Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2947125 C 2023/02/28

(11)(21) 2 947 125

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 2015/04/23
(87) Date publication PCT/PCT Publication Date: 2015/11/26
(45) Date de délivrance/Issue Date: 2023/02/28
(85) Entrée phase nationale/National Entry: 2016/10/26
(86) N° demande PCT/PCT Application No.: GB 2015/051201
(87) N° publication PCT/PCT Publication No.: 2015/177501
(30) Priorité/Priority: 2014/05/21 (GB1409089.8)

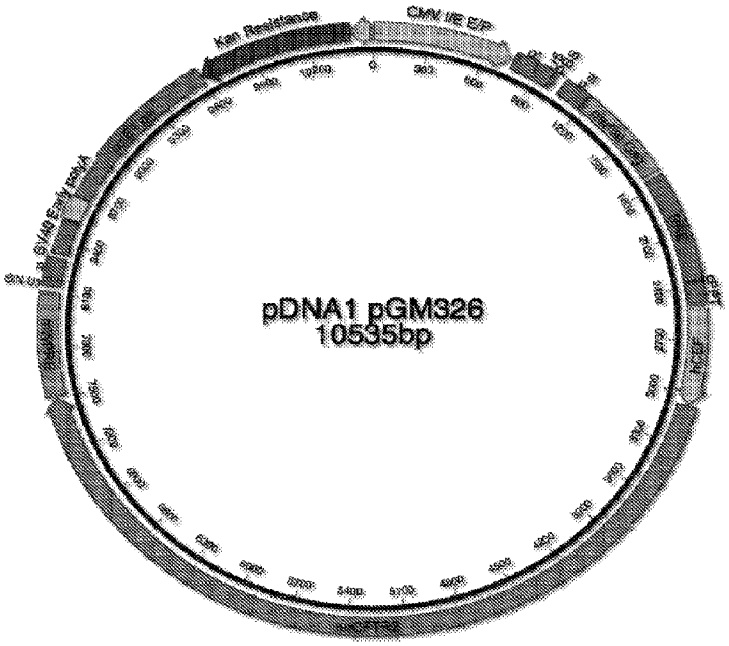
(51) Cl.Int./Int.Cl. *C12N 15/867* (2006.01),
A61K 48/00 (2006.01), *C07K 14/115* (2006.01),
C07K 14/705 (2006.01), *C07K 14/755* (2006.01),
C07K 14/81 (2006.01), *C12N 15/45* (2006.01),
C12N 15/86 (2006.01), *C12N 7/00* (2006.01),
C12N 9/24 (2006.01)

(72) Inventeurs/Inventors:
ALTON, ERIC WALTER FREDERICK WOLFGANG, GB;

GRIESENBACH, UTA, GB;
PYTEL, KAMILA MALGORZATA, GB;
PAUL-SMITH, MICHAEL CHRISTIAN, GB;
...

(73) Propriétaires/Owners:

(54) Titre : VECTEURS LENTIVIRAUX
(54) Title: LENTIVIRAL VECTORS



(57) Abrégé/Abstract:
This invention relates to lentiviral gene transfer vectors pseudotyped with hemagglutinin- neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, comprising a promoter and a transgene; and methods of making the same. The present

(11)(21) 2 947 125

(13) C

(72) Inventeurs(suite)/Inventors(continued): PRINGLE, IAN ANDREW, GB; HYDE, STEPHEN CHARLES, GB; GILL, DEBORAH REBECCA, GB; DAVIES, LEE ADRIAN, GB; BOYD, ALAN CHRISTOPHER, GB; MCLACHLAN, GERARD, GB; INOUE, MAKOTO, JP

(73) Propriétaires(suite)/Owners(continued):ID PHARMA CO., LTD., JP; IP2IP0 INNOVATIONS LIMITED, GB

(74) Agent: AIRD & MCBURNEY LP

(57) Abrégé(suite)/Abstract(continued):
invention also relates to the use of said vectors in gene therapy, particularly for the treatment of respiratory tract diseases such as Cystic Fibrosis (CF).

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 November 2015 (26.11.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/177501 A1

(51) International Patent Classification:
C07K 14/47 (2006.01) *A61K 48/00* (2006.01)
C12N 15/86 (2006.01) *C07K 14/755* (2006.01)

(21) International Application Number:
PCT/GB2015/051201

(22) International Filing Date:
23 April 2015 (23.04.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1409089.8 21 May 2014 (21.05.2014) GB

(71) Applicants: **IMPERIAL INNOVATIONS LIMITED** [GB/GB]; 52 Princes Gate, Exhibition Road, London SW7 2PG (GB). **IMPERIAL COLLEGE OF SCIENCE TECHNOLOGY AND MEDICINE** [GB/GB]; Exhibition Road, South Kensington, London SW7 2AZ (GB). **THE CHANCELLOR, MASTERS AND SCHOLARS OF THE UNIVERSITY OF OXFORD** [GB/GB]; Wellington Square, Oxford OX1 2JD (GB). **THE UNIVERSITY COURT OF THE UNIVERSITY OF EDINBURGH** [GB/GB]; Old College, Southbridge, Edinburgh EH8 9YL (GB). **DNAVEC CORPORATION** [JP/JP]; Techno Park Oho, 6 Ohkubo, Tsukuba, Ibaraki, 300-2611 (JP).

(72) Inventors: **ALTON, Eric Walter Frederick Wolfgang**; National Heart and Lung Institute, Emmanuel Kaye Building, Manresa Road, London Greater London SW3 6LR (GB). **GRIESENBACH, Uta**; 14 Hornbeam Terrace, Carshalton Greater London SM5 1RU (GB). **PYTEL, Kamila Malgorzata**; 7 Hylda Court, 3 - 5 St Albans Road, London Greater London NW5 1RE (GB). **PAUL-SMITH, Michael Christian**; Flat 4, 26 Alexandra Grove, London Greater London N4 2LF (GB). **PRINGLE, Ian Andrew**; Gene Medicine Research Group, Nuffield Division of Laboratory Clinical Sciences, John Radcliffe Hospital (Level 4), University of Oxford, Oxford Oxfordshire OX3 9DU (GB). **HYDE, Stephen Charles**; Gene Medicine Research Group, Nuffield Division of Laboratory Clinical Sciences, John Radcliffe Hospital (Level 4), University of Oxford, Oxford Oxfordshire OX3 9DU (GB). **GILL, De-**

[Continued on next page]

(54) Title: LENTIVIRAL VECTORS

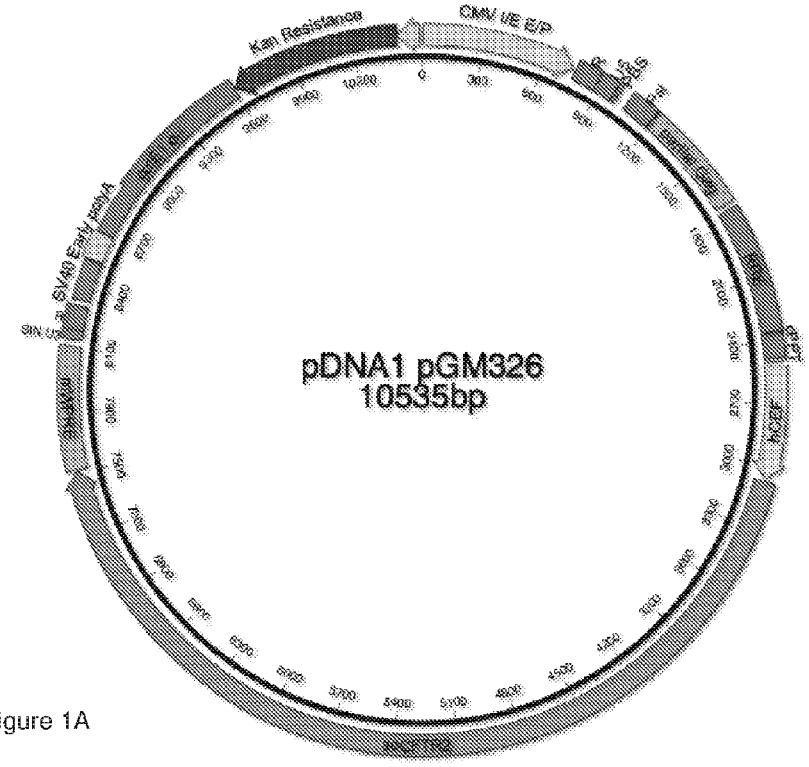


Figure 1A

[Continued on next page]

WO 2015/177501 A1

WO 2015/177501 A1

borah Rebecca; Gene Medicine Research Group, Nuffield Division of Laboratory Clinical Sciences, John Radcliffe Hospital (Level 4), University of Oxford, Oxford Oxfordshire OX3 9DU (GB). **DAVIES, Lee Adrian**; 5 Town Furlong, Appleton, Abingdon Oxfordshire OX13 5JW (GB). **BOYD, Alan Christopher**; 115 Greenbank Road, Edinburgh Lothian EH10 5RR (GB). **MCLACHLAN, Gerard**; 26 Lyne Park, West Linton Peeblesshire EH46 7HP (GB). **INOUE, Makoto**; DNAVEC Corporation, 6 Ohkubo, Tsukuba, Ibaraki 300-2611 (JP).

(74) **Agent**: **MACLEAN, Martin**; Mathys & Squire LLP, The Shard, 32 London Bridge Street, London, Greater London SE1 9SG (GB).

(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, 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, ST, 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, 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:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

(57) **Abstract**: This invention relates to lentiviral gene transfer vectors pseudotyped with hemagglutinin- neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, comprising a promoter and a transgene; and methods of making the same. The present invention also relates to the use of said vectors in gene therapy, particularly for the treatment of respiratory tract diseases such as Cystic Fibrosis (CF).

LENTIVIRAL VECTORS

The present invention relates to lentiviral gene transfer vectors pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, comprising a promoter and a transgene; and methods of making the same. The present invention also relates to the use of said vectors in gene therapy, particularly for the treatment of respiratory tract diseases such as Cystic Fibrosis (CF).

BACKGROUND TO THE INVENTION

Lentiviruses belong to a genus of viruses of the Retroviridae family, and are characterised by a long incubation period. Lentiviruses can deliver a significant amount of viral RNA into the DNA of the host cell and have the unique ability among retroviruses of being able to infect non-dividing cells, so they are one of the most efficient methods of a gene delivery vector.

Lentiviral vectors, especially those derived from HIV-1, are widely studied and frequently used vectors. The evolution of the lentiviral vectors backbone and the ability of viruses to deliver recombinant DNA molecules (transgenes) into target cells have led to their use in many applications. Two possible applications of viral vectors include restoration of functional genes in genetic therapy and *in vitro* recombinant protein production.

Pseudotyping is the process of producing viruses or viral vectors in combination with foreign viral envelope proteins. As such, the foreign viral envelope proteins can be used to alter host tropism or an increased/decreased stability of the virus particles. For example, pseudotyping allows one to specify the character of the envelope proteins. A frequently used protein is the glycoprotein G of the Vesicular stomatitis virus (VSV), short VSV-G.

Efficient and controllable retroviral expression of a transgene is understood to require the presence of intron sequences. However, incorporation of such introns into retroviral vectors involves elaborate and time-consuming methods owing to the multi-step processes employed.

To date, viral gene transfer agents have not been useful for the treatment of diseases, without the transduction of stem cell populations, because of the host adaptive immune response, which prevents successful repeat administration.

Moreover, gene transfer to the airway epithelium has proven more difficult than originally anticipated. For example, the use of lentiviral pseudotypes that require disruption of epithelial integrity to transduce the airways, for example by the use of detergents such as lysophosphatidylcholine or ethylene glycol bis(2-aminoethyl ether)-*N*,*N*,*N'N'*-tetraacetic acid, has been linked to an increased risk of sepsis.

One example of a clinical setting which would benefit from gene transfer to the airway epithelium is treatment of Cystic Fibrosis (CF). CF is a fatal genetic disorder caused by mutations in the CF transmembrane conductance regulator (*CFTR*) gene, which acts as a chloride channel in airway epithelial cells. CF is characterised by recurrent chest infections, increased airway secretions, and eventually respiratory failure. In the UK, the current median age at death is ~25 years. For most genotypes, there are no treatments targeting the basic defect; current treatments for symptomatic relief require hours of self-administered therapy daily. Gene therapy, unlike small molecule drugs, is independent of CFTR mutational class and is thus applicable to all affected CF individuals. However, to date no viral vector has met the requirements for clinical use, and the same applies to other diseases, particularly many other respiratory tract diseases.

In this regard, at least three major problems have been encountered. Gene transfer efficiency is generally poor, at least in part because the respective receptors for many viral vectors appear to be predominantly localised to the basolateral surface of the airway epithelium. Second, penetration of the respiratory tract mucus layer is generally poor. Finally, the ability to administer viral vectors repeatedly, mandatory for the life-long treatment of a self-renewing epithelium, is limited.

Administration of the vectors for clinical application is another pertinent factor. Therefore, viral stability through use of clinically relevant devices (e.g. bronchoscope and nebuliser) must be maintained for treatment efficacy.

Another example of a potential target for gene therapy is α1-antitrypsin (A1AT) deficiency. A1AT deficiency is an inherited disorder that may cause lung disease and liver disease. Symptoms include shortness of breath/wheezing, reduced ability to

exercise, weight loss, recurring respiratory infections, fatigue and rapid heartbeat upon standing. Affected individuals often develop emphysema. About 10-15% percent of patients with A1AT deficiency develop liver disease. Individuals with A1AT deficiency are also at risk of developing a hepatocellular carcinoma.

A1AT is a secreted protein, produced mainly in the liver and then trafficked to the lung, with smaller amounts also being produced in the lung itself. The main function of A1AT is to bind and neutralise neutrophil elastase. A1AT gene therapy is likely to be of therapeutic value in patients with A1AT deficiency, CF and chronic obstructive pulmonary disease (COPD), where increasing or introducing A1AT may improve lung function.

A1AT therapy is also potentially valuable for the treatment of non-respiratory/non-pulmonary diseases, such as type 1 and type 2 diabetes, acute myocardial infarction, rheumatoid arthritis, inflammatory bowel disease, transplant rejection, graft versus host (GvH) disease, multiple sclerosis and infections, particularly viral infections, due to the effect of A1AT deficiency on other tissues/organs, such as the liver and pancreas (see, for example, Lewis Mol. Med. 2012; 18:957-970).

A1AT deficiency is an attractive target disease for gene therapy because the therapeutic threshold levels are well defined. A comparison of A1AT levels in subjects with the risk of developing emphysema/COPD determined a protective threshold level of 11 μM in serum, with levels below 11 μM are used as threshold for initiating protein augmentation therapy where available. A1AT levels in airway lining fluid are only ~10% of serum level, because the lung epithelium constitutes a barrier and the therapeutic threshold in airway surface lining fluid is therefore considered to be 1.1 μM (see Ferraroti *et al.* Thorax. 2012 Aug;67(8):669-74 and Abusriwil & Stockley 2006 Current Opinion in Pulmonary Medicine 12:125-131).

Six FDA-approved commercial formulations of A1AT protein isolated from pooled human blood are in clinical use in the US for the treatment of patients with severe A1AT deficiency (via weekly intravenous injections). Enzyme replacement therapy (ERT) is expensive (~$100,000/year) and although biochemical efficacy for ERT protein augmentation therapy has been proven clinical efficacy has been more difficult to prove.

Date Reçue/Date Received 2021-06-23

A1AT ERT is currently not accessible in all countries and currently not available in the UK. In addition, it is difficult to achieve sufficiently sustained tissue levels using current therapies, which may in part be responsible for the modest clinical efficacy observed so far.

Other attractive targets for gene therapy include cardiovascular diseases and blood disorders, particularly blood clotting deficiencies such as Haemophilia (A and B), von Willebrand disease and Factor VII deficiency.

Haemophilia, particularly Haemophilia A, is an attractive target for gene therapy. Haemophilia A is an inherited bleeding disorder caused by a deficiency or mutation of Factor VIII (FVIII). Its inheritance is sex-linked, with almost all patients being male. Bleeding is typically into the joints. Bleeding into the muscle, mucosal tissue and central nervous system (CNS) is uncommon but can occur. Disease severity is inversely proportional to the level of FVIII: less than 1% (<0.01 IU/ml) results in severe disease, with bleeding after minimal injury; between 1-5% (0.01 IU/ml-0.05 IU/ml) causes moderate disease, with bleeding after mild injury; and greater than 5% (>0.05 IU/ml) causes mild disease, with bleeding only after significant trauma or surgery.

There is accordingly a need for a gene therapy vector that is able to circumvent one or more of the problems described above.

SUMMARY OF THE INVENTION

The present inventors have developed a lentiviral vector, which has been pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, comprising a promoter and a transgene. Typically the backbone of the vector is from a simian immunodeficiency virus (SIV), such as SIV1 or African green monkey SIV (SIV-AGM). Preferably the backbone of a viral vector of the invention is from SIV-AGM. The HN and F proteins function, respectively, to attach to sialic acids and mediate cell fusion for vector entry to target cells. The present inventors have discovered that this specifically F/HN-pseudotyped lentiviral vector can efficiently transduce airway epithelium, resulting in transgene expression sustained for periods beyond the proposed lifespan of airway epithelial cells. Importantly, the present

inventors have also found that re-administration does not result in a loss of efficacy. These features make the vectors of the present invention attractive candidates for treating diseases via their use in expressing therapeutic proteins: (i) within the cells of the respiratory tract; (ii) secreted into the lumen of the respiratory tract; and (iii) secreted into the circulatory system.

The present invention addresses one or more of the above needs by providing lentiviral vectors pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, comprising a promoter and a transgene. In one embodiment, the promoter is preferably a hybrid human CMV enhancer/EF1a (hCEF) promoter. The present invention also provides methods of manufacturing said vectors, compositions comprising said vectors, and uses thereof in therapy.

The vectors of the present invention enable higher and sustained gene expression through efficient gene transfer. The above-identified problems are addressed by the present invention which provides F/HN-pseudotyped lentiviral vectors which are capable of: (i) airway transduction without disruption of epithelial integrity; (ii) persistent gene expression; (iii) lack of chronic toxicity; and (iv) efficient repeat administration. Long term/persistent stable gene expression, preferably at a therapeutically-effective level, may be achieved using repeat doses of a vector of the present invention. Alternatively, a single dose may be used to achieve the desired long-term expression.

By contrast with known lentiviral vectors, the lentiviral vectors of the invention exhibit efficient airway cell uptake, enhanced transgene expression, and suffer no loss of efficacy upon repeated administration.

Thus, advantageously, the lentiviral vectors of the present invention can be used in gene therapy. By way of example, the efficient airway cell uptake properties of the vectors of the invention make them highly suitable for treating respiratory tract diseases. The lentiviral vectors of the invention can also be used in methods of gene therapy to promote secretion of therapeutic proteins. By way of further example, the invention provides secretion of therapeutic proteins into the lumen of the respiratory tract or the circulatory system. Thus, administration of a vector of the invention and its uptake by airway cells may enable the use of the lungs (or nose or airways) as a “factory” to produce a therapeutic protein that is then secreted and enters the general circulation at

therapeutic levels, where it can travel to cells/tissues of interest to elicit a therapeutic effect. In contrast to intracellular or membrane proteins, the production of such secreted proteins does not rely on specific disease target cells being transduced, which is a significant advantage and achieves high levels of protein expression. Thus, other diseases which are not respiratory tract diseases, such as cardiovascular diseases and blood disorders, particularly blood clotting deficiencies, can also be treated by the vectors of the present invention.

As an example, Alpha-1 Antitrypsin (A1AT) is a secreted anti-protease that is produced mainly in the liver and then trafficked to the lung, with smaller amounts also being produced in the lung itself. The main function of A1AT is to bind and neutralise/inhibit neutrophil elastase. Gene therapy with A1AT according to the present invention is relevant to A1AT deficient patient, as well as in other lung diseases such as cystic fibrosis or chronic obstructive pulmonary disease (COPD), and offers the opportunity to overcome some of the problems encountered by enzyme replacement therapy.

The present inventors have previously demonstrated that there is a significant correlation between neutrophil elastase (NE) and A1AT in sputum samples from cystic fibrosis patients, showing that the body produces A1AT in response to NE challenge. The present inventors have also shown that there is a statistically significant correlation between NE and lung clearance index, a marker of small airways disease, implying that increased NE has a negative impact on lung function. As presented herein, the inventors have now surprisingly demonstrated that the lentiviral vectors of the invention can achieve high concentrations of A1AT and long term (at least 90 days) A1AT expression *in vivo*. Thus, gene therapy with A1AT may neutralise NE, improving the lung function of patients with cystic fibrosis and/or COPD (and having a therapeutic effect in other indications as described herein). Accordingly, the present invention relates to the use of a lentiviral vector as described herein for the administration of an A1AT transgene and gene therapy of conditions including, but not limited to, A1AT deficiency, cystic fibrosis and//or COPD. Administration of lentiviral A1AT directly to the nasal epithelium and/or lung may overcome some of the limitations currently faced by enzyme replacement therapy (A1AT isolated from human blood and administered intravenously every week), providing stable, long-lasting expression in the target tissue (lung/nasal epithelium), ease of administration and unlimited availability.

In some embodiments, transduction with a lentiviral vector of the invention leads to secretion of the recombinant protein into the lumen of the lung as well as into the circulation. One benefit of this is that the therapeutic protein reaches the interstitium. In the case of A1AT deficiency, this is advantageous because NE inhibition is also required at this site. A1AT gene therapy may therefore also be beneficial in other disease indications, non-limiting examples of which include type 1 and type 2 diabetes, acute myocardial infarction, ischemic heart disease, rheumatoid arthritis, inflammatory bowel disease, transplant rejection, graft versus host (GvH) disease, multiple sclerosis, liver disease, cirrhosis, vasculitides and infections, such as bacterial and/or viral infections.

A1AT has numerous other anti-inflammatory and tissue-protective effects, for example in pre-clinical models of diabetes, graft versus host disease and inflammatory bowel disease. The production of A1AT in the lung and/or nose following transduction according to the present invention may, therefore, be more widely applicable, including to these indications.

Other examples of diseases that may be treated with gene therapy of a secreted protein according to the present invention include cardiovascular diseases and blood disorders, particularly blood clotting deficiencies such as haemophilia (A and B), von Willebrand disease and Factor VII deficiency.

In some embodiments, Haemophilia A may be treated according to the present invention. Disease severity is inversely proportional to the level of FVIII, and an increase in FVIII of 2-5% (0.02-0.05 IU/ml) is enough to be therapeutically effective.

In some embodiments the nose is a preferred production site for a therapeutic protein using a gene therapy vector of the invention for at least one of the following reasons: (i) extracellular barriers such as inflammatory cells and sputum are less pronounced in the nose; (ii) ease of vector administration; (iii) smaller quantities of vector required; and (iv) ethical considerations. Thus, transduction of nasal epithelial cells with a lentiviral vector of the invention may result in efficient (high-level) and long-lasting expression of the therapeutic transgene of interest.

The vectors of the present invention enable long term gene expression, resulting in long term expression of a therapeutic protein. As described herein, the phrases “long term

expression", "sustained expression" and "persistent expression" are used interchangeably. Long term expression according to the present invention means expression of a therapeutic gene and/or protein, preferably at therapeutic levels, for at least 45 days, at least 60 days, at least 90 days, at least 120 days, at least 180 days, at least 250 days, at least 360 days, at least 450 days, at least 730 days or more. Preferably long term expression means expression for at least 90 days, at least 120 days, at least 180 days, at least 250 days, at least 360 days, at least 450 days, at least 720 days or more, more preferably at least 360 days, at least 450 days, at least 720 days or more. This long term expression may be achieved by repeated doses or by a single dose.

Repeated doses may be administered twice-daily, daily, twice-weekly, weekly, monthly, every two months, every three months, every four months, every six months, yearly, every two years, or more. Dosing may be continued for as long as required, for example, for at least six months, at least one year, two years, three years, four years, five years, ten years, fifteen years, twenty years, or more, up to for the lifetime of the patient to be treated.

Lentiviral vectors, such as those of the invention, can integrate into the genome of transduced cells and lead to long-lasting expression, making them suitable for transduction of stem/progenitor cells. In the lung, several cell types with regenerative capacity have been identified as responsible for maintaining specific cell lineages in the conducting airways and alveoli. These include basal cells and submucosal gland duct cells in the upper airways, Clara cells and neuroendocrine cells in the bronchiolar airways, bronchioalveolar stem cells in the terminal bronchioles and type II pneumocytes in the alveoli. Therefore, and without being bound by theory, it is believed that the vectors of the present invention bring about long term gene expression of the transgene of interest by introducing the transgene into one or more long-lived airway epithelial cells or cell types, such as basal cells and submucosal gland duct cells in the upper airways, Clara cells and neuroendocrine cells in the bronchiolar airways, bronchioalveolar stem cells in the terminal bronchioles and type II pneumocytes in the alveoli.

