus, bacteria, fungus, or parasite which, in some embodiments, is a pathogenic agent. In some embodiments, an agent (e.g., virus, bacterium, fungus, parasite) infects and, in some embodiments, causes disease in, at least one mammalian or avian species, e.g., human, non-human primate, bovine, ovine, equine, caprine, and/or porcine species. In some embodiments, a pathogen is intracellular during at least part of its life cycle. In some embodiments, a pathogen is extracellular. It will be appreciated that an antigen that originates from a particular source may, in various embodiments, be isolated from such source, or produced using any appropriate means (e.g., recombinantly, synthetically, etc.), e.g., for purposes of using the antigen, ***e.g.,*** to identify, generate, test, or use an antibody thereto). An antigen may be modified, e.g., by conjugation to another molecule or entity ***(e.g.*** , an adjuvant), chemical or physical denaturation, etc. In some embodiments, an antigen is an envelope protein, capsid protein, secreted protein, structural protein, cell wall protein or polysaccharide, capsule protein or polysaccharide, or enzyme. In some embodiments an antigen is a toxin, *e.g., a* bacterial toxin.

Exemplary viruses include, e.g., Retroviridae *(e.g.,* lentiviruses such as human immunodeficiency viruses, such as HIV-I); Caliciviridae (e.g. strains that cause gastroenteritis); Togaviridae *(e.g.,* equine encephalitis viruses, rubella viruses); Flaviridae (e.g. dengue viruses, encephalitis viruses, yellow fever viruses, hepatitis C virus); Coronaviridae (e.g. coronaviruses); Rhabdoviridae (e.g. vesicular stomatitis viruses, rabies viruses); Filoviridae (e.g. Ebola viruses); Paramyxoviridae (e.g. parainfluenza viruses, mumps virus, measles virus, respiratory syncytial virus); Orthomyxoviridae (e.g. influenza viruses); Bunyaviridae (e.g. Hantaan viruses, bunga viruses, phleboviruses and Nairo viruses); Arenaviridae (hemorrhagic fever viruses); Reoviridae (erg., reoviruses, orbiviurses and rotaviruses); Birnaviridae; Hepadnaviridae (Hepatitis B virus); Parvoviridae (parvoviruses); Papovaviridae (papilloma viruses, polyoma viruses); Adenoviridae; Herpesviridae (herpes simplex virus (HSV) 1 and 2, varicella zoster virus, cytomegalovirus (CMV), EBV, KSV); Poxviridae (variola viruses, vaccinia viruses, pox viruses); and Picomaviridae (e.g. polio viruses, hepatitis A virus; enteroviruses, human coxsackie viruses, rhinoviruses, echoviruses).

Exemplary bacteria include, e.g., *Helicobacter pylori, Borellia burgdorferi, Legionella pneumophilia,* Mycobacteria (e.g., *M. tuberculosis, M. avium, M, intracellulare, M. kansasii, M. gordonae), Staphylococcus aureus, Neisseria gonorrhoeae, Neisseria meningitidis, Listeria monocytogenes, Streptococcus pyogenes* (Group A Streptococcus), *Streptococcus agalactiae* (Group B Streptococcus), Streptococcus (viridans group), *Streptococcus faecalis, Streptococcus bovis,* Streptococcus (anaerobic sps.), *Streptococcus pneumoniae, Campylobacter sp., Enterococcus sp.,* Chlamydia sp., *Haemophilus influenzae, Bacillus anthracis, Corynebacterium diphtheriae, Erysipelothrix rhusiopathiae, Clostridium perfringens, Clostridium tetani, Enterobacter aerogenes, Klebsiella pneumoniae, Pasturella multocida, Bacteroides sp., Fusobacterium nucleatum, Streptobacillus moniliformis, Treponema pallidum, Treponema pertenue,* Leptospira, *Actinomyces israelii* and *Francisella tularensis.*

Exemplary fungi include, e.g., Aspergillus, such as *Aspergillus flavus, Aspergillus fumigatus, Aspergillus niger,* Blastomyces, such as *Blastomyces dermatitidis,* Candida, such as *Candida albicans, Candida glabrata, Candida guilliermondii, Candida krusei, Candida* *parapsilosis, Candida tropicalis,* Coccidioides, such as *Coccidioides immitis,* Cryptococcus, such as *Cryptococcus neoformans,* Epidermophyton, Fusarium, Histoplasma, such as *Histoplasma capsulatum,* Malassezia, such as *Malassezia furfur,* Microsporum, Mucor, Paracoccidioides, such as *Paracoccidioides brasiliensis,* Penicillium, such as *Penicillium marneffei,* Pichia, such as *Pichia anomala, Pichia guilliermondii,* Pneumocystis, such as *Pneumocystis carinii,* Pseudallescheria, such as *Pseudallescheria boydii,* Rhizopus, such as *Rhizopus oryzae,* Rhodotorula , such as *Rhodotorula rubra,* Scedosporium , such as *Scedosporium apiospermum,* Schizophyllum, such as *Schizophyllum commune,* Sporothrix, such as *Sporothrix schenckii,* Trichophyton, such as *Trichophyton mentagrophytes,* Trichophyton rubrum, *Trichophyton verrucosum, Trichophyton violaceutn,* Trichosporon, such as *Trichosporon asahii, Trichosporon cutaneum, Trichosporon inkin,* and *Trichosporon mucoides.*

In some embodiments, an antigen is a tumor antigen (TA). In general, a tumor antigen can be any antigenic substance produced by tumor cells (e.g., tumorigenic cells or in some embodiments tumor stromal ceils, e.g., tumor-associated cells such as cancer-associated fibroblasts). In many embodiments, a tumor antigen is a molecule (or portion thereof) that is differentially expressed by tumor cells as compared with non-tumor cells. Tumor antigens may include, e.g., proteins that are normally produced in very small quantities and are expressed in larger quantities by tumor cells, proteins that are normally produced only in certain stages of development, proteins whose structure (e.g., sequence or post-translational modification(s)) is modified due to mutation in tumor cells, or normal proteins that are (under normal conditions) sequestered from the immune system. Tumor antigens may be useful in, e.g., identifying or detecting tumor cells (e.g., for purposes of diagnosis and/or for purposes of monitoring subjects who have received treatment for a tumor, e.g., to test for recurrence) and/or for purposes of targeting various agents (e.g., therapeutic agents) to tumor cells. For example, in some embodiments, a chimeric antibody is provided, comprising an antibody of antibody fragment that binds a tumor antigen, and conjugated via click chemistry to a therapeutic agent, for example, a cytotoxic agent. In some embodiments, a TA is an expression product of a mutated gene, e.g., an oncogene or mutated tumor suppressor gene, an overexpressed or aberrantly expressed cellular protein, an antigen encoded by an oncogenic virus (e.g., HBV; HCV; herpesvirus family members such as EBV, KSV; papilloma virus, etc.), or an oncofetal antigen. Oncofetal antigens are normally produced in the early stages of embryonic development and largely or completely disappear by the time the immune system is fully developed. Examples are alphafetoprotein (AFP, found, e.g., in germ cell tumors and hepatocellular carcinoma) and carcinoembryonic antigen (CEA, found, e.g., in bowel cancers and occasionally lung or breast cancer). Tyrosinase is an example of a protein normally produced in very low quantities but whose production is greatly increased in certain tumor cells (e.g., melanoma cells). Other exemplary TAs include, e.g., CA-125 (found, e.g., in ovarian cancer); MUC-1 (found, e.g., in breast cancer); epithelial tumor antigen (found, e.g., in breast cancer); melanoma-associated antigen (MAGE; found, e.g., in malignant melanoma); prostatic acid phosphatase (PAP, found in prostate cancer),. In some embodiments, a TA is at least in part exposed at the cell surface of tumor cells. In some embodiments, a tumor antigen comprises an abnormally modified polypeptide or lipid, e.g., an aberrantly modified cell surface glycolipid or glycoprotein. It will be appreciated that a TA may be expressed by a subset of tumors of a particular type and/or by a subset of cells in a tumor.

Table 1 below lists exemplary proteins of interest and their amino acid sequences:

Table 1: Sequences of Exemplary Proteins of Interest:

Tissue plasminogen activator (lrtf)	Chain A: TTCCGLRQY (SEQ ID NO: 5) Chain B: IKGGLFADIASHPWQAAIFAKHHRRGGERFLCGGILISSCWILS AAHCFQQQQQEEEEEERRRRRFFFFFPPPPPPHHLTVILGRTYR VVPGEEEQKFEVEKYIVHKEFDDDTYDNDIALLQLKSSSSSD DDDDSSSSSSSSSSRRRRRCAQESSVVRTVCLPPADLQLPDWT ECELSGYGKHEALSPFYSERLKEAHVRLYPSSRCTTTSSSQQQ HLLNRTVTDNMLC AGDTTTRRRSS SNNNLHD ACQGD SGGPL VCLNDGRMTLVGIISWGLGCGGQQKDVPGVYTKVTNYLDW IRDNMRP (SEQ ID NO: 4)
Factor IX	Chain A: VVGGEDAKPGQFPWQVVLNGKVDAFCGGSIVNEKWIVTAA HCVEETTGVKITVVAGEHNIEETEHTEQKRNVIRIIPHHNYNN NAAAAAAINKYNHDIALLELDEPLVLNSYVTPICIADKEYTTT NNNIIIFLKFGSGYVSGWGRVFHKGRSALVLQYLRVPLVDRA TCLRSTKFTIYNNMFCAGGFFHEGGGRRDSCQGDSGGPHVTE VEGTSFLTGIISWGEECAAMMKGKYGIYTKVSRYVNWIKEK TKLT (SEQ ID NO: 6) Chain B: MTCNIKNGRCEQFCKNSADNKVVCSCTEGYRLAENQKSCEP AVPFPCGRVSVSQTSK (SEQ ID NO: 7)
Glucocerebrosidase	EFARPCIPKSFGYSSVVCVCNATYCDSFDPPALGTFSRYESTR SGRRMELSMGPIQANHTGTGLLLTLQPEQKFQKVKGFGGAM TDAAALNILALSPPAQNLLLKSYFSEEGIGYNIIRVPMASCDFS IRTYTYADTPDDFQLHNFSLPEEDTKLKIPLIHRALQLAQRPV

	SLLASPWTSPTWLKTNGAVNGKGSLKGQPGDIYHQTWARYF VKFLDAYAEHKLQFWAVTAENEPSAGLLSGYPFQCLGFTPE HQRDFIARDLGPTLANSTHHNVRLLMLDDQRLLLPHWAKVV LTDPEAAKYVHGIAVHWYLDFLAPAKATLGETHRLFPNTML FASEACVGSKFWEQSVRLGSWDRGMQYSHSIITNLLYHVVG WTDWNLALNPEGGPNWVRNFVDSPIIVDITKDTFYKQPMFY HLGHFSKFIPEGSQRVGLVASQKNDLDAVALMHPDGSAVVV VLNRSSKDVPLTIKDPAVGFLETISPGYSIHTYLWHRQ (SEQ ID NO: 8)
alpha galactosidase A	LDNGLARTPTMGWLHWERFMCNLDCQEEPDSCISEKLFMEM AELMVSEGWKDAGYEYLCIDDCWMAPQRDSEGRLQADPQR FPHGIRQLANYVHSKGLKLGIYADVGNKTCAGFPGSFGYYDI DAQTFADWGVDLLKFDGCYCDSLENLADGYKHMSLALNRT GRSIVYSCEWPLYMWPFQKPNYTEIRQYCNHWRNFADIDDS WKSIKSILDWTSFNQERIVDVAGPGGWNDPDMLVIGNFGLS WNQQVTQMALWAIMAAPLFMSNDLRHISPQAKALLQDKDV IAINQDPLGKQGYQLRQGDNFEVWERPLSGLAWAVAMINRQ EIGGPRSYTIAVASLGKGVACNPACFITQLLPVKRKLGFYEWT SRLRSHINPTGTVLLQLENTM (SEQ ID NO: 9)
arylsulfatase-A (iduronidase, a-L-)	RPPNIVLIFADDLGYGDLGCYGHPSSTTPNLDQLAAGGLRFT DFYVPVSLPSRAALLTGRLPVRMGMYPGVLVPSSRGGLPLEE VTVAEVLAARGYLTGMAGKWHLGVGPEGAFLPPHQGFHRF LGIPYSHDQGPCQNLTCFPPATPCDGGCDQGLVPIPLLANLSV EAQPPWLPGLEARYMAFAHDLMADAQRQDRPFFLYYASHH THYPQFSGQSFAERSGRGPFGDSLMELDAAVGTLMTAIGDLG LLEETLVIFTADNGPETMRMSRGGCSGLLRCGKGTTYEGGVR EPALAFWPGHIAPGVTHELASSLDLLPTLAALAGAPLPNVTL DGFDLSPLLLGTGKSPRQSLFFYPSYPDEVRGVFAVRTGKYK AHFFTQGSAHSDTTADPACHASSSLTAHEPPLLYDLSKDPGE NYNLLGATPEVLQALKQLQLLKAQLDAAVTFGPSQVARGED PALQICCHPGCTPRPACCHCP (SEQ ID NO: 10)
arylsulfatase B (N-acetylgalactos-amine-4-sulfatase) (lfsu)	SRPPHLVFLLADDLGWNDVGFHGSRIRTPHLDALAAGGVLL DNYYTQPLTPSRSQLLTGRYQIRTGLQHQIIWPCQPSCVPLDE KLLPQLLKEAGYTTHMVGKWHLGMYRKECLPTRRGFDTYF GYLLGSEDYYSHERCTLIDALNVTRCALDFRDGEEVATGYK NMYSTNIFTKRAIALITNHPPEKPLFLYLALQSVHEPLQVPEE YLKPYDFIQDKNRHHYAGMVSLMDEAVGNVTAALKSSGLW NNTVFIFSTDNGGQTLAGGNNWPLRGRKWSLWEGGVRGVG FVASPLLKQKGVKNRELIHISDWLPTLVKLARGHTNGTKPLD GFDV\VTCTISEGSPSPRIELLHNIDPNFVDSSPCSAFNTSVHAAI RHGNWKLLTGYPGCGYWFPPPSQYNVSEIPSSDPPTKTLWLF DIDRDPEERHDLSREYPHIVTKLLSRLQFYHKHSVPVYFPAQD PRCDPKATGVWGPWM (SEQ ID NO: 11)
beta-hexosaminidase A (2gjx)	LWPWPQNFQTSDQRYVLYPNNFQFQYDVSSAAQPGCSVLDE AFQRYRDLLFGTLEKNVLVVSVVTPGCNQLPTLESVENYTLT INDDQCLLLSETVWGALRGLETFSQLVWKSAEGTFFINKTEIE DFPRFPHRGLLLDTSRHYLPLSSILDTLDVMAYNKLNVFHWH LVDDPSFPYESFTFPELMRKGSYNPVTHIYTAQDVKEVIEYAR LRGIRVLAEFDTPGHTLSWGPGIPGLLTPCYSGSEPSGTFGPV NPSLNNTYEFMSTFFLEVSSVFPDFYLHLGGDEVDFTCWKSN PEIQDFMRKKGFGEDFKQLESFYIQTLLDIVSSYGKGYVVWQ EVFDNKVKIQPDTIIQVWREDIPVNYMKELELVTKAGFRALL SAPWYLNRISYGPDWKDFYVVEPLAFEGTPEQKALVIGGEAC MWGEYVDNTNLVPRLWPRAGAVAERLWSNKLTSDLTFAYE

	RLSHFRCELLRRGVQAQPLNVGFCEQEFEQ (SEQ ID NO: 12)
Hexosaminidase A and B (2gjx)	CHAIN A: LWPWPQNFQTSDQRYVLYPNNFQFQYDVSSAAQPGCSVLDE AFQRYRDLLFGTLEKNVLVVSVVTPGCNQLPTLESVENYTLT INDDQCLLLSETVWGALRGLETFSQLVWKSAEGTFFINKTEIE DFPRFPHRGLLLDTSRHYLPLSSILDTLDVMAYNKLNVFHWH LVDDPSFPYESFTFPELMRKGSYNPVTHIYTAQDVKEVIEYAR LRGIRVLAEFDTPGHTLSWGPGIPGLLTPCYSGSEPSGTFGPV NPSLNNTYEFMSTFFLEVSSVFPDFYLHLGGDEVDFTCWKSN PEIQDFMRKKGFGEDFKQLESFYIQTLLDIVSSYGKGYVVWQ EVFDNKVKIQPDTIIQVWREDIPVNYMKELELVTKAGFRALL SAPWYLNRISYGPDWKDFYVVEPLAFEGTPEQKALVIGGEAC MWGEYVDNTNLVPRLWPRAGAVAERLWSNKLTSDLTFAYE RLSHFRCELLRRGVQAQPLNVGFCEQEFEQ (SEQ ID NO: 13) Chain B: PALWPLPLSVKMTPNLLHLAPENFYISHSPNSTAGPSCTLLEE AFRRYHGYIFGTQVQQLLVSITLQSECDAFPNISSDESYTLLV KEPVAVLKANRVWGALRGLETFSQLVYQDSYGTFTINESTII DSPRFSHRGILIDTSRHYLPVKIILKTLDAMAFNKFNVLHWHI VDDQSFPYQSITFPELSNKGSYSLSHVYTPNDVRMVIEYARLR GIRVLPEFDTPGHTLSWGKGQKDLLTPCYSDSFGPINPTLNTT YSFLTTFFKEISEVFPDQFIHLGGDEVEFKCWESNPKIQDFMR QKGFGTDFKKLESFYIQKVLDIIATINKGSIVWQEVFDDKAKL APGTIVEVWKDSAYPEELSRVTASGFPVILSAPWYLDLISYGQ DWRKYYKVEPLDFGGTQKQKQLFIGGEACLWGEYVDATNL TPRLWPRASAVGERLWSSKDVRDMDDAYDRLTRHRCRMVE RGIAAQPLYAGYCN (SEQ ID NO: 14) Chain C: PALWPLPLSVKMTPNLLHLAPENFYISHSPNSTAGPSCTLLEE AFRRYHGYIFGTQVQQLLVSITLQSECDAFPNISSDESYTLLV KEPVAVLKANRVWGALRGLETFSQLVYQDSYGTFTINESTII DSPRFSHRGILIDTSRHYLPVKIILKTLDAMAFNKFNVLHWHI VDDQSFPYQSITFPELSNKGSYSLSHVYTPNDVRMVIEYARLR GIRVLPEFDTPGHTLSWGKGQKDLLTPCYSLDSFGPINPTLNT TYSFLTTFFKEISEVFPDQFIHLGGDEVEFKCWESNPKIQDFM RQKGFGTDFKKLESFYIQKVLDIIATINKGSIVWQEVFDDKAK LAPGTIVEVWKDSAYPEELSRVTASGFPVILSAPWYLDLISYG QDWRKYYKVEPLDFGGTQKQKQLFIGGEACLWGEYVDATN LTPRLWPRASAVGERLWSSKDVRDMDDAYDRLTRHRCRMV ERGIAAQPLYAGYCN (SEQ ID NO: 15) Chain D: LWPWPQNFQTSDQRYVLYPNNFQFQYDVSSAAQPGCSVLDE AFQRYRDLLFGTLEKNVLVVSVVTPGCNQLPTLESVENYTLT INDDQCLLLSETVWGALRGLETFSQLVWKSAEGTFFINKTEIE DFPRFPHRGLLLDTSRHYLPLSSILDTLDVMAYNKLNVFHWH LVDDPSFPYESFTFPELMRKGSYNPVTHIYTAQDVKEVIEYAR LRGIRVLAEFDTPGHTLSWGPGIPGLLTPCYSGSEPSGTFGPV NPSLNNTYEFMSTFFLEVSSVFPDFYLHLGGDEVDFTCWKSN PEIQDFMRKKGFGEDFKQLESFYIQTLLDIVSSYGKGYVVWQ EVFDNKVKIQPDTIIQVWREDIPVNYMKELELVTKAGFRALL SAPWYLNRISYGPDWKDFYVVEPLAFEGTPEQKALVIGGEAC MWGEYVDNTNLVPRLWPRAGAVAERLWSNKLTSDLTFAYE RLSHFRCELLRRGVQAQPLNVGFCEQEFEQ (SEQ ID NO: 16)
phenylalanine hydroxylase (PAH) (lj8u)	VPWFPRTIQELDRFANQILSYGAELDADHPGFKDPVYRARRK QFADIAYNYRHGQPIPRVEYMEEEKKTWGTVFKTLKSLYKT HACYEYNHIFPLLEKYCGFHEDNIPQLEDVSQFLQTCTGFRLR

	PVAGLLSSRDFLGGLAFRVFHCTQYIRHGSKPMYTPEPDICHE LLGHVPLFSDRSFAQFSQEIGLASLGAPDEYIEKLATIYWFTV EFGLCKQGDSIKAYGAGLLSSFGELQYCLSEKPKLLPLELEKT AIQNYTVTEFQPLYYVAESFNDAKEKVRNFAATIPRPFSVRY DPYTQRIEVL (SEQ ID NO: 17)
Cathepsin A	APDQDEIQRLPGLAKQPSFRQYSGYLKSSGSKHLHYWFVESQ KDPENSPVVLWLNGGPGCSSLDGLLTEHGPFLVQPDGVTLEY NPYSWNLIANVLYLESPAGVGFSYSDDKFYATNDTEVAQSNF EALQDFFRLFPEYKNNKLFLTGESYAGIYIPTLAVLVMQDPS MNLQGLAVGNGLSSYEQNDNSLVYFAYYHGLLGNRLWSSL QTHCCSQNKCNFYDNKDLECVTNLQEVARIVGNSGLNIYNL YAPCAGGVPSHFRYEKDTVVVQDLGNIFTRLPLKRMWHQAL LRSGDKVRMDPPCTNTTAASTYLNNPYVRKALNIPEQLPQW DMCNFLVNLQYRRLYRSMNSQYLKLLSSQKYQILLYNGDVD MACNFMGDEWFVDSLNQKMEVQRRPWLVKYGDSGEQIAGF VKEFSHIAFLTIKGAGHMVPTDKPLAAFTMFSRFLNKQPY (SEQ ID NO: 18)
G-CSF	LPQSFLLKCLEQVRKIQGDGAALQEKLCATYKLCHPEELVLL GHSLGIPWAPLLAGCLSQLHSGLFLYQGLLQALEGISPELGPT LDTLQLDVADFATTIWQQMEELGMMPAFASAFQRRAGGVL VASHLQSFLEVSYRVLRHLA (SEQ ID NO: 19)
GM-CSF	EHVNAIQEARRLLNLSRDTAAEMNETVEVISEMFDLQEPTCL QTRLELYKQGLRGSLTKLKGPLTMMASHYKQHCPPTPETSC ATQIITFESFKENLKDFLLVIP (SEQ ID NO: 20)
Interferon alfa-2	CDLPQTHSLGSRRTLMLLAQMRKISLFSCLKDRHDFGFPQEE FGNQFQKAETIPVLHEMIQQIFNLFSTKDSSAAWDETLLDKFY TELYQQLNDLEACVIQGVGVTETPLMKEDSILAVRKYFQRIT LYLKEKKYSPCAWEVVRAEIMRSFSLSTNLQESLRSKE (SEQ ID NO: 21)
Interferon beta-1	MSYNLLGFLQRSSNFQCQKLLWQLNGRLEYCLKDRMNFDIP EEIKQLQQFQKEDAALTIYEMLQNIFAIFRQDSSSTGWNETIV ENLLANVYHQINHLKTVLEEKLEKEDFTRGKLMSSLHLKRY YGRILHYLKAKEYSHCAWTIVRVEILRNFYFINRLTGYLRN (SEQ ID NO: 22)
Interferon gamma-1b	MQDPYVKEAENLKKYFNAGHSDVADNGTLFLGILKNWKEE SDRKIMQSQIVSFYFKLFKNFKDDQSIQKSVETIKEDMNVKFF NSNKKKRDDFEKLTNYSVTDLNVQRKAIDELIQVMAELGAN VSGEFVKEAENLKKYFNDNGTLFLGILKNWKEESDRKIMQS QIVSFYFKLFKNFKDDQSIQKSVETIKEDMNVKFFNSNKKKR DDFEKLTNYSVTDLNVQRKAIHELIQVMAELSPAA (SEQ ID NO: 23)
IL-2 (1M47)	STKKTQLQLEHLLLDLQMILNGrNNYKNPKLTRIvILTFKFYMP KKATELKHLQCLEEELKPLEEVLNLAQNFHLRPRDLISNINVI VLELKGFMCEYADETATIVEFLNRWITFCQSIISTLT (SEQ ID NO: 24)
IL-1 (2nvh)	APVRSLNCTLRDSQQKSLVMSGPYELKALHLQGQDMEQQV VFSMSFVQGEESNDKIPVALGLKEKNLYLSCVLKDDKPTLQL ESVDPKNYPKKKMEKRFVFNKIEINNKLEFESAQFPNWYISTS QAENMPVFLGGTKGGQDITDFTMQFVS (SEQ ID NO: 25)
TNF-alpha (4tsv)	DKPVAHVVANPQAEGQLQWSNRRANALLANGVELRDNQLV VPIEGLFLIYSQVLFKGQGCPSTHVLLTHTISRIAVSYQTKVNL LSAIKSPCQRETPEGAEAKPWYEPIYLGGVFQLEKGDRLSAEI NRPDYLDFAESGQVYFGIIAL (SEQ ID NO: 26)
TNF-beta (lymphotoxin) (ltnr)	KPAAHLIGDPSKQNSLLWRANTDRAFLQDGFSLSNNSLLVPT SGIYFVYSQVVFSGKAYSPKATSSPLYLAHEVQLFSSQYPFHV PLLSSQKMVYPGLQEPWLHSMYHGAAFQLTQGDQLSTHTDG

	IPHLVLSPSTVFFGAFAL (SEQ ID NO: 27)
Erythropoietin	APPRLICDSRVLERYLLEAKEAEKITTGCAEHCSLNEKITVPD TKVNFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQALLV KSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISNSDA ASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR (SEQ ID NO: 28)
Insulin	Chain A: GIVEQCCTSICSLYQLENYCN (SEQ ID NO: 29) Chain B: FVNQHLCGSHLVEALYLVCGERGFFYTPK (SEQ ID NO: 30)
Growth hormone (GH) (Somatotropin) (lhuw)	FPTIPLSRLADNAWLRADRLNQLAFDTYQEFEEAYIPKEQIHS FWWNPQTSLCPSESIPTPSNKEETQQKSNLELLRISLLLIQSWL EPVQFLRSVFANSLVYGASDSNVYDLLKDLEEGIQTLMGRLE ALLKNYGLLYCFNKDMSKVSTYLRTVQCRSVEGSCGF (SEQ ID NO: 31)
Follicle-stimulating hormone (FSH)	CHHRICHCSNRVFLCQESKVTEIPSDLPRNAIELRFVLTKLRVI QKGAFSGFGDLEKIEISQNDVLEVIEADVFSNLPKLHEIRIEKA NNLLYINPEAFQNLPNLQYLLISNTGIKHLPDVHKIHSLQKVL LDIQDNINIHTIERNSFVGLSFESVILWLNKNGIQEIHNCAFNG TQLDELNLSDNNNLEELPNDVFHGASGPVILDISRTRIHSLPSY GLENLKKLRARSTYNLKKLPTLE (SEQ ID NO: 32)
Leptin (lax8)	IQKVQDDTKTLIKTIVTRJNDILDFIPGLHPILTLSKMDQTLAV YQQILTSMPSRNVIQISNDLENLRDLLHVLAFSKSCHLPEASG LETLDSLGGVLEASGYSTEVVALSRLQGSLQDMLWQLDLSP GC (SEQ ID NO: 33)
Insulin-like growth factor (or somatomedin) (Iwqj)	PETLCGAELVDALQFVCGDRGFYFNKPTGYGSSSRRAPQTGI VDECCFRSCDLRRLEMYCAP (SEQ ID NO: 34)
Adiponectin (lc28)	Chain A: MYRSAFSVGLETRVTVPNVPIRFTKIFYNQQNHYDGSTGKFY CNIPGLYYFSYHITVYMKDVKVSLFKKDKAVLFTYDQYQEN VDQASGSVLLHLEVGDQVWLQVYYADNVNDSTFTGFLLYH DT (SEQ ID NO: 35) Chain B: MYRSAFSVGLPNVPIRFTKIFYNQQNHYDGSTGKFYCNIPGL YYFSYHITVYMKDVKVSLFKKDKVLFTYDQYQEKVDQASGS VLLHLEVGDQVWLQVYDSTFTGFLLYHD (SEQ ID NO: 36) Chain C: MYRSAFSVGLETRVTVPIRFTKIFYNQQNHYDGSTGKFYCNIP GLYYFSYHITVDVKVSLFKKDKAVLFTQASGSVLLHLEVGD QVWLQNDSTFTGFLLYHD (SEQ ID NO: 37)
Factor VIII (aka antihemophilic factor) (2r7e)	Chain A: ATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFN TSVVYKKTLFVEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDT VVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREKE DDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLV KDLNSGLIGALLVCREGSLAKEKTQTLHKFILLFAVFDEGKS WHSETKNAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVY WHVIGMGTTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQT LLMDLGQFLLFCHISSHQHDGMEAYVKVDSCPEEPQFDDDN SPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLAPDDRS YKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGI LGPLLYGEVGDTLLIIFKNQASRPYNIYPHGITDVRPLYSRRLP KGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYS SFVNMERDLASGLIGPLLICYKESVDQRGNQIMSDKRNVILFS VFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSING YVFDSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHK MVYEDTLTLFPFSGETVFMSMENPGLWILGCHNSDFRNRGM

	TALLKVSSCDKNTGDYYEDSYED (SEQ ID NO: 38) Chain B: RSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSVPQ FKKVVFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNI MVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETK TYFWKVQHHMAPTKDEFDCKAWAYSSDVDLEKDVHSGLIG PLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENM ERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQ DQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNL YPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNK CQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINA WSTKEPFSWIKVDLLAPMIIHGIKTQGARQKFSSLYISQFIIMY SLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARY IRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQIT ASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQ VDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWT LFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWV HQIALRMEVLGCEAQDLY (SEQ ID NO: 39)
Human serum albumin (lao6)	Chain A: SEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEV TEFAKTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEM ADCCAKQEPERNECFLQHKDDNPNLPRLVRPEVDVMCTAFH DNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQ AADKAACLLPKLDELRDEGKASSAKQRLKCASLQKFGERAF KAWAVARLSQRFPKAEFAEVSKLVTDLTKVHTECCHGDLLE CADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVEN DEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYA RRHPDYSVVLLLRLAKTYETTLEKCCAAADPHECYAKVFDE FKPLVEEPQNLIKQNCELFEQLGEYKFQNALLVRYTKKVPQV STPTLVEVSRNLGKVGSKCCKHPEAKRMPCAEDYLSVVLNQ LCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPKE FNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQ LKAVMDDFAAFVEKCCKADDKETCFAEEGKKLVAASQAA (SEQ ID NO: 40) Chain B: SEVAHRFKDLGEENFKALVLIAFAQYLQQCPFEDHVKLVNEV TEFAKTCVADESAENCDKSLHTLFGDKLCTVATLRETYGEM ADCCAKQEPERNECFLQHKDDNPNLPRLVRPEVDVMCTAFH DNEETFLKKYLYEIARRHPYFYAPELLFFAKRYKAAFTECCQ AADKAACLLPKLDELRDEGKASSAKQRLKCASLQKFGERAF KAWAVARLSQRFPKAEFAEVSKLVTDLTKVHTECCHGDLLE CADDRADLAKYICENQDSISSKLKECCEKPLLEKSHCIAEVEN DEMPADLPSLAADFVESKDVCKNYAEAKDVFLGMFLYEYA RRHPDYSVVLLLRLAKTYETTLEKCCAAADPHECYAKVFDE FKPLVEEPQNLIKQNCELFEQLGEYKFQNALLVRYTKKVPQV STPTLVEVSRNLGKVGSKCCKHPEAKRMPCAEDYLSVVLNQ LCVLHEKTPVSDRVTKCCTESLVNRRPCFSALEVDETYVPKE FNAETFTFHADICTLSEKERQIKKQTALVELVKHKPKATKEQ LKAVMDDFAAFVEKCCKADDKETCFAEEGKKLVAASQAA (SEQ ID NO: 41)
Hemoglobin (lbzO)	Chain A: VLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTK TYFPHFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNAL SALSDLFIAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVH ASLDKFLASVSTVLTSKYR (SEQ ID NO: 42) Chain B:

	VHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQR FFESFGDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNL KGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFT PPVQAAYQKVVAGVANALAHKYH (SEQ ID NO: 43)

Any of the proteins of interest as described herein may be modified, which may comprise (i) one or more nucleophilic residues such as glycine at the N-terminus *(e.g.,* between 1 and 10 residues) and, optionally, a cleavage recognition sequence, *e.g., a* protease cleavage recognition sequence that masks the nucleophilic residue(s); or (ii) a sortase recognition motif at or near the C-terminus. In some embodiments, the target protein comprises both (i) and (ii). Such modified proteins can be used in the methods of protein conjugation as described herein.