Accordingly, the lentiviral vectors of the invention may transduce one or more cells or cell lines with regenerative potential within the lung (including the airways and respiratory tract) to achieve long term gene expression. In a preferred embodiment the

lentiviral vector of the invention transduces basal cells, such as those in the upper airways/respiratory tract. Basal cells have a central role in processes of epithelial maintenance and repair following injury. In addition, basal cells are widely distributed along the human respiratory epithelium, with a relative distribution ranging from 30% (larger airways) to 6% (smaller airways).

The lentiviral vectors of the invention may be used to transduce isolated and expanded stem/progenitor cells *ex vivo* prior administration to a patient. Preferably, the lentiviral vectors of the invention are used to transduce cells within the lung (or airways/respiratory tract) *in vivo*.

The vectors of the present invention enable high levels of gene expression, resulting in high levels (preferably therapeutic levels) of expression of a therapeutic protein. Expression may be measured by any appropriate method (qualitative or quantitative, preferably quantitative), and concentrations given in any appropriate unit of measurement, for example ng/ml. A high level of expression according to the present invention may mean expression of a therapeutic gene and/or protein at a concentration of at least 10 ng/ml, at least 20 ng/ml, at least 30 ng/ml, at least 40 ng/ml, at least 50 ng/ml, at least 60 ng/ml, at least 70 ng/ml, at least 80 ng/ml, at least 90 ng/ml, at least 100 ng/ml, at least 200 ng/ml, at least 300 ng/ml, at least 400 ng/ml, at least 500 ng/ml, at least 600 ng/ml, at least 700 ng/ml, at least 800 ng/ml, at least 900 ng/ml, at least 1,000 ng/ml, at least 2,000 ng/ml, at least 3,000 ng/ml, at least 4,000 ng/ml, at least 5,000 ng/ml, at least 10,000, at least 15,000 ng/ml, at least 20,000 ng/ml or more. Therapeutic expression may be defined using these same values.

The lentiviral vectors of the present invention typically provide high expression levels of a transgene when administered to a patient. The terms high expression and therapeutic expression are used interchangeably herein.

A high level of expression according to the present invention may mean expression of a therapeutic gene and/or protein at a concentration of at least about 100nM, at least about 200nM, at least about 300nM, at least about 400nM, at least about 500nM, at least about 600nM, at least about 700nM, at least about 800nM, at least about 900nM, at least about 1μM, at least about 1.1μM, at least about 1.2μM, at least about 1.3μM, at least about 1.4μM, at least about 1.5μM, at least about 2μM, at least about 3μM, at least

about 4μM, at least about 5μM, at least about 6μM, at least about 7μM, at least about 8μM, at least about 9μM, at least about 10μM, at least about 11μM, at least about 12μM, at least about 13μM, at least about 14μM, at least about 15μM, at least about 20μM, at least about 25μM, at least about 30μM, at least about 40μM, at least about 50μM, at least about 75μM, or at least about 100μM or more. Therapeutic expression may be defined using these same values.

A high level of expression according to the present invention may mean expression of a therapeutic gene (typically measured by mRNA expression) at least about 1%, at least about 2%, at least about 3%, at least about 4%, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 15%, at least about 20% or more compared with the expression level of the corresponding endogenous (defective) mRNA. Therapeutic expression may be defined using these same values. For example, a typical expression level of endogenous CFTR mRNA may be quantified in terms of the number of copies of the mRNA per lung cell, for example one copy of the endogenous CFTR mRNA per lung cell, two copies of the endogenous CFTR mRNA per lung cell, three copies of the endogenous CFTR mRNA per lung cell, four copies of the endogenous CFTR mRNA per lung cell, five copies of the endogenous CFTR mRNA per lung cell, or more, preferably two copies of the endogenous CFTR mRNA per lung cell. The expression of the therapeutic gene of the invention, such as a functional CFTR gene, may be quantified relative to the endogenous gene, such as the endogenous (dysfunctional) CFTR genes in terms of mRNA copies per cell or any other appropriate unit.

A high level of expression according to the present invention may mean expression of a therapeutic gene and/or protein at a concentration of at least about 0.5%, at least about 1%, at least about 2%, at least about 3%, at least about 4%, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or more compared with the wild type level of the therapeutic gene and/or protein, wherein the wild type level is the level in a normal individual without the disease. In some embodiments, wild type expression is given as 100%, with any improvement in gene expression measured relative to that. As a non-

limiting example, if in a normal individual without the disease the expression of the functional gene is given as 100%, and in an individual with the disease, the expression of the functional gene is 0%, a therapeutic level of expression of the gene or protein may be at least about 0.5%, at least about 1%, at least about 2%, at least about 3%, at least about 4%, at least about 5%, at least about 6%, at least about 7%, at least about 8%, at least about 9%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 40%, at least about 50%, or more compared with the wild type level of the therapeutic gene and/or protein. As another non-limiting example, if in a normal individual without the disease the expression of the functional gene is given as 100%, and in an individual with the disease, the expression of the functional gene is 50%, a therapeutic level of expression of the gene or protein may be at least about 55%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or more compared with the wild type level of the therapeutic gene and/or protein.

For secreted proteins such as A1AT, typically the concentration in the lung or epithelial lining fluid (as measured using BAL) is approximately ten times that in serum. As a non-limiting example, if the concentration of secreted protein in the lung or epithelial lining fluid is in the region of 750 ng/ml, the serum concentration of the protein is in the region of 75 ng/ml.

Expression levels of a therapeutic gene and/or protein of the invention may be measured in the lung tissue, epithelial lining fluid and/or serum/plasma as appropriate. A high and/or therapeutic expression level may therefore refer to the concentration in the lung, epithelial lining fluid and/or serum/plasma.

As a non-limiting example, a therapeutic expression level of CFTR is typically 1-5% of the therapeutic CFTR mRNA compared with the expression level of the endogenous (defective) CFTR mRNA.

As another non-limiting example, a therapeutic expression level of A1AT is typically at least about 1μM in the epithelial lining fluid, and/or at least about 0.1μM in the serum. In a preferred embodiment, a therapeutic expression level of A1AT in the epithelial lining fluid is at least about 1.1 μM, and/or a therapeutic serum expression level of A1AT

according to the present invention is at least about 11μM. As another non-limiting example, a therapeutic expression level of A1AT in the epithelial lining fluid (ELF, i.e. the fluid lining the airways and airspaces in the lungs) is 70μg/ml (compared with a "normal" target level of ATT (A1AT) in the ELF of 200μg/ml).

As another non-limiting example, a therapeutic expression level of FVIII protein is typically at least about 1-3% or at least about 1-6% of the expression level in a normal individual who does not suffer from haemophilia.

The therapeutic gene included in the vector of the invention may be modified to facilitate expression. For example, the gene sequence may be in CpG-depleted and/or codon-optimised form to facilitate gene expression. Standard techniques for modifying the gene sequence in this way are known in the art.

The promoter included in the vector of the invention may be specifically selected and/or modified to further refine regulation of expression of the therapeutic gene. Again, suitable promoters and standard techniques for their modification are known in the art. As a non-limiting example, a number of suitable (CpG-free) promoters suitable for use in the present invention are described in Pringle *et al.* (J. Mol. Med. Berl. 2012, 90(12): 1487-96).

The vector of the invention may be modified to allow shut down of gene expression. Standard techniques for modifying the vector in this way are known in the art. As a non-limiting example, Tet-responsive promoters are widely used.

The vectors of the present invention also demonstrate remarkable resistance to shear forces with only modest reduction in transduction ability when passaged through clinically-relevant delivery devices such as bronchoscopes, spray bottles and nebulisers.

In one embodiment, the invention provides F/HN lentiviral vectors comprising a promoter and a transgene, having no intron positioned between the promoter and the transgene. In one embodiment, the vector of the present invention is delivered to cells of the respiratory tract. In embodiment, the lentivirus is SIV. In one embodiment, the promoter is a hybrid human CMV enhancer/EF1a (hCEF) promoter. Typically said promoter of the invention lacks the intron corresponding to nucleotides 570-709 and the exon

Date Reçue/Date Received 2021-06-23

corresponding to nucleotides 728-733 of the hCEF promoter. A preferred example of an hCEF promoter sequence of the invention is provided by SEQ ID NO: 6. The promoter may be a CMV promoter. An example of a CMV promoter sequence is provided by SEQ ID NO: 17. Other promoters for transgene expression are known in the art and their suitability for the lentiviral vectors of the invention determined using routine techniques known in the art. Non-limiting examples of other promoters include UbC and UCOE. As described herein, the promoter may be modified to further regulate expression of the transgene of the invention.

In one embodiment, the transgene encodes a CFTR. An example of a *CFTR* cDNA is provided by SEQ ID NO: 7.

In one embodiment, the transgene encodes an A1AT. An example of an *A1AT* transgene is provided by SEQ ID NO: 15, or by the complementary sequence of SEQ ID NO: 26. SEQ ID NO: 15 is a codon-optimized CpG depleted *A1AT* transgene designed by the present inventors to enhance translation in human cells. Such optimisation has been shown to enhance gene expression by up to15-fold. Thus, in one embodiment, the invention provides a polynucleotide comprising or consisting of the nucleotide sequence of SEQ ID NO: 15. Variants of same sequence (as defined herein) which possess the same technical effect of enhancing translation compared with the unmodified (wild-type) *A1AT* gene sequence are also encompassed by the present invention. The invention further provides a polypeptide encoded by said *A1AT* transgene, as exemplified by the polypeptide of SEQ ID NO: 27, plasmids (particularly vector genome plasmids as defined herein) and lentiviral vectors comprising said *A1AT* transgene. In a preferred embodiment, aspects of the invention relating to *A1AT* gene therapy according to the present invention use the *A1AT* transgene sequence of SEQ ID NO: 15.

In one embodiment, the transgene encodes a FVIII. Examples of a *FVIII* transgene are provided by SEQ ID NOs: 16 and 30, or by the respective complementary sequences of SEQ ID NO: 28 and 31.

Lentiviral vectors suitable for use in the present invention include Human immunodeficiency virus (HIV), Simian immunodeficiency virus (SIV), Feline immunodeficiency virus (FIV), Equine infectious anaemia virus (EIAV), and Visna/maedi

virus. In one embodiment of the invention, an SIV vector is used, preferably SIV-AGM. In another embodiment, an HIV vector is used.

The lentiviral vectors of the present invention are pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus. In one embodiment, the respiratory paramyxovirus is a Sendai virus (murine parainfluenza virus type 1).

In one embodiment of the invention, the lentiviral vector is integrase-competent (IC). In an alternative embodiment, the lentiviral vector is integrase-deficient (ID).

In another embodiment of the invention, the transgene of the invention is any one or more of *DNAH5, DNAH11, DNAI1,* and *DNAI2*, or other known related gene.

In one embodiment of the invention, the respiratory tract epithelium is targeted for delivery of the vector. In this embodiment, the transgene encodes Alpha-1 Antitrypsin (A1AT), Surfactant Protein B (SFTPB), or Granulocyte-Macrophage Colony-Stimulating Factor (GM-CSF). In another embodiment, the transgene encodes a monoclonal antibody (mAb) against an infectious agent. In one embodiment, transgene encodes anti-TNF alpha. In a further embodiment, the transgene encodes a therapeutic protein implicated in an inflammatory, immune or metabolic condition.

In one embodiment of the invention, the vector is delivered to the cells of the respiratory tract to allow production of proteins to be secreted into circulatory system. In this embodiment, the transgene encodes for Factor VII, Factor VIII, Factor IX, Factor X, Factor XI and/or von Willebrand's factor. Such a vector may be used in the treatment of diseases, particularly cardiovascular diseases and blood disorders, preferably blood clotting deficiencies such as Haemophilia. In another embodiment, the transgene encodes a monoclonal antibody (mAb) against an infectious agent. In one embodiment, the transgene encodes a protein implicated in an inflammatory, immune or metabolic condition, such as, lysosomal storage disease.

In accordance with the invention, there is provided an F/HN-SIV lentiviral vector comprising an hCEF promoter and a *CFTR* transgene, having no intron positioned between the promoter and the transgene. Similarly, there is no intron between the promoter and the transgene in the vector genome (pDNA1) plasmid (for example, pGM326 as described herein, illustrated in Figure 1A and with the sequence of SEQ ID NO: 1).

The invention also provides an F/HN-SIV lentiviral vector comprising an hCEF promoter and an *A1AT* transgene, having no intron positioned between the promoter and the transgene. Such a lentiviral vector may be produced by the method described herein, using a plasmid carrying the *A1AT* transgene and a promoter. Similarly, there is no intron between the promoter and the *A1AT* transgene in the vector genome (pDNA1) plasmid. An exemplary sequence of such a plasmid is given in SEQ ID NO: 9 (F/HN-SIV-hCEF-soA1AT, illustrated in Figure 15A).

The invention also provides an F/HN-SIV lentiviral vector comprising (i) an hCEF promoter or a CMV promoter; and (ii) an *FVIII* transgene; wherein no intron is positioned between the promoter and the transgene. Such a lentiviral vector may be produced by the method described herein, using a plasmid carrying the *FVIII* transgene and a promoter. Similarly, there is no intron between the promoter and the *FVIII* transgene in the vector genome (pDNA1) plasmid. Exemplary sequences of such plasmids are given in SEQ ID NO: 11 to 14 (illustrated in Figure 22A to E).

The lentiviral vector as described above comprises a transgene. The transgene comprises a nucleic acid sequence encoding a gene product, e.g., a protein.

For example, in one embodiment, the nucleic acid sequence encoding a *CFTR, A1AT or FVIII* comprises (or consists of) a nucleic acid sequence having at least 90% (such as at least 90, 92, 94, 95, 96, 97, 98, 99 or 100%) sequence identity to the *CFTR , A1AT or FVIII* nucleic acid sequence respectively. In a further embodiment, the nucleic acid sequence encoding *CFTR, A1AT or FVIII* comprises (or consists of) a nucleic acid sequence having at least 95% (such as at least 95, 96, 97, 98, 99 or 100%) sequence identity to the *CFTR, A1AT or FVIII* nucleic acid sequence respectively. In one embodiment, the nucleic acid sequence encoding *CFTR* is provided by SEQ ID NO: 7, the nucleic acid sequence encoding A1AT is provided by SEQ ID NO: 15, or by the complementary sequence of SEQ ID NO: 26 and/or the nucleic acid sequence encoding FVIII is provided by SEQ ID NO: 16 and 30, or by the respective complementary sequences of SEQ ID NO: 28 and 31, or variants thereof.

The term "polypeptide" as used herein also encompasses variant sequences. Thus, the polypeptide encoded by the transgene of the invention may have at least 90% (such as at least 90, 92, 94, 95, 96, 97, 98, 99 or 100%) sequence identity to a functional CFTR,

A1AT or FVIII polypeptide sequence respectively. In a further embodiment, the amino acid sequence of the CFTR, A1AT or FVIII transgene comprises (or consists of) an amino acid sequence having at least 95% (such as at least 95, 96, 97, 98, 99 or 100%) sequence identity to the functional CFTR, A1AT or FVIII polypeptide sequence respectively. In one embodiment, the amino acid sequence of the A1AT protein encoded by the transgene of the invention comprises (or consists of) the amino acid sequence of SEQ ID NO: 27, or variants thereof. Preferably said variant A1AT proteins of the invention have at least 90% (such as at least 90, 92, 94, 95, 96, 97, 98, 99 or 100%), more preferably at least 95% or more sequence identity with SEQ ID NO: 27.

In one embodiment, the nucleic acid sequence encoding CFTR, *A1AT or FVIII* comprises (or consists of) the *CFTR, A1AT or FVII* complementary DNA sequence respectively. In one embodiment, the *CFTR, A1AT or FVIII* transgene is a sequence-optimised *CFTR, A1AT or FVIII* (*soCFTR2, soA1AT or FVIII*). An example is provided by SEQ ID NOS: 7, 15 and 16 respectively. An exemplary complementary sequence-optimised *A1AT* sequence is given by SEQ ID NO: 26. Exemplary complementary sequence optimised *FVIII* sequences are given by SEQ ID NOs: 28 and 31.

In one embodiment of the invention, the F/HN vector transgene expression is driven by cytomegalovirus (CMV) promoter. In another embodiment, the vector transgene expression is driven by elongation factor 1a (EF1a) promoter. In a preferred embodiment, the vector transgene expression is driven by hybrid human CMV enhancer/EF1a (hCEF) promoter. In one embodiment, the hCEF promoter has all CG dinucleotides replaced with any one of AG, TG or GT. Thus, in one embodiment, the hCEF promoter is CpG-free.

In one embodiment, the lentiviral vector may be produced using the F/HN-SIV-hCEF-soCFTR2-IC plasmid. In this embodiment, *CFTR* is expressed under control of the hCEF promoter. This lentiviral vector may be described as comprising F/HN-SIV-hCEF-soCFTR2-IC, as it comprises the SIV F/HN elements, as well as an integrase competent expression cassette comprising *CFTR* under the control of the hCEF promoter. This lentiviral vector of the invention is capable of producing long-lasting, repeatable, high-level expression in airway cells without inducing an undue immune response. Consequently, the invention provides an efficient means of *in vivo* gene therapy, for example, *CFTR* gene transfer into the CF lung for the treatment of CF lung disease.

In a preferred embodiment, the lentiviral vector may be produced using the F/HN-SIV-hCEF-soA1AT plasmid. In this embodiment, *A1AT* is expressed under control of the hCEF promoter. This lentiviral vector may be described as comprising F/HN-SIV-hCEF-soA1AT, as it comprises the SIV F/HN elements, as well as an expression cassette comprising *A1AT* under the control of the hCEF promoter. This lentiviral vector of the invention is capable of producing long-lasting, repeatable, high-level expression in airway cells without inducing an undue immune response. Consequently, the invention provides an efficient means of *in vivo* gene therapy, for example, *A1AT* gene transfer into a patient's lung or nose for the production of A1AT which is then secreted into the circulatory system (as described herein). Thus, this vector and other vectors of the invention comprising the *A1AT* transgene may be used in the treatment of A1AT deficiency, or other indications as described herein.

In another preferred embodiment, the lentiviral vectors may be produced using the F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-co and/or F/HN-SIV-hCEF-HFVIII-N6-co plasmids. HFVIII refers to human FVIII. In this embodiment, *FVIII* is expressed under control of the hCEF or CMV promoter. These lentiviral vectors may be described as comprising F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-co and F/HN-SIV-hCEF-HFVIII-N6-co respectively, as they comprise the SIV F/HN elements, as well as an expression cassette comprising *FVIII* under the control of the hCEF/CMV promoter. Viral vector products produced using the F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-co and/or F/HN-SIV-hCEF-HFVIII-N6-co plasmids are also known as vGM126, vGM127, vGM142 and vGM129 (see Figure 22). These lentiviral vectors of the invention are capable of producing long-lasting, repeatable, high-level expression in airway cells without inducing an undue immune response. Consequently, the invention provides an efficient means of *in vivo* gene therapy, for example, *FVIII* gene transfer into a patient's lung or nose for the production of FVIII which is then secreted into the circulatory system (as described herein). Thus, these vectors and other vectors of the invention comprising the *FVIII* transgene may be used in the treatment of haemophilia, or other indications as described herein.

The lentiviral vectors of the invention do not contain an intron between the promoter and the transgene. Similarly, the vector genome plasmids of the invention (used to generate said lentiviral vectors as described herein) also do not contain an intron between the

promoter and the transgene. The invention therefore provides, in one embodiment, no intron between the hCEF promoter and the coding sequences to be expressed. In one preferred embodiment, the coding sequence to be expressed is a *CFTR, A1AT and/or FVIII* nucleic acid sequence.

In one embodiment, the vectors of the invention comprise central polypurine tract (cPPT) and the Woodchuck hepatitis virus posttranscriptional regulatory elements (WPRE). In one embodiment, the WPRE sequence is provided by SEQ ID NO: 8.

In one embodiment the vector of the invention is used for gene therapy. In one embodiment the disease to be treated is CF. In another embodiment of the invention, the disease to be treated is Primary Ciliary Dyskinesia (PCD). In one embodiment, the vector is used to treat acute lung injury. In one embodiment of the invention, the disease to be treated is Surfactant Protein B (SP-B) deficiency, Alpha 1-antitrypsin Deficiency (A1AD), Pulmonary Alveolar Proteinosis (PAP), Chronic obstructive pulmonary disease (COPD). In another embodiment, the disease is an inflammatory, immune or metabolic condition.

The disease to be treated may be a cardiovascular disease or blood disorder, particularly a blood clotting deficiency. Thus, in some embodiments, the disease to be treated is Haemophilia A, Haemophilia B, or Haemophilia C, Factor VII deficiency and/or von Willebrand disease. In yet another embodiment, the disease to be treated is an inflammatory disease, infectious disease or metabolic condition, such as, lysosomal storage disease.

Non-limiting examples of diseases which may be treated using A1AT gene therapy according to the present invention include type 1 and type 2 diabetes, acute myocardial infarction, ischemic heart disease, rheumatoid arthritis, inflammatory bowel disease, transplant rejection, graft versus host (GvH) disease, multiple sclerosis, liver disease, cirrhosis, vasculitides and infections, such as bacterial and/or viral infections.

In one aspect of the invention, the vector can effectively treat a disease by providing a transgene for the correction of the disease. For example, inserting a functional copy of the *CFTR* gene to ameliorate or prevent lung disease in CF patients, independent of the underlying mutation.

In another embodiment of the invention, a lentiviral production method is provided. In this embodiment, the method of the invention is a scalable GMP-compatible method. Thus, the method of the invention allows the generation of high titre purified F/HN vectors.

The method of the invention comprises the following steps:

(a) growing cells in suspension;

(b) transfecting the cells with one or more plasmids;

(c) adding a nuclease;

(d) harvesting the lentivirus;

(e) adding trypsin; and

(f) purification.

In one embodiment of the method of the invention, the one or more plasmids provide the vector genome, the Gag-Pol, Rev, F and HN. Thus, there can be five plasmids for each of the vector genome, the Gag-Pol, Rev, F and HN, respectively. In the preferred 5 plasmid method of the invention, the vector genome plasmid encodes all the genetic material that is packaged into final lentiviral vector, including the transgene. Typically only a portion of the genetic material found in the vector genome plasmid ends up in the virus. The vector genome plasmid may be designated herein as "pDNA1". The other four plasmids are manufacturing plasmids encoding the Gag-Pol, Rev, F and HN proteins. These plasmids may be designated "pDNA2a", "pDNA2b", "pDNA3a" and "pDNA3b" respectively.

In one embodiment of the invention, the lentivirus is SIV, such as SIV1, preferably SIV-AGM. In one embodiment, the F and HN proteins are derived from a Paramyxovirus, such as Sendai virus. In one embodiment, the vector genome plasmid (pDNA1) comprises the transgene and the transgene promoter.

In a specific embodiment relating to *CFTR*, the five plasmids are characterised by Figures 1A-1E, thus pDNA1 is the pGM326 plasmid of Figure 1A, pDNA2a is the pGM299 plasmid of Figure 1B, pDNA2b is the pGM299 plasmid of Figure 1C, pDNA3a is the pGM301 plasmid of Figure 1D and pDNA3b is the pGM303 plasmid of Figure 1E. In this embodiment, the final CFTR containing lentiviral vector may be referred to as

vGM058 (see the Examples). The vGM058 vector is a preferred embodiment of the invention.

In an embodiment relating to *A1AT*, the five plasmids may be characterised by Figures 15A (thus plasmid pDNA1 may be pGM407) and 1B-E (as above for the specific CFTR embodiment).

In an embodiment relating to *FVIII*, the five plasmids may be characterised by Figures 22C-F (thus plasmid pDNA1 may be pGM411, pGM412, pGM413 or pGM414) and 1B-E.

In these embodiments of the invention, the plasmid as defined in Figure 1A is represented by SEQ ID NO: 1; the plasmid as defined in Figure 1B is represented by SEQ ID NO: 2; the plasmid as defined in Figure 1C is represented by SEQ ID NO: 3; the plasmid as defined in Figure 1D is represented by SEQ ID NO: 4; the plasmid as defined in Figure 1E is represented by SEQ ID NO: 5; the plasmid as defined in Figure 15A is represented by SEQ ID NO: 9 and the F/HN-SIV-CMV-HFVIII-V3, F/HN-SIV-hCEF-HFVIII-V3, F/HN-SIV-CMV-HFVIII-N6-co and/or F/HN-SIV-hCEF-HFVIII-N6-co plasmids as defined in Figure 22B are represented by SEQ ID NOs: 11 to 14 respectively.

In the 5 plasmid method of the invention all five plasmids contribute to the formation of the final lentiviral vector. During manufacture of the lentiviral vector, the vector genome plasmid (pDNA1) provides the enhancer/promoter, Psi, RRE, cPPT, mWPRE, SIN LTR, SV40 polyA (see Figure 1A), which are important for viral manufacture. Using pGM326 as a non-limiting example of a pDNA1, the CMV enhancer/promoter, SV40 polyA, colE1 Ori and KanR are involved in manufacture of the lentiviral vector of the invention (e.g. vGM058), but are not found in the final viral vector. The RRE, cPPT (central polypurine tract), hCEF, soCFTR2 (transgene) and mWPRE from pGM326 are found in the final viral vector. SIN LTR (long terminal repeats, SIN/IN self inactivating) and Psi (packaging signal) may be found in the final viral vector.

For other lentiviral vectors of the invention, corresponding elements from the other vector genome plasmids (pDNA1) are required for manufacture (but not found in the final vector), or are present in the final viral vector.

The F and HN proteins from pDNA3a and pDNA3b (preferably Sendai F and HN proteins) are important for infection of target cells with the final lentiviral vector, i.e. for entry of a patients epithelial cells (typically lung or nasal cells as described herein). The products of the pDNA2a and pDNA2b plasmids are important for virus transduction, i.e. for inserting the lentiviral DNA into the host's genome. The promoter, regulatory elements (such as WPRE) and transgene are important for transgene expression within the target cell(s).

In one embodiment, steps (a)-(f) are carried out sequentially. In one embodiment, the cells are HEK293 cells or 293T/17 cells. In one embodiment, the cells are grown in animal-component free (serum-free) media. In one embodiment, the transfection is carried out by the use of PEIPro™. In one embodiment, the nuclease is an endonuclease, for example, Benzonase®. In one embodiment, the trypsin activity is provided by an animal origin free, recombinant enzyme such as TrypLE Select™.

In one embodiment of the invention, the addition of the nuclease is at the pre-harvest stage. In an alternative embodiment, the addition of the nuclease is at the post-harvest stage. In another embodiment, the addition of trypsin is at the pre-harvest stage. In another embodiment, the addition of the trypsin is at the post-harvest stage.

In one embodiment, the purification step comprises a chromatography step. In this embodiment, mixed-mode size exclusion chromatography (SEC) is used. In one embodiment, anion exchange chromatography is used. In this embodiment, no salt gradient is used for the elution step.

In one embodiment, this method is used to produce the lentiviral vectors of the invention. In this embodiment, the vector of the invention comprises the *CFTR, A1AT and/or FVIII* gene. In an alternative embodiment, the vector of the invention comprises any of the above-mentioned genes, or the genes encoding the above-mentioned proteins.

In one embodiment of the method of the invention, any combination of one or more of the specific plasmid constructs provided by Figures 1A-1E, Figure 15A and/or Figure 22C-22F is used to provide a vector of the invention.

The invention further provides a method of treating a disease, the method comprising administering a lentiviral vector of the invention to a subject. In this embodiment, the invention provides a lentiviral vector of the invention for use in treatment of a lung disease. In one embodiment, disease is a chronic disease. In a specific embodiment, a method of treating CF is provided. In other embodiments, a method of treating Primary Ciliary Dyskinesia (PCD), Surfactant Protein B (SP-B) deficiency, Alpha 1-antitrypsin Deficiency (A1AD), Pulmonary Alveolar Proteinosis (PAP), Chronic obstructive pulmonary disease (COPD) is provided. In another embodiment, the disease is an inflammatory, immune or metabolic condition.

In another embodiment, the disease may be a cardiovascular disease or blood disorder, particularly a blood clotting deficiency, such as Haemophilia A, Haemophilia B, Haemophilia C, Factor VII deficiency and/or von Willebrand disease, an inflammatory disease, infectious disease or metabolic condition, such as, lysosomal storage disease.

The disease may be type 1 and type 2 diabetes, acute myocardial infarction, ischemic heart disease, rheumatoid arthritis, inflammatory bowel disease, transplant rejection, graft versus host (GvH) disease, multiple sclerosis, liver disease, cirrhosis, vasculitides and infections, such as bacterial and/or viral infections.

The lentiviral vectors of the invention may be administered in any dosage appropriate for achieving the desired therapeutic effect. Appropriate dosages may be determined by a clinician or other medical practitioner using standard techniques and within the normal course of their work. Non-limiting examples of suitable dosages include $1x10^8$ transduction units (TU), $1x10^9$ TU, $1x10^{10}$ TU, $1x10^{11}$ TU or more.

The invention also provides compositions comprising the lentiviral vectors described above, and a pharmaceutically-acceptable carrier. Non-limiting examples of pharmaceutically acceptable carriers include water, saline, and phosphate-buffered saline. In some embodiments, however, the composition is in lyophilized form, in which case it may include a stabilizer, such as bovine serum albumin (BSA). In some embodiments, it may be desirable to formulate the composition with a preservative, such as thiomersal or sodium azide, to facilitate long-term storage.

The vectors of the invention may be administered by any appropriate route. It may be desired to direct the compositions of the present invention (as described above) to the respiratory system of a subject. Efficient transmission of a therapeutic/prophylactic composition or medicament to the site of infection in the respiratory tract may be achieved by oral or intra-nasal administration, for example, as aerosols (e.g. nasal sprays), or by catheters. Typically the lentiviral vectors of the invention are stable in clinically relevant nebulisers, catheters and aerosols, etc.

Formulations for intra-nasal administration may be in the form of nasal droplets or a nasal spray. An intra-nasal formulation may comprise droplets having approximate diameters in the range of 100-5000 µm, such as 500-4000 µm, 1000-3000 µm or 100-1000 µm. Alternatively, in terms of volume, the droplets may be in the range of about 0.001-100 µl, such as 0.1-50 µl or 1.0-25 µl, or such as 0.001-1 µl.

The aerosol formulation may take the form of a powder, suspension or solution. The size of aerosol particles is relevant to the delivery capability of an aerosol. Smaller particles may travel further down the respiratory airway towards the alveoli than would larger particles. In one embodiment, the aerosol particles have a diameter distribution to facilitate delivery along the entire length of the bronchi, bronchioles, and alveoli. Alternatively, the particle size distribution may be selected to target a particular section of the respiratory airway, for example the alveoli. In the case of aerosol delivery of the medicament, the particles may have diameters in the approximate range of 0.1-50 µm, preferably 1-25 µm, more preferably 1-5 µm.

Aerosol particles may be for delivery using a nebulizer (e.g. via the mouth) or nasal spray. An aerosol formulation may optionally contain a propellant and/or surfactant.

As used herein, the terms “nucleic acid sequence” and “polynucleotide” are used interchangeably and do not imply any length restriction. As used herein, the terms “nucleic acid” and “nucleotide” are used interchangeably. The terms “nucleic acid sequence” and “polynucleotide” embrace DNA (including cDNA) and RNA sequences. The terms “transgene” and “gene” are also used interchangeably and both terms encompass fragments or variants thereof encoding the target protein.

The transgenes of the present invention include nucleic acid sequences that have been removed from their naturally occurring environment, recombinant or cloned DNA

isolates, and chemically synthesized analogues or analogues biologically synthesized by heterologous systems.

The polynucleotides of the present invention may be prepared by any means known in the art. For example, large amounts of the polynucleotides may be produced by replication in a suitable host cell. The natural or synthetic DNA fragments coding for a desired fragment will be incorporated into recombinant nucleic acid constructs, typically DNA constructs, capable of introduction into and replication in a prokaryotic or eukaryotic cell. Usually the DNA constructs will be suitable for autonomous replication in a unicellular host, such as yeast or bacteria, but may also be intended for introduction to and integration within the genome of a cultured insect, mammalian, plant or other eukaryotic cell lines.

The polynucleotides of the present invention may also be produced by chemical synthesis, e.g. by the phosphoramidite method or the tri-ester method, and may be performed on commercial automated oligonucleotide synthesizers. A double-stranded fragment may be obtained from the single stranded product of chemical synthesis either by synthesizing the complementary strand and annealing the strand together under appropriate conditions or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

When applied to a nucleic acid sequence, the term "isolated" in the context of the present invention denotes that the polynucleotide sequence has been removed from its natural genetic milieu and is thus free of other extraneous or unwanted coding sequences (but may include naturally occurring 5' and 3' untranslated regions such as promoters and terminators), and is in a form suitable for use within genetically engineered protein production systems. Such isolated molecules are those that are separated from their natural environment.

In view of the degeneracy of the genetic code, considerable sequence variation is possible among the polynucleotides of the present invention. Degenerate codons encompassing all possible codons for a given amino acid are set forth below:

Amino Acid	**Codons**	**Degenerate Codon**
Cys	TGC TGT	TGY

Ser	AGC AGT TCA TCC TCG TCT	WSN
Thr	ACA ACC ACG ACT	ACN
Pro	CCA CCC CCG CCT	CCN
Ala	GCA GCC GCG GCT	GCN
Gly	GGA GGC GGG GGT	GGN
Asn	AAC AAT	AAY
Asp	GAC GAT	GAY
Glu	GAA GAG	GAR
Gln	CAA CAG	CAR
His	CAC CAT	CAY
Arg	AGA AGG CGA CGC CGG CGT	MGN
Lys	AAA AAG	AAR
Met	ATG	ATG
Ile	ATA ATC ATT	ATH
Leu	CTA CTC CTG CTT TTA TTG	YTN
Val	GTA GTC GTG GTT	GTN
Phe	TTC TTT	TTY
Tyr	TAC TAT	TAY
Trp	TGG	TGG
Ter	TAA TAG TGA	TRR
Asn/ Asp		RAY
Glu/ Gln		SAR
Any		NNN

One of ordinary skill in the art will appreciate that flexibility exists when determining a degenerate codon, representative of all possible codons encoding each amino acid. For example, some polynucleotides encompassed by the degenerate sequence may encode variant amino acid sequences, but one of ordinary skill in the art can easily identify such variant sequences by reference to the amino acid sequences of the present invention.

A "variant" nucleic acid sequence has substantial homology or substantial similarity to a reference nucleic acid sequence (or a fragment thereof). A nucleic acid sequence or fragment thereof is "substantially homologous" (or "substantially identical") to a reference sequence if, when optimally aligned (with appropriate nucleotide insertions or deletions) with the other nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 70%, 75%, 80%, 82, 84, 86, 88, 90, 92, 94, 96, 98 or 99% of the nucleotide bases. Methods for homology determination of nucleic acid sequences are known in the art.

Alternatively, a "variant" nucleic acid sequence is substantially homologous with (or substantially identical to) a reference sequence (or a fragment thereof) if the "variant" and the reference sequence they are capable of hybridizing under stringent (e.g. highly stringent) hybridization conditions. Nucleic acid sequence hybridization will be affected by such conditions as salt concentration (e.g. NaCl), temperature, or organic solvents, in addition to the base composition, length of the complementary strands, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. Stringent temperature conditions are preferably employed, and generally include temperatures in excess of 30°C, typically in excess of 37°C and preferably in excess of 45°C. Stringent salt conditions will ordinarily be less than 1000 mM, typically less than 500 mM, and preferably less than 200 mM. The pH is typically between 7.0 and 8.3. The combination of parameters is much more important than any single parameter.

Methods of determining nucleic acid percentage sequence identity are known in the art. By way of example, when assessing nucleic acid sequence identity, a sequence having a defined number of contiguous nucleotides may be aligned with a nucleic acid sequence (having the same number of contiguous nucleotides) from the corresponding portion of a nucleic acid sequence of the present invention. Tools known in the art for determining nucleic acid percentage sequence identity include Nucleotide BLAST.

One of ordinary skill in the art appreciates that different species exhibit "preferential codon usage". As used herein, the term "preferential codon usage" refers to codons that are most frequently used in cells of a certain species, thus favouring one or a few representatives of the possible codons encoding each amino acid. For example, the amino acid threonine (Thr) may be encoded by ACA, ACC, ACG, or ACT, but in mammalian host cells ACC is the most commonly used codon; in other species, different codons may be preferential. Preferential codons for a particular host cell species can be introduced into the polynucleotides of the present invention by a variety of methods known in the art. Introduction of preferential codon sequences into recombinant DNA can, for example, enhance production of the protein by making protein translation more efficient within a particular cell type or species.

Thus, in one embodiment of the invention, the nucleic acid sequence is codon optimized for expression in a host cell.

A "fragment" of a polynucleotide of interest comprises a series of consecutive nucleotides from the sequence of said full-length polynucleotide. By way of example, a "fragment" of a polynucleotide of interest may comprise (or consist of) at least 30 consecutive nucleotides from the sequence of said polynucleotide (e.g. at least 35, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800 850, 900, 950 or 1000 consecutive nucleic acid residues of said polynucleotide). A fragment may include at least one antigenic determinant and/or may encode at least one antigenic epitope of the corresponding polypeptide of interest. Typically a fragment as defined herein retains the same function as the full-length polynucleotide or polypeptide.

DESCRIPTION OF THE DRAWINGS

The present invention will now be described by way of example only with reference to the accompanying drawings, in which:

Figure 1 shows exemplary plasmids utilised in the invention. Figures 1A-1E show schematic drawings of plasmids used for production of the vectors of the invention. In one embodiment of the invention, Figure 1A provides a tool of the invention.

Figure 2 shows the duration of F/HN-SIV transgene expression in mouse nasal tissue, which was perfused with $4x10^8$ TU F/HN-SIV-CMV-EGFP and EGFP expression

determined at the indicated number of days post-treatment. A negative control is shown where nasal tissue was perfused with vector diluent (PBS).

Figure 3 demonstrates consistency of F/HN-SIV transgene expression, by showing at 1 year post-treatment. EGFP expression was determined in 10 independent mice at 360 days post-treatment.

Figure 4 shows cellular distribution of F/HN-SIV transgene expression. EGFP expression was determined in histological sections of the mouse nasal cavity (2 mm from tip of nose) at 30 days post-treatment. EGFP positive cells produce a white punctate signal.

Figure 5 shows cell types transduced by F/HN-SIV treatment of the mouse nose. 69% of the cells transduced in the mouse nasal cavity were ciliated respiratory epithelial cells. Other cell types transduced included neuronal cells in the olfactory epithelium (21%) and squamous cells (7%).

Figure 6 shows repeat administration of F/HN-SIV to the mouse nose. Mouse nasal tissue which was transduced (as Figure 1) with one dose of F/HN-SIV-CMV-Lux or, two doses of F/HN-SIV-CMV-EGFP followed by one dose of F/HN-SIV-CMV-Lux at 28 day intervals. Thus, repeat administration of F/HN-SIV to the mouse nose does not alter gene expression levels. Transgene expression is compared to a leading non-viral gene transfer formulation (CMV-Lux plasmid complexed with GL67A).

Figure 7 displays transduction of human air liquid interface (ALI) respiratory cell cultures. Human ALI cultures were transduced with F/HN-SIV-Lux at the indicated multiplicity of infection (MOI) and imaged for Lux expression at 5 days post-treatment.

Figure 8 demonstrates that F/HN-SIV can direct functional CFTR expression. HEK293T cells were transduced with F/HN-SIV-CMV-EGFP-CFTR at 500 MOI and CFTR functional activity was determined by iodide efflux. F/HN-SIV-CMV-EGFP served as a negative control.

Figure 9 exhibits that F/HN-SIV efficiently transduces sheep & human primary lung cells and mouse lung. Figure 9A shows that transduction of human nasal brushing cells (MOI 250) and human and sheep lung slices cultured *ex vivo* (1E7 TU/slice) with F/HN-SIV-CMV-Lux results in substantial luciferase transgene expression 24-48 hours post-transduction. Figures 9B and 9C display transduction of (~$1x10^5$) primary human CF lung cells cultured at the air-liquid interface (CF hALIs) with (3E7 TU) F/HN-SIV-

soCFTR2 vectors containing CMV- and hCEF transgene promoters. (B) Vector copy number (copies of pro-viral DNA per copy of endogenous CFTR DNA) at 6-8 days post-transduction. (C) CFTR mRNA expression level (%VE: copies of CFTR mRNA per copy of endogenous CFTR mRNA x 100) at 6-8 days post-transduction. The horizontal line in (C) represents target expression level of 5% VE - thought to represent the therapeutic threshold. Following *in vivo* delivery of F/HN-SIV-EGFPLux vectors containing CMV-, EF1aS and hCEF promoters in integrase defective (ID) or integrase competent form (IC or no label) airway cells transgene expression was determined in the nasal (D) and lung (E) murine epithelium (n=6-10/group). Time course of luciferase transgene expression was monitored by repeated in vivo bioluminescence imaging and as normalised to delivered dose. Figure 9F shows representative bioluminescence images following *in vivo* murine transduction at day 14 post transduction. Figure 9G portrays representative bioluminescence images following *in vitro* transduction of non-CF hALI at day 5-6 post transduction. Figure 9H represents EGFP expression at 14 days post transduction in the murine nasal epithelium following delivery of 1.6E8 TU of F/HN-SIV-hCEF-EGFPLux (vGM020). EGFP visualised by immunohistochemistry, nuclei visualised by DAPI. Figure 9I shows time-course of luciferase transgene expression in non-CF ALIs was monitored by repeated bioluminescence imaging and was normalised to delivered dose.

Figure 10 shows that F/HN-SIV efficiently transduces sheep lung i*n vivo.* Figure A shows that to model virus delivery to the sheep lung, we instilled (3x100µL aliquots over ~5 minutes) acriflavine to a proximal airway under direct bronchoscopic visualisation. The distribution of the acriflavine can be appreciated by the orange colouration of the dissected airway at postmortem. Note the acriflavine is largely restricted to the conducting airways and absent from the alveolar regions. Arrow indicates the approximate site of instillation. Numbers on ruler are cm. Figure 10B is a diagrammatic representation of the sheep lung (trachea centre/top). Green circle represents region in (A). In (B), line indicates passage of bronchoscope to deliver 3x100µL aliquots of 2.2E9 TU/mL (6.6E8 TU total) F/HN SIV CMV- EGFPLux to n=3 individual sheep (animal codes T121, T156 & T251). At seven days post-delivery, 5-6 tissue sample blocks were taken at post-mortem at ~1 cm intervals from the site of instillation. Blocks were divided into 2-3 approximately equivalent samples and analysed for transgene expression by (C) luciferase assays normalised to protein content; and (D) quantitative RT-PCR normalised to endogenous CFTR mRNA levels. Horizontal line in (C) represents highest

luciferase activity noted in any sample treated with a non-viral gene transfer vector, and (D) target expression level of 5% VE - thought to represent the therapeutic threshold.

Figure 11 depicts the production and purification of F/HN SIV Vectors. F/HN SIV Vectors were produced by 5 plasmid (pDNA) PEI-mediated transient transfection of 293T cells grown in suspension at 1L scale in pH controlled WAVE Bioreactors (GE), using scalable methods of the invention. Vectors were clarified by depth/end-filtration (GE/Pall), contaminating nucleic acids were removed with Benzonase® (Merck), vectors were activated with TrypLE Select™ (Life Technology), purified and concentrated by anion exchange membrane chromatography (Pall) and tangential-flow filtration (Spectrum). All process vessels, containers and columns were single-use cGMP compliant. All reagents except plasmid DNA were animal-free cGMP compliant. Data from a variety of vector configurations (transgene promoter, transgene, integrase status) are shown. Physical and functional titres were determined using Q-PCR. (A) Physical titre from initial clarified harvest and final purified product. Median process yield is ~44%. (B) Functional titre of final product. Median functional titre is ~$2x10^9$ TU/mL product and ~$3x10^9$ TU/L bioreactor volume. Target productivity and yields were exceeded. Lower yielding CFTR vectors utilise CMV transgene promoter. Higher yielding CFTR vectors utilise EF1aS and hCEF transgene promoters. (C) Final product particle: infectivity ratio is tightly clustered and similar to values from other high quality vector manufacturers (Oxford BioMedica, BlueBirdBio). Median particle: infectivity ratio is ~300. Product consistency has been achieved with the use of "Design of Experiments" methods supporting ultimate transition to QbD-based regulatory agency approval of manufacturing process. (D) Final product functional titre is strongly correlated with initial physical titre indicating non-saturating process conditions with no vector concentration limiting process steps - suggests purification scale-up will be efficient.

Figure 12 displays insertion site (IS) profiling and survival of transduced mice. Figures 12A-F provide a comparison of IS profiles from F/HN-SIV transduced mouse lung and VSV-G-HIV vector transduced mouse retina (Bartholomae *et al,* Mol Ther (2011) 19:703-710). IS profiles for (A,C,E) were derived from deep sequencing of lung DNA from two mice transduced with F/HN-SIV, and for (B,D,F) from retinal VSV-G-HIV IS sequences (Schmidt laboratory). (A,B) Aggregated IS sites (lung, 2862; retina, 262) plotted on karyograms generated via the Ensembl genome browser. (C,D) IS distances to transcription start sites (TSS). Numbers of IS in each distance bin are shown above bars. The sum exceeds the total number of IS analysed because a typical IS is located

near to several TSS. Graphs were generated by use of the UCSC genome browser and the GREAT genome analyser (great.stanford.edu). (E,F) QuickMap comparison between random (diamond) and observed (square) insertion frequencies per chromosome. (G) Survival of mice treated with progenitor F/HN-SIV vector (virus vector manufactured in accordance with known method using adherent cells: black line, 24 months of data) and current generation vector (GTC: dark grey line, 8 months of data) compared with buffer treated mice (light grey line, 24 months of data). Data aggregated from various experiments involving mice treated with buffer or ~1E7 TU virus by nasal sniffing.

Figure 13 shows hCEF mediated respiratory transgene expression – using lentivirus gene transfer with a CpG rich transgene – transgene expression (average radiance p/s/cm^2/sr) against days post dose. High levels of *Gaussia* luciferase reporter gene expression compared with the control (naïve lung) was observed in both the lung (square) and nose (circle) for at least 168 days after dosing.

Figure 14 shows the effect of transduction of human intestinal organoids with a CFTR lentiviral vector (vGM058) of the invention. The left-hand panel shows that forskolin induced swelling was significantly ($p<0.001$) reduced in vGM058 transduced organoids. In the right-hand panel, A549 cells were transduced with vGM058 or a control virus and CFTR function quantified using a radioactive iodide-efflux assay. Significant ($p<0.05$) levels of CFTR-mediated iodide efflux were detected in vGM058 transduced cells.

Figures 15A shows a schematic drawings of a plasmid used for production of the *A1AT* vectors of the invention. In one embodiment of the invention, Figure 15A provides a tool of the invention. Figure 15B shows a control plasmid encoding the *Gaussia* luciferase reporter gene.

Figure 16 shows that F/HN-SIV efficiently transduces human primary lung cells in ALI culture. In particular, transduction of human ALI cultures results in substantial luciferase transgene expression for at least 80 days post-transduction. Each point represents the mean value of RLU/µl in the media from n = 6 ALIs at the timepoint shown. Vertical bars represent the standard error of the mean.

Figure 17 shows gene expression following transduction of human lung slices (n=6 per group) with SIV1 hCEF-sogLux (Figure 17A) and SIV1 hCEF-sohAAT (Figure 17B). High levels of expression were observed. Each point represents the mean value of RLU/µl in the media from n = 6 lung slices at the timepoint shown. Vertical bars represent the standard error of the mean.

Figure 18 shows long term expression (>12 months) of *Gaussia* luciferase following lentiviral-mediated gene transfer (SIV1 hCEF-soGLux) *in vivo*. A: lung tissue homogenate; B: broncho-alveolar lavage (BAL) fluid; C: serum. Each point represents the mean value of RLU/ µl in one group of animals (n = 5 or 6 per group) harvested at the timepoint shown. Vertical bars represent standard error of the mean.

Figure 19 shows high levels of expression of A1AT expression following lentiviral-mediated gene transfer (SIV1 hCEF-sohAAT) *in vivo*. Each point represents one animal. Horizontal bars represent the median of each group.

Figure 21 shows the level of A1A1 in the epithelial lining fluid following lentiviral mediated gene transfer (SIV1 hCEF-sohAAT) *in vivo*. lentiviral mediated gene transfer (SIV1 hCEF-sohAAT) *in vivo*.

Figure 22A shows schematic drawings of *FVIII* cDNA constructs used for production of the FVIII vectors of the invention. Figure 22B shows viral vectors of the invention. In one embodiment of the invention, Figure 22B provides a tool of the invention. Figures 22C-F show schematic drawings of pDNA1 plasmids used for production of the *FVIII* vectors of the invention. In one embodiment of the invention, Figures 22C-F provide tools of the invention.

Figure 23 shows HEK293T transduction efficiency with vGM142 (SIV-F/HN-FVIII-N6-co) for batch 1 (A) and batch 2 (B). The graphs show FVIII activity for each MOI at 48 and 72 hours post-transduction. Each symbol represents an independent experiment (n=5-6 experiments). The horizontal bar indicates group mean +/- SD. Analysis was performed using One-way Anova (GraphPad Prism) with multiple comparisons to untreated control ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

Figure 24 shows the assessment of vGM142 in an *in vivo* system (murine model). A: lung; B: BAL fluid; C: plasma. The graphs show the level of hFVIII (as a percentage of the normal level) against the different treatments for Groups 1 to 3 (Groups 1 and 2 – 10 day treatment, 3 doses/week of 100µl vector per mouse (total amount of doses was equal 3 per animal); Group 3 – 28 day treatment, 3 doses/week of 100µl vector per mouse (total amount of doses was equal 12 per animal)). Each symbol represents an individual mouse (Group 1 (n=4) treated with total vector dose of $1.4x10^6$ TTU/mouse; Group 2 (n=3) treated with total vector dose of $1.57x10^8$ TTU/mouse and Group 3 (n=4) treated with total vector dose of $3.36x10^8$ TTU/mouse). The horizontal bar indicates mean FVIII:Ag levels +/- SD. Analysis was performed using One-way Anova (GraphPad Prism) with multiple comparisons between treated groups **** $p<0.0001$.