One of skill in the art will be aware that certain proteins, *e.g.,* secreted eukaryotic *(e.g.,* mammalian) proteins, often undergo intracellular processing *(e.g.,* cleavage of a secretion signal prior to secretion and/or removal of other portion(s) that are not required for biological activity), to generate a mature form. Such mature, biologically active versions of target proteins are used in certain embodiments of the present disclosure.

Other proteins of interest may be found in, e.g., USSN 10/773,530; 11/531,531; USSN 11/707,014; 11/429,276; 11/365,008, all of which are incorporated by reference herein. The invention encompasses application of the inventive methods to any of the proteins described herein and any proteins known to those of skill in the art.

When the agent is not a protein, such an agent can be conjugated to a protein backbone, which can be ligated to a surface sortaggable protein in the presence of a sortase. Non-protein agents include, but are not limited to a lipid, a carbohydrate, a small molecule such as a click chemistry handle or a chemotherapeutic agent, a detectable label *(e.g.,* an imaging agent), or a chemotherapeutic agent. Additional agents suitable for use in embodiments of the present disclosure will be apparent to the skilled artisan.

In some examples, the non-protein agent is antigenic such that it can elicit immune responses. Such antigenic agents may comprise, for example, a polysaccharide, a carbohydrate, a lipid, a nucleic acid, or combination thereof, and may derived from a pathogen, e.g., those described herein.

The term "conjugated" or "conjugation" refers to an association of two molecules, for example, two proteins or a protein and an agent, e.g., a small molecule, with one another in a way that they are linked by a direct or indirect covalent or non-covalent interaction. In certain embodiments, the association is covalent, and the entities are said to be "conjugated" to one another. In some embodiments, a protein is post-translationally conjugated to another molecule, for example, a second protein, a small molecule, a detectable label, a click chemistry handle, or a binding agent, by forming a covalent bond between the protein and the other molecule after the protein has been formed, and, in some embodiments, after the protein has been isolated. In some embodiments, two molecules are conjugated via a linker connecting both molecules. For example, in some embodiments where two proteins are conjugated to each other to form a protein fusion, the two proteins may be conjugated via a polypeptide linker, *e.g.,* an amino acid sequence connecting the C-terminus of one protein to the N-terminus of the other protein. In some embodiments, two proteins are conjugated at their respective C-termini, generating a C-C conjugated chimeric protein. In some embodiments, two proteins are conjugated at their respective N-termini, generating an N-N conjugated chimeric protein. In some embodiments, conjugation of a protein to a peptide is achieved by transpeptidation using a sortase. See, *e.g.,* Ploegh et al., International PCT Patent Application, PCT7US2010/000274, filed February 1, 2010, published as WO/2010/087994 on August 5, 2010, and Ploegh et al., International Patent Application PCT/US2011/033303, filed April 20, 2011, published as WO/2011/133704 on October 27, 2011, the entire contents of each of which are incorporated herein by reference, for exemplary sortases, proteins, recognition motifs, reagents, and methods for sortase-mediated transpeptidation.

Click chemistry is a chemical philosophy introduced by K. Barry Sharpless of The Scripps Research Institute, describing chemistry tailored to generate covalent bonds quickly and reliably by joining small units comprising reactive groups together (see H. C. Kolb, M. G. Finn and K. B. Sharpless (2001). Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angewandte Chemie International Edition* 40 (11): 2004-2021. Click chemistry does not refer to a specific reaction, but to a concept including, but not limited to, reactions that mimic reactions found in nature. In some embodiments, click chemistry reactions are modular, wide in scope, give high chemical yields, generate inoffensive byproducts, are stereospecific, exhibit a large thermodynamic driving force > 84 kJ/mol to favor a reaction with a single reaction product, and/or can be carried out under physiological conditions. In some embodiments, a click chemistry reaction exhibits high atom economy, can be carried out under simple reaction conditions, use readily available starting materials and reagents, uses no toxic solvents or use a solvent that is benign or easily removed (preferably water), and/or provides simple product isolation by non-chromatographic methods (crystallisation or distillation).

A click chemistry handle can be a reactant, or a reactive group, that can partake in a click chemistry reaction. For example, a strained alkyne, ***e.g.,*** a cyclooctyne, is a click chemistry handle, since it can partake in a strain-promoted cycloaddition. In general, click chemistry reactions require at least two molecules comprising click chemistry handles that can react with each other. Such click chemistry handle pairs that are reactive with each other are sometimes referred to herein as partner click chemistry handles. For example, an azide is a partner click chemistry handle to a cyclooctyne or any other alkyne. Exemplary click chemistry handles suitable for use according to some aspects of this invention are described herein, for example, in Tables A and B. Other suitable click chemistry handles are known to those of skill in the art. For two molecules to be conjugated via click chemistry, the click chemistry handles of the molecules have to be reactive with each other, for example, in that the reactive moiety of one of the click chemistry handles can react with the reactive moiety of the second click chemistry handle to form a covalent bond. Such reactive pairs of click chemistry handles are well known to those of skill in the art and include, but are not limited to, those described in Table 2:

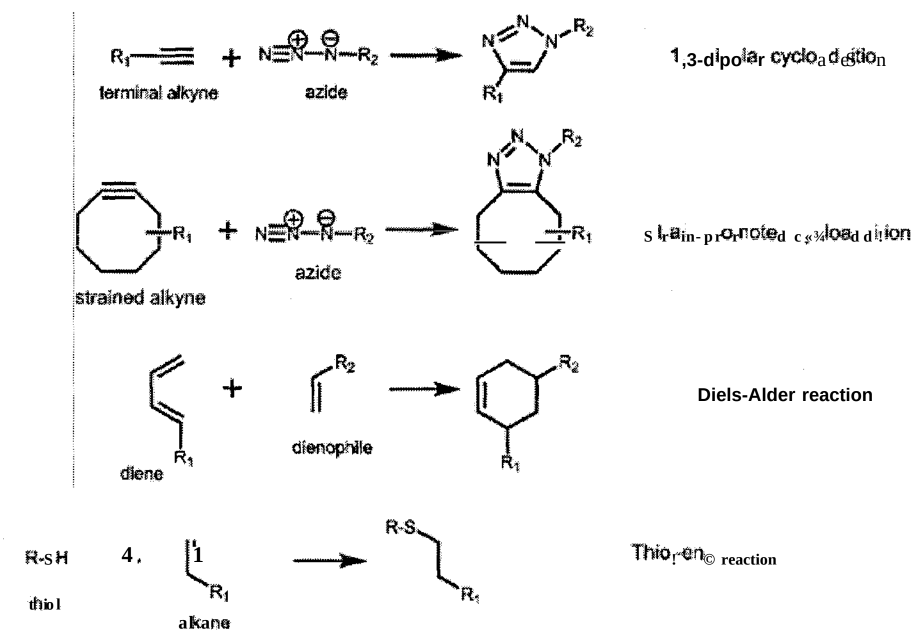


Table 2 provides examples of click chemistry handles and reactions. R, Rl, and R2 may represent any molecule comprising a sortase recognition motif. In some embodiments, each ocurrence of R, Rl, and R2 is independently RR-LPXT (SEQ ID NO: 50)-[X]y-, or -[X]y-LPXT (SEQ ID NO: 50)-RR, wherein each occurrence of X independently represents any amino acid residue, each occurrence of y is an integer between 0 and 10, inclusive, and each occurrence of RR independently represents a protein or an agent *(e.g.,* a protein, peptide, a detectable label, a binding agent, a small molecule, etc.), and, optionally, a linker.

In some embodiments, click chemistry handles are used that can react to form covalent bonds in the absence of a metal catalyst. Such click chemistry handles are well known to those of skill in the art and include the click chemistry handles described in Becer, Hoogenboom, and Schubert, Click Chemistry beyond Metal-Catalyzed Cycloaddition, *Angewandte Chemie International Edition* (2009) 48: 4900 - 4908. See Table 3 below.

Table 3: Exemplary click chemistry handles and reactions.

	Reagent A	Reagent B	Mechanism	Notes on reaction[a]	Reference
0	azide	alkyne	Cu-catalyzed [3+2] azide-alkyne cycloaddition (CuAAC)	2 h at 60°C in H_20	PI
1	azide	cyclooctyne	strain-promoted [3+2] azide-alkyne cycloaddition (SPAAC)	1 h at RT	[6-8, 10, 11]
2	azide	activated alkyne	[3+2] Huisgen cycloaddition	4 h at 50°C	[12]
3	azide	electron-deficient al-kyne	[3+2] cycloaddit(tion	12 h at RT in H_20	[13]
4	azide	aryne	[3+2] cycloaddition	4 h at RT in THF with crown ether or 24 h at RT in CH_3CN	[14, 15]
5	tetrazine	alkene	Diels-Alder retro-[4+2] cycloaddition	40 min at 2S°C (100% yield) N_2 is the only by-product	[36-38]
6	tetrazole	alkene	1,3-dipotar cycloaddition (photoclick)	few min UV irradiation and then overnight at 4°C	[39,40]
7	dithioester	diene	hetero-Diels-Alder cycloaddition	10 min at RT	[43]
8	anthracene	maleimide	[4+2] Diels-Alder reaction	2 days at reflux in toluene	[41]
9	thiol	alkene	radical addition (thio click)	30 min UV (quantitative conv.) or 24 h UV irradiation (> 96%)	[19-23]
10	thiol	enone	Michael addition	24 h at RT in CH_3CN	[27]
11	thiol	maleimide	Michael addition	1 h at 40°C in THF or 16 h at RT in dioxane	[24-26]
12	thiol	*para*-fluoro	nucteophilic substitution	overnight at RT in DMF or 60 min at 40°C in DMF	[32]
13	amine	pora-fluoro	nudeophilic substitution	20 min MW at 95°C in NMP as solvent	[30]

[a] RT = room temperature, DMF = *N,N*-dimethylformamide, NMP = N-methylpyrotidone, THF =tetrahydrofuran, CH_3CN = acetonitrile.

A detectable label is a moiety that has at least one element, isotope, or functional group incorporated into the moiety which enables detection of the molecule, *e.g., a* protein or peptide, or other entity, to which the label is attached. Labels can be directly attached *(i.e.,* via a bond) or can be attached by a linker (such as, for example, an optionally substituted alkylene; an optionally substituted alkenylene; an optionally substituted alkynylene; an optionally substituted heteroalkylene; an optionally substituted heteroalkenylene; an optionally substituted heteroalkynylene; an optionally substituted arylene; an optionally substituted heteroarylene; or an optionally substituted acylene, or any combination thereof, which can make up a linker). It will be appreciated that the label may be attached to or incorporated into a molecule, for example, a protein, polypeptide, or other entity, at any position. In general, a detectable label can fall into any one (or more) of five classes: a) a label which contains isotopic moieties, which may be radioactive or heavy isotopes, including, but not limited to, 2H, 3H, 13C, 14C, 15N, 18F, 31P, 32P, 35S, 67Ga, 76Br, 99mTc (Tc-99m), 11lln, 1231, 1251, 1311, 153Gd, 169Yb, and 186Re; b) a label which contains an immune moiety, which may be antibodies or antigens, which may be bound to enzymes (*e.g.,* such as horseradish peroxidase); c) a label which is a colored, luminescent, phosphorescent, or fluorescent moieties (*e.g.,* such as the fluorescent label fluorescein-isothiocyanate (FITC); d) a label which has one or more photo affinity moieties; and e) a label which is a ligand for one or more known binding partners (*e.g.,* biotin-streptavidin, FK506-FKBP). In certain embodiments, a label comprises a radioactive isotope, preferably an isotope which emits detectable particles, such as β particles. In certain embodiments, the label comprises a fluorescent moiety. In certain embodiments, the label is the fluorescent label fluorescein-isothiocyanate (FITC). In certain embodiments, the label comprises a ligand moiety with one or more known binding partners. In certain embodiments, the label comprises biotin. In some embodiments, a label is a fluorescent polypeptide (*e.g.,* GFP or a derivative thereof such as enhanced GFP (EGFP)) or a luciferase (*e.g.,* a firefly, Renilla, or Gaussia luciferase). It will be appreciated that, in certain embodiments, a label may react with a suitable substrate (*e.g.,* a luciferin) to generate a detectable signal. Non-limiting examples of fluorescent proteins include GFP and derivatives thereof, proteins comprising fluorophores that emit light of different colors such as red, yellow, and cyan fluorescent proteins. Exemplary fluorescent proteins include, *e.g.,* Sirius, Azurite, EBFP2, TagBFP, mTurquoise, ECFP, Cerulean, TagCFP, mTFPl, mUkGl, mAGl, AcGFPl, TagGFP2, EGFP, mWasabi, EmGFP, TagYPF, EYFP, Topaz, SYFP2, Venus, Citrine, mKO, mK02, mOrange, mOrange2, TagRFP, TagRFP-T, mStrawberry, mRuby, mCherry, mRaspberry, mKate2, mPlum, mNeptune, T- Sapphire, mAmetrine, mKeima. See, e.g., Chalfie, M. and

Kain, SR (eds.) Green fluorescent protein: properties, applications, and protocols Methods of biochemical analysis, v. 47 Wiley-Interscience, Hoboken, N.J., 2006; and Chudakov, DM, et al., Physiol Rev. 90(3):1 103-63, 2010, for discussion of GFP and numerous other fluorescent or luminescent proteins. In some embodiments, a label comprises a dark quencher, *e.g.,* a substance that absorbs excitation energy from a fluorophore and dissipates the energy as heat.

"Small molecule" refer to molecules, whether naturally-occurring or artificially created *(e.g.,* via chemical synthesis) that have a relatively low molecular weight. Typically, a small molecule is an organic compound *(i.e.,* it contains carbon). A small molecule may contain multiple carbon-carbon bonds, stereocenters, and other functional groups *(e.g.,* amines, hydroxyl, carbonyls, heterocyclic rings, *etc.).* In some embodiments, small molecules are monomeric and have a molecular weight of less than about 1500 g/mol. In certain embodiments, the molecular weight of the small molecule is less than about 1000 g/mol or less than about 500 g/mol. In certain embodiments, the small molecule is a drug, for example, a drug that has already been deemed safe and effective for use in humans or animals by the appropriate governmental agency or regulatory body.

Any of the non-protein agent described herein can be conjugated to a protein backbone via methods known to those skilled in the art.

(c) Sortase-Catalyzed Transpeptidation Reaction

Sortase-catalyzed transacylation reactions, and their use in transpeptidation (sometimes also referred to as transacylation) for protein engineering are well known to those of skill in the art (see, *e.g.,* Ploegh et al., WO/2010/087994, and Ploegh et al., WO/201 1/133704, the entire contents of which are incorporated herein by reference). In general, the transpeptidation reaction catalyzed by sortase results in the conjugation of a first protein containing a C-terminal sortase recognition motif, *e.g.,* LPXTX (SEQ ID NO: 80; wherein each occurrence of X independently represents any amino acid residue), with a second protein comprising an N-terminal sortase acceptor peptide, *e.g.,* one or more N-terminal glycine residues. In some embodiments, the sortase recognition motif is a sortase recognition motif described herein. In certain embodiments, the sortase recognition motif is LPXT (SEQ ID NO: 50) motif or LPXTG (SEQ ID NO: 1).

The sortase transacylation reaction provides means for efficiently linking an acyl donor with a nucleophilic acyl acceptor. This principle is widely applicable to many acyl donors and a multitude of different acyl acceptors. Previously, the sortase reaction was employed for ligating proteins and/or peptides to one another, ligating synthetic peptides to recombinant proteins, linking a reporting molecule to a protein or peptide, joining a nucleic acid to a protein or peptide, conjugating a protein or peptide to a solid support or polymer, and linking a protein or peptide to a label. Such products and processes save cost and time associated with ligation product synthesis and are useful for conveniently linking an acyl donor to an acyl acceptor. However, the modification and functionalization of proteins on the surface of viral particles via sortagging, as provided herein, has not been described previously.

Sortase-mediated transpeptidation reactions (also sometimes referred to as transacylation reactions) are catalyzed by the transamidase activity of sortase, which forms a peptide linkage (an amide linkage), between an acyl donor compound and a nucleophilic acyl acceptor containing an NH2-CH2-moiety. In some embodiments, the sortase employed to carry out a sortase-mediated transpeptidation reaction is sortase A (SrtA). However, it should be noted that any sortase, or transamidase, catalyzing a transacylation reaction can be used in some embodiments of this invention, as the invention is not limited to the use of sortase A.

Typically, an agent of interest *(e.g.,* a protein or a non-protein agent conjugated to a protein backbone) can be ligated to the surface of a cell expressing sortaggable surface protein by contacting the protein agent or protein backbone to which an agent of interest is conjugated with the cell under suitable conditions allowing for occurrence of the transpeptidation reaction. In some examples, the agent of interest can be incubated with a sortase first under suitable conditions for a suitable period of time allowing for cleavage at the sortase recognition site. The mixture is then in contact with the cells under suitable conditions such that the agent of interest is conjugated to the surface sortaggable protein on the cells.

In some instances, the cell, such as a CD4+ progenitor cell or an enucleated red blood cell, expresses a fusion protein that comprises a sortase recognition motif at the C-terminus *(e.g.,* a type II red cell transmembrane protein fused to a sortase recognition motif at the C-terminus). A protein of interest or a protein backbone conjugated with an agent of interest can be conjugated to the C-terminus of the fusion protein in the presence of a sortase. In other instances, the cell expresses a fusion protein comprising a membrane protein and an acceptor peptide *(e.g.,* a glycine polymer) fused to the N-terminus of the membrane protein *(e.g.,* a type I or type III transmembrane protein). A protein to be conjugated with the cell-surface fusion protein comprises a sortase recognition motif as described above at the C-terminal and can be conjugated to the N-terminus of the surface fusion protein.

The use of a specific sortase in a conjugation method as described herein would depend on the sortase recognition motif included either in the surface sortaggable protein on the cells or in the protein to be conjugated to the surface protein. This is well within the knowledge of those skilled in the art. See, e.g., descriptions above regarding recognition sites for various sortases.

In some embodiments, CD34+ or enucleated red blood cells derived therefrom can express a plurality of different surface sortaggable proteins *(e.g., 2,* 3, or more). In some embodiments, specific modification of one or more of the plurality of surface sortaggable proteins involves the use of different sortases, each specifically recognizing a different sortase recognition motif included in the one or more sortaggable proteins. For example, a first sortaggable protein can be modified with $SrtA_{aureus}$, recognizing the C-terminal sortase recognition motif LPETGG (SEQ ID NO: 107) and the N-terminal sortase recognition motif **(G)n,** and a second sortaggable can be modified with $SrtA_{pyogenes}$, recognizing the C-terminal sortase recognition motif LPETAA (SEQ ID NO: 108) and the N-terminal sortase recognition motif $(A)_n$. The sortases in this example recognize their respective recognition motif but do not recognize the other sortase recognition motif to a significant extent, and, thus, "specifically" recognize their respective recognition motif. In some embodiments, a sortase binds a sortase recognition motif specifically if it binds the motif with an affinity that is at least 5-fold, 10-fold, 20-fold, 50-fold, 100-fold, 200-fold, 500-fold, 1000-fold, or more than 1000-fold higher than the affinity that the sortase binds a different motif Such a pairing of orthogonal sortases and their respective recognition motifs, *e.g.,* of the orthogonal sortase A enzymes $SrtA_{aureus}$ and $SrtA_{pyogenes}$, can be used to site-specifically conjugate two different moieties onto two different surface sortaggable proteins.

In other embodiments, sortagging of a plurality of different proteins is achieved by sequentially contacting a cell comprising the different sortaggable surface proteins with a first sortase recognizing a sortase recognition motif of a first sortaggable protein and a suitable first protein of interest, and then with a second sortase recognizing a sortase recognition motif of a second sortaggable protein and a second protein of interest, and so forth. Alternatively, the cell may be contacted with a plurality of sortases in parallel, for example, with a first sortase recognizing a sortase recognition motif of a first sortaggable protein and a suitable first protein of interest, and with a second sortase recognizing a sortase recognition motif of a second target protein and a second protein of interest, and so forth.

For example, in some embodiments, a first sortaggable protein, *e.g., a* type I transmembrane protein fused to an oligoglycine, is modified using sortase A from *Staphylococcus aureus* (**SrtA**$_{aureus}$), and a second sortaggable protein, *e.g.,* a type II transmembrane protein fused to a suitable sortase recognition motif, is modified using sortase A from *Streptococcus pyogenes* (**SrtA**$_{pyogenes}$). **SrtA**$_{aureus}$ recognizes the motif LPXTG (SEQ ID NO: 1), in which X is any amino acid residue, and uses oligoglycine as the acceptor peptide. Differently, **SrtA**$_{pyogenes}$ recognizes the motif LPXTA (SEQ ID NO: 2), in which X is any amino acid residue, and uses oligoalanine as the acceptor peptide.

Any sortases that recognize sufficiently different sortase recognition motifs with sufficient specificity are suitable for sortagging of a plurality of sortaggable proteins expressed on red blood cells. The respective sortase recognition motifs can be inserted into the sortaggable proteins or the proteins to be conjugated to the sortaggable proteins using recombinant technologies known to those of skill in the art. In some embodiments, suitable sortase recognition motifs may be present in a wild type membrane protein. The skilled artisan will understand that the choice of a suitable sortase for the conjugation of a given protein may depend on the sequence of the sortaggable protein, *e.g.,* on whether or not the sortaggable protein comprises a sequence at its C-terminus or its N-terminus that can be recognized as a substrate by any known sortase. In some embodiments, use of a sortase that recognizes a naturally-occurring C-terminal or N-terminal recognition motif is preferred since further engineering of the target protein can be avoided.

IV, Uses of Red Blood Cells as a Carrier for in vivo Delivery of Agents of Interest

Any of the genetically modified CD34+ progenitor cells and enucleated red blood cells (including those obtained from the in vitro culturing process described herein), which can also be genetically modified as described, are within the scope of the present disclosure.

Enucleated red blood cells having a surface modification of an agent of interest as , described herein can be administered to a subject in need thereof *(e.g.,* a human patient) for various purposes, *e.g.,* treating a specific disease when the agent of interest is a therapeutic agent, detecting the presence of specific cell types when the agent of interest is capable of recognizing the target cells, and eliciting desired immune responses when the agent of interest is immunogenic. Any suitable delivery route can be used in the methods described herein, ***e.g.,*** cell infusion.

In some examples, the red blood cells are delivered to the same subject from which the cells are originally obtained. For example, peripheral blood cells can be obtained from a subject such as human patient, expanded in vitro, genetically modified such that they express sortaggable surface proteins, differentiated into enucleated red blood cells, and conjugated with an agent of interest. The modified enucleated red blood cells can then be administered to the same subject.

In other examples, the $CD34^+$ progenitor cells are obtained from a suitable subject and, after being differentiated and genetically modified as described herein, the resultant enucleated red blood cells are administered to a different subject, which preferably is immunocompatible with the donor *(**e.g.,*** administration of the enucleated blood cells would not elicit undesirable immune responses).

The enucleated red blood ceils can be administered to a subject who has or suspected of having a condition associated with red blood cell deficiency, for example, anemia, blood loss due to surgery or trauma. The enucleated red blood cells can also be administered to a subject in need of a treatment or diagnosis that can be achieved by the agent of interest conjugated on the surface of the enucleated red blood cells. For example, the enucleated red blood cells may be conjugated to an cancer antigen or an antigen derived from a pathogen. Such cells can be delivered to a subject in need *(**e.g.,*** a human subject having or at risk for cancer or infection by the pathogen) for either preventive or therapeutic treatment.

In another example, the nucleated red blood cells can be conjugated to an enzyme effective in treating a disease or disorder associated with deficiency of the enzyme, e.g., Fabry disease, Gaucher disease, Pompe disease, adenosine deaminase deficiency, asparaginase deficiency, porphyria, hemophilia, and hereditary angioedema.

In other examples, the enucleated red blood cells carry a clotting or coagulation factor,*(e*.g., factor VII, Vila, VIII or IX) and may be delivered to a human subject having or suspected of having conditions associated with abnormal blood clotting or coagulation.

V. Kits

Some aspects of the present disclosure provide kits useful for the genetic modification and surface modification of red blood cells via sortagging. Such a kit can comprise one or more expression vectors encoding one or more fusion protein each comprising a red blood cell membrane protein and a peptide as described herein. If a kit comprises multiple expression vectors as described, they can include sequences encoding different sortase recognition motifs. In some embodiments, the different sortase recognition motifs are recognized by orthogonal sortases, for example, one by $SrtA_{aureus}$ and another by $SrtA_{pyogenes}$.

Alternatively or in addition, the kits described herein can comprise one or more of the medium components for use in the *in vitro* culturing process for producing enucleated red blood cells, *e.g.*, one or more of the cytokines used therein, and one or more of the media used therein, and/or other components for cell culture.

Further, the kit can comprise one or more suitable sortases. Typically, the sortase comprised in the kit recognizes a sortase recognition motif encoded by a nucleic acid comprised in the kit. In some embodiments, the sortase is provided in a storage solution and under conditions preserving the structural integrity and/or the activity of the sortase. In some embodiments, where two or more orthogonal sortase recognition motifs are encoded by the nucleic acid(s) comprised in the kit, a plurality of sortases is provided, each recognizing a different sortase recognition motif encoded by the nucleic acid(s). In some embodiments, the kit comprises $SrtA_{aureus}$ and/or $SrtA_{pyogenes}$.

In some embodiments, the kit further comprises an agent of interest, e.g., a protein of interest or a protein backbone conjugated with a non-protein agent. In some embodiments, the protein of interest or the protein backbone comprises a sortase recognition motif, which may be compatible with the peptide in the fusion protein encoded by a nucleic acid in the kit in that both motifs can partake in a sortase-mediated transpeptidation reaction catalyzed by the same sortase. For example, if the kit comprises a nucleic acid encoding a fusion membrane protein comprising a $SrtA_{aureus}$ N-terminal recognition sequence, the kit may also comprise $SrtA_{aureus}$ and a protein of interest or a protein backbone , which will comprise the C-terminal sortase recognition motif.

In some embodiments, the kit further comprises a buffer or reagent useful for carrying out a sortase-mediated transpeptidation reaction, for example, a buffer or reagent described in the Examples section.

Without further elaboration, it is believed that one skilled in the art can, based on the above description, utilize the present invention to its fullest extent. The following specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever. All publications cited herein are incorporated by reference for the purposes or subject matter referenced herein.