EXAMPLES

The invention is now described with reference to the Examples below. These are not limiting on the scope of the invention, and a person skilled in the art would be appreciate that suitable equivalents could be used within the scope of the present invention. Thus, the Examples may be considered component parts of the invention, and the individual aspects described therein may be considered as disclosed independently, or in any combination.

Example 1: Cell culture

HEK293T, Freestyle 293F (Life Technologies, Paisley, UK) and 293T/17 cells (CRL-11268; ATCC, Manassas, VA) were maintained in Dulbecco's minimal Eagle's medium (Invitrogen, Carlsbad, CA) containing 10% fetal bovine serum and supplemented with penicillin (100 U/ml) and streptomycin (100 µg/ml) or Freestyle™ 293 Expression Medium (Life Technologies).

Example 2: Plasmid construction

pCAGGS-Fct4 and pCAGGS-SIVct+HN were constructed as follows:

(i) Plasmid SIVct/HN contains the gene encoding the cytoplasmic tail of SIVagm TMP (reversed) fused to the ectodomain and transmembrane regions of SeV HN protein. Three oligonucleotide pairs were synthesized: pair 1, 5′-TCGAGATGTGGTCTGAGTTAAAAATCAGGAGCAACGACGGAGGTGAAGGACCAGA

CGCCAACGACCC-3′ (SEQ ID NO: 18) and 5′-CCGGGGGTCGTTGGCGTCTGGTCCTTCACCTCCGTCGTTGCTCCTGATTTTTAACTCAGACCACATC-3′ (SEQ ID NO: 19); pair 2, 5′-CCGGGGAAAGGGGGTGCAACACATCCATATCCAGCCATCTCTACCTGTTTATGGACAGA-3′ (SEQ ID NO: 20) and 5′-ACCCTCTGTCCATAAACAGGTAGAGATGGCTGGATATGGATGTGTTGCACCCCTTTCC-3′ (SEQ ID NO: 21); and pair 3, 5′-GGGTTAGGTGGTTGCTGATTCTCTCATTCACCCAGTGGG-3′ (SEQ ID NO: 22) and 5′-GATCCCCACTGGGTGAATGAGAGAATCAGCAACCACCTA-3′ (SEQ ID NO: 23). These oligonucleotide pairs were annealed and cloned into the *Xho*I and *Bam*HI sites of pBluescript KS+ (Stratagene) to yield pKS+SIVct. pCAGGS-SIVct/HN was constructed by cloning the 160-bp *Xho*I-*Dra*III fragment from pKS+SIVct and a 1.5-kbp *Dra*III-*Bsu*36I fragment from pCAGGS-HN, which carries the wild-type HN gene (HNwt), in the *Xho*I site of pCAGGS vector, into the *Xho*I and *Bsu*36I sites of pCAGGS. This plasmid was constructed so that the cytoplasmic tail of the HN protein was replaced with the cytoplasmic tail of SIVagm TMP.

For construction of pCAGGS-SIVct+HN, the genes encoding the cytoplasmic tail of SIVagm TMP and the N terminus of HN protein were first amplified by PCR from pCAGGS-SIVct/HN with the primer pair 5′-GAGACTCGAGATGTGGTCTGAGTTAAAAATCAGG-3′ (SEQ ID NO: 24) and 5′-AGAGGTAGACCAGTACGAGTCACGTTTGCCCCTATCACCATCCCTAACCCTCTGTCATAAAC-3′ (SEQ ID NO: 25).The resulting PCR fragment was cloned into the *Xho*I and *Acc*I sites of pKS+SIVct to generate pKS+SIVct-H. Then a *Xho*I-*Dra*III fragment from pKS+SIVct-H and a *Dra*III-*Bsu*36I fragment from pCAGGS-HN were cloned into the *Xho*I and *Bsu*36I sites of pCAGGS to yield pCAGGS-SIVct+HN.

The cPPT and WPRE sequences were inserted in the SIV-derived gene transfer plasmid. An example of the WPRE sequence used is provided in SEQ ID NO: 8.

The plasmid pGM101 contains the colE1 origin of replication, kanamycin resistance gene and promoter and was created by synthetic gene synthesis (GeneArt, Regensburg, Germany; now LifeTechnologies Ltd).

The hybrid CMV/SIV R U5 LTR, partial Gag, RRE, cPPT, SIN U3 and R sequences from pBS/CG2-Rc/s-CMV-D U (Nakajima *et al.* 2000 Human Gene Therapy 11:1863) were amplified by PCR and inserted along with the hCEF enhancer/Promoter sequence amplified by PCR from pGM169 (Hyde *et al.* Nature Biotechnology 26:549) and the soCFTR2 cDNA isolated from pGM169 on a NheI-ApaI restriction enzyme fragment into pGM101 to create pGM326.

The CMV enhancer/chicken beta actin promoter along with associated exon/intron sequences, SIV GagPol and RRE sequences and the SV40 polyA/origin of replication were amplified by PCR from pCAGGS/Sagm-gtr (Nakajima *et al.* 2000 Human Gene Therapy 11:1863) to create pGM297. The CMV enhancer/ promoter along with associated exon/intron sequences and SV40 polyA sequence from pCI (Promega, Madison, WI, USA) were isolated on a BglII-BamHI restriction enzyme fragment and the SIV Rev sequence derived from pCAGGS/Sagm-gtr amplified by PCR were inserted into pGM101 to create pGM299.

The CMV enhancer/chicken beta actin promoter along with associated exon/intron sequences, the Fct4 cDNA and SV40 polyA/origin from pCAGGS-Fct4 were isolated on a SalI-HindIII restriction enzyme fragment by a combination of gene synthesis, PCR and restriction enzyme fragment recombination and inserted into pGM101 to create pGM301.

The CMV enhancer/chicken beta actin promoter along with associated exon/intron sequences, the SIVct+HN cDNA and SV40 polyA/origin from pCAGGS-SIVct+HN were isolated on a SalI-HindIII restriction enzyme fragment by a combination of gene synthesis, PCR and restriction enzyme fragment recombination and inserted into pGM101 to create pGM303.

Other pGM plasmids of the invention were made using standard techniques and in accordance with the above disclosure.

Throughout these plasmid DNA assembly approaches, restriction enzymes and PCR polymerases were supplied by New England Biolabs (Ipswich, MA, USA) and DNA purification reagents were supplied by Qiagen (Limburg, Netherlands).

Example 3: Production of SIV vector

Four plasmid system:

Replication-defective self-inactivating SIV vector was constructed with minor modifications. Briefly, the SeV-F/HN-pseudotyped SIV vector was produced by transfecting 293T/17 cells (15 cm diameter culture dishes) with four plasmids complexed to Lipofectamine/Plus reagents (Invitrogen) according to the manufacturer's recommendations [Plasmid-1: 10 μg SIV-derived transfer plasmid carrying a GFP, a luciferase (lux) reporter gene, or a GFP-CFTR fusion construct, Plasmid-2: 3 μg packaging plasmid, Plasmid-3: 2 μg pCAGGS-Fct4, Plasmid 4: 2 μg pCAGGS-SIVct+HN; Figures 1A-E show schematic drawings of plasmids used for the production of vectors of the invention]. The VSV-G pseudotyped SIV vector was produced using a similar protocol, but a pVSV-G plasmid (2 μg; Clontech, Mountain View, CA) was used instead of pCAGGS-Fct4 and pCAGGS-SIVct+HN. At 12 hours after transfection the culture medium was replaced with 30 ml serum-free Dulbecco's modified Eagle medium containing 5 mmol/l sodium butyrate. Sodium butyrate stimulates the vector production to inhibit histone deacetylase. The culture supernatant containing the SIV vector was harvested 48 hours after transfection, filtered through a 0.45 μm filter membrane, and further concentrated by high-speed centrifugation (20,000 g for 4 hours at 4 °C, Avanti JA18 rotor; Beckman Coulter, Brea, CA). The vector pellets were suspended in PBS (Invitrogen) to 100- to 200-fold concentration and stored at −80 °C.

Five plasmid system (preferred):

SeV-F/HN-pseudotyped SIV vector was produced by transfecting HEK293T or 293T/17 cells cultured in FreeStyle™ 293 Expression Medium with a mixture of five plasmids with the following characteristics: pDNA1 (for example pGM326; Figure 1A) encodes the lentiviral vector mRNA; pDNA2a (for example pGM297; Figure 1B) encodes SIV Gag and Pol proteins; pDNA2b (for example pGM299: Figure 1C) encodes SIV Rev proteins; pDNA3a (for example pGM301; Figure 1D) encodes the Sendai virus-derived Fct4 protein [Kobayashi et al., 2003 J. Virol. 77:2607]; and pDNA3b (for example pGM303; Figure 1E) encodes the Sendai virus-derived SIVct+HN [Kobayashi et al., 2003 J. Virol. 77:2607] complexed with PEIpro (Polyplus, Illkirch, France).

Cell culture media was supplemented at 12-24 post-transfection with sodium butyrate. Sodium butyrate stimulates vector production via inhibiting histone deacetylase resulting

in increasing expression of the SIV and Sendai virus fusion protein components encoded by the five plasmids. Cell culture media was supplemented at 44-52 hours and/or 68-76 hours post-transfection with 5 units/mL Benzonase Nuclease (Merck Millipore, Nottingham, UK). The culture supernatant containing the SIV vector was harvested 68-76.5 hours after transfection, and clarified by filtration through a 0.45 µm membrane. The SIV vector is treated by digestion with a protease containing trypsin activity – for example an animal origin free, recombinant enzyme such as TrypLE Select™. Subsequently, SIV vector is typically further purified and concentrated by anion-exchange chromatography and/or tangential flow filtration ultra-filtration/dia-filtration.

This same method was used to generate lentiviral vectors comprising the *A1AT* and *FVIII* transgenes, with the plasmids of Figure 15A and Figure 22B replacing that of Figure 1A to provide the appropriate transgene (see Examples 15 and 20).

Example 4: Vector titration

Method 1:

The particle titre was determined using real-time reverse transcriptase-PCR. Virus RNA was purified using a QIAamp viral RNA mini-kit (QIAGEN, Strasse, Germany), and reverse transcribed using Superscript II (Invitrogen). The QuantiTect probe PCR system (QIAGEN) and primers for amplifying 131 nucleotides (bp) spanning the WPRE sequence (forward primer: 5′-ggatacgctgctttaatgcc-3′, reverse primer: 5′-acgccacgttgcctgacaac-3′) were used according to the manufacturer's protocol in an ABI PRISM 7700 Sequence Detector System (PE Applied Biosystems, Foster City, CA). SIV gene transfer plasmid DNA (3×10^4 to 2×10^6 molecules) was used as standard.

Transduction units (TU/ml) were determined by transducing 293T cells with serial dilutions of vector stock and quantification of transduced cells by GFP fluorescence (for F/HN-SIV-GFP and VSV-G-SIV-GFP) or staining with anti-luciferase antibody (for F/HN-SIV-lux).

Method 2 (preferred):

The particle titre (VP/mL) was typically determined using real-time reverse transcriptase-PCR. Virus RNA was purified using a QIAamp viral RNA mini-kit (QIAGEN, Strasse, Germany), and reverse transcribed using reverse transcriptase (Life Technologies).

TaqMan quantitative PCR system (Life Technologies) using primers amplifying a portion of the WPRE sequence in an ABI PRISM 7700 Sequence Detector System (Life Technologies). In vitro transcribed WPRE RNA molecules were used as quantitative standards.

Transduction units (TU/mL) were determined by transducing 293T/17 or Freestyle 293F cells with serial dilutions of SIV vector and quantification of WPRE containing provirus DNA by TaqMan quantitative PCR system (Life Technologies) using primers amplifying a portion of the WPRE sequence in an ABI PRISM 7700 Sequence Detector System (Life Technologies). Plasmid DNA molecules containing WPRE sequences were used as quantitative standards.

Example 5: Generation of basal cells-enriched tEC cultures

Murine tracheal epithelial cells (tEC) were isolated as follows. C57BL/6N Mice were culled and the tracheas were excised from the larynx to the main bronchial branches using sterile surgical instruments. The tissues were placed in a tube containing cold Ham's F-12 medium with 100 U/ml penicillin (P), 100 µg/ml streptomycin (S) and 2.5 mg/ml amphotericin B (A) (Ham's F12/PSA medium) and kept on ice. In a sterile tissue culture hood, the tracheas were cleaned from adherent muscles and connective tissue, cut longitudinally to expose the internal respiratory epithelium, and placed in 0.15% pronase solution in F-12 medium (~5 ml in 15 ml tube). Tissue digestion was performed overnight (15-18 hr) at 4 ºC. To block the enzymatic reaction, 10% fetal bovine serum (FBS) was added to the tissue digest. After gently inverting the tube to detach more cells, the tracheas were placed into a new tube containing 10% FBS/Ham's F-12/PS solution, and inverted as before. This step was repeated two more times. The content of the four tubes was pooled together and centrifuged at 500g for 10 min at 4°C. The pellet was re-suspended in DNase solution (0.5 mg/ml crude pancreatic DNase plus 10 mg/ml BSA in FBS/Ham's F-12/PS solution, about 200 µl/trachea), incubated on ice for 5 minutes, and centrifuged as before.

After removing the supernatant, tEC were resuspended in Progenitor Cell Targeted (PCT) medium (CnT-17, CELLnTEC, Bern, Switzerland), an antibiotics and antimycotics-free formulation specifically designed for human and mouse airways progenitor cells isolation and proliferation. tEC were then plated in a Primaria tissue culture dish (Becton Dickinson Labwere, Franklin Lakes, NJ, USA) and incubated for 3-4 hr in 5% CO_2 at 37

ºC. Non-adherent cells were collected and centrifuged at 500g for 5 min at 4°C and counted in a haemocytometer. To generate basal cells-enriched cultures, tEC were suspended in PCT medium and seeded on a Nunclon™Δ plate (Nunc A/S, Roskilde, Denmark), coated with 50 µg/ml type 1 rat tail collagen at a recommended seeding density of $4x10^3$ cells/cm^2. tEC were also cultured in a control basic medium, containing DMEM/Ham's F12 supplemented with L-glutamine (4 mM), HEPES (15 mM) and NaHCO3 (3.4 mM). Plates were incubated at 37°C with 5% CO_2. To determine the proportion of basal cells in the tEC population before and after expansion in PCT medium, cytospin preparations were stained with the anti-KRT5 antibody and the appropriate secondary antibody.

The pool of tEC produced comprised both fully differentiated cells (Clara and ciliated cells) and basal cells. Once seeded onto permeable membranes tEC were able to generate an air liquid interface (ALI) culture system as a result of basal cell proliferation and redifferentiation into secretory and ciliated cells. To establish an alternative and rapid protocol for basal cell expansion, two-dimensional (2D) tEC cultures using a commercially available proprietary Progenitor Cell Targeted (PCT) medium, specifically formulated to support the proliferation of airway progenitor cells while maintaining them in an undifferentiated status were assessed. As a negative control, tEC were exposed to a basic media formulation without addition of specific growth factors. tEC seeded on collagen-coated plastic surfaces ($4x10^3$ cells/cm^2) and exposed to PCT medium were able to grow rapidly and became confluent within 5-8 days whereas tEC exposed to the growth factor-deficient control medium were unable to adhere and propagate. To establish whether the use of PCT medium resulted in a substantial enrichment of the basal cell population, tEC were harvested at approximately 80% confluence (n=6 wells), fixed and treated with an anti-keratin 5 (Krt5) antibody, a specific marker of basal cells. Freshly isolated tEC were used as controls (n=3 unique preparations). The proportion of Krt5 positive basal cells after expansion in PCT medium was higher (78±1.4%) than in freshly isolated pools of tEC (33±0.6%), demonstrating that murine airway basal cells can be selectively and rapidly expanded from a mixed pool of tEC using a commercial medium.

Example 6: *Ex vivo* transduction of basal cells-enriched tEC cultures with F/HN-SIV-GFP

To determine whether the F/HN-SIV vector can effectively transduce basal cells ex vivo, tEC prepared in Example 5 were grown to approximately 70% confluence over 7 days in PCT medium and transduced with F/HN-SIV carrying a green fluorescent protein reporter gene (F/HN-SIV-GFP) at an MOI 100 and incubated at 37°C with 5% CO_2 for 3 days. tEC derived from wild-type and GFP transgenic animals were cultured under the same conditions and used as negative (no viral transduction) and positive control groups, respectively (n=3-6 wells/group).

To quantify the proportion of GFP-positive cells, basal cells-enriched tEC cultures were detached with the enzyme accutase (CELLnTEC), re-suspended in PBS/1%BSA and subjected to FACS analysis, counting an average of 20.277±2.478 cells/group. The F/HN-SIV vector transduced 26%±0.9% of basal cells-enriched tEC ($p<0.0001$ when compared to untransduced controls).

To assess whether transduced GFP-expressing cells were basal cells, three days post-infection cells were double stained with antibodies against Krt5 and GFP. Immunofluorescence staining of cultured cells showed that approximately 40% of Krt5-expressing cells also expressed the GFP reporter gene, showing that the F/HN-SIV vector can transduce progenitor basal cells *ex vivo*.

Example 7: *In vivo* administration to the mouse nose

C57BL/6N mice (female, 6–8 weeks) were used. Mice were anesthetized, placed horizontally on their backs onto a heated board, and a thin catheter (<0.5 mm outer diameter) was inserted ~2.5 mm from the tip of nose into the left nostril. Using a syringe pump (Cole-Parmer, Vernon Hills, IL), vector (100 µl) was then slowly perfused onto the nasal epithelium (1.3 µl/min) for 75 minutes. Despite perfusion of virus into the left nostril, we routinely observe transfection in both left and right nostrils, which is due to dispersion of the solutions throughout the entire nasal cavity. PBS and VSV-G-SIV transduced mice preconditioned with 1% lysophosphatidylcholine as described by Limberis *et al.*, 2002, were used as controls. At indicated time points (3–360 days after transduction), mice were culled to visualize GFP expression. As shown in Figure 2, GFP expression was observed for at least 449 days post-transduction, whereas the negative control showed no GFP expression. As shown in Figure 3, transgene expression with the F/NH-SIV vector was consistent, with observable GFP expression at least 360 days post-transduction in 10 independently tested mice.

Similarly, as shown in Figure 13, transduction with an F/HN-SIV vector of the invention comprising an hCEF promoter resulted in long-term expression of a CpG rich reporter gene (luciferase). High levels of expression relative to the control were observed in both the lung and the nose for at least 169 days post-transduction.

In the repeat administration experiments groups of mice were transduced with either one dose of F/HN-SIV-lux (single-dose group), or two doses of F/HN-SIV-GFP (day 0, day 28), followed by F/HN-SIV-lux on day 56 (repeat-dose group). Importantly, mice receiving F/HN-SIV-lux (single-dose group) and F/HN-SIV-lux on day 56 (repeat-dose group) were of similar age and were transduced at the same time. Gene expression was analysed 30 days after F/HN-SIV-lux administration. For comparison, mice were transfected with the cationic lipid GL67A complexed to a luciferase reporter gene as previously described (Griesenbach, U. *et al., Methods Mol Biol*. 2008; 433:229-42) and luciferase expression was measured 2 days after transfection.

As shown in Figure 6, repeat administration of F/HN-SIV to the mouse nose does not alter gene expression levels. Transgene expression is compared to a leading non-viral gene transfer formulation (CMV-Lux plasmid complexed with GL67A).

Insertion site profiling was conducted on transduced mice, and survival time investigated as set out in the description of Figure 12 above. Transduction using the F/HN-SIV vector of the invention was not observed to have any adverse effect on mouse survival compared with an existing F/HN-SIV vector or negative (buffer only) control (see Figure 12G).

Example 8: Induced regeneration of nasal epithelial cells by polidocanol treatment

Nasal epithelial cells were stripped by polidocanol treatment according to the method described (Borthwick *et al., Am J Respir Cell Mol Biol*. 2001 Jun; 24(6):662-70), with some modification. In brief, mice were anesthetized and 10 µl polidocanol (2%) was administered to the nose as a bolus by "nasal sniffing". To confirm the stripping and regeneration of nasal epithelial cells, nasal tissue was perfused with 10 µl of 2% (vol/vol in PBS) polidocanol (nonaethylene glycol mono-dodecyl ether; SIGMA, St Louis, MO) and histological analysis undertaken 24 hours and 7 days after treatment (n = 3/group).

To analyse transduction of possible progenitor or stem cells, we first administered F/HN-SIV-GFP (4×10^8 TU/mouse) vector to mouse nasal epithelium. Seven days after transduction, nasal tissue was perfused with 10 µl of 2% (vol/vol in PBS) polidocanol, and this treatment was repeated again 3 weeks later. Histological sections were analysed 58 days after vector administration (30 days after the last polidocanol treatment).

Example 9: Bioluminescent imaging

Mice were injected intraperitoneally with 150 mg/kg of D-luciferin (Xenogen, Alameda, CA) 10 minutes before imaging and were anesthetized with isoflurane. Bioluminescence (photons/s/cm2/sr) from living mice was measured using an IVIS50 system (Xenogen) at a binning of 4 for 10 minutes, using the software programme Living Image (Xenogen). For anatomical localization a pseudocolor image representing light intensity (blue: least intense, red: most intense) was generated using Living Image software and superimposed over the grayscale reference image. To quantify bioluminescence in the nose, photon emission in a defined area (red box) was measured by marking a standardized area for quantification. The size of the red box was kept constant and was placed over the heads of the animals as indicated in the figure. Importantly, the areas were marked using the grayscale reference image to avoid bias.

Example 10: Tissue preparation for histological assessment of GFP expression and/or basal cell detection

Mice were culled and the skin was removed. The head was cut at eye level and skin, jaw, tongue, and the soft tip of the nose were carefully removed. For in situ imaging of GFP expression in the nasal cavity, GFP fluorescence was detected using fluorescence stereoscopic microscopy (Leica, Ernst Leitz Optische Werke, Germany). Subsequently, the tissue was fixed in 4% paraformaldehyde (pH 7.4) overnight at room temperature and was then submerged in 20% EDTA (pH 7.5 for 5 days) for decalcification. The EDTA solution was changed at least every second day. After decalcification, the tissue was incubated in 15% sucrose overnight at room temperature and was then embedded in Tissue Mount (Chiba Medical, Soka, Japan). Ten micrometer sections were cut at six different positions in each mouse head (~0–6 mm from the tip of nasal bone). GFP expression was observed using a fluorescent microscope (Leica). Quantification and identification of cell types were carried out on six levels per mouse using a ×40 or ×63 objective. Prolonged image exposure was necessary to capture the structure of the

nasal epithelium using fluorescent microscopy. This led to pixel saturation of GFP-positive cells and caused GFP-positive cells to appear almost white rather than the common green appearance that we, and others, observe under higher magnification.

Cellular distribution of F/NH-SIV transgene expression was investigated in histological sections. Specifically, EGFP expression was determined in histological sections of the mouse nasal cavity (2mm from the tip of the nose) at 30 days post treatment. Figure 4 shows the location of EGFP expression (Figure 4, white punctate signal).

Figure 5 shows the cell types transduced by /NH-SIV transgene treatment of the mouse nose. 69% of the cells transduced in the mouse nasal cavity were ciliated respiratory epithelial cells. Other transduced cell types included neuronal cells in the olfactory epithelium (21%) and squamous cells (7%).

To detect basal cells following polidocanol treatment horseradish peroxidase (HRP)-based immunostaining was performed using the Envision kit (Dako, Glostrup, Denmark). Briefly, slides were treated with 0.6% hydrogen peroxide in methanol for 15 min, washed in tap water and incubated with 1.5% normal goat serum (Abcam) for 30 min. Slides were then incubated with a rabbit polyclonal anti-Cytokeratin 5 antibody (1:500) (Abcam) for 1 hr following a Goat anti-rabbit IgG conjugated to HRP (provided with the kit) for 30 min. Sections were then washed in PBS and incubated with the peroxidase substrate 3-amino-9-ethylcarbazole (AEC) (provided with the kit) for 5 min. Finally, slides were washed in distilled H_2O, counterstained with aqueous Harris' hematoxylin (BDH) for 15 seconds, washed in tap water, and then in distilled H_2O.

Immunofluorescence detection of GFP-positive transduced nasal epithelial cells and Krt5 positive basal cells was performed using the following primary and secondary antibodies: goat polyclonal anti-GFP antibody (1:250) (Abcam), rabbit monoclonal anti-KRT5 antibody (1:500) (Abcam), Alexa Fluor 488 donkey anti-Goat IgG (1:200) (Invitrogen, Paisley, UK) and Alexa Fluor 594 goat anti-rabbit IgG (1:200) (Invitrogen). To improve antibody performance, sections were subjected to heat-mediated antigen retrieval in citrate buffer (10 mM citric acid, 0.05% Tween20, pH 6.0) for 20 min on a water bath at 100 ºC. Stained sections were mounted in ProLong® Gold Antifade Reagent with DAPI (Invitrogen) and analysed with a confocal microscope as before (all Zeiss). GFP-positive basal cells (identification based on morphology and location within

the epithelial layer) were quantified on a total of 13 sections/mouse. Sections that displayed putative GFP positive basal cells were selected for double staining with the anti-KRT5 and anti-GFP antibodies to confirm the basal cell phenotype.

Example 11: Transduction of ALI cultures

Fully differentiated airway epithelial cells grown as ALI cultures were purchased from Epithelix (Geneva, Switzerland). ALIs were transfected with F/HN-SIV-lux at a multiplicity of infection ranging from ~25 to ~300 TU/cell. After 6 hours, the virus was removed and ALIs were incubated for 10–26 days. The basolateral medium was changed every 48 hours during this incubation period. At specified time points, the ALIs were lysed in 100 µl reporter lysis buffer and luciferase expression was quantified using the Luciferase Assay System (Promega, Southampton, UK) according to the manufacturer's instructions. The total protein content of the cultures was quantified using the BioRad protein assay kit (BioRad, Hemel Hempstead, UK). Each sample was assayed in duplicate. Luciferase expression was then presented as relative light units/mg total protein. For bioluminescence imaging 100 µg luciferin in PBS were added to the apical membrane.

As shown in Figure 7, cells in ALI cultures were successfully transduced with F/HN-SIV-Lux, as evidenced by luciferase expression. Luciferase expression was greater at MOI 250 compared with MOI 25.

Example 12: Iodide efflux assay

HEK293T cells were transfected with F/HN-SIV-GFP-CFTR or an F/HN-SIV-GFP control virus at a multiplicity of infection of 500 TU/cell and cultured for 2 days. CFTR chloride channel activity was assayed by measuring the rate of ^{125}iodide efflux as previously described (Derand, R., *et al.*, 2003). The ^{125}iodide efflux rates were normalized to the time of forskolin/IBMX addition (time 0). Curves were constructed by plotting rates of ^{125}iodide efflux against time. To reflect the cumulative levels of ^{125}iodide efflux following agonist-stimulation, all comparisons are based on areas under the time-^{125}iodide efflux curves. The area under the curve was calculated by the trapezium rule. Experiments were carried out in duplicate (n = 6 wells/group/experiment).

As shown in Figure 8, F/HN-SIV can direct functional CFTR expression, with a relative ^{125}iodide efflux of 0.65. In contrast, the GL67A plasmid vector achieved a lower ^{125}iodide efflux value of 0.33.

Example 13: Transduction of sheep and human primary lung cells and mouse lung

F/HN-SIV efficiently transduces sheep & human primary lung cells and mouse lung.

F/HN-SIV-CMV-Lux was used to transduce human nasal brushings (MOI 250) and human and sheep lung slices cultured *ex vivo* ($1x10^7$ TU/slice). As shown in Figure 9A, transduction of both the human nasal brushing cells and human and sheep lung slices resulted in substantial luciferase transgene expression (average values in the region of $2x10^2$ RLU/mg protein for the human nasal brushings, $1x10^7$ RLU/mg protein for the human lung slices and 2×10^7 RLU/mg protein for the sheep lung slices) 24-48 hours post-transduction.

Primary human CF lung cells cultured at the air-liquid interface (CF hALIs, ~$1x10^5$) were transduced with (3×10^7 TU) F/HN-SIV-soCFTR2 vectors containing CMV- and hCEF transgene promoters. Vector copy number (copies of pro-viral DNA per copy of endogenous CFTR DNA) was measured at 6-8 days post-transduction. Both the CMW and hCEF promoters were able to achieve a vector copy number of at least 1×10^1 (Figure 9B).

CFTR mRNA expression level (%VE: copies of CFTR mRNA per copy of endogenous CFTR mRNA x 100) at 6-8 days post-transduction was also measured. The horizontal dotted line in Figure 9C represents a target expression level of 5% VE, which is thought to represent the therapeutic threshold. Both the F/HN-SIV-soCFTR-CMV and F/HN-SIV-soCFTR2-hCEF induced expression significantly above this target (in the region of 40583±10687 and 18509±13588 respectively, mean ± SD, n=4).

Following *in vivo* delivery of F/HN-SIV-EGFPLux vectors containing CMV, EF1a and hCEF promoters in integrase defective (ID) or integrase competent form (IC or no label) airway cells transgene expression was determined in the nasal (Figure 9D) and lung (Figure 9E) murine epithelium (n=6-10/group). The time course of luciferase transgene expression was monitored by repeated *in vivo* bioluminescence imaging and was normalised to delivered dose. Four of the five vectors tested (vGM012 CMW, vGM014

EF1a, vGM020 hCEF and vGM076 hCEF ID) achieved expression in the nose above the target level for the whole time course of the experiment. The fifth vector, vGM074 CMV ID, achieved expression in the nose above the accepted expression level for the whole time course of the experiment.

Two of the five vectors tested (vGM014 EF1a and vGM020 hCEF) achieved expression in the lung above the target level for the whole time course of the experiment. One vector, vGM012 CMW, achieved expression in the lung above the accepted expression level for the whole time course of the experiment.

Bioluminescence was detected following *in vivo* murine transduction at day 14 post transduction. Representative images are shown in Figure 9F. The vGM020 hCEF vector achieved the highest level of *in vivo* expression out of the five vectors tested.

Bioluminescence was also detected following *in vitro* transduction of non-CF hALI at day 5-6 post transduction. Representative images are shown Figure 9G. Again, the vGM020 hCEF vector achieved the highest level of expression out of the five vectors tested.

Figure 9H shows EGFP expression at 14 days post transduction in the murine nasal epithelium following delivery of 1.6×10^{8} TU of F/HN-SIV-hCEF-EGFPLux (vGM020), as visualised by immunohistochemistry (nuclei stained with DAPI).

The time-course of luciferase transgene expression in non-CF ALIs was monitored by repeated bioluminescence imaging and was normalised to the delivered dose. As shown in Figure 9I, the vGM014 EF1a and vGM020 hCEF vectors achieved the highest level of expression.

F/HN-SIV also efficiently transduces sheep lung i*n vivo.* Acriflavine was instilled (3x100µL aliquots over ~5 minutes) to a proximal airway under direct bronchoscopic visualisation. The distribution of the acriflavine can be appreciated by the orange colouration of the dissected airway at postmortem (Figure 10A).

The acriflavine was largely restricted to the conducting airways and absent from the alveolar regions. The arrow in Figure 10A indicates the approximate site of instillation. The numbers on the ruler are in cm.

Figure 10B is a diagrammatic representation of the sheep lung (trachea centre/top). The circle represents the region in (Figure 10A). In Figure 10B, the arrow indicates passage of bronchoscope to deliver 3x100µL aliquots of 2.2E9 TU/mL (6.6E8 TU total) F/HN SIV CMV- EGFPLux to n=3 individual sheep (animal codes T121, T156 & T251). At seven days post-delivery, 5-6 tissue sample blocks were taken at post-mortem at ~1 cm intervals from the site of instillation.

The sample blocks were divided into 2-3 approximately equivalent samples and analysed for transgene expression, the results of which are shown in Figure 10C as luciferase assays normalised to protein content; and In Figure 10D as quantitative RT-PCR normalised to endogenous CFTR mRNA levels. The horizontal line in Figure 10C represents the highest luciferase activity noted in any sample treated with a non-viral gene transfer vector, and in Figure 10D the target expression level of 5% VE (thought to represent the therapeutic threshold, see above). For each treatment group the average was higher than the non-viral vector comparison (Figure 10C), and also achieved expression above the target 5% VE threshold.

Example 14: CFTR expression and function measured using human CF intestinal organoids

Human CF intestinal organoids were generated as described by Dekkers JF *et al*, (Nature Medicine 2013, 19(7): 939-945). Briefly, intestinal biopsies were washed in EDTA containing solutions to dissociate crypt cells. Crypt cells were then transduced with vGM058 (approximately $1x10^7$ transduction units) or a control virus (n=3 wells/condition) and embedded in Matrigel and allowed to form organoids for 3-4 days. CFTR function was assessed by exposing the organoids to forskolin (approximately 5 µM forskolin) which increases intracellular cAMP levels and thereby activates CFTR. In response the CFTR activation the organoids increase chloride transport which leads to water uptake and swelling. Organoids (minimum 10/well) were directly analysed by confocal live-cell microscopy (LSM710, Zeiss, ×5 objective). Forskolin-stimulated organoid swelling was automatically quantified using Volocity imaging software (Improvision). The total organoid area (xy plane) increase relative to that at t = 0 of forskolin treatment was calculated and averaged. Forskolin induced swelling was significantly ($p<0.001$) increased in vGM058 transduced organoids compared to controls

(see Figure 14A). Significant levels of CFTR-mediated iodide efflux ($p<0.05$) were also detected in vGM058 transduced cells (Figure 14B) using the iodide efflux assay disclosed herein (see Example 12).

Example 15: Generation of lentiviral vectors for A1AT

Lentiviral vectors were prepared using the SIV backbone and 5 plasmid method described above in Examples 2 and 3 (for the CFTR lentivirus) and using the hCEF promoter as described herein. Two separate lentiviral constructs were generated: one with a human alpha-1-antitrypsin (hAAT) transgene; one with a *Gaussia luciferase* (Glux) transgene (see Figure 15A and B, SEQ ID NOs: 9 and 10 respectively). The cDNAs contained within these vectors were codon-optimised and CpG-depleted.

Example 16: Air-liquid interface (ALI) culture using lux reporter gene lentiviral vector

Fully differentiated wild-type human ALI cultures (MucilAir) were purchased from Epithelix SARL (Geneva, CH). ALIs were cultured at 37°C and 5% CO_2 and the basolateral culture medium changed every 2-3 days. The culture medium was stored at -20°C until further analysis.

ALIs were transduced (on day 0) by pipetting 100µl of virus ($1x10^7$ Tagman transfection units (TTU) onto the apical surface. The virus was removed after 4 hours incubation at 37°C, and the basolateral medium replaced.

At indicated timepoints post-transduction, the apical surface of the ALIs was washed by incubating with sterile PBS for one hour. The washings were removed and stored at -20°C until further analysis.

A *Gaussia* luciferase assay (New England Biolabs, Ipswich, USA) was performed according to manufacturer's recommendations. 15µl of sample was analysed in duplicate, and luminescence determined in an Appliskan plate reader. Glux expression was expressed as RLU/µl fluid (RUL = relative light units).

Figure 16 provides the results, with each point representing the mean value of RLU/µl in the media from n=6 ALIs at the time point shown (standard error indicated by error bars).

Lentiviral-mediated gene transfer in human air-liquid interfaces resulted in the long-term expression of secreted reporter protein *Gaussia* luciferase.

Example 17: Transgene expression in lung slice cultures using A1AT and lux reporter gene lentiviral vectors

Precision-cut human lung slices were prepared as described in Moreno L *et al*, Respir Res 2006 Aug 21;7:111. Lung slices were placed in 12-well tissue culture plates (1 slice per well) in 1ml of media and incubated at 37°C and 5% CO_2. The media was changed daily and stored at -20°C until further analysis.

On day 0 lung slices (n=6 per group) were transduced with SIV hCEF-sogLux (1 x 10^6 TTU) or SIV1 hCEF-sohAAT (2 x 10^6 TTU) virus diluted in medium to a final volume of 1000µl and incubated for 4 hours. After the incubation, medium was replaced and stored at -20°C until further analysis.

Gaussia luciferase expression was determined as described above (Example 16). As shown in Figure 17A, high levels of expression of secreted reporter protein *Gaussia* luciferase followed lentiviral-mediated gene transfer in human lung slices. AAT (also referred to herein as A1AT) expression was determined using a sandwich ELISA (Abcam, Cambridge, UK), performed according to the manufacturer's recommendations. 50µl of sample was assayed in duplicate, and measured on a microplate reader at 450nm. As shown in Figure 17B, high levels of expression of alpha-1-antitrypsin (AAT/A1AT) followed lentiviral-mediated gene transfer in human lung slices.

Example 18: *In vivo* administration of A1AT and lux reporter gene lentiviral vectors to the mouse nose

Mouse lung transduction:

Female C57BL/6 mice (Charles River, UK) were anaesthetised with isoflurane and given 100ul of virus by nasal instillation as described in Xenariou S *et al*, Gene Ther 2007 May;14(9): 768-75. Animals were given between 1 and 5 doses and observed daily for signs of toxicity.

For *Gaussia* luciferase, female C57BL/6 mice were anaesthetised on day 0 using isoflurane and given a single 100ul dose of the SIV1 hCEF-soGLux virus ($1x10^6$ TTU) by nasal instillation . Control animals were treated with DMEM (tissue culture medium), the main constituent of the viral preparation used in the study.

For A1AT, female C57bl/6 mice (n = 5 per group) were treated with 3 doses of SIV1 hCEF-sohAAT at 10-day intervals (100µl per dose, $6.8x10^7$ TTU; total dose $2.4x10^8$ TTU). Control animals were instilled with 100ul of sterilised PBS (the main constituent of the lentivirus production batch used in the study) at each dosing point.

10 days after the third dose, animals were sacrificed and lung tissue homogenate, broncho-alveolar lavage fluid and serum analysed for AAT expression.

In addition, long term expression of A1AT was investigated. On days 1 to 5 of the experiment, C57bl/6 mice were treated with 100µl of SIV1 hCEF-sohAAT by nasal instillation (5 doses of 4×10^5 TTU, i.e. $2x10e^6$ TTU per animal in total). Control animals were instilled with 100ul of DMEM (tissue culture medium), the main constituent of the lentivirus production batch used in the study. Animals were sacrificed at various timepoints post-transduction and lung tissue homogenate, broncho-alveolar lavage fluid and serum were analysed for AAT expression.

Mouse tissue collection:

Mice were sacrificed at the indicated time-points post transduction. Blood was collected by puncturing the left ventricle, and centrifuged at $760g_{av}$ for 10 minutes to prepare serum. Serum was subsequently frozen at -80°C.

A bronco-alveolar lavage (BAL) was performed by dissecting the neck, inserting a cannula into the trachea and securing it in place with suture thread. 500 µl of PBS was instilled into the lung, and aspirated three times to obtain thorough washing of the epithelial lining. The sample was immediately snap frozen in liquid nitrogen, and stored at -80°C for further analysis.

Lungs were then dissected and snap-frozen in liquid nitrogen, and subsequently homogenised in lysing matrix D tubes (MP Biomedicals), centrifuged in a FastPrep

machine (ThermoFisher Scientific, Waltham, MA, USA) at 4m/s for 45 seconds, and stored at -80 °C for further analysis.

AAT (A1AT) and *Gaussia luciferase* (Glux) expression was determined as described in Examples 16 and 17 above. *In vivo* transduction of mouse airway cells with a single dose of the lux reporter gene lentiviral vector of Example 15 resulted in long-term expression (at least 12 months) of the secreted reporter protein *Gaussia* luciferase, in lung homogenate (Figure 18A), bronco-alveolar lavage fluid (BAL, Figure 18B) and serum (Figure 18C).

High levels of expression of A1AT were observed in lung homogenate, BAL and serum following lentiviral-mediated transfer of the AAT (A1AT) gene *in vivo* (Figure 19), with over a 100-fold increase in ATT (A1AT) expression in the lung homogenate and BAL observed compared with the corresponding negative (PBS) controls. A significant increase (at least one order of magnitude) in ATT (A1AT) expression was also observed in the serum.

In addition, long-term expression (at least 90 days) of alpha-1-antitrypsin was observed in lung homogenate (Figure 20A), BAL (Figure 20B) and serum (Figure 20C) following lentiviral-mediated gene transfer of the AAT (A1AT) gene *in vivo*.

Example 19: Urea assay

C57bl/6 mice (n = 5 per group) were treated with 3 doses of SIV1 hCEF-sohAAT at 10-day intervals (100µl per dose, $6.8x10^7$ TTU; total dose $2.4x10^8$ TTU). Control animals were instilled with 100ul of sterilised PBS (the main constituent of the lentivirus production batch used in the study) at each dosing point.

10 days after the third dose, animals were sacrificed and lung tissue homogenate, broncho-alveloar lavage fluid and serum analysed for A1AT expression.

A urea assay (Abcam, Cambridge, UK) was performed according to the manufacturer's instructions.

Firstly, serial dilutions of murine serum and BAL fluid samples were prepared and analysed to determine the appropriate dilution to use in further experiments.

Secondly, corresponding serum and BAL fluid samples from single mice (n =14) were analysed to calculate the fold-difference between urea concentration in serum and BAL fluid, equivalent to the dilutional effect of BAL on epithelial lining fluid (as per Rennard SI *et al*, J Appl Physiol (1985). 1986 Feb;60(2):532-8). The mean dilution of BAL was 41-fold (range 24-88).

Taking into account this dilutional effect, the concentration of ATT (A1AT) in the epithelial lining fluid was calculated. Specifically, the concentration of AAT in the broncho-alveolar lavage fluid was multiplied by the dilution factor, to provide an estimate of the 'true' AAT concentration in epithelial lining fluid.

A "protective" target level of ATT (A1AT) in the epithelial lining fluid (ELF, i.e. the fluid lining the airways and airspaces in the lungs) is 70µg/ml (compared with a "normal" target level of ATT (A1AT) in the ELF of 200µg/ml). As shown in Figure 21, therapeutic levels of alpha-1-antitrypsin in epithelial lining fluid followed ATT (A1AT) lentiviral-mediated gene transfer *in vivo*.

Example 20: Generation of lentiviral vectors for FVIII

Four different FVIII lentiviral vectors were prepared using the SIV backbone and 5 plasmid method described above in Examples 2 and 3 (for the CFTR lentivirus) and 15 (for the A1AT lentivirus). The promoter-transgene plasmids have SEQ ID NOs: 11 to 14 respectively.

The SIV sequence was identical to the CFTR constructs (Examples 2 and 3) except for the promoter and cDNA. The human cytomegalovirus promotor (CMV) or tissue specific hCEFI promotor/enhancer was used as indicated (Figure 22) to drive expression of FVIII transgenes.

SIV-F/HN-FVIII-N6-co contained the wild type human FVIII cDNA from which the BDD domain has been deleted and replaced with codon optimised 226 amino acid 6N-glycosylation fragment.

SIV-FVIII-V3 contains the wild type human FVIII cDNA from which the 226 amino acid glycosylation site has been deleted and replaced with 17 amino acid peptide which expresses 6N-glycosylation triples within the B domain (McIntosh et al., Blood 2013 121(17); 3335-3344).

Example 21: Quantification of hFVIII antigen and activity levels in *in vivo* and *in vitro* models

Human FVIII antigen levels in a murine model were quantified by enzyme-linked immunosorbent assay (ELISA) according to the manufacturer's protocol. Briefly, plasma, BAL and lung were analysed for the presence of FVIII antigen using Asserachrom (FVIII:Ag) Elisa (Stago Diagnostics, France).

Samples were diluted 1:2 and incubated on a mouse monoclonal anti-human factor VIII fragment-coated 96-well plate for 2 hours at room temperature. Following washing, anti-mouse secondary antibody coupled with peroxidase was added to the plate and incubation was carried out for 2 hours at RT. hFVIII:Ag levels were determined spectrophotometric at 450 nm using TMB substrate (data not shown).

Another ELISA assay was used to evaluate FVIII activity in an *in vitro* HEK293T model (FVIII:C, Affinity Biological, Canada). Supernatants were collected 48 and 72 hours after HEK293T transduction with SIV-F/HN-FVIII-N6 or SIV-F/HN-FVIII-N3. FVIII activity was evaluated by following the manufacturer's instructions using 50µl supernatants assayed in duplicate. As a negative control the supernatant from untreated HEK293T cells was tested. hFVIII activity was calculated from a standard curve generated using a series of dilutions of normal human pooled plasma (13th British Standard for blood coagulation Factor VIII concentrate, Human; NIBSC).

HEK293T cells were transduced with two different batches of vGM142 (Batch 1 - $5.9x10^8$ TTU/ml and Batch 2 – $2.8x10^8$ TTU/ml). HEK293T cells were transduced with vGM142 Batch 1 vector (Figure 23A) and vGM142 Batch 2 vector (Figure 23B) at 3 different MOIs (MOI 1; 10; 100), and collected 48 and 72 hours post-transduction.

As is clear from Figure 23, increasing FVIII activity was observed with increasing MOI for both Batch 1 and Batch 2 of vGM142 at both 48 and 72 hours post-transduction. Furthermore, FVIII activity increased from 48 hours to 72 hours for each MOI tested.

Example 22: *In vivo* administration of FVIII and lux reporter gene lentiviral vectors to the mouse nose

Mouse lung transduction:

All animal procedures were performed in accordance with the conditions and limitation of the UK Home Office Project and Personal licence regulations under the Animal Scientific Procedure Act (1986).

Wild type C57BL/6 female mice aged 6-8 weeks old (Charles River, UK were anaesthetised using isofluorane and given 100µl of virus in Dulbecco's phosphate-buffered saline (D-PBS), as described previously (Griesenbach *et al.*, 2012) and the presence of FVIII antigen was assessed.

In two experiments (Group 1 and 2) mice received 3 doses (every other day) of SIV-F/HN-FVIII-N6 (vGM142) and were culled 10 days after the first dose. Group 1 (n=4) were treated with a total vector dose of $1.4x10^6$ TTU/mouse. Group 2 (n=3) were treated with a total vector dose of $1.57x10^8$ TTU/mouse

In one experiment (Group 3) mice were treated with 12 doses (every other day) of SIV-F/HN-FVIII-N6 (vGM142) and culled 28 days after the first dose. Group 3 (n=4) were treated with a total vector dose of $3.36x10^8$ TTU/mouse)

Plasma, BAL fluid and Lung were collected (as described in Example 14). Briefly, the mice were sacrificed at the indicated time-points post transduction. Blood was then collected from heart into the 3.2 trisodium citrate anticoagulant collection tubes, before being centrifuged at 2000 – 2500 x g to obtain plasma. BAL fluid was collected by applying 3 consecutive installations of PBS (500µl) into mouse lung at room temperature. Supernatants were stored at – 80°C. Lungs were collected and stored at – 80°C prior to tissue homogenisation.

The presence of FVIII expression was then assessed. FVIII levels were assessed in lung tissue homogenates (Figure 24A), BAL fluid (Figure 24B) and plasma (Figure 24C) collected separately in 3 independent experiments at 10 and/or 28 days post SIV-F/HN-FVIII-N6 treatment. Analysis was performed using One-way Anova (GraphPad Prism) with multiple comparisons between treated groups (**** $p<0.0001$).

As is clear from Figure 24A, all three treatment groups produced an observable increase in hFVIII levels within the lung tissue compared with the corresponding control (D-PBS). The 28 day treatment of Group 3 resulted in a significant increase in hFVIII expression compared with the 10 day treatments of Groups 1 and 2. Similar results were observed for the BAL fluid samples (Figure 24B), although in these samples there was also a significant increase in hFVIII levels in Group 2 compared with Group 1. Group 3 treatment resulted in a significant increase in hFVIII levels in plasma (Figure 24C).

Example 23: Manufacturing method for lentiviral vectors in accordance with the invention

HEK293 cells are grown in suspension, in Freestyle Expression Media (chemically defined, animal & protein-free), and the cell count is monitored. Glucose concentration is determined and titrated to ~35mM. A transfection mixture of pDNA/PEIPro™ is prepared and the cells are transfected at 0.33mg pDNA/1E9 cells.

Cell count is monitored again and further Freestyle Expression Media is added. Glucose concentration is again determined and titrated to ~35mM. 5u/mL of Benzonase® can be added and three-stage inline virus clarification is carried out. Benzonase® is added followed by TrypLE Select™. The virus is cooled to 0°C and kept on wet ice for all subsequent steps. After filtering any non-virus particulate matter (mPES 0.45µm filter), the virus is loaded onto Mustang® Q XT (3mL membrane / L clarified virus) followed by washing with 0.15M NaCl Tris pH7.5 and elution with 1.0M NaCl Tris pH7.5. The virus fraction is collected and diluted to 0.1-0.2 initial volume with Freestyle media. TrypLE Select™ can be added here if not added above and Benzonase® can be also be added at this stage in addition to or instead of above.

Spectrum Tangential flow filtration (TFF) is carried out (UF to ~0.1-0.05 initial volume = HV; DF retentate x5 HV against formulation buffer; UF to ~0.001 - 0.002 initial volume) and retentate is collected. A second TFF step may be carried out and a smaller TFF unit

for DF and/or final UF can be used. Additional steps can include mixed-mode/SEC and 0.45µm or 0.2µm sterile filtration.