EXAMPLES

Example 1: *In Vitro* Production of Human Mature Enucleated Red Blood Cells

Human peripheral blood G-CSF mobilized hematopoietic stem/progenitor cells enriched for CD34+ are purchased from Fred Hutchinson Cancer Research Center (FHCRC), Seattle. Cells are thawed according to the FHCRC protocol. The human CD^{34+} (hCD^{34+}) blood cells were differentiated into mature enucleated red blood cells by a method composed of 4 phases in 21 days: (a) Expansion, (b) Differentiation I ("Dif I"), (c) Differentiation II ("Dif II"), and Differentiation III ("Dif III") illustrated in Figure 1, and described below.

In the Expansion phase, the human CD^{34+} blood cells were cultured in Expansion medium (StemspanSFEM, CCIOO cytokine mix, which includes Flt-3 ligant, SCF, IL-3 and IL-6, and 2% penicillin-streptomycin) at 10^5 cells/mL from Day 1-4. After expansion, cells are subsequently cultured in IMDM-based erythroid differentiation medium supplemented with different cytokines in Dif I, II and III. The medium base for all three differentiation phases comprises: IMDM, 15% FBS, 2mM glutamine, 1% BSA, 500 μg/mL holo human transferrin, 10 μg/mL recombinant human insulin and 2% penicillin-streptomycin. Day 5-9, cells are cultured in Dif I medium, which contains erythroid medium base, 2μM Dexamethasone, IμM b-estradiol, 5 ng/mL IL-3, 100 ng/niL SCF and 6U Epo. Day 10-13, ceils are grown in Dif II medium containing erythroid medium base, 50 ng/mL SCF and 6U Epo. Day 14-21, cells were cultured in fibronectin-coated plates in Dif III medium, which is erythroid medium base supplemented with 2U Epo. Cell numbers re-seeded in the beginning of Dif I, II and III were 10^5, $2x10^5$, and $3x10^5$/mL, respectively.

The red blood cells at various time points during the above-noted 4-phrase process were subjected to Benzidine-Giemsa staining and the cell morphology was observed under a microscope. As shown in Figure 2, decrease of cell sizes was observed upon erythroid maturation and enucleated red blood cells were observed at day 8 of Dif. III. The sizes of the enucleated red blood cells obtained from the 4-phrase method described were found to be similar to those of normal human reticulocytes or red blood cells. Table 4 below.

Table 4. Diameters and Areas of Red Cell Clones

Cell#	Diameter (µm)	Area (μm^2)	Cell#	Diameter (µm)	Area (μm^2)
1	6.81	39.79	**6**	6.95	43.01
2	6.91	38.37	**7**	6.05	37.62
3	5.15	23.12	**8**	7	38.34
4	6.74	38.14	**9**	7.99	56.24
5	6.02	28.16	**10**	7.63	46.1
			Mean	**6.73**	**38.89**

· Cell Nos. 1-5: Enucleated red blood cells obtained from the differentiation process described above

* Cell Nos. 6-10: normal human reticulocytes or red blood cells

The hemoglobin contents of the red blood cells at Dif I, Dif II, and Dif III were determined by Drabkin's reagent. The results indicate that the enucleated blood cells obtained at the end of the differentiation process is around 30 pg/cell.

The expression levels of various red blood cell surface proteins, including CD235a (Glycophorin A), c-Kit, and CD71, were examined. FACS analysis indicated that CD235a expression was elevated during red erythropoiesis (red blood cell differentiation). The expression levels of CD235a versus those of c-kit or CD71 (transferrin) at different stages of the differentiation process are shown in Figures 4 and 5. Hoechst and glycophorin A staining indicated that around 40-50% of the red blood cells underwent enucleation at the end of the differentiation process. Figure 5. Enucleation% = 32.5/(37.8+32.5) = 46.2%.

The cell numbers during the differentiation process were counted and the results indicate that the cell number doubled almost every day during Dif I to early Dif. III. Figures 6 and 7. The expression levels of globin genes, as well as other genes (hc-kit, hGATA2, hGATAl, hGYPA, hALAS2, and hSLCAl) during hCD^{34+} cell differentiation were also examined by FACS analysis. The results are shown in Figures 8 and 9.

In sum, the study described above indicates that a four-phase hCD^{34+} cell culture and differentiation process has been successfully developed. This process yielded about 17,000-32,000-fold cell expansion (14-15 doubling) after a 21-day culture and around 40-50% of the cells were found to undergo enucleation. The enucleated red blood cells obtained from this culturing process are similar to normal human reticulocytes or normal red blood cells in cell size, hemoglobin content (~ 30 pg/cell), and cell surface marker expression.

Example 2: Conjugation of Functional Probes to Engineered Red Blood Cells via Sortase Reaction

RBCs lack a nucleus and at their mature stage cannot be genetically modified. Therefore erythroid precursors were genetically engineered to express sortase-modifiable proteins that are retained on the plasma membrane of mature RBCs. In this study, two sortase-modifiable membrane proteins were expressed in erythroid progenitors, Kell and Glycophorin A (GPA):

(1) The blood group antigen Kell is a type II membrane protein with an extracellularly exposed C-terminus and was selected as a target for C-terminal labeling.

(2) GPA, a type I membrane protein with its N-terminus extracellularly disposed, is the most abundant protein on the RBC surface and was chosen for N-terminal modification.

Materials and Methods

(i) Plasmids

The coding sequences of human, mouse glycophorin A (GPA) and Kell were obtained from NCBI ref sequence NM_002099.6, NM_010369.3, and NM_000420.2, respectively. The Myc-tag coding sequence (GAGCAGAAACTCATCTCAGAAGAGGATCTG; SEQ ID NO: 109) was inserted between the signal peptide and the mature GPA for detection of cell surface expression. To attach the sortase tag at the N terminus of GPA, five glycine (SEQ ID NO: 3) or three glycine tag coding sequences (GGTGGCGGAGGTGGA; SEQ ID NO: 110 or GGTGGCGGA) were inserted immediately after the signal peptide. The full length of engineered human and mouse GPAs were synthesized by Genscript, and subsequently cloned into MSCV retroviral vectors. Human Kell was cloned into MSCV retroviral vector extended at its C-terminus with the coding sequence for Myc tag (GAACAAAAACTTATTTCTGAAGAAGATCTG; SEQ ID NO: 112), LPETGG (SEQ ID NO: 107) (CTGCCAGAAACTGGTGGA; SEQ ID NO: 113), followed by HA epitope tag (TACCCATACGACGTCCCAGACTACGCT; SEQ ID NO: 114).

(ii) Protein expression and purification

Both, sortase A from *Staphylococcus aureus* Δ59 mutated for Kcat improvement(l) (P94R, D160N, D165A, K190E, K196T) and Ca^{2+} independent activity (Hirakawa et al., *Biotechnol. Bioeng* 109(12):2955-2961) (E105K, E108Q) in pET30 and sortase A from *Streptococcus pyogenes* were expressed and purified as described previously. Guimaraes et al., (2013) *Nat Protoc* 8(9): 1787-1799. VHH7 C-terminally extended with LPETGGHHHHHH (SEQ ID NO: 115) in pHEN was expressed and purified as described previously(4). Design and synthesis of sortase probes are been described previously as well. Guimaraes et al, 2013; and Theile et al., (2013) *Nat. Protoc.* 8(9): 1800-1 807.

(iii) Sortase labeling reaction

For labeling of GPA N-terminus with biotin, 30 µï of 500 µM *S. aureus* sortase, 1 mM K(biotin)LPRTGG (SEQ ID NO: 116) peptide in 83 mM Tris-HCl pH 7.5, 250 mM NaCl buffer was pre-incubated on ice for 15 minutes and added to 70 µï of ~ lxlO ^{6} cultured cells (reticulocytes or HEK293T cells) or up to 5x1 0^{7} red blood cells (10 µï of whole blood) in DMEM (pre-washed with DMEM). For labeling GPA N-terminus with VHH7, 50 µï of 100 µM *S. aureus* sortase, 100 µM VHH-LPETG-His6 (SEQ ID NO: 126) in 50 mM Tris-HCl pH 7.5, 150 mM NaCl buffer was pre-incubated on ice for 15 minutes and added to 50 µï of up to 5x1 0^{7} red blood cells in DMEM. For labeling Kell C-terminus with biotin, 30 µï of 500 µM *S. aureus* sortase, 1.4 mM GGGK(biotin) (SEQ ID NO: 47) peptide in 83 mM Tris-HCl pH 7.5, 250 mM NaCl buffer was added to 70 µï of ~ 1x106 cultured cells (reticulocytes or HEK293T cells) or up to 5x107 red blood cells in DMEM (pre-washed with DMEM). All sortase and cell mixtures were incubated on ice for 30 minutes with occasional gentle mixing. They were spun at 2000 RPM for 2 minutes at 4 °C to remove buffer/DMEM and washed 3 times with 1 ml of ice-cold PBS. For sequential dual labeling of GPA and Kell, 30 µï of 333 µM *S. pyogenes* sortase, 1.33 mM K(biotin)LPETAA (SEQ ID NO: 45) peptide in 83 mM TrisHCl, 250 mM NaCl, 16.7 mM CaC12 buffer was pre-incubated at 37 °C for 15 minutes, added to 70 µï of lxlO ^{6} HEK293T cells in DMEM (pre-washed with DMEM) and incubated at 37 °C for 25 minutes with occasional gentle mixing. Cells were spun down at 2000 RPM for 2 minutes at 4 °C, washed 2 times with 1 ml of ice-cold PBS and finally re-suspended into 70 µï of ice-cold DMEM. 30 µï of 333 µM *S. aureus* sortase,1.67 mM GGGK(Alexa647; SEQ ID NO: 117) peptide in 83 mM Tris-HCl pH 7.5, 250 mM NaCl was added to the cells and incubated for 30 minutes on ice with occasional gentle mixing. Cells were washed 3 times with 1 ml of ice-cold PBS.

(iv) Western blotting

When sortase-labeling whole blood, red blood were first lysed with 500 µl of ammonium chloride solution (Stemcell Technologies) via a 5 minute incubation on ice and spun down at 14000 RPM for 10 minutes at 4 °C. Membranes were washed once with 500 µl of PBS. Cells were solubilized in 25 mM Tris-HCl pH7.5, 0.5 % NP40, 5 mM $MgCl_2$, 150 mM NaCl supplemented with a Complete mini protease inhibitor cocktail tablet (Roche), vortexed, incubated on ice for 10 minutes, and spun at 14,000 RPM for 10 minutes at 4 °C. Supernatant was incubated with SDS sample buffer and boiled for 8 minutes. After SDS-PAGE, proteins were transferred onto PVDF membrane and immunoblotted for biotin (using streptavidin-HRP (GE Healthcare)), myc tag (Cell Signaling #2040), HA tag (Roche 3F10), hemoglobin, or CD19 (Cell Signaling #3574).

(v) Transfection

Virus production: Six million 293T cells were split and plated on 10 cm plates one day before transfection in antibiotic free Dulbecco's Modified Eagle Medium (DMEM) with added 15% Fetal Bovine Serum (FBS) and 2 mM L-Glutamine (Invitrogen). On day 0, 10 ug plasmids together with 5 ug packaging vector were added to the 293T 10-cm plate with Fugene 6 (Promega). Six-eight hours later, new DMEM with added 15% FBS, 2 mM L-Glutamine, and 1x Pen Strep (Invitrogen) replaced the old medium. On day 1, the supernatant containing fresh virus was collected, filtered through 0.45 µm filter (Millipore), and then immediately used to infect the murine erythroid progenitors. GPA & Kell expression: HEK293T cells were plated on 10 cm dish and transfected with either 15 µg of Gn-myc-h/mGPA constructs in retroviral vectors, or 7.5 µg of hKell-LPETG (SEQ ID NO: 44)-HA and 7.5 µg of 3G-myc-hGPA both in retroviral vectors (for sequential dual sortase labeling) using TransIT transfection reagent according to manufacturer's recommendations (Mirus). Cells were analyzed for protein expression and sortase-labeled 24 hours after cellular transfection.

(vi) Isolation of erythroid progenitors from murine E14.5 fetal liver cells

After etherization by carbon dioxide, Day 14.5 pregnant C57BL/6J mice were dissected, and the embryos were isolated. The entire fetal livers were carefully separated and placed in Phosphate Buffered Saline (PBS) with 2% Fetal Bovine Serum (FBS) and 100 uM EDTA. Single fetal liver cell suspensions were obtained after triturating and filtering through a 70 μηι filter (BD). Mature red blood cells among fetal liver cell suspension were lysed after incubation with Ammonium Chloride Solution (Stemcell) for ten minutes. Nucleated cells were collected by centrifugation at 1500 RPM for 5 minutes, and resuspended in PBS. Using BD Biotin Mouse Lineage Panel (559971) and BD Streptavidin Particles Plus DM (557812), we purified lineage negative fetal liver cells, which were enriched for erythroid progenitors (more than 90%).

(viii) Viral infection and culture of murine erythroid progenitors

After purification, lineage negative fetal liver cells were prepared at a concentration of 10 million cells per ml. Ten microliters of the cells were plated in each well of a 24-well plate, together with 1 ml virus-containing supernatant with 5 ug/ml polybrene. The plate was spun at 2000 RPM for 90 min at 37 degree. Immediately following spin-infection, the virus-containing supernatant was aspirated and replaced with erythroid maintenance medium (StemSpan-SFEM (StemCell Technologies) supplemented with recombinant mouse stem cell factor (100 ng/ml SCF, R&D), recombinant mouse IGF-1 (40 ng/ml, R&D), dexamethasone (100 nM, Sigma), and erythropoietin (2 u/ml, Amgen)). The next morning, the infection rate was examined by flow cytometry by checking for the population of GFP positive cells. It was typically more than 95%. The cells were then cultured in Epo-only erythroid differentiation medium (Iscove modified Dulbecco's medium (IMDM) containing 15% FBS (Stemcell), 1% detoxified bovine serum albumin (BSA) (Stemcell), 500 μg/mL holo-transferrin (Sigma-Aldrich), 0.5 U/mL Epo (Amgen), 10 μg/mL recombinant human insulin (Sigma-Aldrich), 2 mM L-glutamine (Invitrogen), and 1× Pen Strep (Invitrogen)) for 48 hours.

(ix) Flow cytometry

The desired cells, washed them once by PBS, and resuspended them at a density of 5-10 M/ml in PBS with 1 μg/ml PI for FACS sorting or analysis. During enucleation analysis, the cells were first stained with Anti-Mouse TER-1 19 antibody (eBioscience, 14-5921) and Hoechst (Sigma) for 15 minutes at room temperature. For detecting the surface expression or labelling of Myc-tag, HA-tag or Biotin-labelling, the cells were stained with anti-Myc antibody (Cell signalling, 3739), anti-HA antibody (Thermo Fisher Scientific, NC9843881) or anti-biotin antibody (eBioscience, 12-9895-82), respectively, for 30 minutes at room temperature.

(x) Viral infection of bone marrow cells

Bone marrow in femur and tibias from C57BL/6J mice was isolated using 23G needle and cultured at density of 2x106 cells/ml for 18 hours in DMEM supplemented with 15% fetal calf serum, 2 mM L-Glutamine (Invitrogen), 1x Pen/Strep (Invitrogen), 20 ng/ml IL-3 (Peprotech), 50 ng/ml SCF (Peprotech), and 50 ng/ml IL-6 (Peprotech) in 6-well plates. Cells were infected by incubating 4x106 cells in 500 µï of retrovirus-containing media, 500 µï DMEM, and 5 µg/ml polybrene and spinning the cells at 2500 RPM for 1.5 hours at room temperature and further incubating them in a $C0_2$ incubator for 5 hours. Cells were returned to the above IL-3/SCF/IL-6 supplemented media and further incubated for 16 hours.

(xi) Irradiation procedure and mouse fetal liver transplantation

B6.SJL-i¾ɔ*rc*a *Pe*p3b/BoyJ mice (The Jackson Lab) were subjected to total body irradiation with 1050 cGy in a Gammacell 40 irradiator chamber 1 day before transplantation. Mouse fetal liver cells were harvested and prepared as mentioned above. Following retroviral infection, cells were cultured in erythroid maintenance media for 18 hours. Alternatively, mouse bone marrow cells were harvested and retrovirus transduced as described above. Infected cells were then washed 2 times in sterile PBS and resuspended in sterile PBS at 2-5 million cells/mL. IOOµI of these mouse stem and progenitor cells were then retro-orbitally injected into the lethally irradiated mice. Starting from four weeks, mature red cells were extracted from the irradiated mice for analysis and sortagging.

(xii) Cytospin preparation and immunofluorescence

After mature red blood cells and *in vitro* differentiated reticulocytes were sortagged with biotin, they were washed twice in cold IX PBS. 50,000 sorted, biotinylated mature red blood cells or unsorted, *in vitro* differentiated reticulocytes were centrifuged onto Poly-L-Lysine coated slides for 5 minutes at 400rpm (Cytospin 3, Thermo Shandon). Samples were air dried, fixed in 4% paraformaldehyde for 30 minutes at room temperature and blocked for 1 hour in blocking buffer (2% BSA + 2% Donkey Serum in PBS). Cells were then incubated with primary antibodies: PE-conjugated anti-biotin (1:1000, eBioscience, 12-9895-82) and

APC-conjugated anti-Terl 19 (1:100, eBioscience, 14-5921) in blocking buffer overnight at 4°C, followed with three washes in cold IX PBS. Finally, cells were mounted with mounting media containing DAPI (Prolong Gold Antifade, Invitrogen) to visualize nuclei in all immunostaining experiments. Visualization was carried out using a Zeiss LSM 700 Laser Scanning Confocal Microscope.

(xiii) Transfusion and survival of eRBC

After mature RBCs were sortagged, they were washed twice in IX PBS and resuspended in RPMI medium. These sortagged RBCs were collected by centrifugation at 1500rpm for 5 minutes and labeled with 5μM CFSE in Hank's Balanced Salt Solution (HBSS) for 8 minutes (Life Technologies). Equivolumes of 10% FBS in HBSS were then added to quench the reaction. RBCs were washed, counted, and resuspended in sterile HBSS for injection. 2.5 billion CFSElabeled RBCs (± 300μ1mouse blood) were then injected intravenously into recipient $CD^{45.1+}$ mice. Normal RBCs served as a control. Since only 20-50% of RBCs expressed engineered hGPA or hKell, the rest of the RBCs were normal and served as an internal control by monitoring the CFSE signal. One hour after transfusion, the first blood sample (20 μï) were collected from retro orbital and labeled as Day 0. The subsequent blood samples with same amount were collected at days 1, 4, 7, and so on, until 28. The blood samples were stained with anti-biotin antibody conjugated with PE (ebioscience), so that eRBCs linked with biotin had strong PE signal during flow cytometry analysis. However, RBCs with hKell-biotin had very low red fluorescent intensity after staining with anti-biotin antibody conjugated with PE. It was difficult to detect it in the presence of strong green fluorescent signal from CFSE. The hKell-RBCs were stortagged with biotin, and transfused into mice without CFSE staining. The hKell-RBCs were monitored by their inherent GFP signal and weak PE signal from anti-biotin antibodies conjugated with PE. The control RBCs and hGPA-RBCs were transfused into a total of six mice in two separate times. The hKell-RBCs were transfused to a total of three mice, with one dying for an unknown reason.

Results

(I) Genetically modified type-I membrane protein human glycophorin A (GPA) was expressed and sortase-labeled on the RBC surface

(a) Expression of Sortaggable Human Glycophorin (hGPA) on RBCs

The expression cassettes of fusion proteins 5Gly (SEQ ID NO: 3)-myc-hGYPA and 3Gly-myc-hGYPA were inserted into two lentiviral vectors EF1 and MSCV (System Biosciences, Inc.) as shown in Figure 10. Both vectors carry GFP as a report gene. The resultant lentiviral vectors for expressing the fusion proteins were transduced into human $CD34^+$ cells in the beginning of Dif I.

Figure 11 shows the expression of the fusion proteins in the human $CD34^+$ cells and red blood cells obtained at the end of Dif. I and Dif II as described in Example 1 above. The expression of 5Gly (SEQ ID NO: 3)-myc-hGYPA fusion protein on the surface of red blood cells transduced by the EF1 vectors was observed at least at the end of Dif. II.

Similar results were observed in red blood cells transduced with the MSCV expression vectors. As shown in Figure 12, expression of MSCV-5Gly (SEQ ID NO: 3)-myc-hGYPA on the surface of red blood cells was observed at the end of Dif. III.

The lentiviral vectors for expressing the 5Gly (SEQ ID NO: 3)-myc-hGYPA and 3Gly-myc-hGYPA fusion proteins were also transduced into K562 cells and expression of both fusion proteins were observed on the surface of the K562 cells.

In sum, expression of sortaggable hGYPA were detected in 50-70% of the cells at the end of Dif I by flow cytometry and Western blotting. hGYPA expression remained constant and last until the end of Dif III.

(b) Conjugate an Agent of Interest to Red Blood Cell Surface in the Presence of a Sortase

To sortag hGYPA, 300 μM sortase A of *Staphylococcus aureus* and 500 μM biotin-LPRTGG (SEQ ID NO: 118) were pre-incubated in a sortase buffer at 37°C for 30 minutes. One million red blood cells expressing sortaggable hGYPA (5Gly (SEQ ID NO: 3)-myc-hGYPA or 3Gly-myc-hGYPA) were then incubated with the sortase-substrate mix at 37°C for 1 hour. The cells were washed with PBS for 3~4 times before detection by flow cytometry or Western blotting for biotin conjugation. Labeling efficiency is normally 80-100%.

As shown in Figure 13, Western blotting results indicate expression of 5Gly (SEQ ID NO: 3)-myc-hGYPA on red blood cells and the conjugation of biotin to the hGYPA. FACS analysis indicated that biotin was conjugated to the surface of red blood cells at day 8 of Dif. III. Figures 14 and 15. This result shows that the surface of mature enucleated red blood cells can be modified via a sortase-catalyzed reaction.

Similarly, sortaggable hGYPA expressed on K562 cells was successfully modified by biotin in the presence of a sortase.

In another experiment, a PE-conjugated peptide comprising the LPXTG (SEQ ID NO: 1) motif was included with sortase A for 45 minutes at 37 °C. The mixture was then incubated with erythrocytes expressing 3Gly-GYPA or 5Gly (SEQ ID NO: 3)-GYPA sortaggable surface protein for 30 minutes at 37 °C. The cells were then subjected to FACS analysis. Results thus obtained indicate that the PE-conjugated peptide was conjugated to both 3Gly-GYPA and 5Gly (SEQ ID NO: 3)-GYPA on the surface of the erythrocytes.

(c) <u>Genetically Modified hGPA was expressed as sortase-labled on Murine PJBCs</u>

Extension of the N-terminus of GPA (GPA) with glycine residues was the minimal modification needed to render GPA a suitable sortase substrate. In the modified version, the N-terminal signal sequence was retained, followed by glycine residues and a myc epitope tag. Cleavage of the signal peptide would yield (Gly)n-myc tag-GPA. Incubation of cells carrying the modified GPA with sortase A from *Staphylococcus aureus* and an LPETG (SEQ ID NO: 44)-based probe leads to conjugation of this probe to the N-terminus of GPA (Fig. *16A).*

Four retroviral constructs were prepared for GPA, encoding products extended at their N-terminus with either 3 or 5 N-terminal glycine residues (SEQ ID NO: 3), using either mouse or human GPA and *in vitro* erythroid differentiation system (Zhang et al., 203, *Blood* 102(12):3938-3946)

as with Kell below, to test GPA expression and modification by sortase. Upon retroviral transduction of progenitors none of the four GPA constructs affected the differentiation process, as in each case we obtained normal numbers of enucleated, Terl 19 positive erythroblasts that expressed the modified GPA on their surface (Fig. 17 and Fig. 18, *A,B).*

These GPA constructs were confirmed as sortase-modifiable by expressing them in HEK293T cells, followed by sortagging with a biotin-containing probe (Fig. 18Q. While high levels of expression of the mouse constructs were observed, no biotinylation of the encoded products was detected. In contrast, human 3G-myc-GPA (3G-myc-hGPA) was readily sortagged and was used for subsequent experiments. Expression of 3Gmyc-hGPA in erythroid progenitors yielded ~ 36% of nucleated erythroblasts and ~ 67% of enucleated reticulocytes that bear the modified GPA on the surface upon *in vitro* differentiation. Almost all of the modified GPA on both nucleated and enucleated cells were sortagged with a biotin-containing probe, as monitored by flow cytometry, immunoblotting, and immunofluorescence (Fig. 19).

As with Kell (see descriptions below), murine fetal liver lineage-negative cells were infected with retroviral vector expressing 3G-myc-hGPA and transplanted them into lethally irradiated mice (Fig. 20). After 4 weeks, these transplanted mice contained 20-50% of Terl 19 positive, discoidshaped mature RBCs expressing the modified GPA (Fig. *16B*; n= 10), which can sortagged with a biotin-containing probe with the efficiency of 85% +/- 5% as determined by flow cytometry (Fig. 16*C*). The observed reduction in gel mobility of 3Gmyc-hGPA upon sortagging with biotin can be used as a readout for reaction efficiency and indicates that most of GPA is modified by sortase (Fig. *16D).* Immunofluorescence microscopy (Fig. *16E)* confirmed that all of these modified red cells indeed have biotin conjugated to their surface. These results show that sortaggable GPA was retained on the surface of mature RBCs and was labeled efficiently in a sortase-catalyzed reaction.

(d) The type-II membrane protein Kell was expressed and sortase-labeled on RBC surface

HEK293T cells were transfected with expression vectors for producing surface fusion proteins hsCD71-LPETG (SEQ ID NO: 44)-HA and hsKell-LPETG (SEQ ID NO: 44)-HA via routine recombinant technology. Stable ceil lines expressing the fusion proteins were established. The cells were incubated with 200 μM sortase and 0.5 mM or 1 mM GGG-biotin and subjected to Western blotting analysis. The results indicate that biotin was successfully conjugated to the fusion proteins.

Extension of Kell C-terminus with the sortase recognition motif LPXTG (SEQ ID NO: 1) was the minimal modification required to render it sortase-modifiable. A retroviral construct encoding human Kell, C-terminally modified by extension with LPETG (SEQ ID NO: 44), was constructed as described herein, followed by a hemagglutinin (HA) epitope tag. A sortase reaction performed on thus modified Kell as described herein using a glycine-based probe leads to conjugation of the probe onto the Cterminus of Kell, with concomitant loss of the HA tag (Fig. *2 1A).*

To test expression and modification of Kell by sortase, an *in vitro* erythroid differentiation system described previously was employed. Zhang et al., 2003. Culture of murine fetal liver-derived progenitors *in vitro* for ~ 48 hours allows ~4 terminal cell divisions and formation of hemoglobin-containing erythroblasts, about 30-50% of which undergo enucleation to yield reticulocytes. Expression of hKell-LPETG (SEQ ID NO: 44)-HA did not inhibit the *in vitro* differentiation process. Almost half of both nucleated erythroblasts and enucleated reticulocytes displayed modified Kell at the cell surface, and a large fraction of these could be sortagged with a biotin-containing probe as shown by flow cytometry, immunoblotting, and immunofluorescence (Fig. 22).

Reticulocytes obtained by *in vitro* differentiation must undergo not only expulsion of remaining organelles, they must also execute the membrane reorganizations that lead to the biconcave disk shape of mature RBCs. To ensure that this maturation step does not lead to loss of modified Keii, murine fetal liver lineage-negative cells were infected with retroviral vectors expressing engineered hKell-LPETG (SEQ ID NO: 44)-HA and transplanted into lethally irradiated mice (Fig. 20). Mature red blood cells were harvested from transplanted mice after 4 weeks and analyzed for the presence of sortase-modifiable Kell on their surface. It was routinely found that ~ 20-50% of mature RBCs in these chimeric mice contained sortaggable human Kell on their surface (Fig. *21B)* and that these cells could be covalently modified with biotin using sortase (Fig. 2 1*C*) with efficiency of 81% +/- 28% as determined by flow cytometry. A high level of conjugation of the biotin probe onto the C-terminus of Kell upon sortagging was observed, as evidenced by complete loss of the HA tag on the immunobiot (Fig. *21D).* Immunofluorescence microscopy (Fig. 21*E)* confirms the presence of biotin conjugated to the surface of red blood cells. These data demonstrates that Kell, extended at its C-terminus with the sortase recognition motif, does not inhibit erythroid differentiation, is retained on the plasma membrane of mature red cells, and can be labeled in a sortagging reaction.

(e) Dual-labeling of RBCs

Using sortases with distinct substrate specificity, it is possible to combine N-terminal and C-terminal labeling strategies (Antos et al., 2009, J. Ameri. Chem. Soc., 131(31): 10800-10801) to generate multi-labeled RBCs. Unlike Sortase A from *Staphylococcus aureus,* Sortase A derived from *Streptococcus pyogenes* recognizes LPXTA (SEQ ID NO: 2) motifs and accepts oligo-alanine probes as nucleophiles. Therefore, the sortase reactions of both enzymes can be performed as orthogonal reactions.

HEK293T cells were infected with 2 constructs: AAA-myc-hGPA and hKell-LPETG (SEQ ID NO: 44)-HA. The cells were sequentially incubated with 2 different types of Sortase A in the presence of either biotin-LPETA (SEQ ID NO: 119; for GPA sortagging via *S. pyogenes* sortase) or GGG-TAMRA (for Kell sortagging via *S. aureus* sortase), giving rise to HEK293T cells containing both the biotinylated GPA and TAMRA-modified Kell (Fig. *16F).* Moreover, mature RBCs with cell surface expression of both 3A-myc-hGPA and hKell-LPETG (SEQ ID NO: 44)-HA were generated. Despite the fact that expression levels of hKell and hGPA on individual RBCs were somewhat variable, hGPA was successfully sortagged with biotin, and hKell with Alexa-647 on the RBC surface (Fig. 23). Dual labeling can thus be used to attach two different functional moieties onto the surface of RBCs.

(f) Survival of engineered RBCs in circulation

To assess whether the process of sortagging RBCs, including incubation with sortase, centrifugation, and washing, affects their *in vivo* survival, wild-type RBCs and wild-type RBCs that had undergone a mock C-terminal sortagging reaction- with or without sortase (no conjugation of a probe) were labeled with CFSE, a fluorescent dye that stably stains cytosol of live cells. The RBCs were then transfused into normal recipient mice; cell survival in circulation was assessed periodically using CFSE fluorescence. No difference was observed in *in vivo* survival of RBCs between the experimental groups (Fig. 24A), indicating that the sortagging procedure does not cause significant damage to the cells, which would lead to their premature removal.

<u>Example 3</u>: Conjugation of an Antibody to Engineered Red Blood Cells for Cell Type-Specific Targeting

One potential application of modified RBCs is to provide them with a targeting moiety that would enable their delivery (and attached or incorporated payload) to particular tissues or cell types. Such targeting may be accomplished by installation of proteins or other entities, such as ligands for specific receptors that can participate in specific recognition of the intended target.

To test conjugation of a functional protein onto the surface of RBCs, a sortaggable alpaca-derived single domain antibody was used. This antibody is specific for murine Class II MHC molecules, VHH7-LPETG (SEQ ID NO: 44), and was covalently attached it to red cells with 3G-myc-hGPA on their surface. As shown by the shift in gel mobility of the 3G-myc-hGPA, the sortagging reaction is stoichiometric (Fig. *25A).*

A B cell - red blood cell binding assay was performed as follows. B cells were isolated from either wild-type C57BL/6 mice or mice knock-out for MHC class II using Dynabeads Mouse CD43 (Untouched B cells) (Life Technologies) according to manufacturer's recommendations. Cells from 1-1.5 spleen were incubated with 5 µg of biotin-labeled anti-murine CD19 antibody (BD Pharmigen) for 30 minutes on ice and washed once with 500 µl of ice-cold PBS. Cells were re-suspended in 500 µl of PBS along with 70 µl of pre-washed magnetic Dynabeads Myone streptavidin T1 (Life Technologies), incubated on a rotating platform for 1 hour at 4 °C, and washed with 500 µl of ice-cold PBS. Resuspended cells (500 µl of PBS) were further incubated with ~ 2x107 red blood cells expressing 3G-myc-hGPA and sortase-labeled with VHH7 for 1 hour at 4 °C, washed 4 times with 1 ml of ice-cold PBS, and finally incubated with SDS sample buffer and boiled for subsequent SDS-PAGE analysis.

Incubation of these modified RBCs with Class II MHC-positive B cells immobilized on beads results in binding of the RBCs to B cells. Specific binding was inferred from the presence of hemoglobin, which co-purifies with B cells after washing. Neither incubation of wild-type B cells with unmodified red cells nor incubation of modified red cells with B cells derived from Class II MHC knock-out animals led to binding of the two cell types (Fig. *25B, C*).

<u>Example 4:</u> Sortagging and Cytosolic Modification of Terminally Differentiated Human Red Blood Cells

To extend the utility of the method described here, whether erythroid progenitor modification, as well as subsequent sortase-mediated cell surface labeling is feasible also for human cells, was investigated.

Terminally differentiated human red blood cells were genetically engineered as follows. Plasmids used to engineer in vzYro-differentiated human RBCs were created by cloning the GGGcontaining human GPA into HIV/MSCV lentiviral vector with the addition of enhanced GFP at its C-terminus to make 3G-myc-hGPA-EGFP. Lentivirus was produced by cotransfection of 293T cells with pVSV-G envelope plasmids and pDelta 8.9 packaging vectors. G-CSF (granulocyte - colony stimulating factor)-mobilized CD34+ peripheral blood stem cells were differentiated *in vitro* in 18 days into hemoglobin-containing reticulocytes using the method as described previously(Zaitsev et al., 2010, *Blood* 115(25):5241-5248). On day 3 of culture, differentiating cells were infected with by incubating $5x10^5$ cells in 2mL of lentiviruscontaining media in the presence of polybrene, and spun at 2500 RPM for 1.5 hours at room temperature, followed by further incubation in a C02 incubator overnight. The day after, the infected cells were washed twice and put into fresh differentiation media. At day 18, the enucleated reticulocytes were collected and subjected to sortase-labeling with a biotin-containing probe as described above. Analyses of the resulting engineered human reticulocytes were carried out by means of flow cytometry using the following antibodies: PE-conjugated anti-biotin (1:1000, eBioscience, 12-9895-82), Hoechst (Sigma), and APC-conjugated (1:100, eBioscience, 17-0087-42).

For the experiments described in this example, an *in vitro* human erythroid differentiation system as described previously. See, e.g., Hu et al., 2013. G-CSF (granulocyte - colony stimulating factor)-mobilized CD34+ peripheral blood stem cells were differentiated *in vitro* in 18 days into hemoglobin-containing reticulocytes, with an enucleation efficiency of -50%. A lentiviral vector that encodes a version of 3G-myc-hGPA fused to enhanced GFP at its C-terminus (3G-myc-GPA-GFP) was constructed. This provides an example of a cytoplasmically expressed protein domain designed to be retained in mature RBC *via* its genetic fusion to GPA. Both the empty (control) vector and the 3G-myc-hGPA-GFP vector encoded a GFP moiety 3' of the IRES sequence (3G-myc-hGPA-GFP-IRES-GFP). When cells are successfully transduced with 3G-myc-hGPA-GFP, which contains an additional enhanced GFP attached to the cytoplasmic domain of GPA, these cells will express a significantly higher level (~2x) of GFP signal, as indicated by flow cytometry (Fig. 26*A*, first panel). Since the molecular weights of GFP and hGPA are 32.7 kDa and 37 kDa respectively, the mobility of the 3G-myc-hGPA-GFP on SDS-PAGE also indicates that the modified GPA proteins are expressed in monomeric and dimeric forms (Fig. 265). Viral transduction of human CD34+ cells with 3G-myc-hGPA-GFP did not affect their differentiation, as similar numbers of enucleated reticulocytes (50-60%) were formed as in cultures infected with an empty (control) vector (Fig. *26A,* middle panel).

All of the enucleated reticulocytes that contain the modified GPA at the surface could be sortaselabeled with a biotin-containing probe, as monitored by flow cytometry (Fig. *26A).* Furthermore, all of the 3G-myc-hGPA-GFP present on the reticulocyte surface was modified by biotin sortagging, as evidenced by the shift in gel mobility of the 3G-mychGPA-GFP (Fig. 265). This experiment therefore establishes the ability to equip mature human RBCs with a cytoplasmically disposed protein of interest, while retaining the ability to selectively target the RBC to an intended target by sortagging of the modified GPA with a targeting moiety of interest.

<u>Example 5:</u> Creation of mKell-LPETG (SEQ ID NO: 44) mice using CRISPP*J* Cas9 technology

An sgRNA sequence for CRISPR7Cas9-dependent double strand break to stimulate Homologous direct repair (HDR) was designed on the last exon of mKel locus as in Table 5, and the oligo DNA as a template to introduce LPETG (SEQ ID NO: 44) into mKel C-terminus was designed as in Table 6.

Table 5

sgRNA	5' to 3' sgRNA template sequence with PAM motif
mKel CT-targeting sgRNA	CTCTGCCCGCTGCAAGCTC<u>TGG</u> (SEQ ID NO: 120)

*Underlined: PAM motif

Table 6

Template Oligo DNA for HDR	5' to 3' Oligo DNA sequence
mKel-LPETG (SEQ ID NO: 44) template Oligo DNA	TGAGCAATACTCCAGATTTTGCCAAACATTTTCATTGTCC ACGTGGGACCCTTCTGAATCCCTCTGCCCGCTGCAAGCT**<u>C GGAGGATCAGGAGGATCATTACCAGAGACAGGAGG</u>**AT GGTAAAACTTGGCTACCAAAGAGACTGATGTAAATGCAT GGGCTGCTTGTGAGTCCATCCTTGAAGTCAAAATAAATCT (SEQ ID NO: 121)

- Boldfaced/Underlined: LPETG (SEQ ID NO: 44)-encoding sequence

Potential off-target effects of the target sequence were explored using the NCBI Mus musculus Nucleotide BLAST. Cas9 mRNA was prepared as described in Wang et al., 2013. T7 promoter and the mKel-CT-targeting sgRNA coding sequence without PAM motif were added just in front of sgRNA generic tail (the part from hairpin region to terminal motif) by PCR amplification using pX330 vector as a template and the primers, mKel T7-sgRNA Fw and sgRNA generic Rv, as shown in Table 7 below.

Table 7

Primers	5' to 3' primer sequence
mKel T7-sgRNA Fw	*TAATACGACTCACTATA*<u>GCTCTGCCCGCTGCAAGCTC</u>gtttagagctagaaatagcaag (SEQ ID NO: 122)
sgRNA generic Rv	TAAGTTATGTAACGGGTAC (SEQ ID NO: 123)
mKel sequencing Fw	ATCTAACCCATCCCTATCACCCTATGG (SEQ ID NO: 124)
mKel sequencing Rv	ATGGAGATGTAGCTGATGAGCAGC (SEQ ID NO: 125)

- Italic: T7 promoter
- Underlined: sgRNA target template
- Lower case: 5'-generic sequence of sgRNA tail

The T7-sgRNA template PCR product for *in vitro* transcription (IVT) was gel purified by column (Promega), and the purified PCR product was used as a template for IVT using MEGAshortscript T7 kit (Life Technologies). Both the Cas9 mRNA and the sgRNA were purified using MEGAclear kit (Life Technologies) and eluted in RNase-free water. Fertilized zygotes were collected from oviducts of superovulated females, and Cas9 mRNAs, sgRNA and template oligo DNA were injected into the cytoplasm at the pronuclear stage. The injected zygotes were transferred at the 2-cell stage into the oviduct of pseudopregnant females. To genotype the mice: mouse-tails were lysed at 55°C for 12 hours using tail lysis buffer and genomic DNA was purified from the lysates by isopropanol precipitation. Genomic DNA in the vicinity of the sgRNA target was amplified by PCR using primers mKel-sequencing Fw and mKel-sequencing Rv, as in table 3 [KOD Xtreme (VWR), Condition: 95°C for 2 min; 35x(98°C for 10 s, 50°C for 30s, 68°C for 30 s); 68°C for 2 min;

hold at 4 °C], and then gel purified. The purified PCR products were sequenced using the same Fw primer.

The transgenic mice constructed using CRISPR/Cas-9 genome editing technology (Jinek et al., 2012, *Science* 337(6096):816-821; and Wang et al, 2013, *Cell* 153(4):910-918) were used to investigate RBC survival in vivo. Such a transgenic mouse contains LPETG (SEQ ID NO: 44) inserted at the C-terminus of the murine endogenous Kell gene (mKell-LPETG; SEQ ID NO: 44). These mice appear normal and reproduce normally. The mKell-LPETG (SEQ ID NO: 44) RBCs from these mice can be labeled with a probe in a sortase-mediated reaction as efficiently as hKell-LPETG (SEQ ID NO: 44)-HA RBCs.

CFSE staining was performed on control RBCs, mKell-LPETG (SEQ ID NO: 44) RBCs, and mKell-LPETG (SEQ ID NO: 44) RBCs sortagged with biotin, transfused them into wild- type recipient mice, and their survival in circulation was monitored. The modification of mKell with LPETG (SEQ ID NO: 44) at its C-terminus did not affect the survival of the modified RBCs *in vivo* (Fig. 245). A slight, but significant difference was observed in survival of mKell-LPETG (SEQ ID NO: 44) RBCs sortagged with a biotin probe. hKell-LPETG (SEQ ID NO: 44) RBCs were also harvested from transplanted mice, sortagged with biotin, and transfused into recipients. Similarly, engineered biotin-labeled Kell-LPETG (SEQ ID NO: 44) RBCs last in circulation more than 28 days albeit with a slight lower survival compared to wild-type RBCs (Fig. 24Q. These results show similar experiments comparing the in vivo survival of normal RBCs to sortase-labeled biotin-3G-myc-hGPA RBCs. The results (Fig. 24D) echoed the trend seen with Kell-LPETG (SEQ ID NO: 44) RBCs; there was a slightly but significantly lower survival rate for sortagged RBCs (Fig. 24*D*). Importantly, the half-life of the engineered RBCs in normal mice, including the ones conjugated with biotin, nevertheless is at least 28 days and thus drastically longer than other RBC-based carriers designed thus far.

OTHER EMBODIMENTS

All of the features disclosed in this specification may be combined in any combination. Each feature disclosed in this specification may be replaced by an alternative feature serving the same, equivalent, or similar purpose. Thus, unless expressly stated otherwise, each feature disclosed is only an example of a generic series of equivalent or similar features.

From the above description, one skilled in the art can easily ascertain the essential characteristics of the present invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, other embodiments are also within the claims.

CLAIMS

What is claimed is:

1. A genetically engineered enucleated blood cell, which expresses on the surface a first fusion protein comprising a first peptide of interest and a first red blood cell membrane protein.

2. The genetically engineered enucleated blood cell of claim 1, wherein the first red blood cell membrane protein is a type I red blood cell transmembrane protein, and the first peptide of interest is fused to the N-terminus of the type I red blood cell transmembrane protein.

3. The genetically engineered enucleated blood cell of claim 2, wherein the type I red blood cell transmembrane protein is glycophorin A (GPA).

4. The genetically engineered enucleated blood cell of claim 1, wherein the first red blood cell membrane protein is a type II transmembrane protein, and the first peptide of interest is fused to the C-terminus of the type II transmembrane protein.

5. The genetically engineered enucleated blood cell of claim 4, wherein the first peptide of interest comprises a first sequence recognizable by a first sortase.

6. The genetically engineered enucleated blood cell of claim 5, wherein the first sortase is a first sortase A.

7. The genetically engineered enucleated blood cell of claim 6, wherein the sequence recognizable by the first sortase A is LPXTG (SEQ ID NO: 1), in which X is any amino acid residue.

8. The genetically engineered enucleated blood cell of any of claims 4-7, wherein the type II red blood cell transmembrane protein is Kell or CD71.

9. The genetically engineered enucleated blood cell of claim 1, wherein the first red blood cell membrane protein is a type III red blood cell transmembrane protein.

10. The genetically engineered enucleated blood cell of claim 9, wherein the type III red blood cell transmembrane protein is GLUT1.

11. The genetically engineered enucleated blood cell of any of claims 1-10, wherein the peptide of interest comprises a protein drug, a vaccine antigen, a fluorescent protein, streptavidin, biotin, an enzyme, or a peptide capable of targeting a cell.

12. The genetically engineered enucleated blood cell of claim 11, wherein the first peptide of interest is a peptide capable of targeting a disease cell.

13. The genetically engineered enucleated blood cell of claim 12, wherein the first peptide of interest is an antibody or a fragment thereof.

14. The genetically engineered enucleated blood cell of claim 13, wherein the antibody is a single domain antibody.

15. The genetically engineered enucleated blood cell of any of claims 1-14, wherein the first peptide of interest is conjugated to a detectable label or a chemotherapeutic agent.

16. The genetically engineered enucleated blood ceil of any of claims 1-B14, wherein the first peptide of interest is conjugated to a lipid, a carbohydrate, a nucleic acid, a binding agent, a click-chemistry handle, or a small molecule.

17. The genetically engineered enucleated blood cell of any of claims 1-16, wherein the first fusion protein further comprises a protein of interest, wherein the protein of interest is fused to the terminus of the first red blood cell membrane protein that is exposed to a cytoplasmic space, and wherein the first peptide of interest is fused to the terminus of the first red blood cell membrane protein that is exposed to an extracellular or luminal space.

18. The genetically engineered enucleated blood cell of claim 17, wherein the first red blood cell membrane protein is a type I membrane protein, wherein the protein of interest is fused to the C-terminus of the type I membrane protein, and wherein the first peptide of interest is fused to the N-terminus of the type I membrane protein.

19. The genetically engineered enucleated blood cell of claim 17, wherein the first red blood cell membrane protein is a type II membrane protein, wherein the protein of interest is fused to the N-terminus of the type II membrane protein, and wherein the first peptide of interest is fused to the C-terminus of the type I membrane protein.

20. The genetically engineered enucleated blood cell of any of claims 17-19, wherein the protein of interest is a cytoplasmic protein.

21. The genetically engineered enucleated blood ceil of claim 20, wherein the protein of interest is a diagnostic agent or a therapeutic agent.

22. The genetically engineered enucleated blood cell of claim 20 or claim 21, wherein the first peptide of interest is a protein drug, a vaccine antigen, a fluorescent protein, streptavidin, biotin, an enzyme, or a peptide capable of targeting a cell.

23. The genetically engineered enucleated blood cell of claim 22, wherein the first peptide of interest is a peptide capable of targeting a disease cell.

24. The genetically engineered enucleated blood cell of any of claims 1-23, which further expresses on the surface a second fusion protein comprising a second peptide of interest and a second red blood cell membrane protein.

25. The genetically engineered enucleated blood cell of claim 24, wherein the first peptide of interest in the first fusion protein comprises a recognizable site or an acceptable peptide of a first sortase and the second peptide of interest in the second fusion protein comprises a recognizable site or an acceptable peptide of a second sortase, and wherein the

first sortase and the second sortase use different recognizable sites and different acceptable peptides.

26. The genetically engineered enucleated blood cell of claim 25, wherein the first peptide of interest comprises the motif of LPXTA (SEQ ID NO: 2), in which X is any amino acid residue, or an oligoalanine; and the second peptide of interest comprises the motif LPXTG (SEQ ID NO: 1), or an oligoglycine.

27. The genetically engineered enucleated blood cell of claim 24, wherein the first fusion protein comprises Kell, the C-terminus of which is fused to a first peptide of interest comprising the motif LPXTG (SEQ ID NO: 1), and the second fusion protein comprises GPA, the C-terminus of which is fused to a second peptide of interest comprising an oligoalanine.

28. The genetically engineered enucleated blood cell of any of claims 24-27, wherein the first fusion protein and the second fusion protein are conjugated to two different functional moieties.

29. A method for delivering an agent into a subject, the method comprising administering to a subject in need thereof the enucleated blood cell of any of claims 11-28.

30. The method of claim 29, wherein the enucleated blood cell is derived from the same subject.

31. A method for conjugating a peptide of interest to the surface of red blood cells, the method comprising:

providing a red blood cell expressing a first fusion protein comprising a first red blood cell membrane protein and a first peptide, and

contacting the red blood cell with a first peptide of interest in the presence of a first sortase under conditions suitable for the sortase to conjugate the peptide of interest to the first peptide in the fusion protein, wherein either the first peptide in the fusion protein or the first peptide of interest comprises a sequence recognizable by the first sortase.

32. The method of claim 31, wherein the sortase is a Sortase A.

33. The method of claim 31 or 32, wherein the sequence recognizable by the sortase is LPXTG (SEQ ID NO: 1), in which X is any amino acid residue.

34. The method of any of claims 31-33, wherein the first red blood cell membrane protein is a type I red blood cell transmembrane protein and the first peptide is an acceptor peptide fused to the N-terminus of the type I red blood cell transmembrane protein, and wherein the first peptide of interest comprises the sequence recognizable by the sortase.

35. The method of claim 34, wherein the type I red blood cell transmembrane protein is giycophorin A (GPA).

36. The method of claim 34 or 35, wherein the acceptor peptide is an oligoglycine.

37. The method of claim 36, wherein the acceptor peptide is an oligoglycine consisting of 1-5 glycine residues.

38. The method of any of claims 31-33, wherein the first red blood cell membrane protein is a type II red blood cell transmembrane protein and the first peptide is fused to the C-terminus of the type II red blood cell transmembrane protein, and wherein the first peptide comprises the sequence recognizable by the sortase.

39. The method of claim 38, wherein the type II red blood cell transmembrane protein is Kell or CD71.

40. The method of any of claims 31-33, wherein the first red blood cell membrane protein is a type III red blood cell transmembrane protein.

41. The method of claim 40, wherein the type III red blood cell transmembrane protein is GLUT1.

42. The method of any of claims 31-41, wherein the first peptide of interest comprises a protein drug, a vaccine antigen, a fluorescent protein, streptavidin, biotin, an enzyme, or a peptide capable of targeting a cell.

43. The method of claim 42, wherein the first peptide of interest is a peptide capable of targeting a disease cell.

44. The method of claim 42, wherein the first peptide of interest is an antibody or a fragment thereof.

45. The method of claim 44, wherein the antibody is a single domain antibody.

46. The method of any of claims 31-45, wherein the first peptide of interest is conjugated to a detectable label or a chemotherapeutic agent.

47. The method of any of claims 31-45, wherein the first peptide of interest is conjugated to a lipid, a carbohydrate, a nucleic acid, a binding agent, a click-chemistry handle, or a small molecule.

48. The method of any of claims 31-47, wherein the first fusion protein further comprises a protein of interest, and wherein the protein of interest is fused to the terminus of the first red blood cell membrane protein that is exposed to a cytoplasmic space and the first peptide of interest is fused to the terminus of the first red blood cell membrane protein that is exposed to an extracellular or luminal space.

49. The method of claim 48, wherein the first red blood cell membrane protein is a type I membrane protein, wherein the first protein of interest is fused to the C-terminus of the type I membrane protein, and wherein the first peptide is fused to the N-terminus of the type I membrane protein.

50. The method of claim 48, wherein the first red blood cell membrane protein is a type II membrane protein, wherein the protein of interest is fused to the N-terminus of the type II membrane protein, and wherein the first peptide is fused to the C-terminus of the type I membrane protein.

51. The method of any of claims 48-50, wherein the protein of interest is a cytoplasmic protein.

52. The method of claim 51, wherein the protein of interest is a diagnostic agent or a therapeutic agent.

53. The method of claim 51 or 52, wherein the first peptide of interest is a protein drug, a vaccine antigen, a fluorescent protein, streptavidin, biotin, an enzyme, or a peptide capable of targeting a cell.

54. The method of claim 53, wherein the first peptide of interest is a peptide capable of targeting a disease cell.

55. The method of any of claims 31-54, the method further comprising:
contacting the red blood cell in the presence of a second sortase under conditions suitable for the second sortase to conjugate a second peptide of interest to a second fusion protein expressed on the surface of the red blood cell, the second fusion protein comprising a second red blood ceil membrane protein and a second peptide to which the second peptide of interest conjugates;

wherein either the second peptide in the fusion protein or the second peptide of interest comprises a sequence recognizable by the second sortase, and

wherein the sequence recognizable by the first sortase differ from the sequence recognizable by the second sortase.

56. The method of claim 55, wherein the first peptide in the first fusion protein comprises a sequence recognizable by the first sortase or an acceptable peptide of the first sortase; and

the second peptide in the second fusion protein comprises a sequence recognizable by the second sortase or an acceptable peptide of the second sortase, and wherein the acceptable peptide of the first sortase is different from the acceptable peptide of the second sortase.

57. The method of claim 56, wherein the first sortase is Sortase A from *Staphylococcus arreus*, and wherein one of the first peptide in the first fusion protein and the first peptide of interest comprises the motif LPXTG (SEQ ID NO: 1), in which X is any amino acid residue, and the other comprises an acceptable peptide, which is an oligoglycin.

58. The method of claim 57, wherein the first fusion protein comprises Kell, the C-terminus of which is fused to the first peptide which comprises the motif LPXTG (SEQ ID NO: 1), and the first peptide of interest comprises an oligoglycine.

59. The method of claim 57 or claim 58, wherein the second sortase is Sortase A from *Streptococcuspyogenes*, and wherein one of the second peptide in the second fusion protein and the second peptide of interest comprises the motif LPXTA (SEQ ID NO: 2), in which X is any amino acid residue, and the other comprises an acceptable peptide, which is an oligoalanine.

60. The method of claim 59, wherein the second fusion protein comprises GPA, the N-terminus of which is fused to the second peptide which comprises an oligoglycine, and the second peptide of interest comprises the motif LPXTG (SEQ ID NO: 1), in which X is any amino acid residue.

61. The method of any of claims 55-60, wherein the first and second peptides of interest comprises or are conjugated to two different functional moieties.

62. A method for producing human enucleated red blood cells, the method comprising:

(i) providing a population of human $CD34^+$ progenitor cells;

(ii) expanding the population of human $CD34^+$ progenitor cells in a first medium for 0-6 days, wherein the first medium comprises Flt-3 ligand, stem cell factor (SCF), interleukin 3 (IL-3), and interleukin 6 (IL-6);

(iii) differentiating the expanded human $CD34^+$ peripheral blood cells obtained from step (ii) in a second medium for 4-7 days, wherein the second medium comprises dexamethasone, β-estradiol, IL-3, SCF, and erythropoietin (EPO);

(iv) differentiating the cells obtained from step (iii) in a third medium for 3~5 days, wherein the third medium comprises SCF and EPO;

(v) differentiating the cells obtained from step (iv) in a fourth medium for 4~12 days to produce human enucleated red blood cells, wherein the fourth medium comprises EPO; and

(vi) collecting the human enucleated red blood cells obtained from step (v).

63. The method of claim 62, wherein in step (ii), the population of human $CD34^+$ peripheral blood cells is cultured in the first medium for 4 days.

64. The method of claim 62 or claim 63, wherein in step (iii), the expanded human $CD34^+$ progenitor cells are differentiated in the second medium for 5 days.

65. The method of any of claims 62-64, wherein in step (iv), the cells are differentiated in the third medium for 4 days.

66. The method of any of claims 62-65, wherein in step (v), the cells are differentiated in the fourth medium for 9 days.

67. The method of any of claims 62-66, wherein the total time period for steps (ii)-(v) ranges from 11-25 days.

68. The method of claim 67, wherein the total time period for steps (ii)-(v) is 21 days.

69. The method of any of claims 62-68, wherein step (v) is performed in a culture container having surface coating of an extracellular matrix component.

70. The method of any of claims 62-69, wherein one or more of the second, third, and fourth media further comprise holo human transferrin and insulin.

71. The method of claim 70, wherein the second medium comprises 250-1 500 μg/ml holo human transferrin, 5-20 μg/ml insulin, 100 nM~5 μM dexamethasone, 0.1-5 μM β-estradiol, 1-10 ng/ml IL-3, 10-500 ng/ml SCF, and 2-10 U EPO.

72. The method of claim 71, wherein the second medium comprises approximately 500 μg/ml holo human transferrin, approximately 10 μg/ml insulin, approximately 2 μM dexamethasone, approximately 1 μM β-estradiol, approximately 5 ng/ml IL-3, approximately 100 ng/ml SCF, and approximately 6 U EPO.

73. The method of any of claims 70-72, wherein the second medium is free of Fit-3 ligand, IL-6, or both.

74. The method of any of claims 70-73, wherein the third medium comprises 250-1 500 μg/ml holo human transferrin, 5-20 μg/ml insulin, 10-100 ng/ml SCF, and 2-10 U EPO.

75. The method of claim 74, wherein the third medium comprises approximately 500 μg/ml holo human transferrin, approximately 10 μg/ml insulin, approximately 50 ng/ml SCF, and approximately 6 U EPO.

76. The method of claim 70, 74, or 75, wherein the third medium is free of Flt-3 ligang, IL-6, dexamethasone, β-estradiol, IL-3, or a combination thereof.

77. The method of any of claims 70-76, wherein the fourth medium comprises 250-1500 μg/ml holo human transferrin, 5-20 μg/ml insulin, and 0.5-3 U EPO.

78. The method of claim 77, wherein the fourth medium comprises approximately 500 μg/ml holo human transferrin, approximately 10 μg/ml insulin, and approximately 2 U EPO.

79. The method of claim 70, 77, or 78, wherein the fourth medium is free of Fit-3 ligand, IL-6, dexamethasone, β-estradiol, IL-3, SCF, or a combination thereof.

80. The method of any of claims 62-79, wherein the human $CD34^+$ progenitor cells express a fusion protein comprising a red blood cell membrane protein and a peptide of interest.

81. The method of claim 80, wherein the fusion protein comprises a type I red blood cell transmembrane protein fused to an acceptor peptide at the N-terminus of the type I red blood cell transmembrane protein.

82. The method of claim 81, wherein the type I red blood cell transmembrane protein is glycophorin A.

83. The method of claim 81 or 82, wherein the acceptor peptide is an oiigogiycine.

84. The method of claim 83, wherein the acceptor peptide is an oliogoglycine that consists of 1-5 glycine residues.

85. The method of claim 80, wherein the fusion protein comprises a type II red blood cell transmembrane protein fused to a peptide comprising a sequence recognizable by a sortase at the C-terminus of the type II red blood cell transmembrane protein.

86. The method of claim 85, wherein the sortase is a sortase A.

87. The method of claim 86, wherein the sequence recognizable by the sortase A is LPXTG (SEQ ID NO: 1), in which X is any amino acid residue.

88. The method of any of claims 85-87, wherein the type II red blood cell transmembrane protein is Kell or CD71 .

89. The method of claim 80, wherein the membrane protein is a type III red blood cell transmembrane protein.

90. The method of claim 89, therein the type III red blood cell transmembrane protein is glucose transporter 1 (GLUT1).

91. The method of any of claims 62-90, wherein the human $CD34^+$ progenitor cells are human $CD34^+$ peripheral blood cells.