Figure 11 depicts the production and purification of F/HN SIV Vectors. F/HN SIV Vectors were produced by 5 plasmid (pDNA) PEI-mediated transient transfection of 293T cells grown in suspension at 1L scale in pH controlled WAVE Bioreactors (GE), using scalable methods of the invention. Vectors were clarified by depth/end-filtration (GE/Pall), contaminating nucleic acids were removed with Benzonase® (Merck), vectors were activated with TrypLE Select™ (Life Technology), purified and concentrated by anion exchange membrane chromatography (Pall) and tangential-flow filtration (Spectrum). All process vessels, containers and columns were single-use cGMP compliant. All reagents except plasmid DNA were animal-free cGMP compliant. Data from a variety of vector configurations (transgene promoter, transgene, integrase status) are shown. Physical and functional titres were determined using Q-PCR.

See the results of this exemplary method of the invention are discussed in the description of Figure 11 above.

Key to SEQ ID NOs

SEQ ID NO: 1	Plasmid as defined in Figure 1A (pDNA1 pGM326)
SEQ ID NO: 2	Plasmid as defined in Figure 1B (pDNA2a pGM297)
SEQ ID NO: 3	Plasmid as defined in Figure 1C (pDNA2b pGM299)
SEQ ID NO: 4	Plasmid as defined in Figure 1D (pDNA3a pGM301)
SEQ ID NO: 5	Plasmid as defined in Figure 1E (pDNA3b pGM303)
SEQ ID NO: 6	Exemplified hCEF promoter
SEQ ID NO: 7	Exemplified CFTR transgene (*soCFTR2*)
SEQ ID NO: 8	Exemplified WPRE component (mWPRE)
SEQ ID NO: 9	F/HN-SIV-hCEF-soA1AT plasmid as defined in Figure 15 (pDNA1 pGM407)
SEQ ID NO: 10	F/HN-SIV-hCEF-sogLux plasmid as defined in Figure 15 (pDNA1 pGM358)
SEQ ID NO: 11	F/HN-SIV-CMV-HFVIII-V3 plasmid as defined in Figure 22C (pDNA1 pGM411)
SEQ ID NO: 12	F/HN-SIV-hCEF-HFVIII-V3 plasmid as defined in Figure 22D (pDNA1 pGM413)
SEQ ID NO: 13	F/HN-SIV-CMV-HFVIII-N6-co plasmid as defined in Figure 22E (pDNA1 pGM412)
SEQ ID NO: 14	F/HN-SIV-hCEF-HFVIII-N6-co plasmid as defined in Figure 22F (pDNA1 pGM414)

SEQ ID NO: 15 Exemplified A1AT transgene

SEQ ID NO: 16 Exemplified FVIII transgene (N6)

SEQ ID NO: 17 Exemplified CMV promoter

SEQ ID NO: 18 Primer for the construction of pCAGGS-Fct4

SEQ ID NO: 19 Primer for the construction of pCAGGS-Fct4

SEQ ID NO: 20 Primer for the construction of pCAGGS-Fct4

SEQ ID NO: 21 Primer for the construction of pCAGGS-Fct4

SEQ ID NO: 22 Primer for the construction of pCAGGS-Fct4

SEQ ID NO: 23 Primer for the construction of pCAGGS-Fct4

SEQ ID NO: 24 Primer for the construction of pCAGGS-SIVct+HN

SEQ ID NO: 25 Primer for the construction of pCAGGS-SIVct+HN

SEQ ID NO: 26 Complementary strand to the exemplified A1AT transgene

SEQ ID NO: 27 Exemplified A1A1 peptide

SEQ ID NO: 28 Complementary strand to the exemplified FVIII transgene (N6)

SEQ ID NO: 29 Exemplified FVIII peptide (N6)

SEQ ID NO: 30 Exemplified FVIII transgene (V3)

SEQ ID NO: 31 Complementary strand to the exemplified FVIII transgene (V3)

SEQ ID NO: 32 Exemplified FVIII peptide (V3)

SEQ ID NO: 33 Complementary strand to the exemplified CMV promoter

5

Sequences

SEQ ID NO: 1

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT AGTAATCAAT
181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA
361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAC GGGACTTTCC
481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA
541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT
601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA
661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG GTGGGAGGTC
721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC AGCTTGAGCC
781 TGGGTGTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT CCTTGGCTTA
841 GAAAGCTAAT AAACTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCT CATTGACGCC
901 TCACTCTCTT GAACGGGAAT CTTCCTTACT GGGTTCTCTC TCTGACCCAG GCGAGAGAAA
961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA CAGCTGAGAA
1021 GGCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC TTGGTGAGTA
1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG AGGCCGTAGC
1141 CGTAACTACT CTTGGGCAAG TAGGGCAGGC GGTGGGTACG CAATGGGGGC GGCTACCTCA
1201 GCACTAAATA GGAGACAATT AGACCAATTT GAGAAAATAC GACTTCGCCC GAACGGAAAG
1261 AAAAAGTACC AAATTAAACA TTTAATATGG GCAGGCAAGG AGATGGAGCG CTTCGGCCTC
1321 CATGAGAGGT TGTTGGAGAC AGAGGAGGGG TGTAAAAGAA TCATAGAAGT CCTCTACCCC
1381 CTAGAACCAA CAGGATCGGA GGGCTTAAAA AGTCTGTTCA ATCTTGTGTG CGTGCTATAT
1441 TGCTTGCACA AGGAACAGAA AGTGAAAGAC ACAGAGGAAG CAGTAGCAAC AGTAAGACAA
1501 CACTGCCATC TAGTGGAAAA AGAAAAAAGT GCAACAGAGA CATCTAGTGG ACAAAAGAAA
1561 AATGACAAGG GAATAGCAGC GCCACCTGGT GGCAGTCAGA ATTTTCCAGC GCAACAACAA

1621 GGAAATGCCT GGGTACATGT ACCCTTGTCA CCGCGCACCT TAAATGCGTG
GGTAAAAGCA
1681 GTAGAGGAGA AAAAATTTGG AGCAGAAATA GTACCCATTT TTTTGTTTCA
AGCCCTATCG
1741 AATTCCCGTT TGTGCTAGGG TTCTTAGGCT TCTTGGGGGC TGCTGGAACT
GCAATGGGAG
1801 CAGCGGCGAC AGCCCTGACG GTCCAGTCTC AGCATTTGCT TGCTGGGATA
CTGCAGCAGC
1861 AGAAGAATCT GCTGGCGGCT GTGGAGGCTC AACAGCAGAT GTTGAAGCTG
ACCATTTGGG
1921 GTGTTAAAAA CCTCAATGCC CGCGTCACAG CCCTTGAGAA GTACCTAGAG
GATCAGGCAC
1981 GACTAAACTC CTGGGGGTGC GCATGGAAAC AAGTATGTCA TACCACAGTG
GAGTGGCCCT
2041 GGACAAATCG GACTCCGGAT TGGCAAAATA TGACTTGGTT GGAGTGGGAA
AGACAAATAG
2101 CTGATTTGGA AAGCAACATT ACGAGACAAT TAGTGAAGGC TAGAGAACAA
GAGGAAAAGA
2161 ATCTAGATGC CTATCAGAAG TTAACTAGTT GGTCAGATTT CTGGTCTTGG
TTCGATTTCT
2221 CAAAATGGCT TAACATTTTA AAAATGGGAT TTTTAGTAAT AGTAGGAATA
ATAGGGTTAA
2281 GATTACTTTA CACAGTATAT GGATGTATAG TGAGGGTTAG GCAGGGATAT
GTTCCTCTAT
2341 CTCCACAGAT CCATATCCGC GGCAATTTTA AAAGAAAGGG AGGAATAGGG
GGACAGACTT
2401 CAGCAGAGAG ACTAATTAAT ATAATAACAA CACAATTAGA AATACAACAT
TTACAAACCA
2461 AAATTCAAAA AATTTTAAAT TTTAGAGCCG CGGAGATCTG TTACATAACT
TATGGTAAAT
2521 GGCCTGCCTG GCTGACTGCC CAATGACCCC TGCCCAATGA TGTCAATAAT
GATGTATGTT
2581 CCCATGTAAT GCCAATAGGG ACTTTCCATT GATGTCAATG GGTGGAGTAT
TTATGGTAAC
2641 TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ATGCCCCCTA
TTGATGTCAA
2701 TGATGGTAAA TGGCCTGCCT GGCATTATGC CCAGTACATG ACCTTATGGG
ACTTTCCTAC
2761 TTGGCAGTAC ATCTATGTAT TAGTCATTGC TATTACCATG GGAATTCACT
AGTGGAGAAG
2821 AGCATGCTTG AGGGCTGAGT GCCCCTCAGT GGGCAGAGAG CACATGGCCC
ACAGTCCCTG
2881 AGAAGTTGGG GGGAGGGGTG GGCAATTGAA CTGGTGCCTA GAGAAGGTGG
GGCTTGGGTA
2941 AACTGGGAAA GTGATGTGGT GTACTGGCTC CACCTTTTTC CCCAGGGTGG
GGGAGAACCA
3001 TATATAAGTG CAGTAGTCTC TGTGAACATT CAAGCTTCTG CCTTCTCCCT
CCTGTGAGTT
3061 TGCTAGCCAC CATGCAGAGA AGCCCTCTGG AGAAGGCCTC TGTGGTGAGC
AAGCTGTTCT
3121 TCAGCTGGAC CAGGCCCATC CTGAGGAAGG GCTACAGGCA GAGACTGGAG
CTGTCTGACA
3181 TCTACCAGAT CCCCTCTGTG GACTCTGCTG ACAACCTGTC TGAGAAGCTG
GAGAGGGAGT
3241 GGGATAGAGA GCTGGCCAGC AAGAAGAACC CCAAGCTGAT CAATGCCCTG
AGGAGATGCT
3301 TCTTCTGGAG ATTCATGTTC TATGGCATCT TCCTGTACCT GGGGGAAGTG
ACCAAGGCTG

3361 TGCAGCCTCT GCTGCTGGGC AGAATCATTG CCAGCTATGA CCCTGACAAC
AAGGAGGAGA
3421 GGAGCATTGC CATCTACCTG GGCATTGGCC TGTGCCTGCT GTTCATTGTG
AGGACCCTGC
3481 TGCTGCACCC TGCCATCTTT GGCCTGCACC ACATTGGCAT GCAGATGAGG
ATTGCCATGT
3541 TCAGCCTGAT CTACAAGAAA ACCCTGAAGC TGTCCAGCAG AGTGCTGGAC
AAGATCAGCA
3601 TTGGCCAGCT GGTGAGCCTG CTGAGCAACA ACCTGAACAA GTTTGATGAG
GGCCTGGCCC
3661 TGGCCCACTT TGTGTGGATT GCCCCTCTGC AGGTGGCCCT GCTGATGGGC
CTGATTTGGG
3721 AGCTGCTGCA GGCCTCTGCC TTTTGTGGCC TGGGCTTCCT GATTGTGCTG
GCCCTGTTTC
3781 AGGCTGGCCT GGGCAGGATG ATGATGAAGT ACAGGGACCA GAGGGCAGGC
AAGATCAGTG
3841 AGAGGCTGGT GATCACCTCT GAGATGATTG AGAACATCCA GTCTGTGAAG
GCCTACTGTT
3901 GGGAGGAAGC TATGGAGAAG ATGATTGAAA ACCTGAGGCA GACAGAGCTG
AAGCTGACCA
3961 GGAAGGCTGC CTATGTGAGA TACTTCAACA GCTCTGCCTT CTTCTTCTCT
GGCTTCTTTG
4021 TGGTGTTCCT GTCTGTGCTG CCCTATGCCC TGATCAAGGG GATCATCCTG
AGAAAGATTT
4081 TCACCACCAT CAGCTTCTGC ATTGTGCTGA GGATGGCTGT GACCAGACAG
TTCCCCTGGG
4141 CTGTGCAGAC CTGGTATGAC AGCCTGGGGG CCATCAACAA GATCCAGGAC
TTCCTGCAGA
4201 AGCAGGAGTA CAAGACCCTG GAGTACAACC TGACCACCAC AGAAGTGGTG
ATGGAGAATG
4261 TGACAGCCTT CTGGGAGGAG GGCTTTGGGG AGCTGTTTGA GAAGGCCAAG
CAGAACAACA
4321 ACAACAGAAA GACCAGCAAT GGGGATGACT CCCTGTTCTT CTCCAACTTC
TCCCTGCTGG
4381 GCACACCTGT GCTGAAGGAC ATCAACTTCA AGATTGAGAG GGGGCAGCTG
CTGGCTGTGG
4441 CTGGATCTAC AGGGGCTGGC AAGACCAGCC TGCTGATGAT GATCATGGGG
GAGCTGGAGC
4501 CTTCTGAGGG CAAGATCAAG CACTCTGGCA GGATCAGCTT TTGCAGCCAG
TTCAGCTGGA
4561 TCATGCCTGG CACCATCAAG GAGAACATCA TCTTTGGAGT GAGCTATGAT
GAGTACAGAT
4621 ACAGGAGTGT GATCAAGGCC TGCCAGCTGG AGGAGGACAT CAGCAAGTTT
GCTGAGAAGG
4681 ACAACATTGT GCTGGGGGAG GGAGGCATTA CACTGTCTGG GGGCCAGAGA
GCCAGAATCA
4741 GCCTGGCCAG GGCTGTGTAC AAGGATGCTG ACCTGTACCT GCTGGACTCC
CCCTTTGGCT
4801 ACCTGGATGT GCTGACAGAG AAGGAGATTT TTGAGAGCTG TGTGTGCAAG
CTGATGGCCA
4861 ACAAGACCAG AATCCTGGTG ACCAGCAAGA TGGAGCACCT GAAGAAGGCT
GACAAGATCC
4921 TGATCCTGCA TGAGGGCAGC AGCTACTTCT ATGGGACCTT CTCTGAGCTG
CAGAACCTGC
4981 AGCCTGACTT CAGCTCTAAG CTGATGGGCT GTGACAGCTT TGACCAGTTC
TCTGCTGAGA
5041 GGAGGAACAG CATCCTGACA GAGACCCTGC ACAGATTCAG CCTGGAGGGA
GATGCCCCTG

5101 TGAGCTGGAC AGAGACCAAG AAGCAGAGCT TCAAGCAGAC AGGGGAGTTT GGGGAGAAGA
5161 GGAAGAACTC CATCCTGAAC CCCATCAACA GCATCAGGAA GTTCAGCATT GTGCAGAAAA
5221 CCCCCCTGCA GATGAATGGC ATTGAGGAAG ATTCTGATGA GCCCCTGGAG AGGAGACTGA
5281 GCCTGGTGCC TGATTCTGAG CAGGGAGAGG CCATCCTGCC TAGGATCTCT GTGATCAGCA
5341 CAGGCCCTAC ACTGCAGGCC AGAAGGAGGC AGTCTGTGCT GAACCTGATG ACCCACTCTG
5401 TGAACCAGGG CCAGAACATC CACAGGAAAA CCACAGCCTC CACCAGGAAA GTGAGCCTGG
5461 CCCCTCAGGC CAATCTGACA GAGCTGGACA TCTACAGCAG GAGGCTGTCT CAGGAGACAG
5521 GCCTGGAGAT TTCTGAGGAG ATCAATGAGG AGGACCTGAA AGAGTGCTTC TTTGATGACA
5581 TGGAGAGCAT CCCTGCTGTG ACCACCTGGA ACACCTACCT GAGATACATC ACAGTGCACA
5641 AGAGCCTGAT CTTTGTGCTG ATCTGGTGCC TGGTGATCTT CCTGGCTGAA GTGGCTGCCT
5701 CTCTGGTGGT GCTGTGGCTG CTGGGAAACA CCCCACTGCA GGACAAGGGC AACAGCACCC
5761 ACAGCAGGAA CAACAGCTAT GCTGTGATCA TCACCTCCAC CTCCAGCTAC TATGTGTTCT
5821 ACATCTATGT GGGAGTGGCT GATACCCTGC TGGCTATGGG CTTCTTTAGA GGCCTGCCCC
5881 TGGTGCACAC ACTGATCACA GTGAGCAAGA TCCTCCACCA CAAGATGCTG CACTCTGTGC
5941 TGCAGGCTCC TATGAGCACC CTGAATACCC TGAAGGCTGG GGGCATCCTG AACAGATTCT
6001 CCAAGGATAT TGCCATCCTG GATGACCTGC TGCCTCTCAC CATCTTTGAC TTCATCCAGC
6061 TGCTGCTGAT TGTGATTGGG GCCATTGCTG TGGTGGCAGT GCTGCAGCCC TACATCTTTG
6121 TGGCCACAGT GCCTGTGATT GTGGCCTTCA TCATGCTGAG GGCCTACTTT CTGCAGACCT
6181 CCCAGCAGCT GAAGCAGCTG GAGTCTGAGG GCAGAAGCCC CATCTTCACC CACCTGGTGA
6241 CAAGCCTGAA GGGCCTGTGG ACCCTGAGAG CCTTTGGCAG GCAGCCCTAC TTTGAGACCC
6301 TGTTCCACAA GGCCCTGAAC CTGCACACAG CCAACTGGTT CCTCTACCTG TCCACCCTGA
6361 GATGGTTCCA GATGAGAATT GAGATGATCT TTGTCATCTT CTTCATTGCT GTGACCTTCA
6421 TCAGCATTCT GACCACAGGA GAGGGAGAGG GCAGAGTGGG CATTATCCTG ACCCTGGCCA
6481 TGAACATCAT GAGCACACTG CAGTGGGCAG TGAACAGCAG CATTGATGTG GACAGCCTGA
6541 TGAGGAGTGT GAGCAGAGTG TTCAAGTTCA TTGATATGCC CACAGAGGGC AAGCCTACCA
6601 AGAGCACCAA GCCCTACAAG AATGGCCAGC TGAGCAAAGT GATGATCATT GAGAACAGCC
6661 ATGTGAAGAA GGATGATATC TGGCCCAGTG GAGGCCAGAT GACAGTGAAG GACCTGACAG
6721 CCAAGTACAC AGAGGGGGGC AATGCTATCC TGGAGAACAT CTCCTTCAGC ATCTCCCCTG
6781 GCCAGAGAGT GGGACTGCTG GGAAGAACAG GCTCTGGCAA GTCTACCCTG CTGTCTGCCT

6841 TCCTGAGGCT GCTGAACACA GAGGGAGAGA TCCAGATTGA TGGAGTGTCC
TGGGACAGCA
6901 TCACACTGCA GCAGTGGAGG AAGGCCTTTG GTGTGATCCC CCAGAAAGTG
TTCATCTTCA
6961 GTGGCACCTT CAGGAAGAAC CTGGACCCCT ATGAGCAGTG GTCTGACCAG
GAGATTTGGA
7021 AAGTGGCTGA TGAAGTGGGC CTGAGAAGTG TGATTGAGCA GTTCCCTGGC
AAGCTGGACT
7081 TTGTCCTGGT GGATGGGGGC TGTGTGCTGA GCCATGGCCA CAAGCAGCTG
ATGTGCCTGG
7141 CCAGATCAGT GCTGAGCAAG GCCAAGATCC TGCTGCTGGA TGAGCCTTCT
GCCCACCTGG
7201 ATCCTGTGAC CTACCAGATC ATCAGGAGGA CCCTCAAGCA GGCCTTTGCT
GACTGCACAG
7261 TCATCCTGTG TGAGCACAGG ATTGAGGCCA TGCTGGAGTG CCAGCAGTTC
CTGGTGATTG
7321 AGGAGAACAA AGTGAGGCAG TATGACAGCA TCCAGAAGCT GCTGAATGAG
AGGAGCCTGT
7381 TCAGGCAGGC CATCAGCCCC TCTGATAGAG TGAAGCTGTT CCCCCACAGG
AACAGCTCCA
7441 AGTGCAAGAG CAAGCCCCAG ATTGCTGCCC TGAAGGAGGA GACAGAGGAG
GAAGTGCAGG
7501 ACACCAGGCT GTGAGGGCCC AATCAACCTC TGGATTACAA AATTTGTGAA
AGATTGACTG
7561 GTATTCTTAA CTATGTTGCT CCTTTTACGC TATGTGGATA CGCTGCTTTA
ATGCCTTTGT
7621 ATCATGCTAT TGCTTCCCGT ATGGCTTTCA TTTTCTCCTC CTTGTATAAA
TCCTGGTTGC
7681 TGTCTCTTTA TGAGGAGTTG TGGCCCGTTG TCAGGCAACG TGGCGTGGTG
TGCACTGTGT
7741 TTGCTGACGC AACCCCCACT GGTTGGGGCA TTGCCACCAC CTGTCAGCTC
CTTTCCGGGA
7801 CTTTCGCTTT CCCCCTCCCT ATTGCCACGG CGGAACTCAT CGCCGCCTGC
CTTGCCCGCT
7861 GCTGGACAGG GGCTCGGCTG TTGGGCACTG ACAATTCCGT GGTGTTGTCG
GGGAAATCAT
7921 CGTCCTTTCC TTGGCTGCTC GCCTGTGTTG CCACCTGGAT TCTGCGCGGG
ACGTCCTTCT
7981 GCTACGTCCC TTCGGCCCTC AATCCAGCGG ACCTTCCTTC CCGCGGCCTG
CTGCCGGCTC
8041 TGCGGCCTCT TCCGCGTCTT CGCCTTCGCC CTCAGACGAG TCGGATCTCC
CTTTGGGCCG
8101 CCTCCCCGCA AGCTTCGCAC TTTTTAAAAG AAAAGGGAGG ACTGGATGGG
ATTTATTACT
8161 CCGATAGGAC GCTGGCTTGT AACTCAGTCT CTTACTAGGA GACCAGCTTG
AGCCTGGGTG
8221 TTCGCTGGTT AGCCTAACCT GGTTGGCCAC CAGGGGTAAG GACTCCTTGG
CTTAGAAAGC
8281 TAATAAACTT GCCTGCATTA GAGCTCTTAC GCGTCCCGGG CTCGAGATCC
GCATCTCAAT
8341 TAGTCAGCAA CCATAGTCCC GCCCCTAACT CCGCCCATCC CGCCCCTAAC
TCCGCCCAGT
8401 TCCGCCCATT CTCCGCCCCA TGGCTGACTA ATTTTTTTTA TTTATGCAGA
GGCCGAGGCC
8461 GCCTCGGCCT CTGAGCTATT CCAGAAGTAG TGAGGAGGCT TTTTTGGAGG
CCTAGGCTTT
8521 TGCAAAAAGC TAACTTGTTT ATTGCAGCTT ATAATGGTTA CAAATAAAGC
AATAGCATCA

8581 CAAATTTCAC AAATAAAGCA TTTTTTTCAC TGCATTCTAG TTGTGGTTTG
TCCAAACTCA
8641 TCAATGTATC TTATCATGTC TGTCCGCTTC CTCGCTCACT GACTCGCTGC
GCTCGGTCGT
8701 TCGGCTGCGG CGAGCGGTAT CAGCTCACTC AAAGGCGGTA ATACGGTTAT
CCACAGAATC
8761 AGGGGATAAC GCAGGAAAGA ACATGTGAGC AAAAGGCCAG CAAAAGGCCA
GGAACCGTAA
8821 AAAGGCCGCG TTGCTGGCGT TTTTCCATAG GCTCCGCCCC CCTGACGAGC
ATCACAAAAA
8881 TCGACGCTCA AGTCAGAGGT GGCGAAACCC GACAGGACTA TAAAGATACC
AGGCGTTTCC
8941 CCCTGGAAGC TCCCTCGTGC GCTCTCCTGT TCCGACCCTG CCGCTTACCG
GATACCTGTC
9001 CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA
GGTATCTCAG
9061 TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC GAACCCCCCG
TTCAGCCCGA
9121 CCGCTGCGCC TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGTAAGAC
ACGACTTATC
9181 GCCACTGGCA GCAGCCACTG GTAACAGGAT TAGCAGAGCG AGGTATGTAG
GCGGTGCTAC
9241 AGAGTTCTTG AAGTGGTGGC CTAACTACGG CTACACTAGA AGAACAGTAT
TTGGTATCTG
9301 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT
CCGGCAAACA
9361 AACCACCGCT GGTAGCGGTG GTTTTTTTGT TTGCAAGCAG CAGATTACGC
GCAGAAAAAA
9421 AGGATCTCAA GAAGATCCTT TGATCTTTTC TACGGGGTCT GACGCTCAGT
GGAACGAAAA
9481 CTCACGTTAA GGGATTTTGG TCATGAGATT ATCAAAAAGG ATCTTCACCT
AGATCCTTTT
9541 AAATTAAAAA TGAAGTTTTA AATCAATCTA AAGTATATAT GAGTAAACTT
GGTCTGACAG
9601 TTAGAAAAAC TCATCGAGCA TCAAATGAAA CTGCAATTTA TTCATATCAG
GATTATCAAT
9661 ACCATATTTT TGAAAAAGCC GTTTCTGTAA TGAAGGAGAA AACTCACCGA
GGCAGTTCCA
9721 TAGGATGGCA AGATCCTGGT ATCGGTCTGC GATTCCGACT CGTCCAACAT
CAATACAACC
9781 TATTAATTTC CCCTCGTCAA AAATAAGGTT ATCAAGTGAG AAATCACCAT
GAGTGACGAC
9841 TGAATCCGGT GAGAATGGCA ACAGCTTATG CATTTCTTTC CAGACTTGTT
CAACAGGCCA
9901 GCCATTACGC TCGTCATCAA AATCACTCGC ATCAACCAAA CCGTTATTCA
TTCGTGATTG
9961 CGCCTGAGCG AGACGAAATA CGCGATCGCT GTTAAAAGGA CAATTACAAA
CAGGAATCGA
10021 ATGCAACCGG CGCAGGAACA CTGCCAGCGC ATCAACAATA TTTTCACCTG
AATCAGGATA
10081 TTCTTCTAAT ACCTGGAATG CTGTTTTTCC GGGGATCGCA GTGGTGAGTA
ACCATGCATC
10141 ATCAGGAGTA CGGATAAAAT GCTTGATGGT CGGAAGAGGC ATAAATTCCG
TCAGCCAGTT
10201 TAGTCTGACC ATCTCATCTG TAACATCATT GGCAACGCTA CCTTTGCCAT
GTTTCAGAAA
10261 CAACTCTGGC GCATCGGGCT TCCCATACAA TCGATAGATT GTCGCACCTG
ATTGCCCGAC