92. A human enucleated blood cell population prepared by any of the methods of 62-91.

Figure 1

<table>
<tr><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td><td>16</td><td>17</td><td>18</td><td>19</td><td>20</td><td>21</td><td>22</td></tr>
<tr>
<td colspan="4"><Expansion> (4 days)
• StemSpan SFEM medium
• 1X CC100 cytokine mix (Flt-3 ligand, SCF, IL-3, IL-6)
• 2% Pen/Strep</td>
<td colspan="5"><Differentiation I> (5 days)
• IMDM
• 15% FBS
• 2mM glutamine
• 1% BSA
• 2% Pen/Strep
• holo human transferrin: 500 µg/mL
• rh insulin: 10 µg/mL
• Dexamethasone: 2 µM
• β-estradiol: 1 µM
• IL-3: 5 ng/mL
• SCF 100 ng/mL
• Epo 6U</td>
<td colspan="4"><Differentiation II> (4 days)
• IMDM
• 15% FBS
• 2mM glutamine
• 1% BSA
• 2% Pen/Strep
• holo human transferrin: 500 µg/mL
• rh insulin: 10 µg/mL
• SCF 50 ng/mL
• Epo 6U</td>
<td colspan="9"><Differentiation III> (9 days)
• IMDM
• 15% FBS
• 2mM glutamine
• 1% BSA
• 2% Pen/Strep
• holo human transferrin: 500 µg/mL
• rh insulin: 10 µg/mL
• Epo 2U

* Fibronectin plates</td>
</tr>
</table>

Figure 2

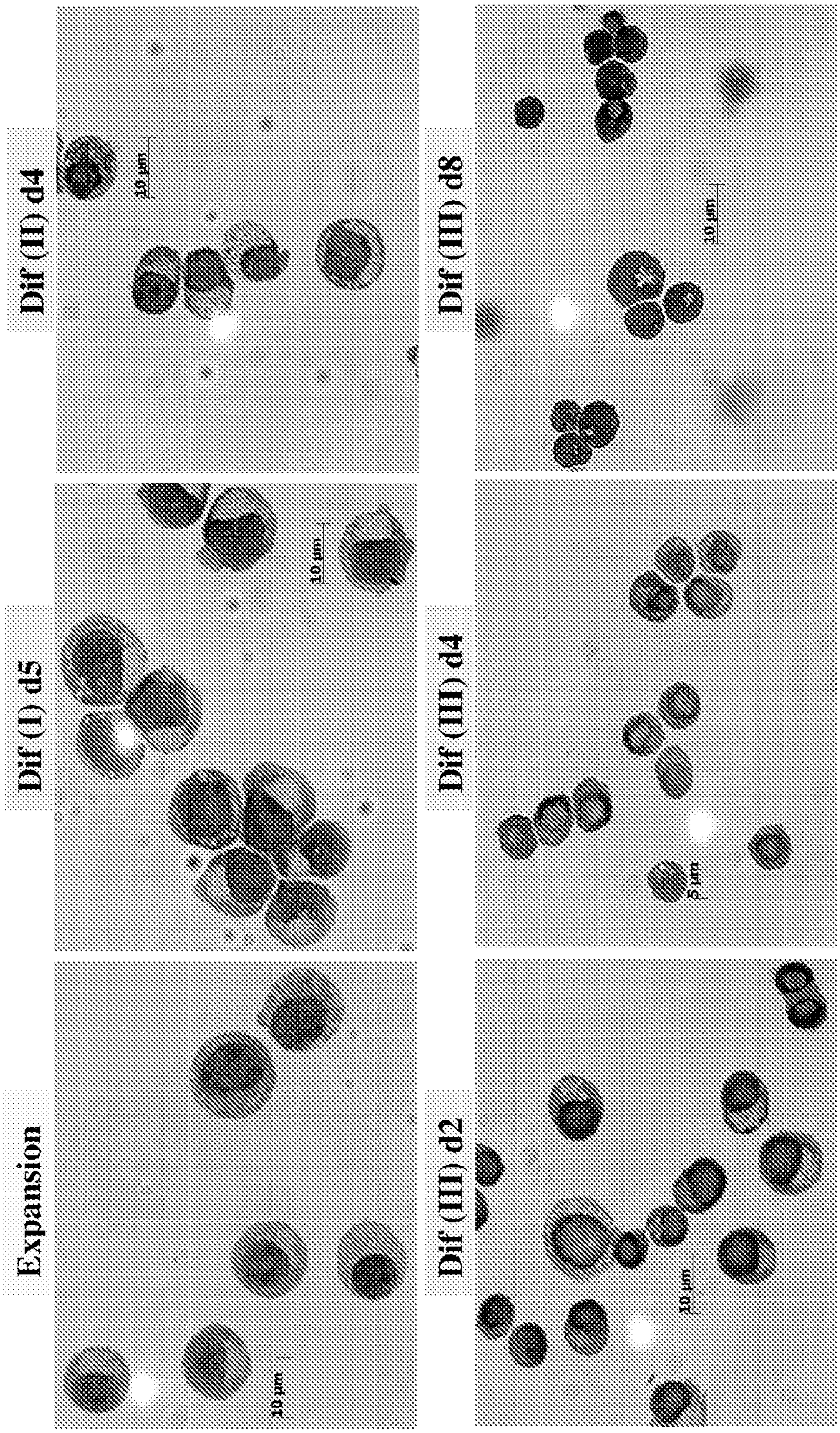


Figure 3

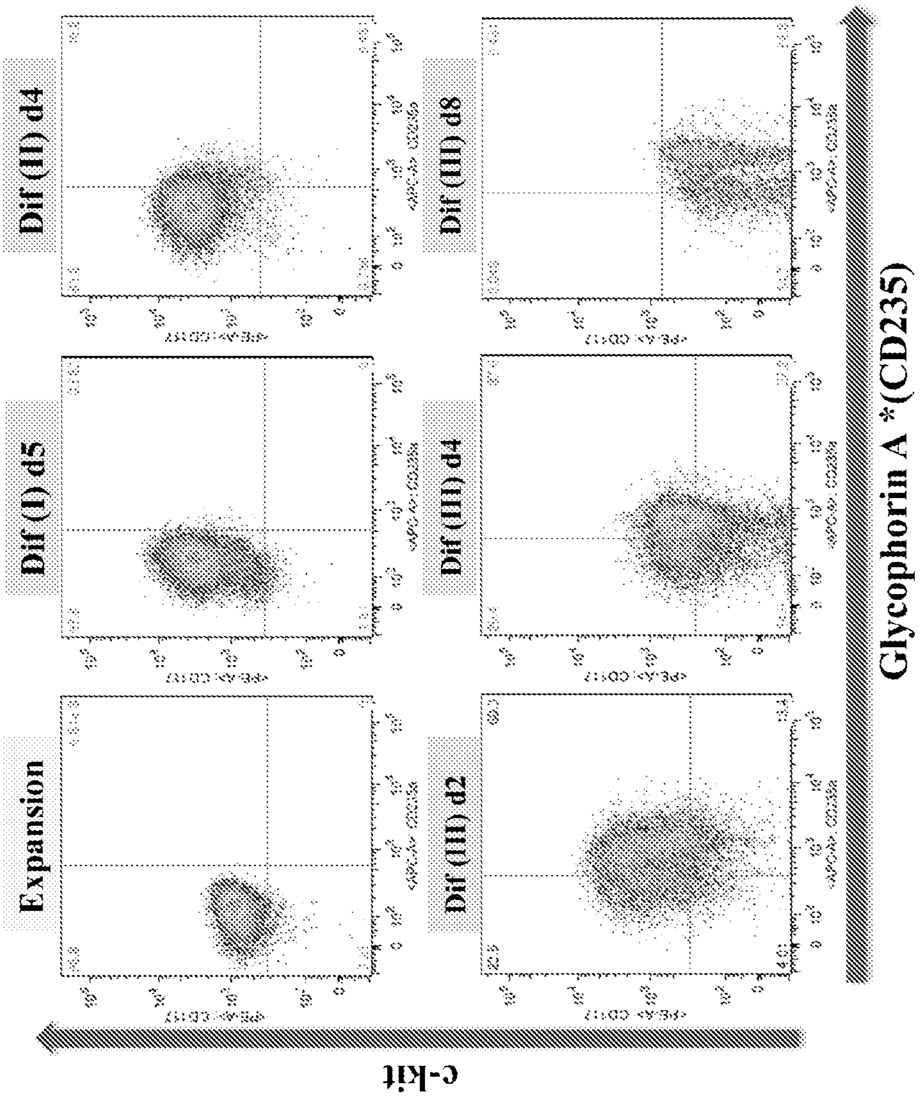

Expansion
Dif (I) d5
Dif (II) d4
Dif (III) d2
Dif (III) d4
Dif (III) d8
c-kit
Glycophorin A *(CD235)

Figure 4

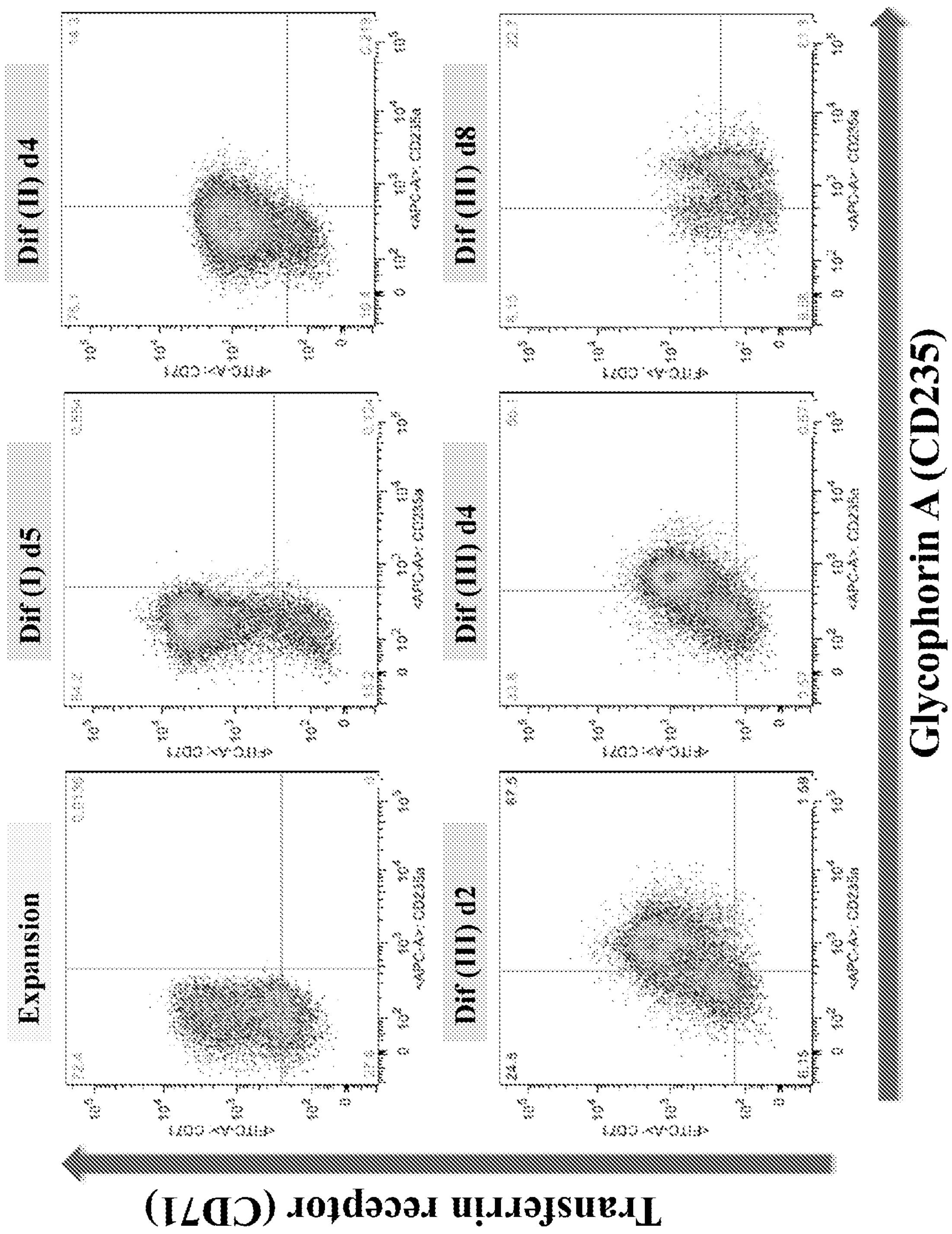

Expansion
Dif (I) d5
Dif (II) d4
Dif (III) d2
Dif (III) d4
Dif (III) d8
Glycophorin A (CD235)
Transferrin receptor (CD71)

Figure 5

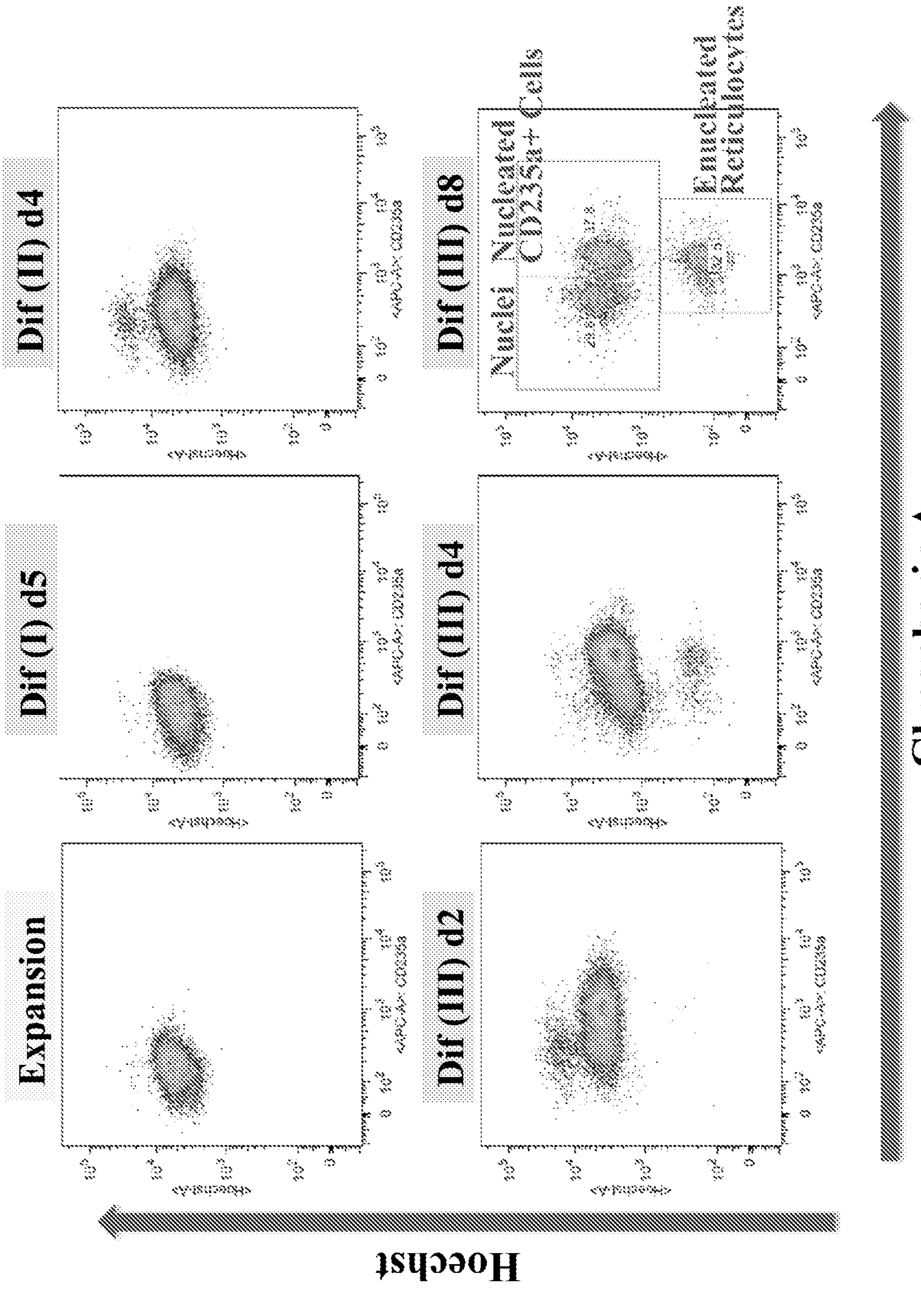
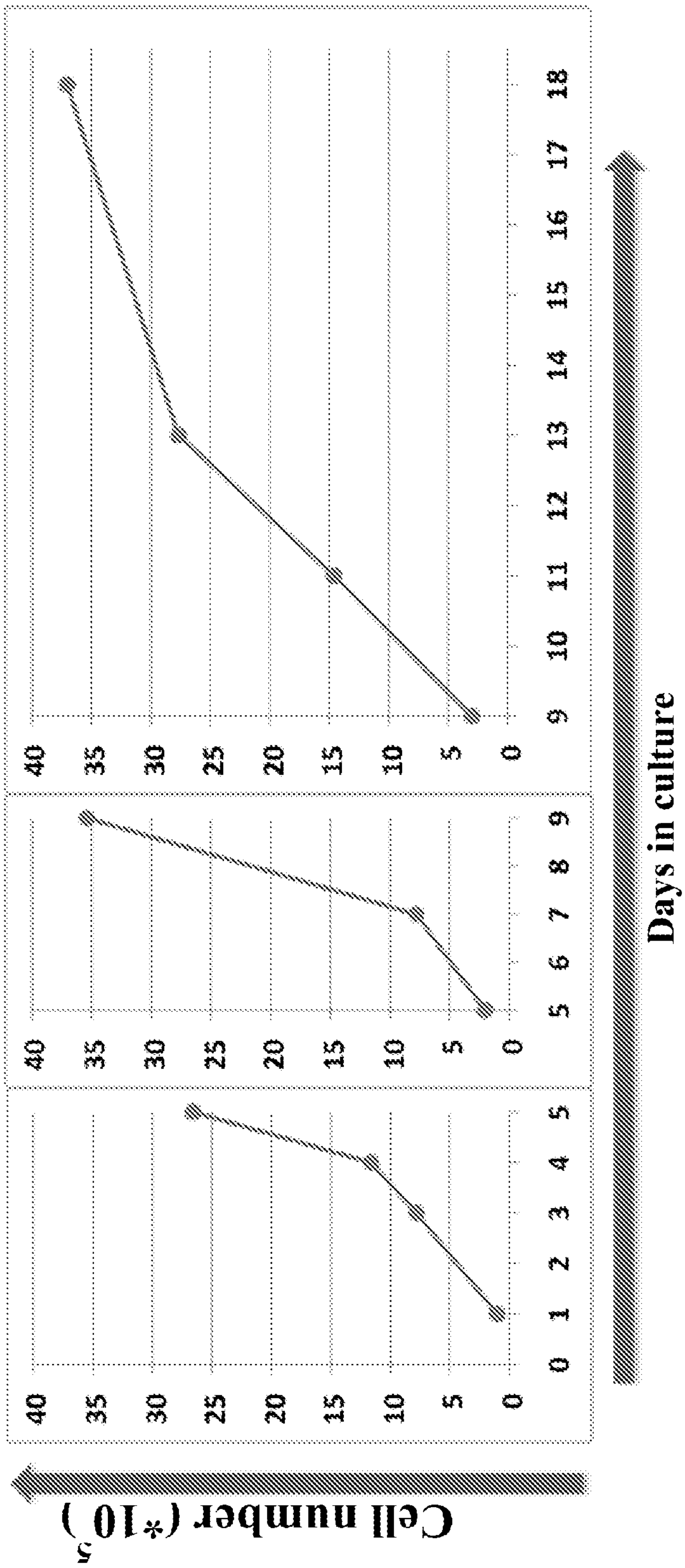
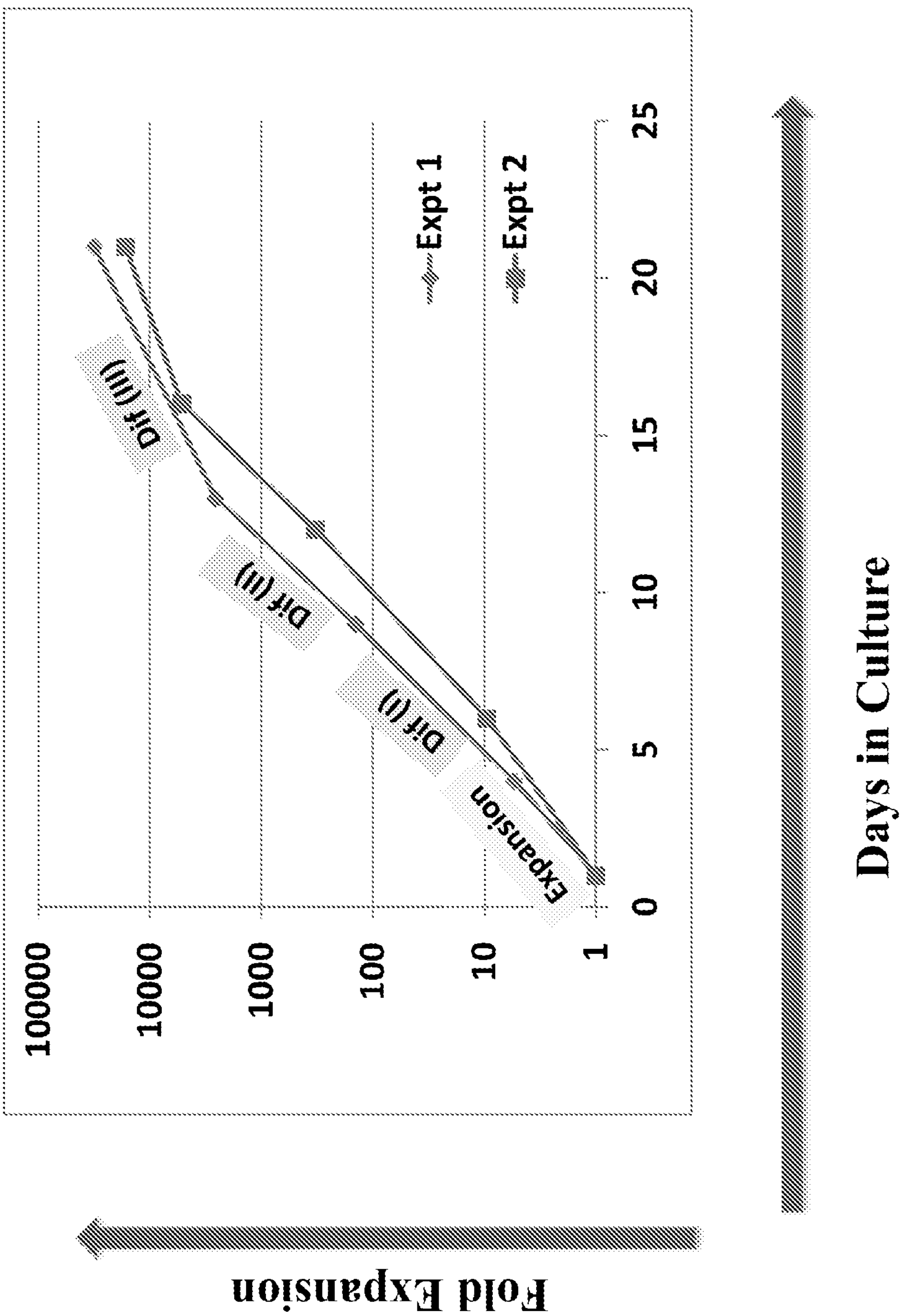

Expansion
Dif (I) d5
Dif (II) d4
Dif (III) d2
Dif (III) d4
Dif (III) d8
Nuclei
Nucleated CD235a+ Cells
Enucleated Reticulocytes
Glycophorin A
Hoechst

Figure 6



	Dif I (5 d)	Dif II (4 d)	Dif III (9 d)	Total
Fold expansion	26.6	17.8	12.3	5836.2

Figure 7



Fold Expansion
Days in Culture
100000
10000
1000
100
10
1
0
5
10
15
20
25
Expansion
Dif (I)
Dif (II)
Dif (III)
Expt 1
Expt 2

Figure 8

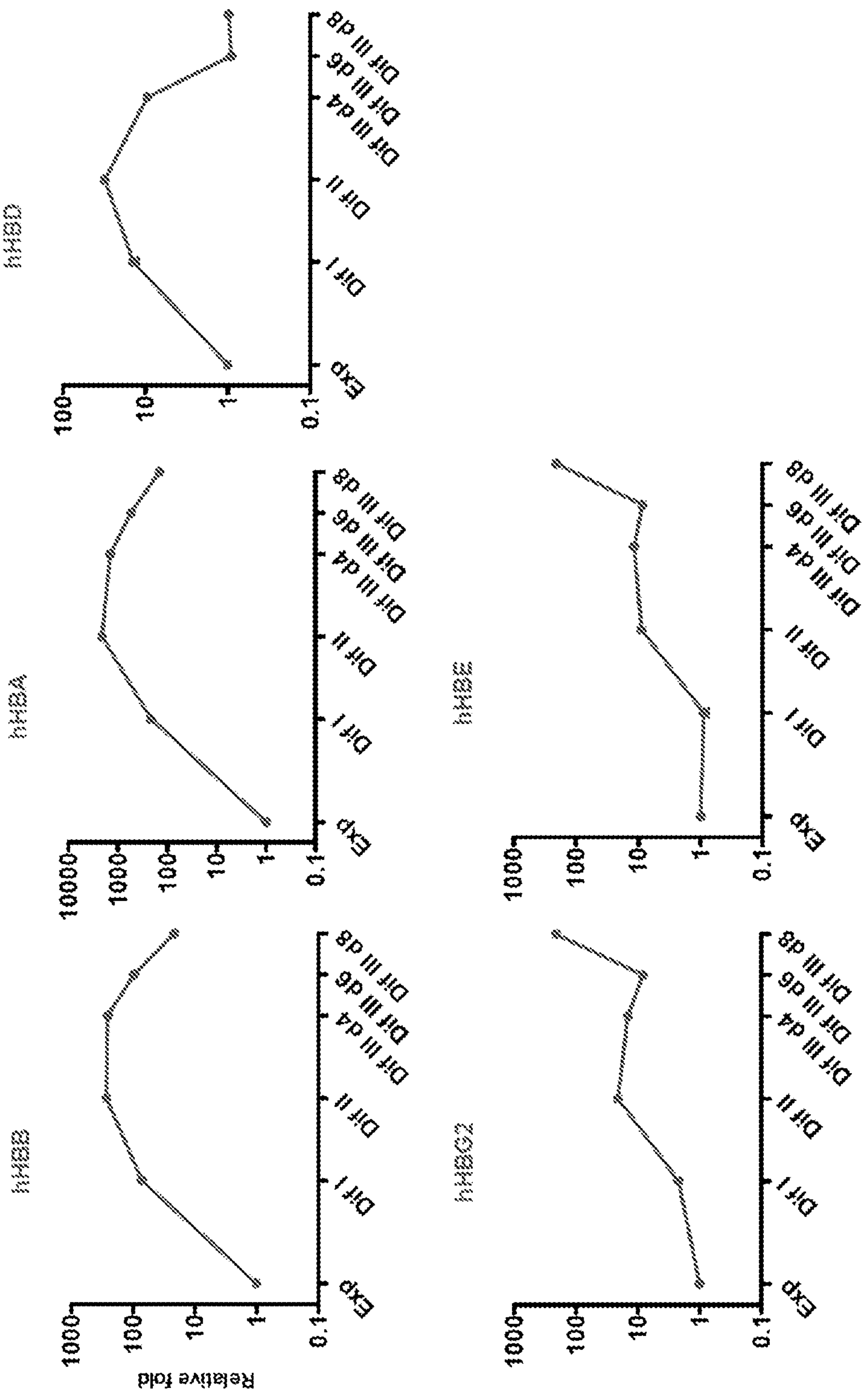

hHBB
hHBA
hHBD
hHBG2
hHBE
Relative fold
1000
100
10
1
0.1
10000
Exp
Dif I
Dif II
Dif III d4
Dif III d6
Dif III d8