10321 ATTATCGCGA GCCCATTTAT ACCCATATAA ATCAGCATCC ATGTTGGAAT TTAATCGCGG
10381 CCTAGAGCAA GACGTTTCCC GTTGAATATG GCTCATAACA CCCCTTGTAT TACTGTTTAT
10441 GTAAGCAGAC AGTTTTATTG TTCATGATGA TATATTTTTA TCTTGTGCAA TGTAACATCA
10501 GAGATTTTGA GACACAACAA TTGGTCGACG GATCC

SEQ ID NO: 2

1 GCTCGAGACT AGTGACTTGG TGAGTAGGCT TCGAGCCTAG TTAGAGGACT AGGAGAGGCC
61 GTAGCCGTAA CTACTCTGGG CAAGTAGGGC AGGCGGTGGG TACGCAATGG GGGCGGCTAC
121 CTCAGCACTA AATAGGAGAC AATTAGACCA ATTTGAGAAA ATACGACTTC GCCCGAACGG
181 AAAGAAAAAG TACCAAATTA AACATTTAAT ATGGGCAGGC AAGGAGATGG AGCGCTTCGG
241 CCTCCATGAG AGGTTGTTGG AGACAGAGGA GGGGTGTAAA AGAATCATAG AAGTCCTCTA
301 CCCCCTAGAA CCAACAGGAT CGGAGGGCTT AAAAAGTCTG TTCAATCTTG TGTGCGTACT
361 ATATTGCTTG CACAAGGAAC AGAAAGTGAA AGACACAGAG GAAGCAGTAG CAACAGTAAG
421 ACAACACTGC CATCTAGTGG AAAAAGAAAA AAGTGCAACA GAGACATCTA GTGGACAAAA
481 GAAAAATGAC AAGGGAATAG CAGCGCCACC TGGTGGCAGT CAGAATTTTC CAGCGCAACA
541 ACAAGGAAAT GCCTGGGTAC ATGTACCCTT GTCACCGCGC ACCTTAAATG CGTGGGTAAA
601 AGCAGTAGAG GAGAAAAAAT TTGGAGCAGA AATAGTACCC ATGTTTCAAG CCCTATCAGA
661 AGGCTGCACA CCCTATGACA TTAATCAGAT GCTTAATGTG CTAGGAGATC ATCAAGGGGC
721 ATTACAAATA GTGAAAGAGA TCATTAATGA AGAAGCAGCC CAGTGGGATG TAACACACCC
781 ACTACCCGCA GGACCCCTAC CAGCAGGACA GCTCAGGGAC CCTCGCGGCT CAGATATAGC
841 AGGGACCACC AGCTCAGTAC AAGAACAGTT AGAATGGATC TATACTGCTA ACCCCCGGGT
901 AGATGTAGGT GCCATCTACC GGAGATGGAT TATTCTAGGA CTTCAAAAGT GTGTCAAAAT
961 GTACAACCCA GTATCAGTCC TAGACATTAG GCAGGGACCT AAAGAGCCCT TCAAGGATTA
1021 TGTGGACAGA TTTTACAAGG CAATTAGAGC AGAACAAGCC TCAGGGGAAG TGAAACAATG
1081 GATGACAGAA TCATTACTCA TTCAAAATGC TAATCCAGAT TGTAAGGTCA TCCTGAAGGG
1141 CCTAGGAATG CACCCCACCC TTGAAGAAAT GTTAACGGCT TGTCAGGGGG TAGGAGGCCC
1201 AAGCTACAAA GCAAAAGTAA TGGCAGAAAT GATGCAGACC ATGCAAAATC AAAACATGGT
1261 GCAGCAGGGA GGTCCAAAAA GACAAAGACC CCCACTAAGA TGTTATAATT GTGGAAAATT
1321 TGGCCATATG CAAAGACAAT GTCCGGAACC AAGGAAAACA AAATGTCTAA AGTGTGGAAA
1381 ATTGGGACAC CTAGCAAAAG ACTGCAGGGG ACAGGTGAAT TTTTTAGGGT ATGGACGGTG

1441 GATGGGGGCA AAACCGAGAA ATTTTCCCGC CGCTACTCTT GGAGCGGAAC CGAGTGCGCC
1501 TCCTCCACCG AGCGGCACCA CCCCATACGA CCCAGCAAAG AAGCTCCTGC AGCAATATGC
1561 AGAGAAAGGG AAACAACTGA GGGAGCAAAA GAGGAATCCA CCGGCAATGA ATCCGGATTG
1621 GACCGAGGGA TATTCTTTGA ACTCCCTCTT TGGAGAAGAC CAATAAAGAC AGTGTATATA
1681 GAAGGGGTCC CCATTAAGGC ACTGCTAGAC ACAGGGGCAG ATGACACCAT AATTAAAGAA
1741 AATGATTTAC AATTATCAGG TCCATGGAGA CCCAAAATTA TAGGGGGCAT AGGAGGAGGC
1801 CTTAATGTAA AAGAATATAA CGACAGGGAA GTAAAAATAG AAGATAAAAT TTTGAGAGGA
1861 ACAATATTGT TAGGAGCAAC TCCCATTAAT ATAATAGGTA GAAATTTGCT GGCCCCGGCA
1921 GGTGCCCGGT TAGTAATGGG ACAATTATCA GAAAAAATTC CTGTCACACC TGTCAAATTG
1981 AAGGAAGGGG CTCGGGGACC CTGTGTAAGA CAATGGCCTC TCTCTAAAGA GAAGATTGAA
2041 GCTTTACAGG AAATATGTTC CCAATTAGAG CAGGAAGGAA AAATCAGTAG AGTAGGAGGA
2101 GAAAATGCAT ACAATACCCC AATATTTTGC ATAAAGAAGA AGGACAAATC CCAGTGGAGG
2161 ATGCTAGTAG ACTTTAGAGA GTTAAATAAG GCAACCCAAG ATTTCTTTGA AGTGCAATTA
2221 GGGATACCCC ACCCAGCAGG ATTAAGAAAG ATGAGACAGA TAACAGTTTT AGATGTAGGA
2281 GACGCCTATT ATTCCATACC ATTGGATCCA AATTTTAGGA AATATACTGC TTTTACTATT
2341 CCCACAGTGA ATAATCAGGG ACCCGGGATT AGGTATCAAT TCAACTGTCT CCCGCAAGGG
2401 TGGAAAGGAT CTCCTACAAT CTTCCAAAAT ACAGCAGCAT CCATTTTGGA GGAGATAAAA
2461 AGAAACTTGC CAGCACTAAC CATTGTACAA TACATGGATG ATTTATGGGT AGGTTCTCAA
2521 GAAAATGAAC ACACCCATGA CAAATTAGTA GAACAGTTAA GAACAAAATT ACAAGCCTGG
2581 GGCTTAGAAA CCCCAGAAAA GAAGGTGCAA AAAGAACCAC CTTATGAGTG GATGGGATAC
2641 AAACTTTGGC CTCACAAATG GGAACTAAGC AGAATACAAC TGGAGGAAAA AGATGAATGG
2701 ACTGTCAATG ACATCCAGAA GTTAGTTGGG AAACTAAATT GGGCAGCACA ATTGTATCCA
2761 GGTCTTAGGA CCAAGAATAT ATGCAAGTTA ATTAGAGGAA AGAAAAATCT GTTAGAGCTA
2821 GTGACTTGGA CACCTGAGGC AGAAGCTGAA TATGCAGAAA ATGCAGAGAT TCTTAAAACA
2881 GAACAGGAAG GAACCTATTA CAAACCAGGA ATACCTATTA GGGCAGCAGT ACAGAAATTG
2941 GAAGGAGGAC AGTGGAGTTA CCAATTCAAA CAAGAAGGAC AAGTCTTGAA AGTAGGAAAA
3001 TACACCAAGC AAAAGAACAC CCATACAAAT GAACTTCGCA CATTAGCTGG TTTAGTGCAG
3061 AAGATTTGCA AAGAAGCTCT AGTTATTTGG GGGATATTAC CAGTTCTAGA ACTCCCGATA
3121 GAAAGAGAGG TATGGGAACA ATGGTGGGCG GATTACTGGC AGGTAAGCTG GATTCCCGAA

3181 TGGGATTTTG TCAGCACCCC ACCTTTGCTC AAACTATGGT ACACATTAAC AAAAGAACCC
3241 ATACCCAAGG AGGACGTTTA CTATGTAGAT GGAGCATGCA ACAGAAATTC AAAAGAAGGA
3301 AAAGCAGGAT ACATCTCACA ATACGGAAAA CAGAGAGTAG AAACATTAGA AAACACTACC
3361 AATCAGCAAG CAGAATTAAC AGCTATAAAA ATGGCTTTGG AAGACAGTGG GCCTAATGTG
3421 AACATAGTAA CAGACTCTCA ATATGCAATG GGAATTTTGA CAGCACAACC CACACAAAGT
3481 GATTCACCAT TAGTAGAGCA AATTATAGCC TTAATGATAC AAAAGCAACA AATATATTTG
3541 CAGTGGGTAC CAGCACATAA AGGAATAGGA GGAAATGAGG AGATAGATAA ATTAGTGAGT
3601 AAAGGCATTA GAAGAGTTTT ATTCTTAGAA AAAATAGAAG AAGCTCAAGA AGAGCATGAA
3661 AGATATCATA ATAATTGGAA AAACCTAGCA GATACATATG GGCTTCCACA AATAGTAGCA
3721 AAAGAGATAG TGGCCATGTG TCCAAAATGT CAGATAAAGG GAGAACCAGT GCATGGACAA
3781 GTGGATGCCT CACCTGGAAC ATGGCAGATG GATTGTACTC ATCTAGAAGG AAAAGTAGTC
3841 ATAGTTGCGG TCCATGTAGC CAGTGGATTC ATAGAAGCAG AAGTCATACC TAGGGAAACA
3901 GGAAAAGAAA CGGCAAAGTT TCTATTAAAA ATACTGAGTA GATGGCCTAT AACACAGTTA
3961 CACACAGACA ATGGGCCTAA CTTTACCTCC CAAGAAGTGG CAGCAATATG TTGGTGGGGA
4021 AAAATTGAAC ATACAACAGG TATACCATAT AACCCCCAAT CTCAAGGATC AATAGAAAGC
4081 ATGAACAAAC AATTAAAAGA GATAATTGGG AAAATAAGAG ATGATTGCCA ATATACAGAG
4141 ACAGCAGTAC TGATGGCTTG CCATATTCAC AATTTTAAAA GAAAGGGAGG AATAGGGGGA
4201 CAGACTTCAG CAGAGAGACT AATTAATATA ATAACAACAC AATTAGAAAT ACAACATTTA
4261 CAAACCAAAA TTCAAAAAAT TTTAAATTTT AGAGTCTACT ACAGAGAAGG GAGAGACCCT
4321 GTGTGGAAAG GACCAGCACA ATTAATCTGG AAAGGGGAAG GAGCAGTGGT CCTCAAGGAC
4381 GGAAGTGACC TAAAGGTTGT ACCAAGAAGG AAAGCTAAAA TTATTAAGGA TTATGAACCC
4441 AAACAAAGAG TGGGTAATGA GGGTGACGTG GAAGGTACCA GGGGATCTGA TAACTAAATG
4501 GCAGGGAATA GTCAGATATT GGATGAGACA AAGAAATTTG AAATGGAACT ATTATATGCA
4561 TCAGCTGGCG GCCGCGAATT CACTAGTGAT TCCCGTTTGT GCTAGGGTTC TTAGGCTTCT
4621 TGGGGGCTGC TGGAACTGCA ATGGGAGCAG CGGCGACAGC CCTGACGGTC CAGTCTCAGC
4681 ATTTGCTTGC TGGGATACTG CAGCAGCAGA AGAATCTGCT GGCGGCTGTG GAGGCTCAAC
4741 AGCAGATGTT GAAGCTGACC ATTTGGGGTG TTAAAAACCT CAATGCCCGC GTCACAGCCC
4801 TTGAGAAGTA CCTAGAGGAT CAGGCACGAC TAAACTCCTG GGGGTGCGCA TGGAAACAAG
4861 TATGTCATAC CACAGTGGAG TGGCCCTGGA CAAATCGGAC TCCGGATTGG CAAAATATGA

4921 CTTGGTTGGA GTGGGAAAGA CAAATAGCTG ATTTGGAAAG CAACATTACG
AGACAATTAG
4981 TGAAGGCTAG AGAACAAGAG GAAAAGAATC TAGATGCCTA TCAGAAGTTA
ACTAGTTGGT
5041 CAGATTTCTG GTCTTGGTTC GATTTCTCAA AATGGCTTAA CATTTTAAAA
ATGGGATTTT
5101 TAGTAATAGT AGGAATAATA GGGTTAAGAT TACTTTACAC AGTATATGGA
TGTATAGTGA
5161 GGGTTAGGCA GGGATATGTT CCTCTATCTC CACAGATCCA TATCCAATCG
AATTCCCGCG
5221 GCCGCAATTC ACTCCTCAGG TGCAGGCTGC CTATCAGAAG GTGGTGGCTG
GTGTGGCCAA
5281 TGCCCTGGCT CACAAATACC ACTGAGATCT TTTTCCCTCT GCCAAAAATT
ATGGGGACAT
5341 CATGAAGCCC CTTGAGCATC TGACTTCTGG CTAATAAAGG AAATTTATTT
TCATTGCAAT
5401 AGTGTGTTGG AATTTTTTGT GTCTCTCACT CGGAAGGACA TATGGGAGGG
CAAATCATTT
5461 AAAACATCAG AATGAGTATT TGGTTTAGAG TTTGGCAACA TATGCCCATA
TGCTGGCTGC
5521 CATGAACAAA GGTTGGCTAT AAAGAGGTCA TCAGTATATG AAACAGCCCC
CTGCTGTCCA
5581 TTCCTTATTC CATAGAAAAG CCTTGACTTG AGGTTAGATT TTTTTTATAT
TTTGTTTTGT
5641 GTTATTTTTT TCTTTAACAT CCCTAAAATT TTCCTTACAT GTTTTACTAG
CCAGATTTTT
5701 CCTCCTCTCC TGACTACTCC CAGTCATAGC TGTCCCTCTT CTCTTATGGA
GATCCCTCGA
5761 CCTGCAGCCC AAGCTTGGCG TAATCATGGT CATAGCTGTT TCCTGTGTGA
AATTGTTATC
5821 CGCTCACAAT TCCACACAAC ATACGAGCCG GAAGCATAAA GTGTAAAGCC
TGGGGTGCCT
5881 AATGAGTGAG CTAACTCACA TTAATTGCGT TGCGCTCACT GCCCGCTTTC
CAGTCGGGAA
5941 ACCTGTCGTG CCAGCGGATC CGCATCTCAA TTAGTCAGCA ACCATAGTCC
CGCCCCTAAC
6001 TCCGCCCATC CCGCCCCTAA CTCCGCCCAG TTCCGCCCAT TCTCCGCCCC
ATGGCTGACT
6061 AATTTTTTTT ATTTATGCAG AGGCCGAGGC CGCCTCGGCC TCTGAGCTAT
TCCAGAAGTA
6121 GTGAGGAGGC TTTTTTGGAG GCCTAGGCTT TTGCAAAAAG CTAACTTGTT
TATTGCAGCT
6181 TATAATGGTT ACAAATAAAG CAATAGCATC ACAAATTTCA CAAATAAAGC
ATTTTTTTCA
6241 CTGCATTCTA GTTGTGGTTT GTCCAAACTC ATCAATGTAT CTTATCATGT
CTGTCCGCTT
6301 CCTCGCTCAC TGACTCGCTG CGCTCGGTCG TTCGGCTGCG GCGAGCGGTA
TCAGCTCACT
6361 CAAAGGCGGT AATACGGTTA TCCACAGAAT CAGGGGATAA CGCAGGAAAG
AACATGTGAG
6421 CAAAAGGCCA GCAAAAGGCC AGGAACCGTA AAAAGGCCGC GTTGCTGGCG
TTTTTCCATA
6481 GGCTCCGCCC CCCTGACGAG CATCACAAAA ATCGACGCTC AAGTCAGAGG
TGGCGAAACC
6541 CGACAGGACT ATAAAGATAC CAGGCGTTTC CCCCTGGAAG CTCCCTCGTG
CGCTCTCCTG
6601 TTCCGACCCT GCCGCTTACC GGATACCTGT CCGCCTTTCT CCCTTCGGGA
AGCGTGGCGC

6661 TTTCTCATAG CTCACGCTGT AGGTATCTCA GTTCGGTGTA GGTCGTTCGC
TCCAAGCTGG
6721 GCTGTGTGCA CGAACCCCCC GTTCAGCCCG ACCGCTGCGC CTTATCCGGT
AACTATCGTC
6781 TTGAGTCCAA CCCGGTAAGA CACGACTTAT CGCCACTGGC AGCAGCCACT
GGTAACAGGA
6841 TTAGCAGAGC GAGGTATGTA GGCGGTGCTA CAGAGTTCTT GAAGTGGTGG
CCTAACTACG
6901 GCTACACTAG AAGAACAGTA TTTGGTATCT GCGCTCTGCT GAAGCCAGTT
ACCTTCGGAA
6961 AAAGAGTTGG TAGCTCTTGA TCCGGCAAAC AAACCACCGC TGGTAGCGGT
GGTTTTTTTG
7021 TTTGCAAGCA GCAGATTACG CGCAGAAAAA AAGGATCTCA AGAAGATCCT
TTGATCTTTT
7081 CTACGGGGTC TGACGCTCAG TGGAACGAAA ACTCACGTTA AGGGATTTTG
GTCATGAGAT
7141 TATCAAAAAG GATCTTCACC TAGATCCTTT TAAATTAAAA ATGAAGTTTT
AAATCAATCT
7201 AAAGTATATA TGAGTAAACT TGGTCTGACA GTTAGAAAAA CTCATCGAGC
ATCAAATGAA
7261 ACTGCAATTT ATTCATATCA GGATTATCAA TACCATATTT TTGAAAAAGC
CGTTTCTGTA
7321 ATGAAGGAGA AAACTCACCG AGGCAGTTCC ATAGGATGGC AAGATCCTGG
TATCGGTCTG
7381 CGATTCCGAC TCGTCCAACA TCAATACAAC CTATTAATTT CCCCTCGTCA
AAAATAAGGT
7441 TATCAAGTGA GAAATCACCA TGAGTGACGA CTGAATCCGG TGAGAATGGC
AACAGCTTAT
7501 GCATTTCTTT CCAGACTTGT TCAACAGGCC AGCCATTACG CTCGTCATCA
AAATCACTCG
7561 CATCAACCAA ACCGTTATTC ATTCGTGATT GCGCCTGAGC GAGACGAAAT
ACGCGATCGC
7621 TGTTAAAAGG ACAATTACAA ACAGGAATCG AATGCAACCG GCGCAGGAAC
ACTGCCAGCG
7681 CATCAACAAT ATTTTCACCT GAATCAGGAT ATTCTTCTAA TACCTGGAAT
GCTGTTTTTC
7741 CGGGGATCGC AGTGGTGAGT AACCATGCAT CATCAGGAGT ACGGATAAAA
TGCTTGATGG
7801 TCGGAAGAGG CATAAATTCC GTCAGCCAGT TTAGTCTGAC CATCTCATCT
GTAACATCAT
7861 TGGCAACGCT ACCTTTGCCA TGTTTCAGAA ACAACTCTGG CGCATCGGGC
TTCCCATACA
7921 ATCGATAGAT TGTCGCACCT GATTGCCCGA CATTATCGCG AGCCCATTTA
TACCCATATA
7981 AATCAGCATC CATGTTGGAA TTTAATCGCG GCCTAGAGCA AGACGTTTCC
CGTTGAATAT
8041 GGCTCATAAC ACCCCTTGTA TTACTGTTTA TGTAAGCAGA CAGTTTTATT
GTTCATGATG
8101 ATATATTTTT ATCTTGTGCA ATGTAACATC AGAGATTTTG AGACACAACA
ATTGTCGACA
8161 TTGATTATTG ACTAGTTATT AATAGTAATC AATTACGGGG TCATTAGTTC
ATAGCCCATA
8221 TATGGAGTTC CGCGTTACAT AACTTACGGT AAATGGCCCG CCTGGCTGAC
CGCCCAACGA
8281 CCCCCGCCCA TTGACGTCAA TAATGACGTA TGTTCCCATA GTAACGCCAA
TAGGGACTTT
8341 CCATTGACGT CAATGGGTGG AGTATTTACG GTAAACTGCC CACTTGGCAG
TACATCAAGT

8401 GTATCATATG CCAAGTACGC CCCCTATTGA CGTCAATGAC GGTAAATGGC CCGCCTGGCA
8461 TTATGCCCAG TACATGACCT TATGGGACTT TCCTACTTGG CAGTACATCT ACGTATTAGT
8521 CATCGCTATT ACCATGGTCG AGGTGAGCCC CACGTTCTGC TTCACTCTCC CCATCTCCCC
8581 CCCCTCCCCA CCCCCAATTT TGTATTTATT TATTTTTTAA TTATTTTGTG CAGCGATGGG
8641 GGCGGGGGGG GGGGGGGGGC GCGCGCCAGG CGGGGCGGGG CGGGGCGAGG GGCGGGGCGG
8701 GGCGAGGCGG AGAGGTGCGG CGGCAGCCAA TCAGAGCGGC GCGCTCCGAA AGTTTCCTTT
8761 TATGGCGAGG CGGCGGCGGC GGCGGCCCTA TAAAAAGCGA AGCGCGCGGC GGGCGGGAGT
8821 CGCTGCGCGC TGCCTTCGCC CCGTGCCCCG CTCCGCCGCC GCCTCGCGCC GCCCGCCCCG
8881 GCTCTGACTG ACCGCGTTAC TCCCACAGGT GAGCGGGCGG GACGGCCCTT CTCCTCCGGG
8941 CTGTAATTAG CGCTTGGTTT AATGACGGCT TGTTTCTTTT CTGTGGCTGC GTGAAAGCCT
9001 TGAGGGGCTC CGGGAGGGCC CTTTGTGCGG GGGGAGCGGC TCGGGGGGTG CGTGCGTGTG
9061 TGTGTGCGTG GGGAGCGCCG CGTGCGGCTC CGCGCTGCCC GGCGGCTGTG AGCGCTGCGG
9121 GCGCGGCGCG GGGCTTTGTG CGCTCCGCAG TGTGCGCGAG GGGAGCGCGG CCGGGGGCGG
9181 TGCCCCGCGG TGCGGGGGGG GCTGCGAGGG GAACAAAGGC TGCGTGCGGG GTGTGTGCGT
9241 GGGGGGGTGA GCAGGGGGTG TGGGCGCGTC GGTCGGGCTG CAACCCCCCC TGCACCCCCC
9301 TCCCCGAGTT GCTGAGCACG GCCCGGCTTC GGGTGCGGGG CTCCGTACGG GGCGTGGCGC
9361 GGGGCTCGCC GTGCCGGGCG GGGGGTGGCG GCAGGTGGGG GTGCCGGGCG GGGCGGGGCC
9421 GCCTCGGGCC GGGGAGGGCT CGGGGGAGGG GCGCGGCGGC CCCCGGAGCG CCGGCGGCTG
9481 TCGAGGCGCG GCGAGCCGCA GCCATTGCCT TTTATGGTAA TCGTGCGAGA GGGCGCAGGG
9541 ACTTCCTTTG TCCCAAATCT GTGCGGAGCC GAAATCTGGG AGGCGCCGCC GCACCCCCTC
9601 TAGCGGGCGC GGGGCGAAGC GGTGCGGCGC CGGCAGGAAG GAAATGGGCG GGGAGGGCCT
9661 TCGTGCGTCG CCGCGCCGCC GTCCCCTTCT CCCTCTCCAG CCTCGGGGCT GTCCGCGGGG
9721 GGACGGCTGC CTTCGGGGGG GACGGGGCAG GGCGGGGTTC GGCTTCTGGC GTGTGACCGG
9781 CGGCTCTAGA GCCTCTGCTA ACCATGTTCA TGCCTTCTTC TTTTTCCTAC AGCTCCTGGG
9841 CAACGTGCTG GTTATTGTGC TGTCTCATCA TTTTGGCAAA GAATT