Figure 9

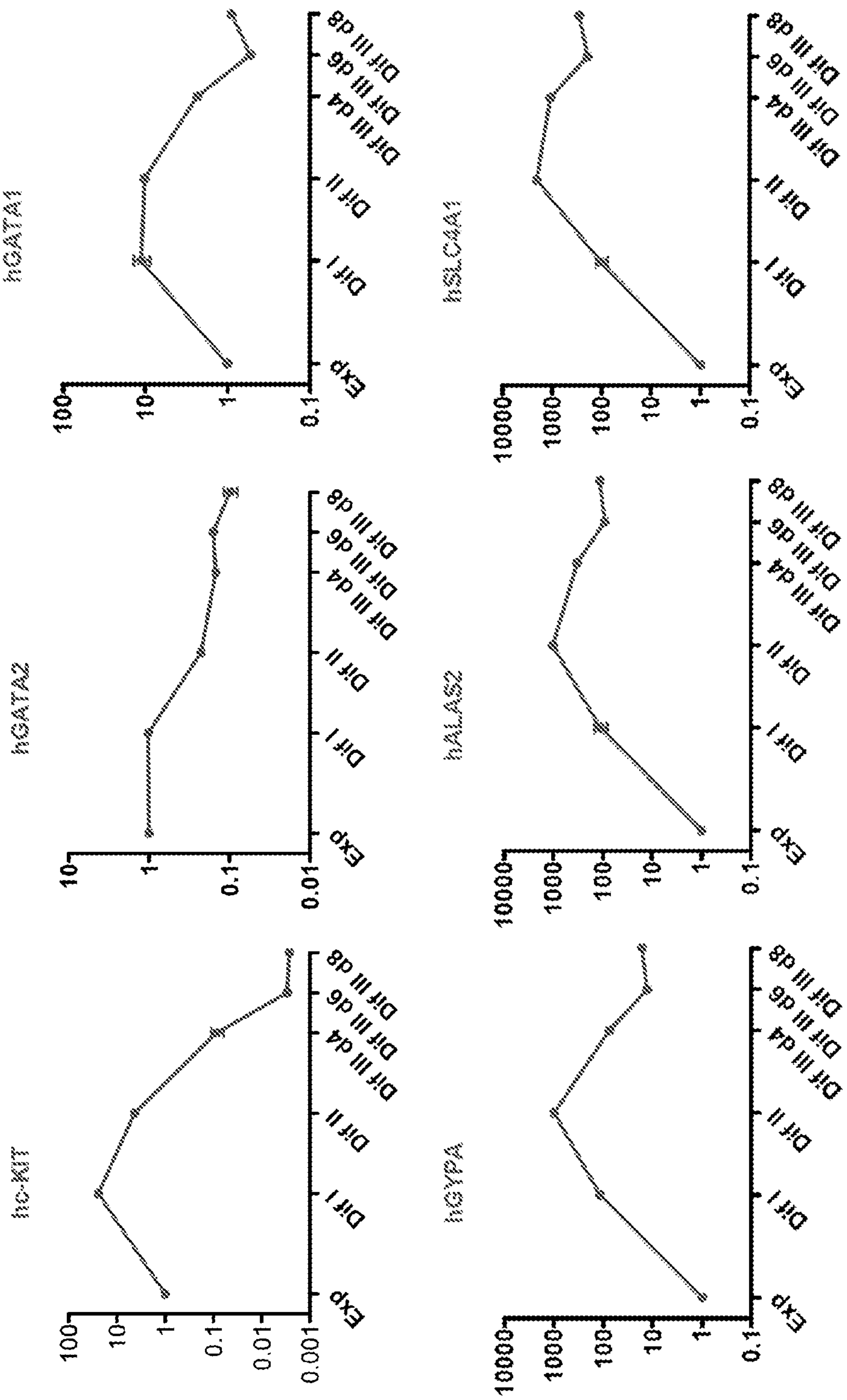
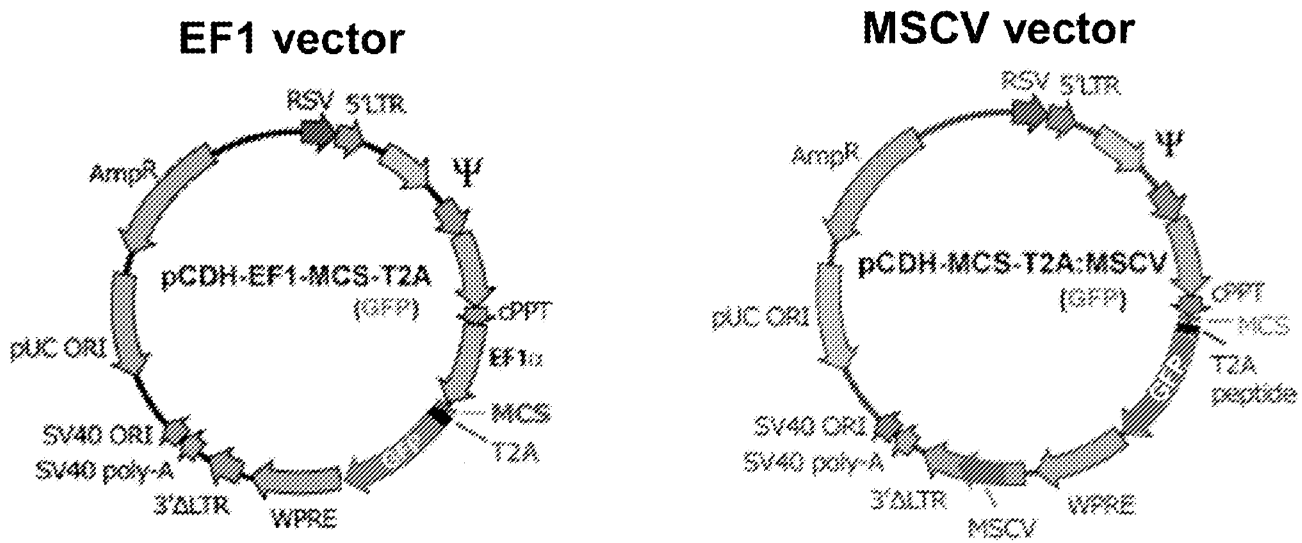

hc-KIT
hGATA2
hGATA1
hGYPA
hALAS2
hSLC4A1
100
10
1
0.1
0.01
0.001
10000
1000
Exp
Dif I
Dif II
Dif III d4
Dif III d6
Dif III d8

Figure 10



Insert at Multiple Cloning Site (MCS) of both vectors:

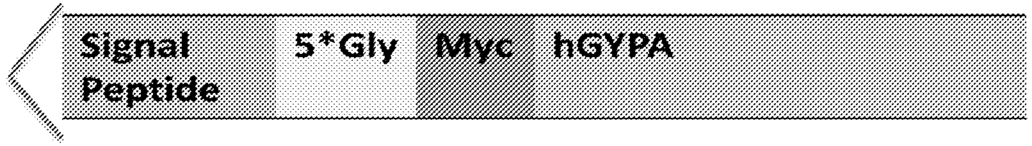
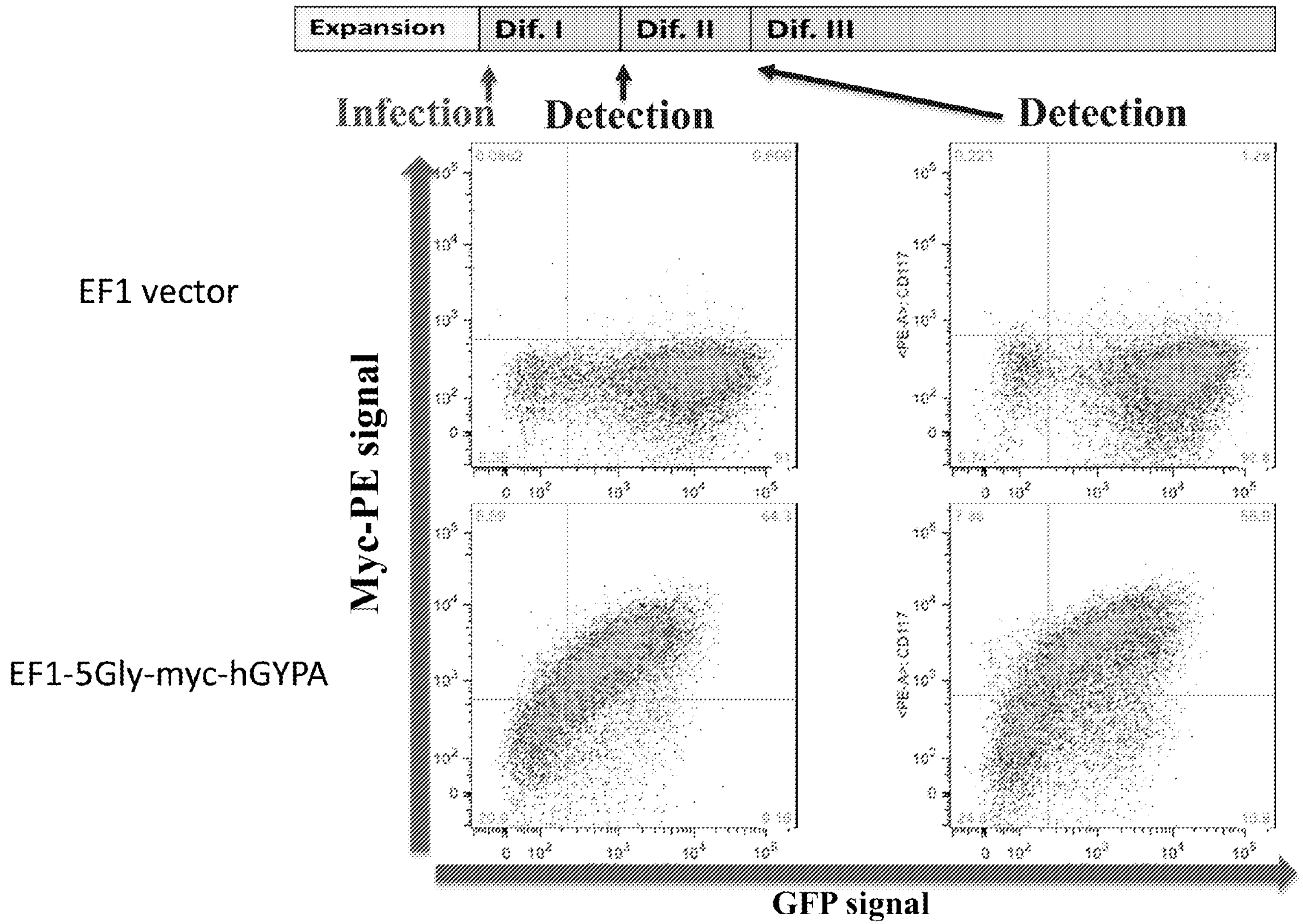


Figure 11



Expansion
Dif. I
Dif. II
Dif. III
Infection
Detection
Detection
EF1 vector
EF1-5Gly-myc-hGYPA
Myc-PE signal
GFP signal

Figure 12

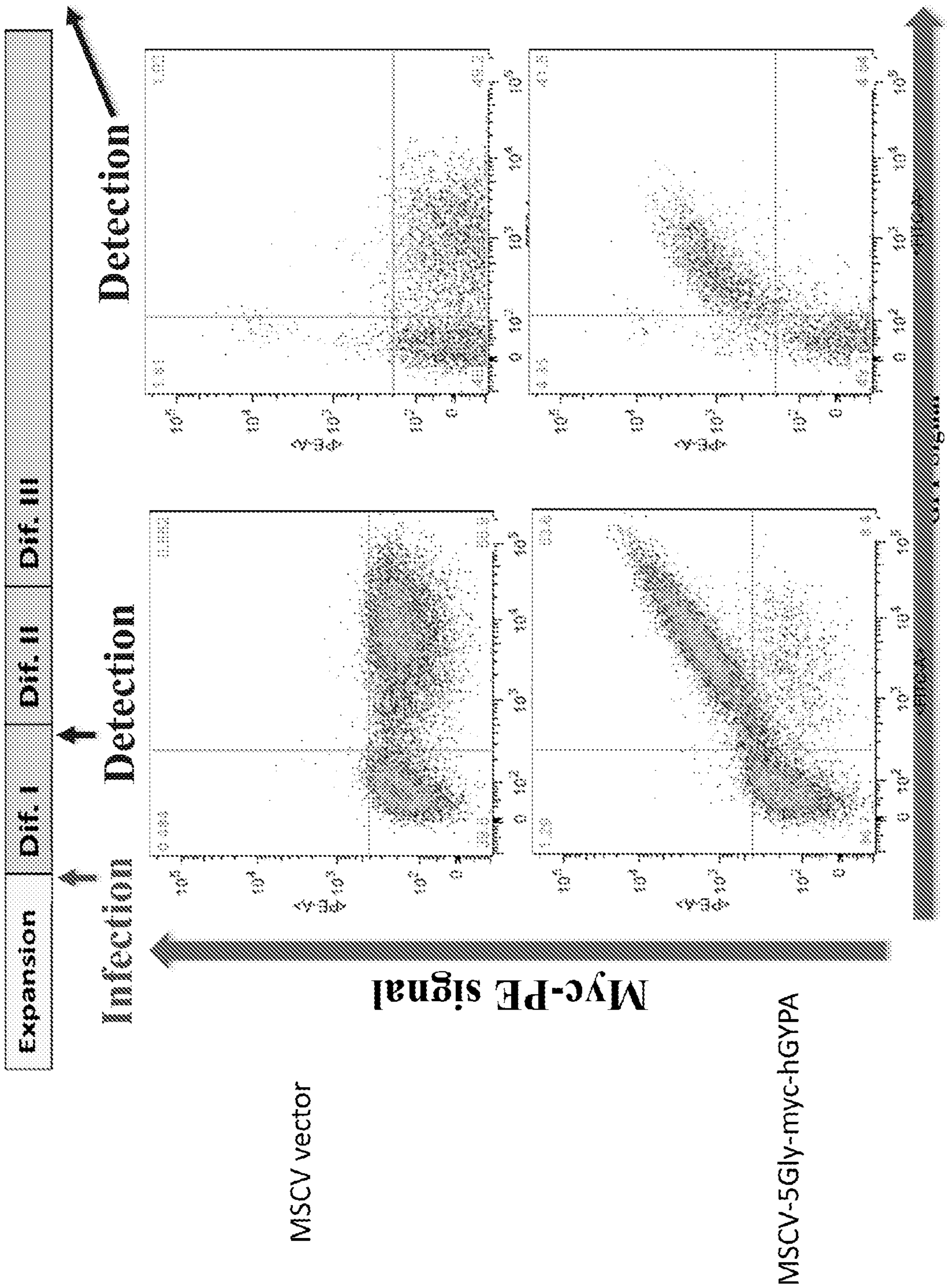

Expansion
Dif. I
Dif. II
Dif. III
Infection
Detection
Detection
Myc-PE signal
MSCV vector
MSCV-5Gly-myc-hGYPA

Figure 13

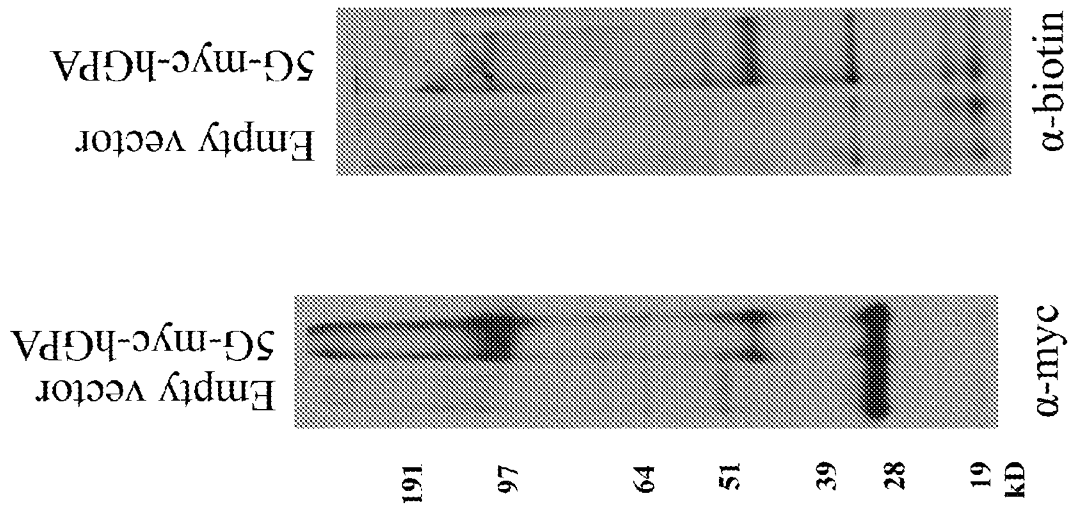

Empty vector
5G-myc-hGPA
α-biotin
Empty vector
5G-myc-hGPA
α-myc
191
97
64
51
39
28
19
kD

Figure 14

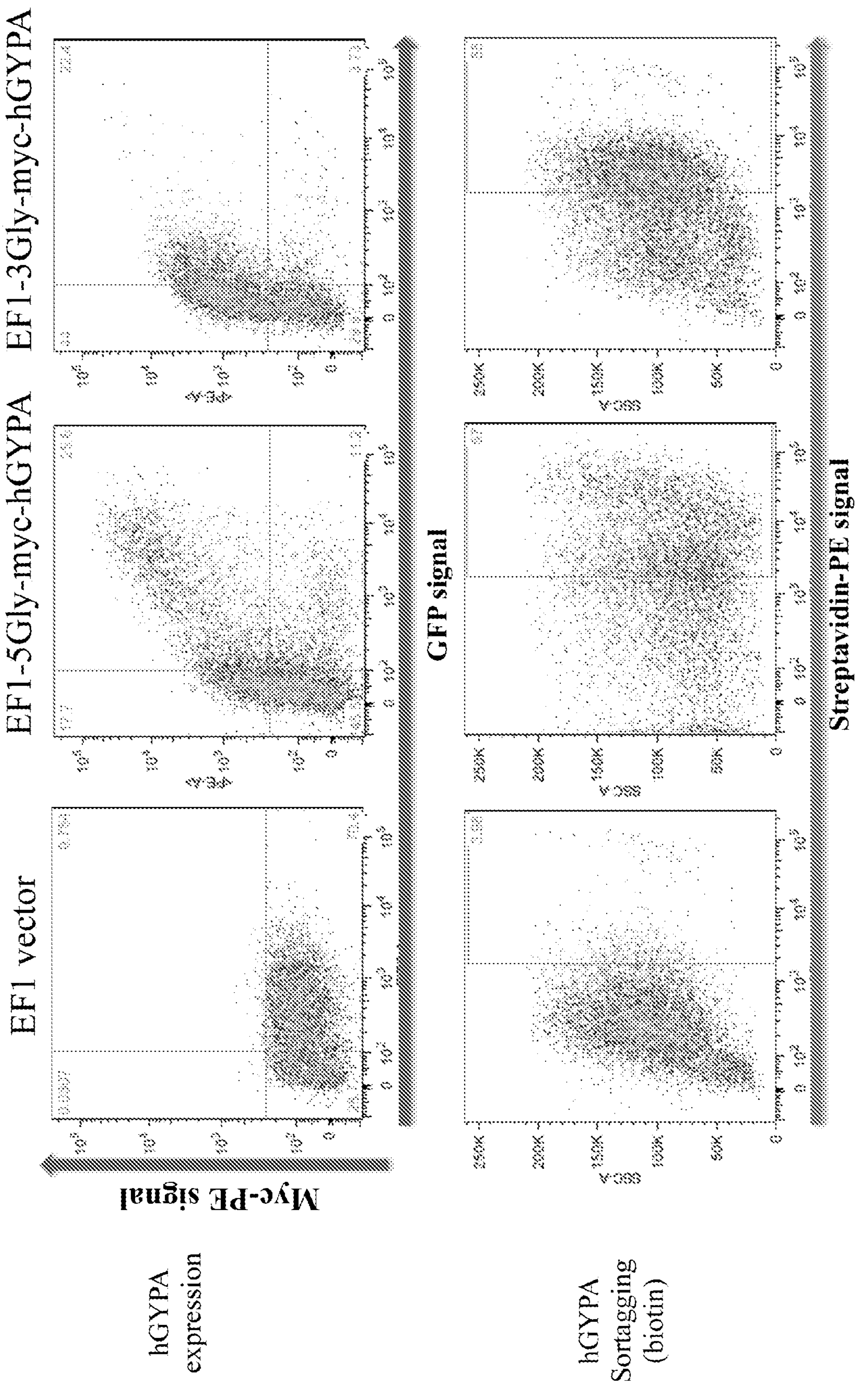

EF1 vector
EF1-5Gly-myc-hGYPA
EF1-3Gly-myc-hGYPA
GFP signal
Myc-PE signal
Streptavidin-PE signal
hGYPA expression
hGYPA Sortagging (biotin)
PE-A
SSC-A

Figure 15

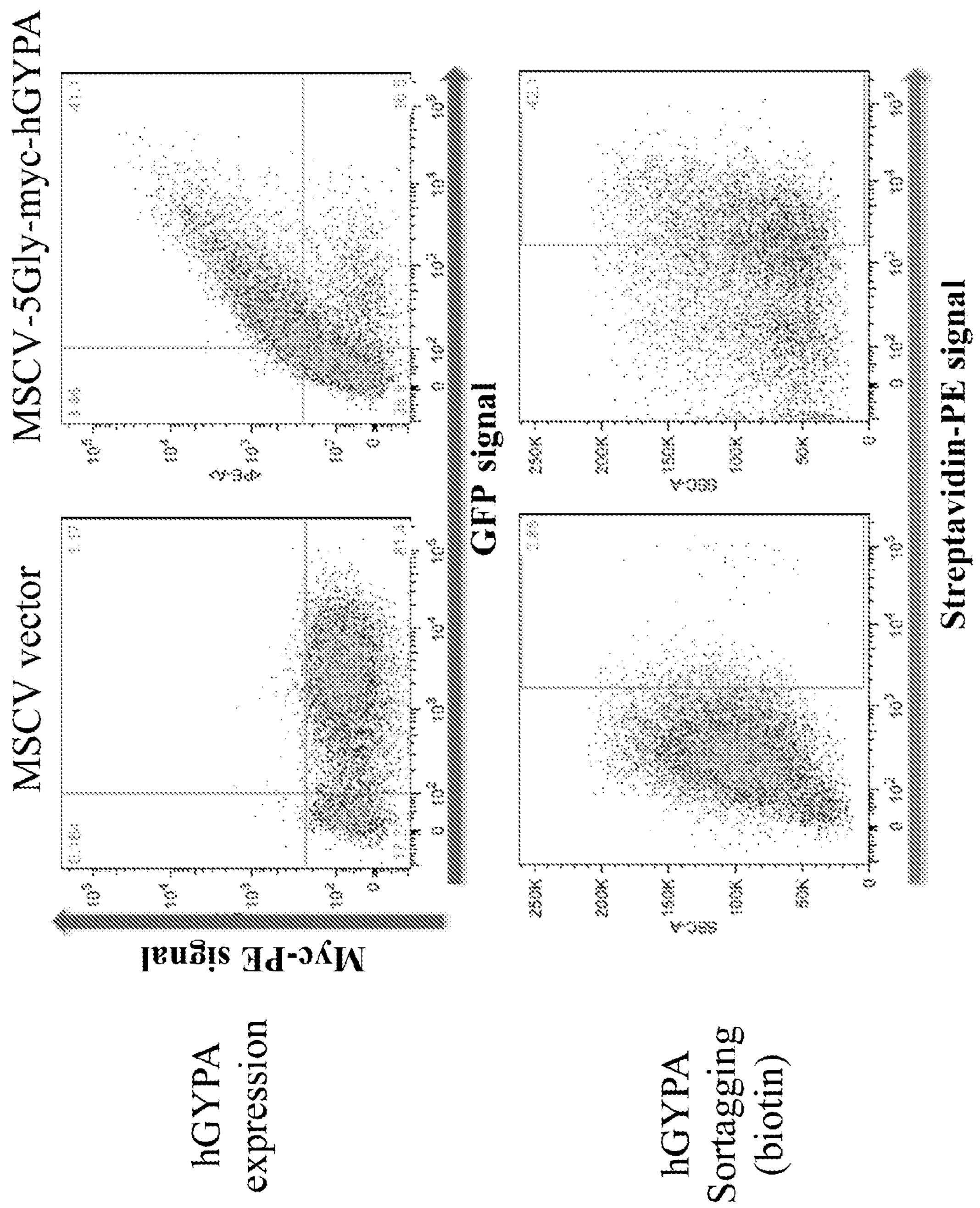

MSCV vector
MSCV-5Gly-myc-hGYPA
hGYPA expression
hGYPA Sortagging (biotin)
Myc-PE signal
GFP signal
Streptavidin-PE signal

Figure 16

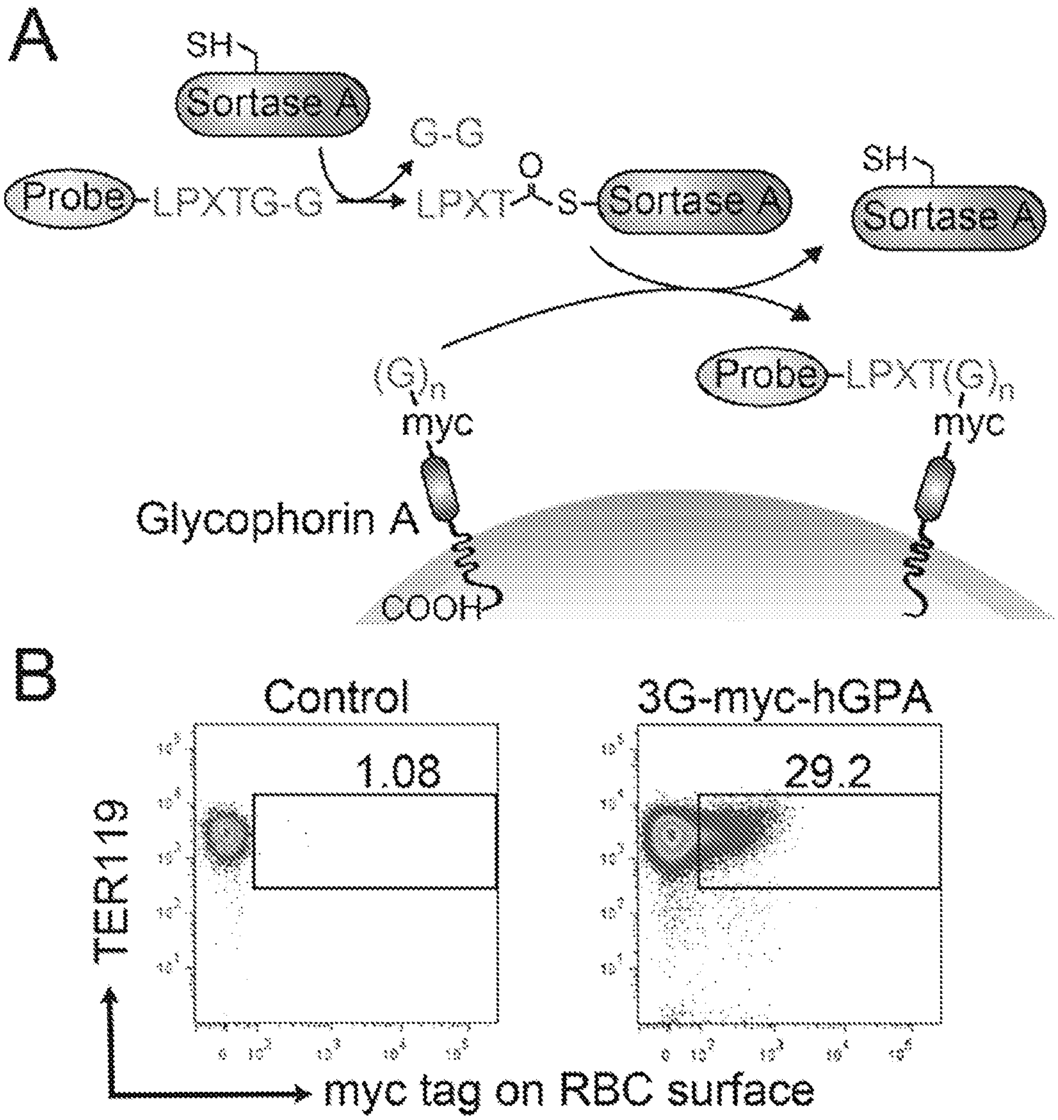
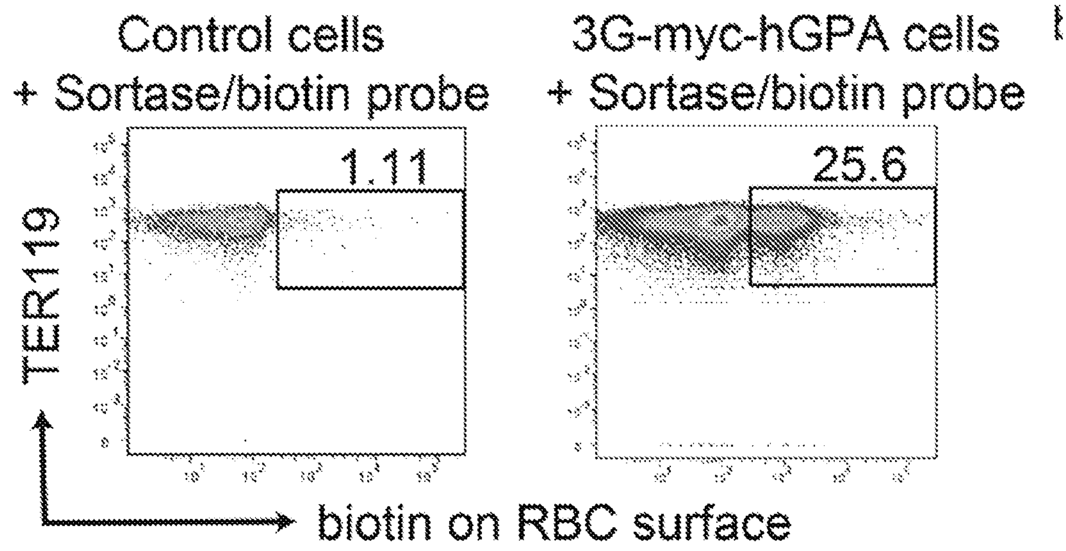

A
SH
Sortase A
G-G
Probe
LPXTG-G
LPXT
O
S
Sortase A
SH
Sortase A
(G)n
myc
Probe
LPXT(G)n
myc
Glycophorin A
COOH
B
Control
3G-myc-hGPA
1.08
29.2
TER119
myc tag on RBC surface

Figure 16 (cont'd)

C



D

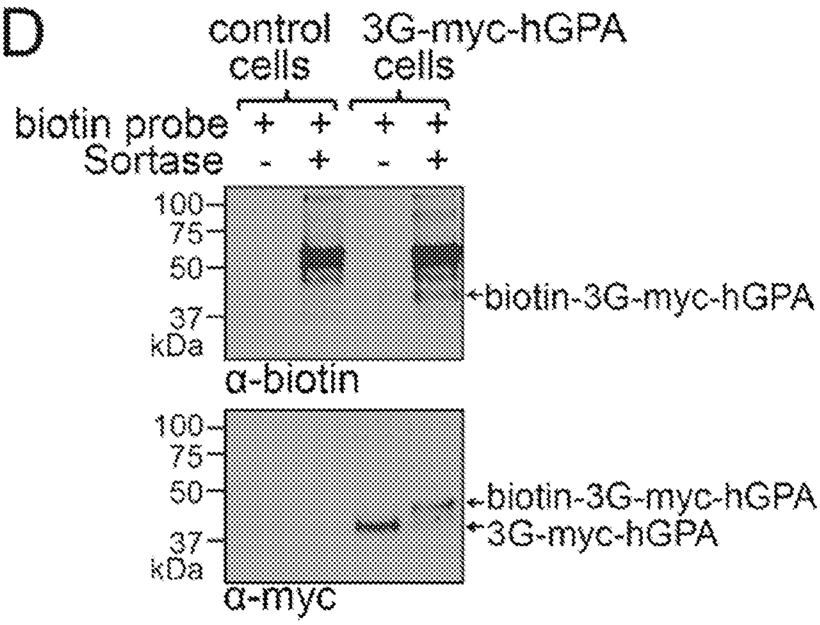


Figure 16 (cont'd)

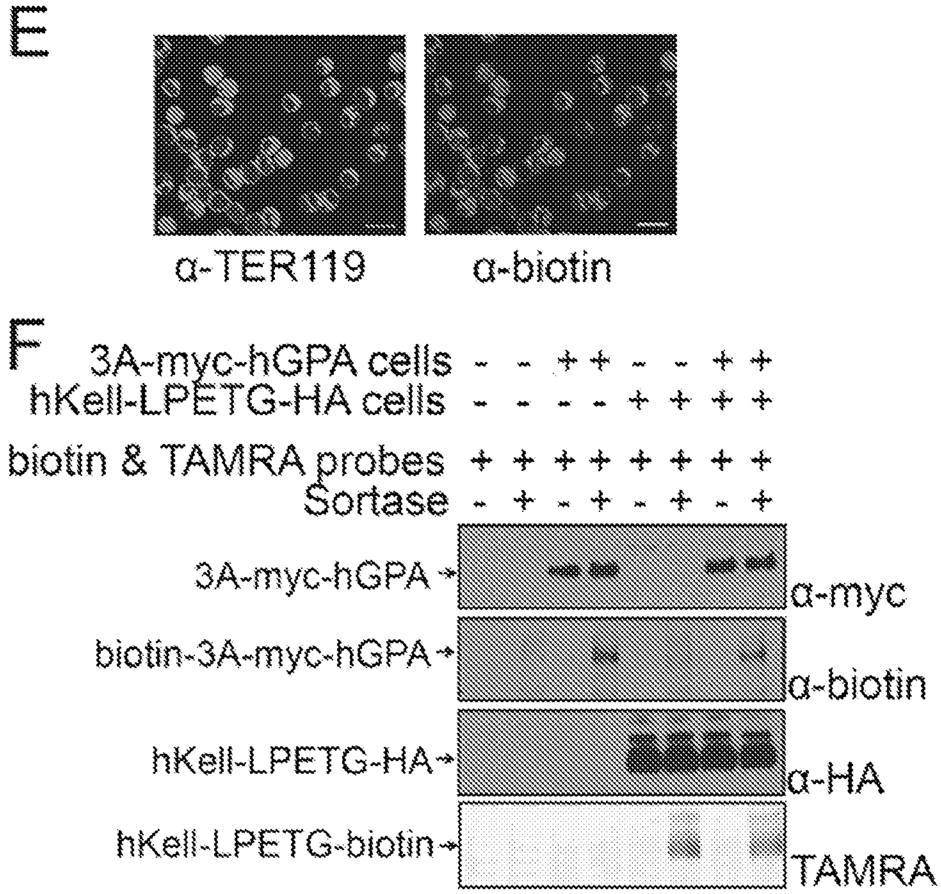


Figure 17

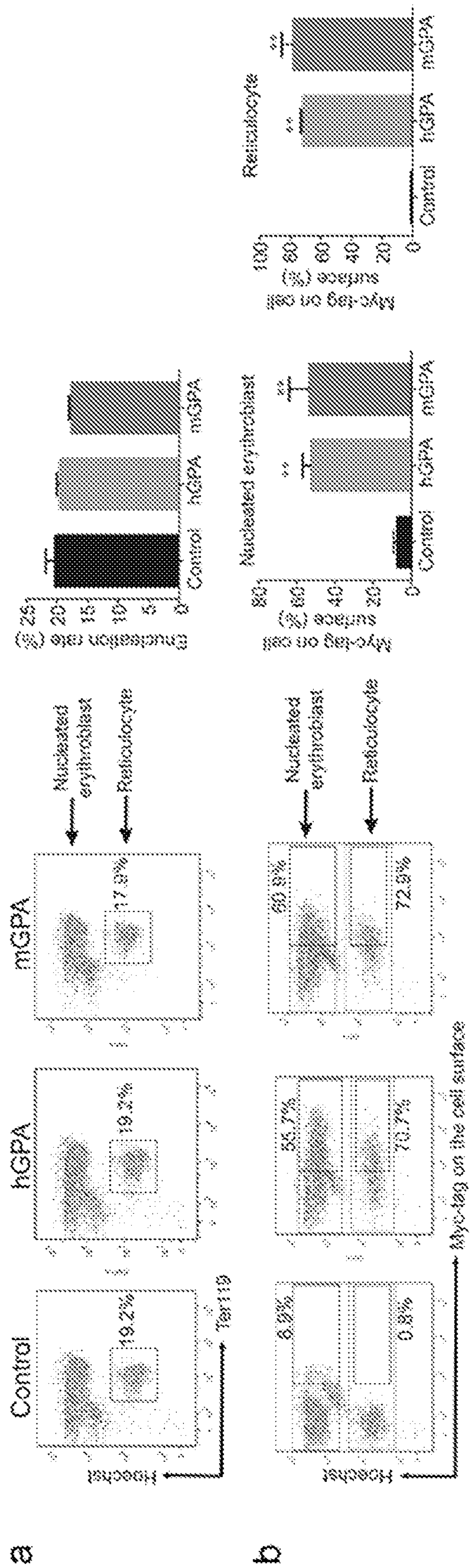

a
b
Control
hGPA
mGPA
Hoechst
Ter119
19.2%
19.2%
17.9%
Nucleated erythroblast
Reticulocyte
6.9%
0.8%
55.7%
70.7%
60.8%
72.9%
Myc-tag on the cell surface
Enucleation rate (%)
Myc-tag on cell surface (%)
Nucleated erythroblast
Reticulocyte
Control
hGPA
mGPA

Figure 17 (cont'd)

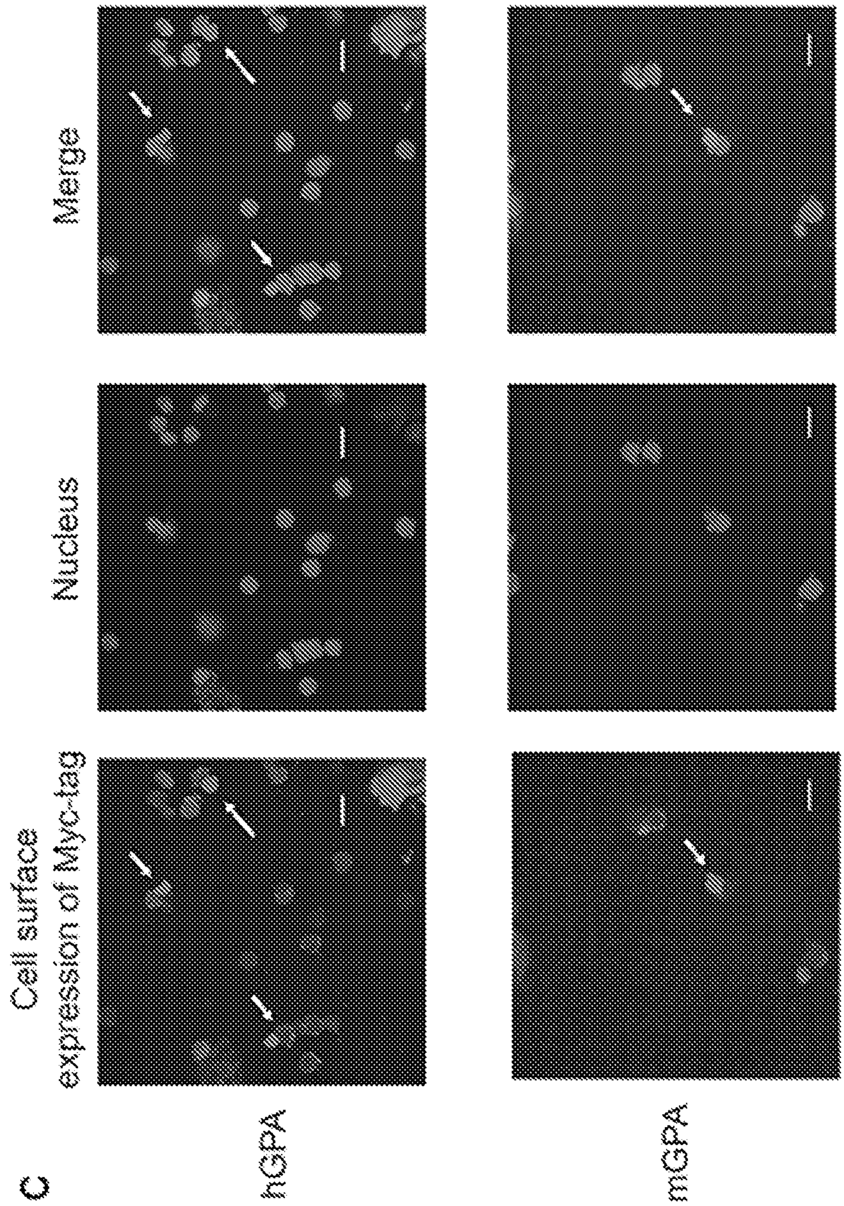


Figure 18

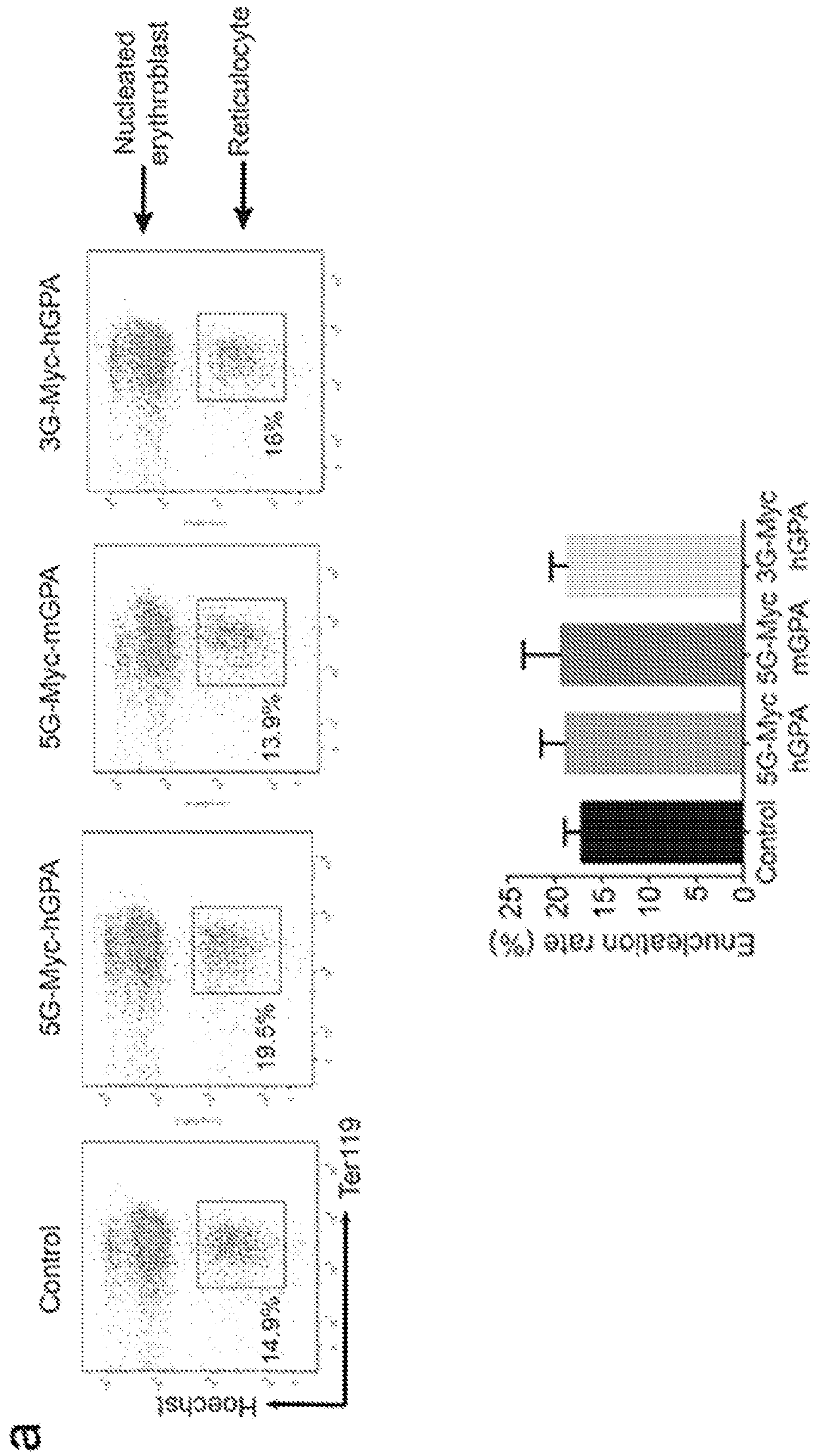

a
Control
5G-Myc-hGPA
5G-Myc-mGPA
3G-Myc-hGPA
Nucleated erythroblast
Reticulocyte
14.9%
19.5%
13.9%
16%
Hoechst
Ter119
Enucleation rate (%)
25
20
15
10
5
0
Control
5G-Myc hGPA
5G-Myc mGPA
3G-Myc hGPA

Figure 18 (cont'd)

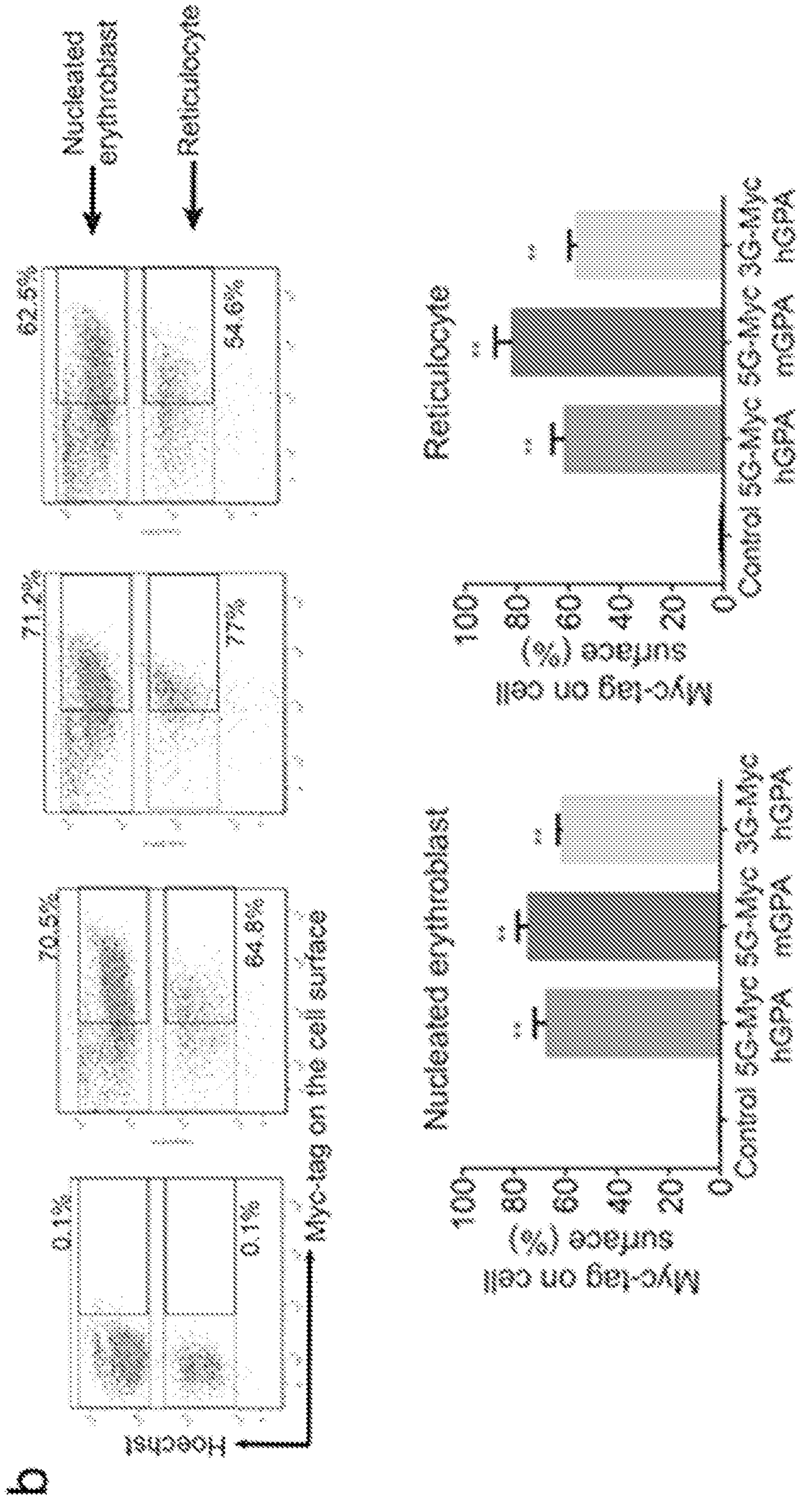


Figure 18 (cont'd)

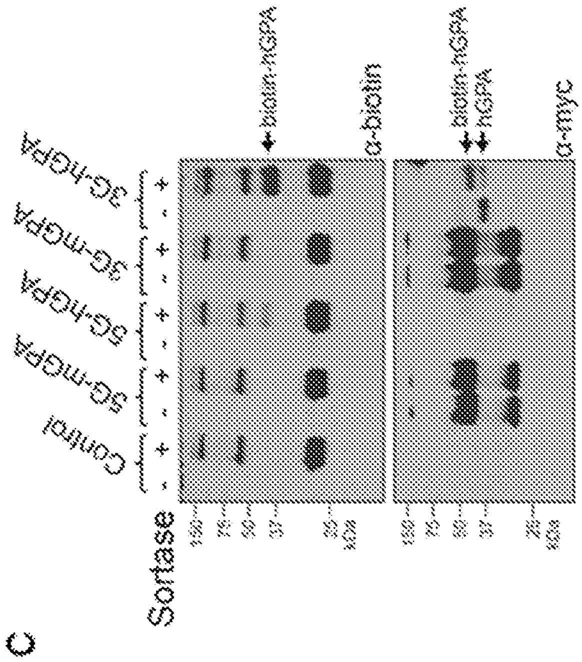


Figure 19

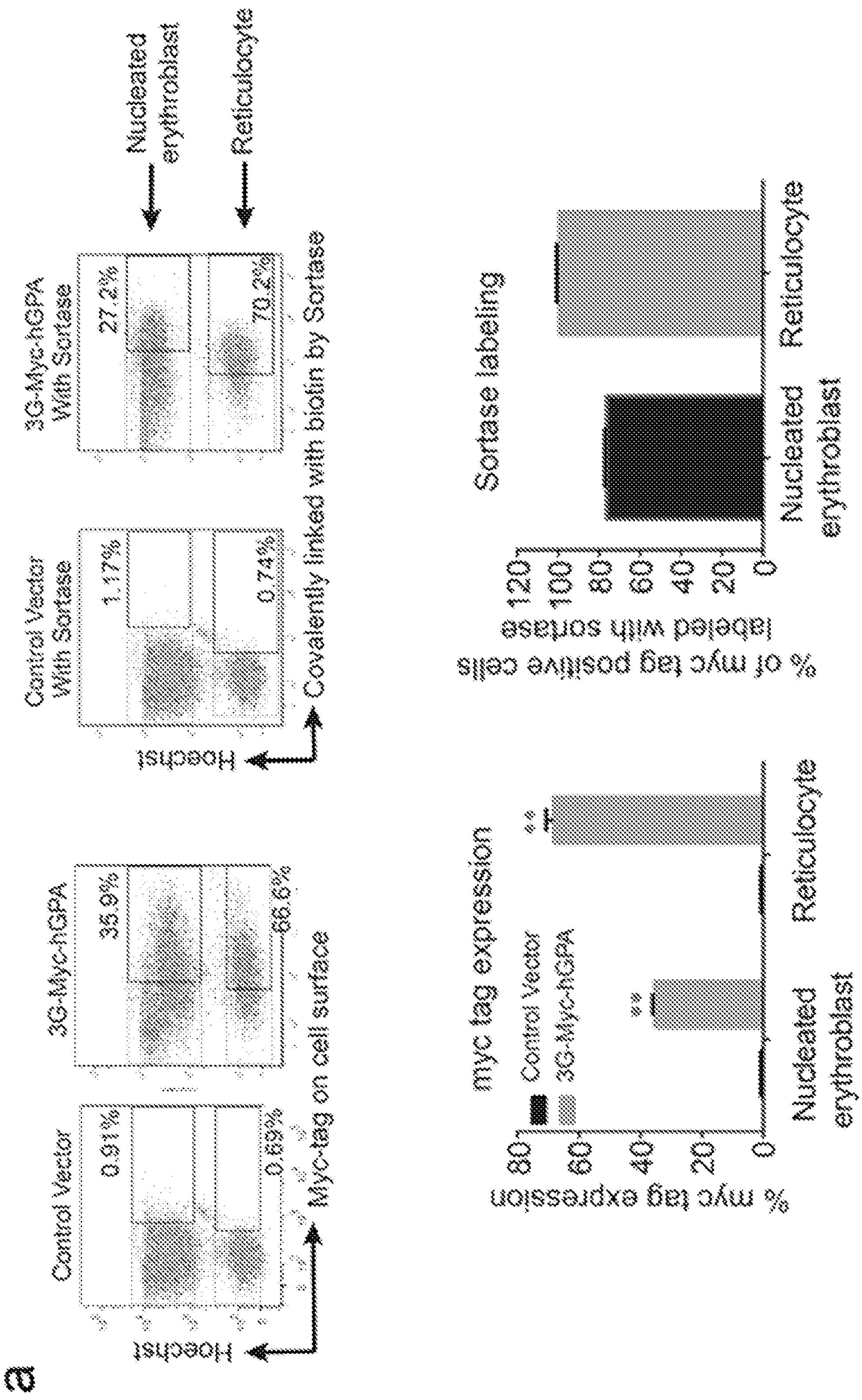

a
Control Vector
3G-Myc-hGPA
0.91%
0.69%
35.9%
66.6%
Hoechst
Myc-tag on cell surface
Control Vector With Sortase
3G-Myc-hGPA With Sortase
1.17%
0.74%
27.2%
70.2%
Covalently linked with biotin by Sortase
Nucleated erythroblast
Reticulocyte
myc tag expression
Control Vector
3G-Myc-hGPA
% myc tag expression
80
60
40
20
0
Nucleated erythroblast
Reticulocyte
Sortase labeling
% of myc tag positive cells labeled with sortase
120
100
80
60
40
20
0
Nucleated erythroblast
Reticulocyte

Figure 19 (cont'd)

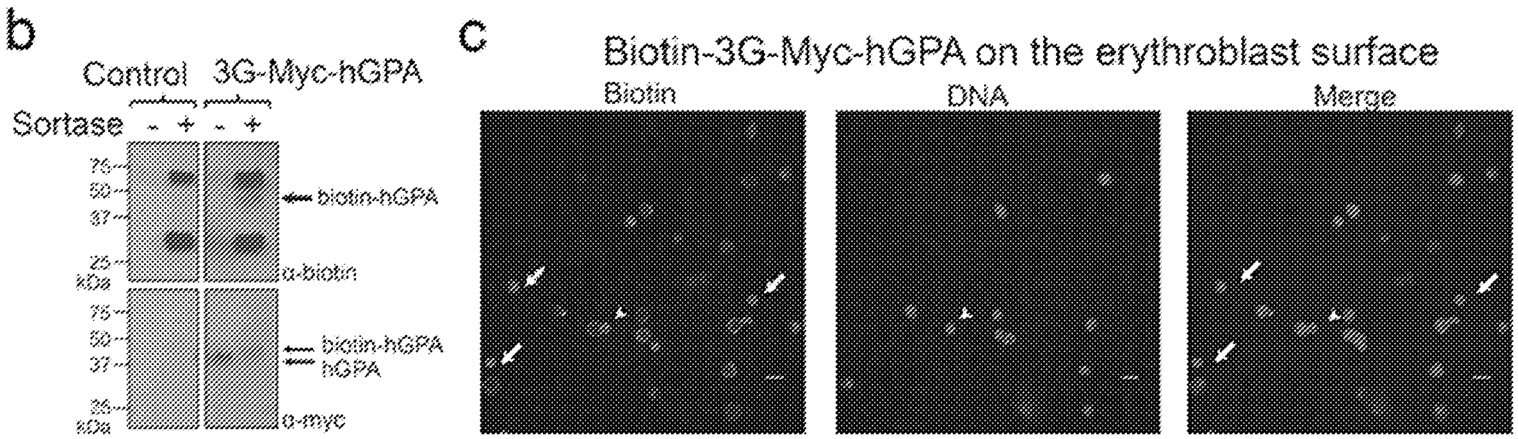


Figure 20

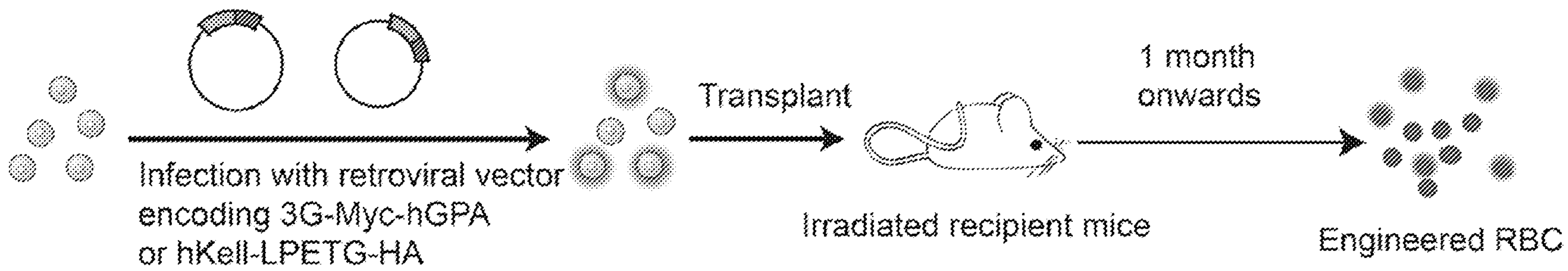

Infection with retroviral vector
encoding 3G-Myc-hGPA
or hKell-LPETG-HA
Transplant
Irradiated recipient mice
1 month
onwards
Engineered RBC

Figure 21

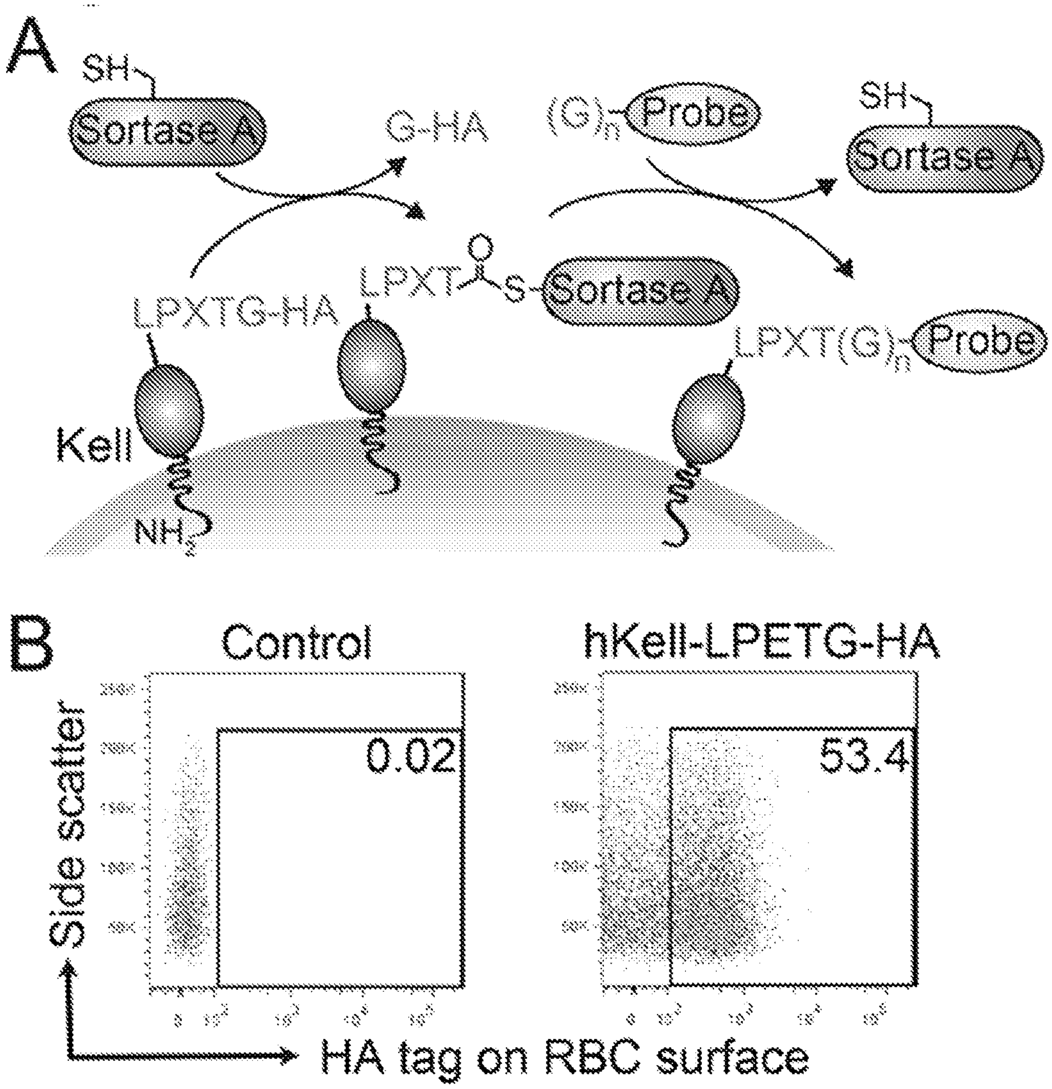
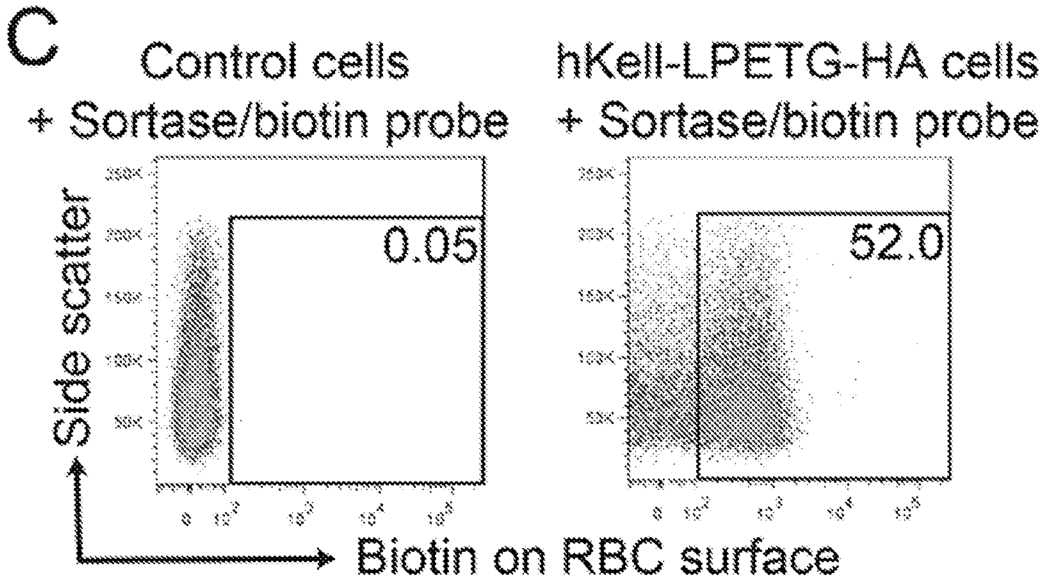
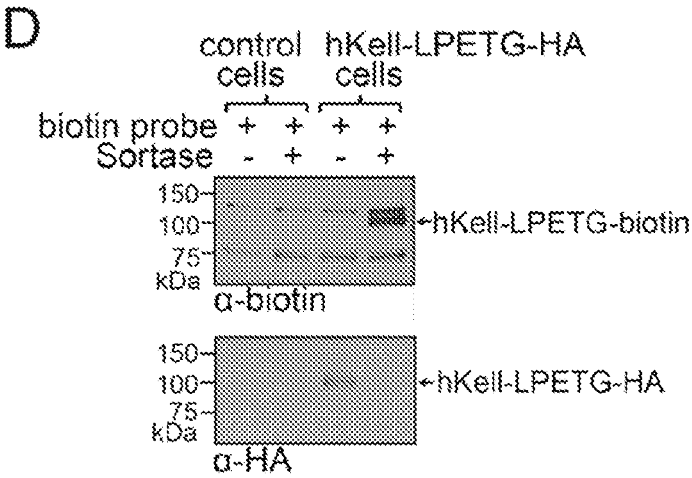

A
SH
Sortase A
G-HA
(G)n
Probe
SH
Sortase A
LPXT
S
Sortase A
LPXTG-HA
LPXT(G)n
Probe
Kell
NH2
B
Control
hKell-LPETG-HA
0.02
53.4
Side scatter
HA tag on RBC surface

Figure 21 (cont'd)





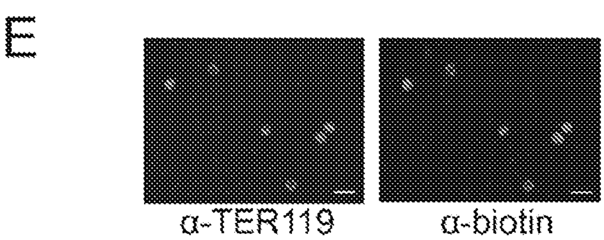


Figure 22

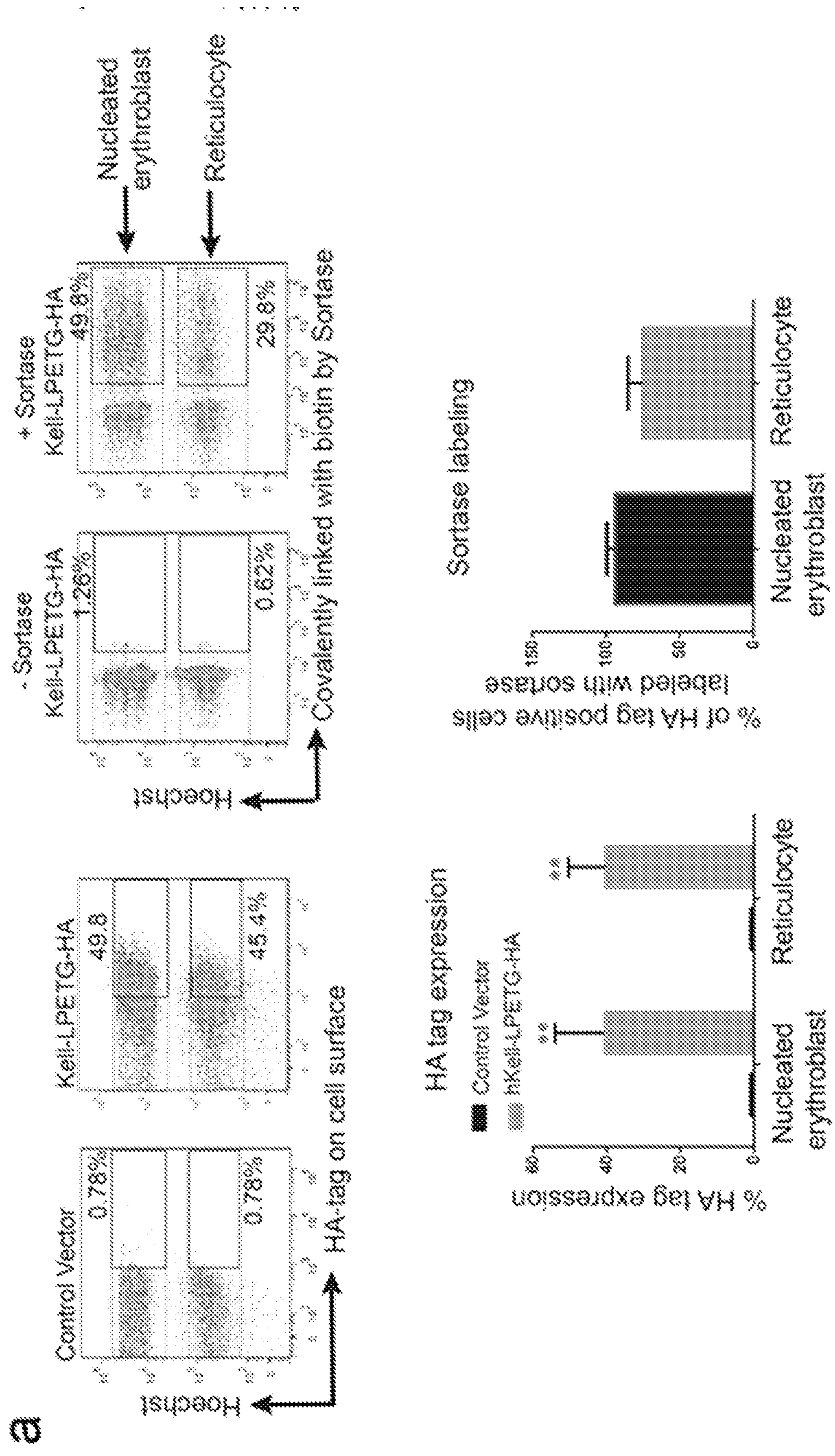

a
Control Vector
0.78%
0.78%
Kell-LPETG-HA
49.8
45.4%
Hoechst
HA-tag on cell surface
- Sortase
Kell-LPETG-HA
1.26%
0.62%
+ Sortase
Kell-LPETG-HA
49.8%
29.8%
Hoechst
Covalently linked with biotin by Sortase
Nucleated erythroblast
Reticulocyte
HA tag expression
Control Vector
hKell-LPETG-HA
% HA tag expression
Nucleated erythroblast
Reticulocyte
Sortase labeling
% of HA tag positive cells labeled with sortase
Nucleated erythroblast
Reticulocyte

Figure 22 (cont'd)

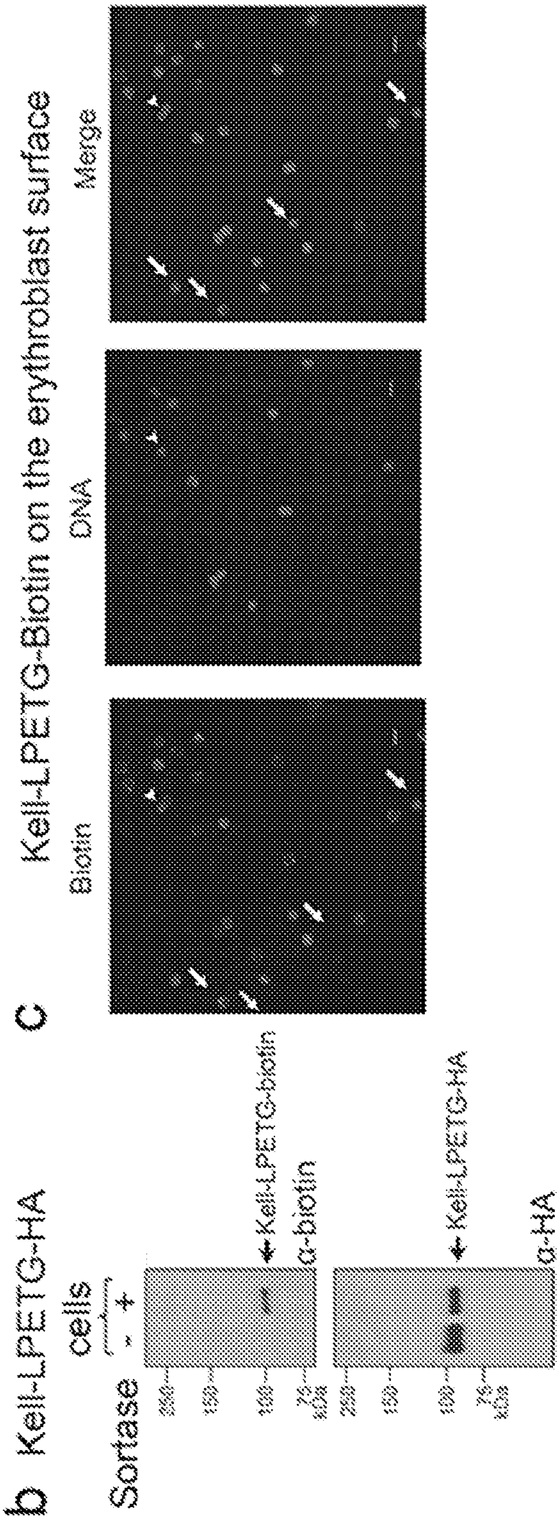


Figure 23

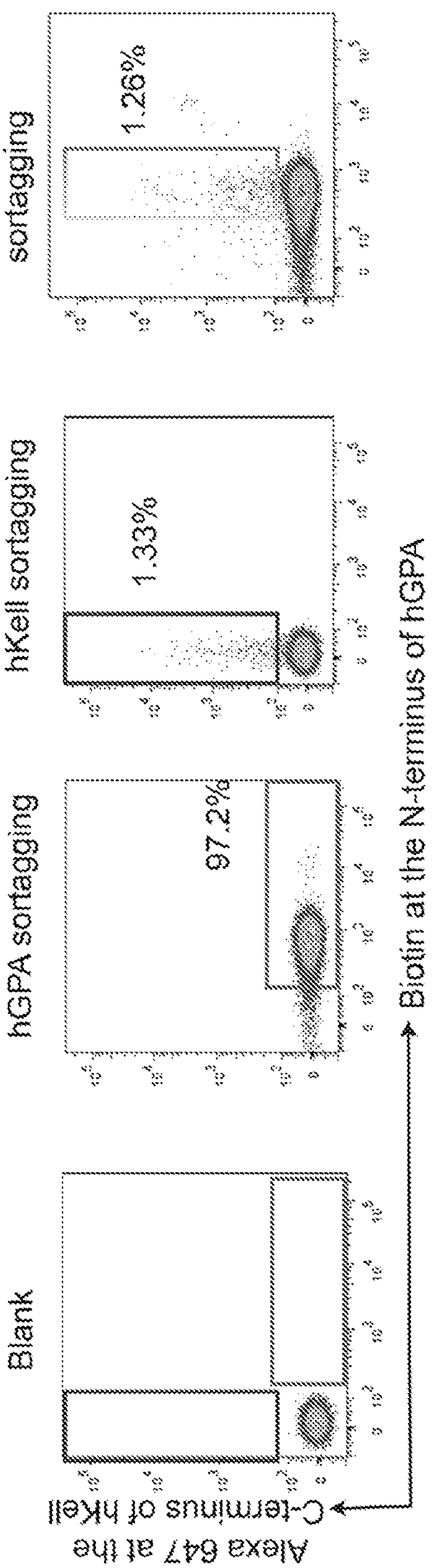

Blank
hGPA sortagging
hKell sortagging
hGPA AND hKell sortagging
97.2%
1.33%
1.26%
Biotin at the N-terminus of hGPA
Alexa 647 at the C-terminus of hKell

Figure 24

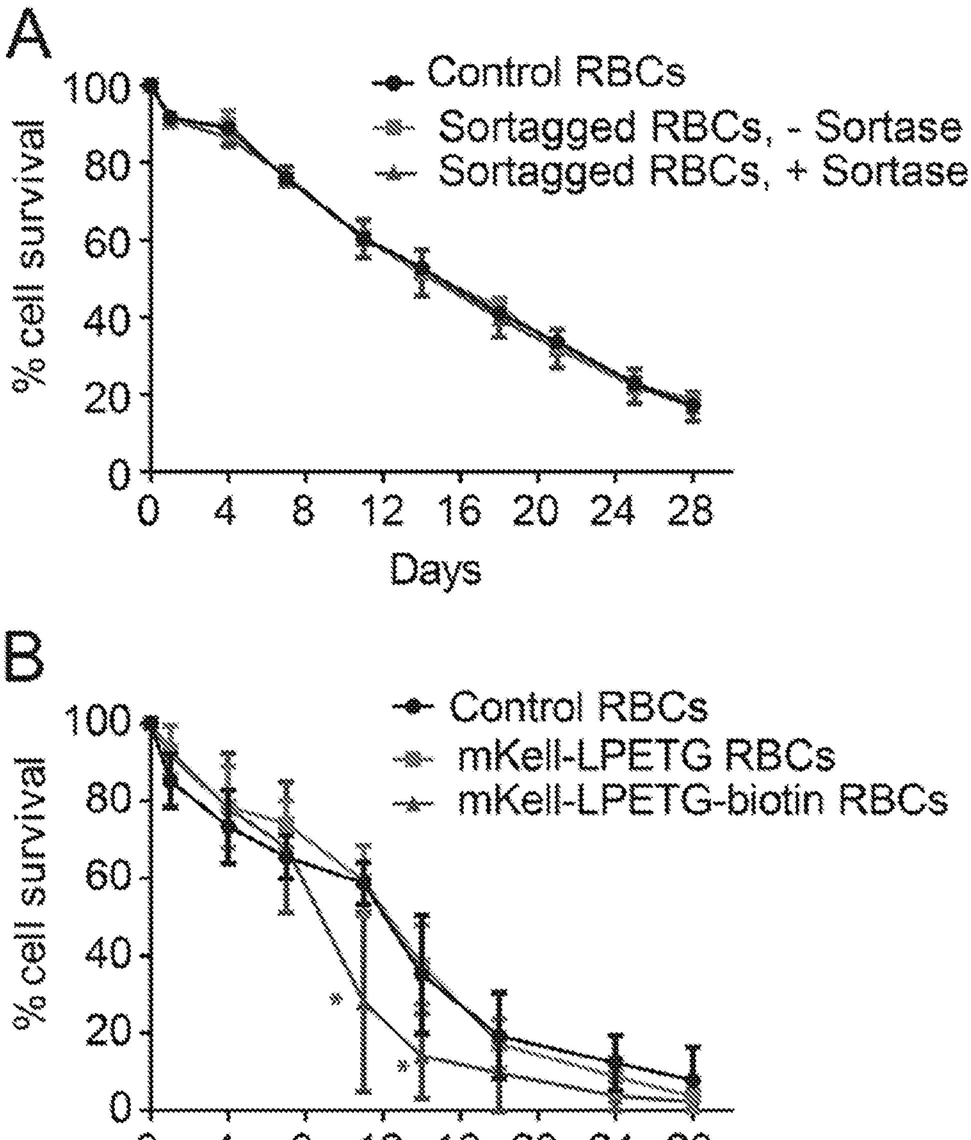
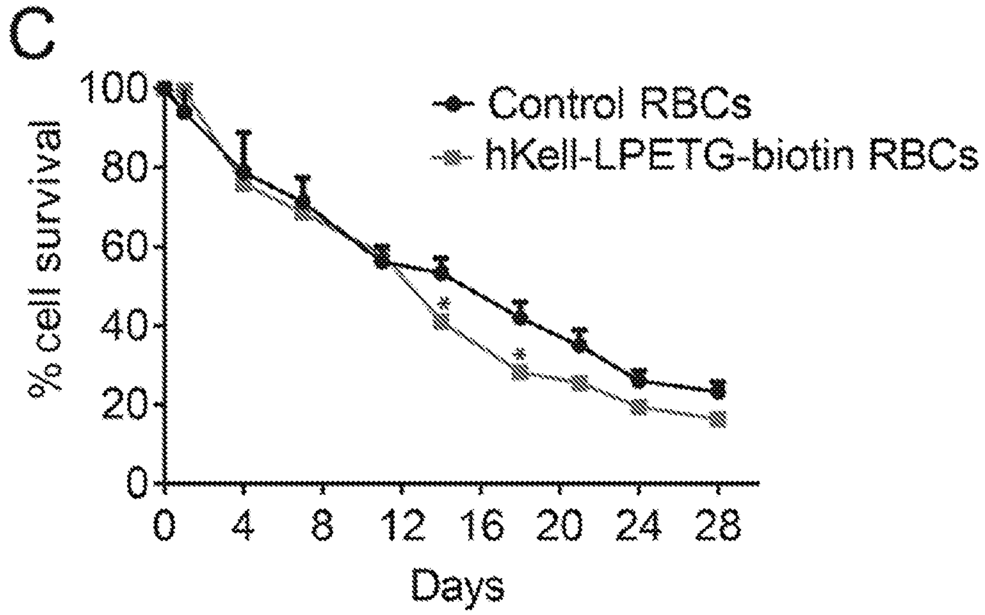

A
100
80
60
40
20
0
% cell survival
0 4 8 12 16 20 24 28
Days
Control RBCs
Sortagged RBCs, - Sortase
Sortagged RBCs, + Sortase
B
100
80
60
40
20
0
% cell survival
0 4 8 12 16 20 24 28
Days
Control RBCs
mKell-LPETG RBCs
mKell-LPETG-biotin RBCs
*
*

Figure 24 (cont'd)



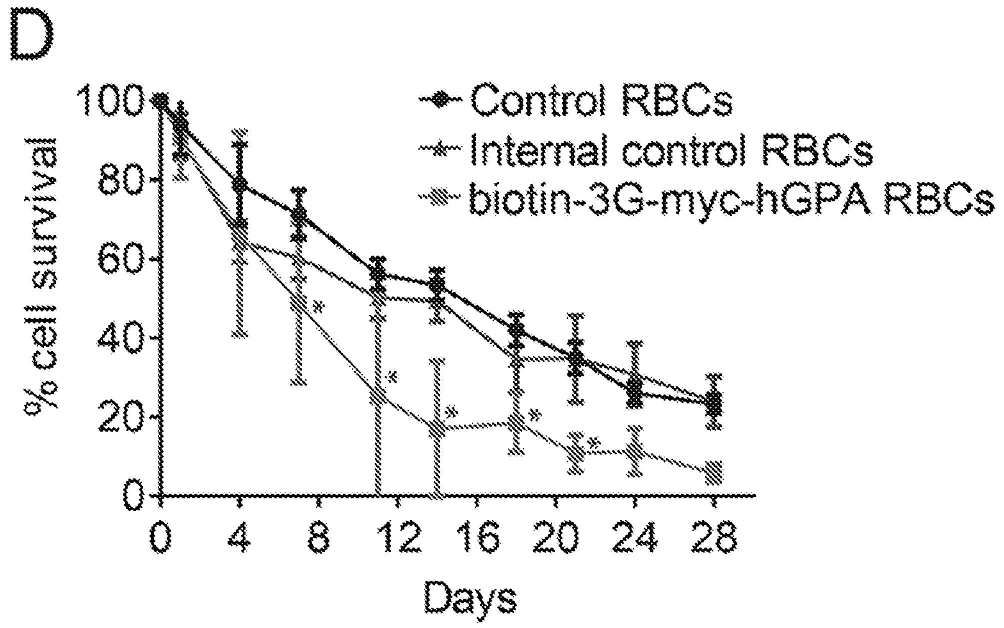


Figure 25

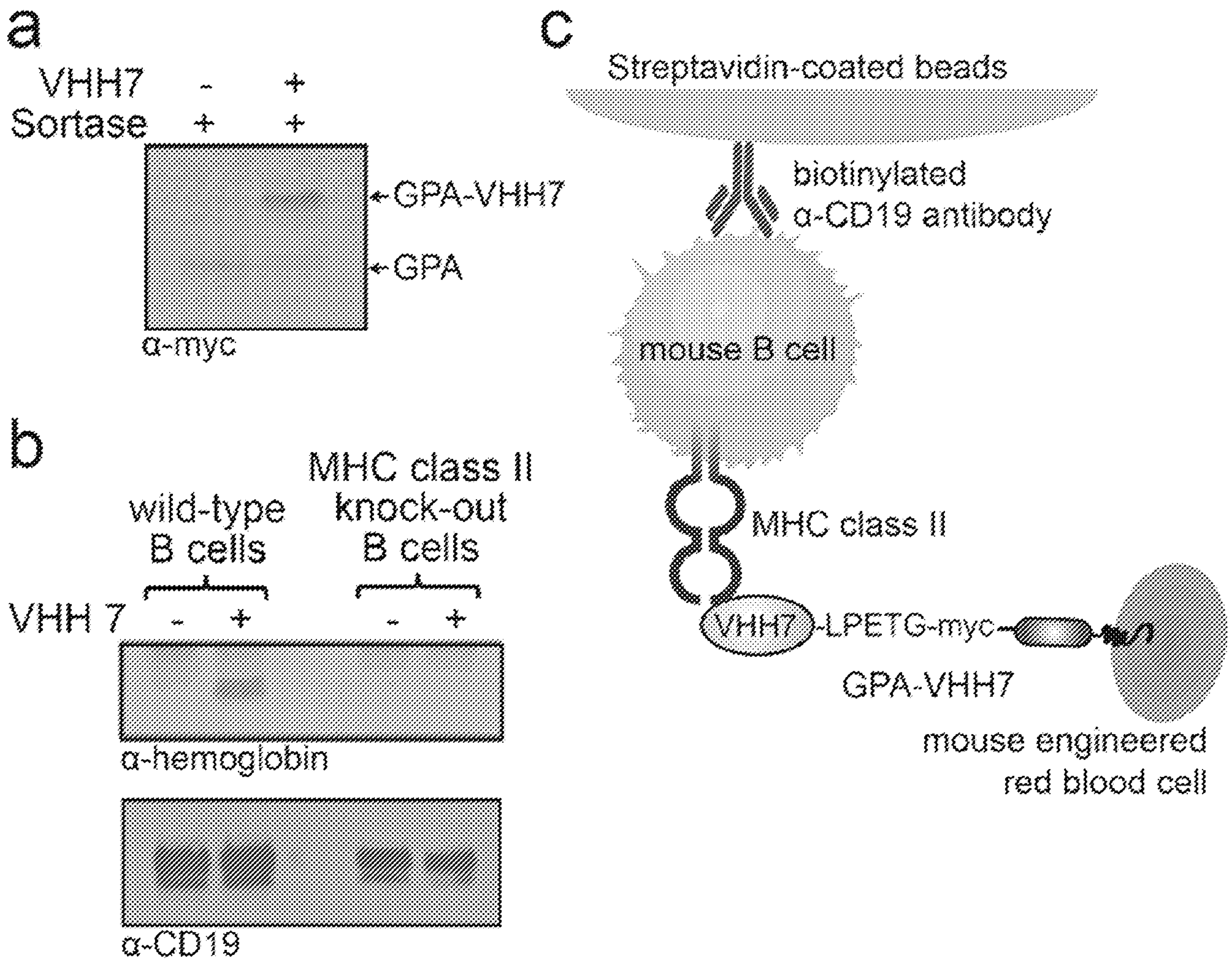

a
VHH7 - +
Sortase + +
GPA-VHH7
GPA
α-myc
b
wild-type B cells
MHC class II knock-out B cells
VHH 7 - + - +
α-hemoglobin
α-CD19
c
Streptavidin-coated beads
biotinylated α-CD19 antibody
mouse B cell
MHC class II
VHH7
LPETG-myc
GPA-VHH7
mouse engineered red blood cell

Figure 26

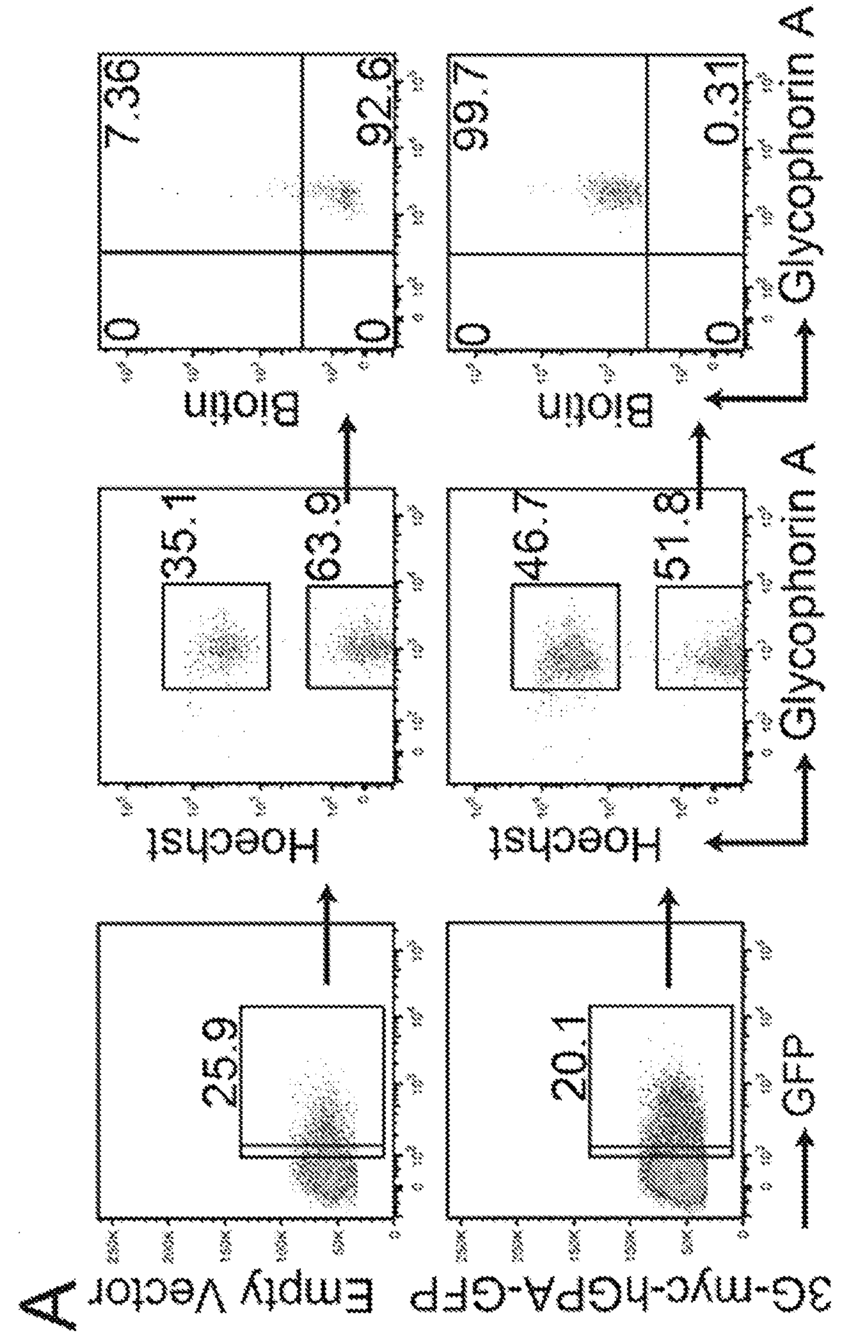

A
Empty Vector
3G-myc-hGPA-GFP
25.9
20.1
GFP
Hoechst
35.1
63.9
46.7
51.8
Glycophorin A
Biotin
0
7.36
92.6
99.7
0.31
Glycophorin A

Figure 26 (cont'd)