SEQ ID NO: 3

1 TCAATATTGG CCATTAGCCA TATTATTCAT TGGTTATATA GCATAAATCA ATATTGGCTA
61 TTGGCCATTG CATACGTTGT ATCTATATCA TAATATGTAC ATTTATATTG GCTCATGTCC
121 AATATGACCG CCATGTTGGC ATTGATTATT GACTAGTTAT TAATAGTAAT CAATTACGGG

181 GTCATTAGTT CATAGCCCAT ATATGGAGTT CCGCGTTACA TAACTTACGG TAAATGGCCC
241 GCCTGGCTGA CCGCCCAACG ACCCCCGCCC ATTGACGTCA ATAATGACGT ATGTTCCCAT
301 AGTAACGCCA ATAGGGACTT TCCATTGACG TCAATGGGTG GAGTATTTAC GGTAAACTGC
361 CCACTTGGCA GTACATCAAG TGTATCATAT GCCAAGTCCG CCCCCTATTG ACGTCAATGA
421 CGGTAAATGG CCCGCCTGGC ATTATGCCCA GTACATGACC TTACGGGACT TTCCTACTTG
481 GCAGTACATC TACGTATTAG TCATCGCTAT TACCATGGTG ATGCGGTTTT GGCAGTACAC
541 CAATGGGCGT GGATAGCGGT TTGACTCACG GGGATTTCCA AGTCTCCACC CCATTGACGT
601 CAATGGGAGT TTGTTTTGGC ACCAAAATCA ACGGGACTTT CCAAAATGTC GTAATAACCC
661 CGCCCCGTTG ACGCAAATGG GCGGTAGGCG TGTACGGTGG GAGGTCTATA TAAGCAGAGC
721 TCGTTTAGTG AACCGTCAGA TCACTAGAAG CTTTATTGCG GTAGTTTATC ACAGTTAAAT
781 TGCTAACGCA GTCAGTGCTT CTGACACAAC AGTCTCGAAC TTAAGCTGCA GAAGTTGGTC
841 GTGAGGCACT GGGCAGGTAA GTATCAAGGT TACAAGACAG GTTTAAGGAG ACCAATAGAA
901 ACTGGGCTTG TCGAGACAGA GAAGACTCTT GCGTTTCTGA TAGGCACCTA TTGGTCTTAC
961 TGACATCCAC TTTGCCTTTC TCTCCACAGG TGTCCACTCC CAGTTCAATT ACAGCTCTTA
1021 AGGCTAGAGT ACTTAATACG ACTCACTATA GGCTAGCCTC GAGAATTCGA TTATGCCCCT
1081 AGGACCAGAA GAAAGAAGAT TGCTTCGCTT GATTTGGCTC CTTTACAGCA CCAATCCATA
1141 TCCACCAAGT GGGGAAGGGA CGGCCAGACA ACGCCGACGA GCCAGGAGAA GGTGGAGACA
1201 ACAGCAGGAT CAAATTAGAG TCTTGGTAGA AAGACTCCAA GAGCAGGTGT ATGCAGTTGA
1261 CCGCCTGGCT GACGAGGCTC AACACTTGGC TATACAACAG TTGCCTGACC CTCCTCATTC
1321 AGCTTAGAAT CACTAGTGAA TTCACGCGTG GTACCTCTAG AGTCGACCCG GGCGGCCGCT
1381 TCGAGCAGAC ATGATAAGAT ACATTGATGA GTTTGGACAA ACCACAACTA GAATGCAGTG
1441 AAAAAAATGC TTTATTTGTG AAATTTGTGA TGCTATTGCT TTATTTGTAA CCATTATAAG
1501 CTGCAATAAA CAAGTTAACA ACAACAATTG CATTCATTTT ATGTTTCAGG TTCAGGGGGA
1561 GATGTGGGAG GTTTTTTAAA GCAAGTAAAA CCTCTACAAA TGTGGTAAAA TCGATAAGGA
1621 TCCGTCGACC AATTGTTGTG TCTCAAAATC TCTGATGTTA CATTGCACAA GATAAAAATA
1681 TATCATCATG AACAATAAAA CTGTCTGCTT ACATAAACAG TAATACAAGG GGTGTTATGA
1741 GCCATATTCA ACGGGAAACG TCTTGCTCTA GGCCGCGATT AAATTCCAAC ATGGATGCTG
1801 ATTTATATGG GTATAAATGG GCTCGCGATA ATGTCGGGCA ATCAGGTGCG ACAATCTATC
1861 GATTGTATGG GAAGCCCGAT GCGCCAGAGT TGTTTCTGAA ACATGGCAAA GGTAGCGTTG

1921 CCAATGATGT TACAGATGAG ATGGTCAGAC TAAACTGGCT GACGGAATTT
ATGCCTCTTC
1981 CGACCATCAA GCATTTTATC CGTACTCCTG ATGATGCATG GTTACTCACC
ACTGCGATCC
2041 CCGGAAAAAC AGCATTCCAG GTATTAGAAG AATATCCTGA TTCAGGTGAA
AATATTGTTG
2101 ATGCGCTGGC AGTGTTCCTG CGCCGGTTGC ATTCGATTCC TGTTTGTAAT
TGTCCTTTTA
2161 ACAGCGATCG CGTATTTCGT CTCGCTCAGG CGCAATCACG AATGAATAAC
GGTTTGGTTG
2221 ATGCGAGTGA TTTTGATGAC GAGCGTAATG GCTGGCCTGT TGAACAAGTC
TGGAAAGAAA
2281 TGCATAAGCT GTTGCCATTC TCACCGGATT CAGTCGTCAC TCATGGTGAT
TTCTCACTTG
2341 ATAACCTTAT TTTTGACGAG GGGAAATTAA TAGGTTGTAT TGATGTTGGA
CGAGTCGGAA
2401 TCGCAGACCG ATACCAGGAT CTTGCCATCC TATGGAACTG CCTCGGTGAG
TTTTCTCCTT
2461 CATTACAGAA ACGGCTTTTT CAAAAATATG GTATTGATAA TCCTGATATG
AATAAATTGC
2521 AGTTTCATTT GATGCTCGAT GAGTTTTTCT AACTGTCAGA CCAAGTTTAC
TCATATATAC
2581 TTTAGATTGA TTTAAAACTT CATTTTTAAT TTAAAAGGAT CTAGGTGAAG
ATCCTTTTTG
2641 ATAATCTCAT GACCAAAATC CCTTAACGTG AGTTTTCGTT CCACTGAGCG
TCAGACCCCG
2701 TAGAAAAGAT CAAAGGATCT TCTTGAGATC CTTTTTTTCT GCGCGTAATC
TGCTGCTTGC
2761 AAACAAAAAA ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG
CTACCAACTC
2821 TTTTTCCGAA GGTAACTGGC TTCAGCAGAG CGCAGATACC AAATACTGTT
CTTCTAGTGT
2881 AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC GCCTACATAC
CTCGCTCTGC
2941 TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC
GGGTTGGACT
3001 CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTCGGGCTG AACGGGGGGT
TCGTGCACAC
3061 AGCCCAGCTT GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT
GAGCTATGAG
3121 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC
GGCAGGGTCG
3181 GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC CTGGTATCTT
TATAGTCCTG
3241 TCGGGTTTCG CCACCTCTGA CTTGAGCGTC GATTTTTGTG ATGCTCGTCA
GGGGGGCGGA
3301 GCCTATGGAA AAACGCCAGC AACGCGGCCT TTTTACGGTT CCTGGCCTTT
TGCTGGCCTT
3361 TTGCTCACAT GGCTCGACAG ATCT

SEQ ID NO: 4

1 ATTGATTATT GACTAGTTAT TAATAGTAAT CAATTACGGG GTCATTAGTT
CATAGCCCAT
61 ATATGGAGTT CCGCGTTACA TAACTTACGG TAAATGGCCC GCCTGGCTGA
CCGCCCAACG

121 ACCCCCGCCC ATTGACGTCA ATAATGACGT ATGTTCCCAT AGTAACGCCA
ATAGGGACTT
181 TCCATTGACG TCAATGGGTG GAGTATTTAC GGTAAACTGC CCACTTGGCA
GTACATCAAG
241 TGTATCATAT GCCAAGTACG CCCCCTATTG ACGTCAATGA CGGTAAATGG
CCCGCCTGGC
301 ATTATGCCCA GTACATGACC TTATGGGACT TTCCTACTTG GCAGTACATC
TACGTATTAG
361 TCATCGCTAT TACCATGGTC GAGGTGAGCC CCACGTTCTG CTTCACTCTC
CCCATCTCCC
421 CCCCCTCCCC ACCCCCAATT TTGTATTTAT TTATTTTTTA ATTATTTTGT
GCAGCGATGG
481 GGGCGGGGGG GGGGGGGGGG CGCGCGCCAG GCGGGGCGGG GCGGGGCAAG
GGGCGGGGCG
541 GGGCGAGGCG GAAAGGTGCG GCGGCAGCCA ATCAAAGCGG CGCGCTCCGA
AAGTTTCCTT
601 TTATGGCGAG GCGGCGGCGG CGGCGGCCCT ATAAAAAGCG AAGCGCGCGG
CGGGCGGGAG
661 TCGCTGCGCG CTGCCTTCGC CCCGTGCCCC GCTCCGCCGC CGCCTCGCGC
CGCCCGCCCC
721 GGCTCTGACT GACCGCGTTA CTCCCACAGG TGAGCGGGCG GGACGGCCCT
TCTCCTCCGG
781 GCTGTAATTA GCGCTTGGTT TAATGACGGC TTGTTTCTTT TCTGTGGCTG
CGTGAAAGCC
841 TTGAGGGGCT CCGGGAGGGC CCTTTGTGCG GGGGGAGCGG CTCGGGGGGT
GCGTGCGTGT
901 GTGTGTGCGT GGGGAGCGCC GCGTGCGGCT CCGCGCTGCC CGGCGGCTGT
GAGCGCTGCG
961 GGCGCGGCGC GGGGCTTTGT GCGCTCCGCA GTGTGCGCGA GGGGAGCGCG
GCCGGGGGCG
1021 GTGCCCCGCG GTGCGGGGGG GGCTGCGAGG GGAACAAAGG CTGCGTGCGG
GGTGTGTGCG
1081 TGGGGGGGTG AGCAGGGGGT GTGGGCGCGT CGGTCGGGCT GCAACCCCCC
CTGCACCCCC
1141 CTCCCCGAGT TGCTGAGCAC GGCCCGGCTT CGGGTGCGGG GCTCCGTACG
GGGCGTGGCG
1201 CGGGGCTCGC CGTGCCGGGC GGGGGGTGGC GGCAGGTGGG GGTGCCGGGC
GGGGCGGGGC
1261 CGCCTCGGGC CGGGGAGGGC TCGGGGGAGG GGCGCGGCGG CCCCCGGAGC
GCCGGCGGCT
1321 GTCGAGGCGC GGCGAGCCGC AGCCATTGCC TTTTATGGTA ATCGTGCGAG
AGGGCGCAGG
1381 GACTTCCTTT GTCCCAAATC TGTGCGGAGC CGAAATCTGG GAGGCGCCGC
CGCACCCCCT
1441 CTAGCGGGCG CGGGGCGAAG CGGTGCGGCG CCGGCAGGAA GGAAATGGGC
GGGGAGGGCC
1501 TTCGTGCGTC GCCGCGCCGC CGTCCCCTTC TCCCTCTCCA GCCTCGGGGC
TGTCCGCGGG
1561 GGGACGGCTG CCTTCGGGGG GGACGGGGCA GGGCGGGGTT CGGCTTCTGG
CGTGTGACCG
1621 GCGGCTCTAG AGCCTCTGCT AACCATGTTC ATGCCTTCTT CTTTTTCCTA
CAGCTCCTGG
1681 GCAACGTGCT GGTTATTGTG CTGTCTCATC ATTTTGGCAA AGAATTCGAT
TGCCATGGCA
1741 ACATATATCC AGAGAGTACA GTGCATCTCA ACATCACTAC TGGTTGTTCT
CACCACATTG
1801 GTCTCGTGTC AGATTCCCAG GGATAGGCTC TCTAACATAG GGGTCATAGT
CGATGAAGGG

1861 AAATCACTGA AGATAGCTGG ATCCCACGAA TCGAGGTACA TAGTACTGAG TCTAGTTCCG
1921 GGGGTAGACT TTGAGAATGG GTGCGGAACA GCCCAGGTTA TCCAGTACAA GAGCCTACTG
1981 AACAGGCTGT TAATCCCATT GAGGGATGCC TTAGATCTTC AGGAGGCTCT GATAACTGTC
2041 ACCAATGATA CGACACAAAA TGCCGGTGCT CCCCAGTCGA GATTCTTCGG TGCTGTGATT
2101 GGTACTATCG CACTTGGAGT GGCGACATCA GCACAAATCA CCGCAGGGAT TGCACTAGCC
2161 GAAGCGAGGG AGGCCAAAAG AGACATAGCG CTCATCAAAG AATCGATGAC AAAAACACAC
2221 AAGTCTATAG AACTGCTGCA AAACGCTGTG GGGGAACAAA TTCTTGCTCT AAAGACACTC
2281 CAGGATTTCG TGAATGATGA GATCAAACCC GCAATAAGCG AATTAGGCTG TGAGACTGCT
2341 GCCTTAAGAC TGGGTATAAA ATTGACACAG CATTACTCCG AGCTGTTAAC TGCGTTCGGC
2401 TCGAATTTCG GAACCATCGG AGAGAAGAGC CTCACGCTGC AGGCGCTGTC TTCACTTTAC
2461 TCTGCTAACA TTACTGAGAT TATGACCACA ATCAGGACAG GGCAGTCTAA CATCTATGAT
2521 GTCATTTATA CAGAACAGAT CAAAGGAACG GTGATAGATG TGGATCTAGA GAGATACATG
2581 GTCACCCTGT CTGTGAAGAT CCCTATTCTT TCTGAAGTCC CAGGTGTGCT CATACACAAG
2641 GCATCATCTA TTTCTTACAA CATAGACGGG GAGGAATGGT ATGTGACTGT CCCCAGCCAT
2701 ATACTCAGTC GTGCTTCTTT CTTAGGGGGT GCAGACATAA CCGATTGTGT TGAGTCCAGA
2761 TTGACCTATA TATGCCCCAG GGATCCCGCA CAACTGATAC CTGACAGCCA GCAAAAGTGT
2821 ATCCTGGGGG ACACAACAAG GTGTCCTGTC ACAAAAGTTG TGGACAGCCT TATCCCCAAG
2881 TTTGCTTTTG TGAATGGGGG CGTTGTTGCT AACTGCATAG CATCCACATG TACCTGCGGG
2941 ACAGGCCGAA GACCAATCAG TCAGGATCGC TCTAAAGGTG TAGTATTCCT AACCCATGAC
3001 AACTGTGGTC TTATAGGTGT CAATGGGGTA GAATTGTATG CTAACCGGAG AGGGCACGAT
3061 GCCACTTGGG GGGTCCAGAA CTTGACAGTC GGTCCTGCAA TTGCTATCAG ACCCGTTGAT
3121 ATTTCTCTCA ACCTTGCTGA TGCTACGAAT TTCTTGCAAG ACTCTAAGGC TGAGCTTGAG
3181 AAAGCACGGA AAATCCTCTC GGAGGTAGGT AGATGGTACA ACTCAAGAGA GACTGTGATT
3241 ACGATCATAG TAGTTATGGT CGTAATATTG GTGGTCATTA TAGTGATCAT CATCGTGCTT
3301 TATAGACTCA GAAGGTGAAA TCACTAGTGA ATTCACTCCT CAGGTGCAGG CTGCCTATCA
3361 GAAGGTGGTG GCTGGTGTGG CCAATGCCCT GGCTCACAAA TACCACTGAG ATCTTTTTCC
3421 CTCTGCCAAA AATTATGGGG ACATCATGAA GCCCCTTGAG CATCTGACTT CTGGCTAATA
3481 AAGGAAATTT ATTTTCATTG CAATAGTGTG TTGGAATTTT TTGTGTCTCT CACTCGGAAG
3541 GACATATGGG AGGGCAAATC ATTTAAAACA TCAGAATGAG TATTTGGTTT AGAGTTTGGC

3601 AACATATGCC CATATGCTGG CTGCCATGAA CAAAGGTTGG CTATAAAGAG GTCATCAGTA
3661 TATGAAACAG CCCCCTGCTG TCCATTCCTT ATTCCATAGA AAAGCCTTGA CTTGAGGTTA
3721 GATTTTTTTT ATATTTTGTT TTGTGTTATT TTTTTCTTTA ACATCCCTAA AATTTTCCTT
3781 ACATGTTTTA CTAGCCAGAT TTTTCCTCCT CTCCTGACTA CTCCCAGTCA TAGCTGTCCC
3841 TCTTCTCTTA TGGAGATCCC TCGACCTGCA GCCCAAGCTT GGCGTAATCA TGGTCATAGC
3901 TGTTTCCTGT GTGAAATTGT TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA
3961 TAAAGTGTAA AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAATT GCGTTGCGCT
4021 CACTGCCCGC TTTCCAGTCG GGAAACCTGT CGTGCCAGCG GATCCGCATC TCAATTAGTC
4081 AGCAACCATA GTCCCGCCCC TAACTCCGCC CATCCCGCCC CTAACTCCGC CCAGTTCCGC
4141 CCATTCTCCG CCCCATGGCT GACTAATTTT TTTTATTTAT GCAGAGGCCG AGGCCGCCTC
4201 GGCCTCTGAG CTATTCCAGA AGTAGTGAGG AGGCTTTTTT GGAGGCCTAG GCTTTTGCAA
4261 AAAGCTAACT TGTTTATTGC AGCTTATAAT GGTTACAAAT AAAGCAATAG CATCACAAAT
4321 TTCACAAATA AAGCATTTTT TTCACTGCAT TCTAGTTGTG GTTTGTCCAA ACTCATCAAT
4381 GTATCTTATC ATGTCTGTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCGTTCGGC
4441 TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG
4501 ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG
4561 CCGCGTTGCT GGCGTTTTTC CATAGGCTCC GCCCCCCTGA CGAGCATCAC AAAAATCGAC
4621 GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG
4681 GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT
4741 TTCTCCCTTC GGGAAGCGTG GCGCTTTCTC ATAGCTCACG CTGTAGGTAT CTCAGTTCGG
4801 TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT
4861 GCGCCTTATC CGGTAACTAT CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATCGCCAC
4921 TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT GCTACAGAGT
4981 TCTTGAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGAAC AGTATTTGGT ATCTGCGCTC
5041 TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA
5101 CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT TACGCGCAGA AAAAAAGGAT
5161 CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAACTCAC
5221 GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT
5281 AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA AACTTGGTCT GACAGTTAGA

5341 AAAACTCATC GAGCATCAAA TGAAACTGCA ATTTATTCAT ATCAGGATTA
TCAATACCAT
5401 ATTTTTGAAA AAGCCGTTTC TGTAATGAAG GAGAAAACTC ACCGAGGCAG
TTCCATAGGA
5461 TGGCAAGATC CTGGTATCGG TCTGCGATTC CGACTCGTCC AACATCAATA
CAACCTATTA
5521 ATTTCCCCTC GTCAAAAATA AGGTTATCAA GTGAGAAATC ACCATGAGTG
ACGACTGAAT
5581 CCGGTGAGAA TGGCAACAGC TTATGCATTT CTTTCCAGAC TTGTTCAACA
GGCCAGCCAT
5641 TACGCTCGTC ATCAAAATCA CTCGCATCAA CCAAACCGTT ATTCATTCGT
GATTGCGCCT
5701 GAGCGAGACG AAATACGCGA TCGCTGTTAA AAGGACAATT ACAAACAGGA
ATCGAATGCA
5761 ACCGGCGCAG GAACACTGCC AGCGCATCAA CAATATTTTC ACCTGAATCA
GGATATTCTT
5821 CTAATACCTG GAATGCTGTT TTTCCGGGGA TCGCAGTGGT GAGTAACCAT
GCATCATCAG
5881 GAGTACGGAT AAAATGCTTG ATGGTCGGAA GAGGCATAAA TTCCGTCAGC
CAGTTTAGTC
5941 TGACCATCTC ATCTGTAACA TCATTGGCAA CGCTACCTTT GCCATGTTTC
AGAAACAACT
6001 CTGGCGCATC GGGCTTCCCA TACAATCGAT AGATTGTCGC ACCTGATTGC
CCGACATTAT
6061 CGCGAGCCCA TTTATACCCA TATAAATCAG CATCCATGTT GGAATTTAAT
CGCGGCCTAG
6121 AGCAAGACGT TTCCCGTTGA ATATGGCTCA TAACACCCCT TGTATTACTG
TTTATGTAAG
6181 CAGACAGTTT TATTGTTCAT GATGATATAT TTTTATCTTG TGCAATGTAA
CATCAGAGAT
6241 TTTGAGACAC AACAATTGGT CGAC

SEQ ID NO: 5

1 ATTGATTATT GACTAGTTAT TAATAGTAAT CAATTACGGG GTCATTAGTT
CATAGCCCAT
61 ATATGGAGTT CCGCGTTACA TAACTTACGG TAAATGGCCC GCCTGGCTGA
CCGCCCAACG
121 ACCCCCGCCC ATTGACGTCA ATAATGACGT ATGTTCCCAT AGTAACGCCA
ATAGGGACTT
181 TCCATTGACG TCAATGGGTG GAGTATTTAC GGTAAACTGC CCACTTGGCA
GTACATCAAG
241 TGTATCATAT GCCAAGTACG CCCCCTATTG ACGTCAATGA CGGTAAATGG
CCCGCCTGGC
301 ATTATGCCCA GTACATGACC TTATGGGACT TTCCTACTTG GCAGTACATC
TACGTATTAG
361 TCATCGCTAT TACCATGGTC GAGGTGAGCC CCACGTTCTG CTTCACTCTC
CCCATCTCCC
421 CCCCCTCCCC ACCCCCAATT TTGTATTTAT TTATTTTTTA ATTATTTTGT
GCAGCGATGG
481 GGGCGGGGGG GGGGGGGGGG CGCGCGCCAG GCGGGGCGGG GCGGGGCGAG
GGGCGGGGCG
541 GGGCGAGGCG GAAAGGTGCG GCGGCAGCCA ATCAAAGCGG CGCGCTCCGA
AAGTTTCCTT
601 TTATGGCGAG GCGGCGGCGG CGGCGGCCCT ATAAAAAGCG AAGCGCGCGG
CGGGCGGGAG

661 TCGCTGCGCG CTGCCTTCGC CCCGTGCCCC GCTCCGCCGC CGCCTCGCGC CGCCCGCCCC
721 GGCTCTGACT GACCGCGTTA CTCCCACAGG TGAGCGGGCG GGACGGCCCT TCTCCTCCGG
781 GCTGTAATTA GCGCTTGGTT TAATGACGGC TTGTTTCTTT TCTGTGGCTG CGTGAAAGCC
841 TTGAGGGGCT CCGGGAGGGC CCTTTGTGCG GGGGGAGCGG CTCGGGGGGT GCGTGCGTGT
901 GTGTGTGCGT GGGGAGCGCC GCGTGCGGCT CCGCGCTGCC CGGCGGCTGT GAGCGCTGCG
961 GGCGCGGCGC GGGGCTTTGT GCGCTCCGCA GTGTGCGCGA GGGGAGCGCG GCCGGGGGCG
1021 GTGCCCCGCG GTGCGGGGGG GGCTGCGAGG GGAACAAAGG CTGCGTGCGG GGTGTGTGCG
1081 TGGGGGGGTG AGCAGGGGGT GTGGGCGCGT CGGTCGGGCT GCAACCCCCC CTGCACCCCC
1141 CTCCCCGAGT TGCTGAGCAC GGCCCGGCTT CGGGTGCGGG GCTCCGTACG GGGCGTGGCG
1201 CGGGGCTCGC CGTGCCGGGC GGGGGGTGGC GGCAGGTGGG GGTGCCGGGC GGGGCGGGGC
1261 CGCCTCGGGC CGGGGAGGGC TCGGGGGAGG GGCGCGGCGG CCCCCGGAGC GCCGGCGGCT
1321 GTCGAGGCGC GGCGAGCCGC AGCCATTGCC TTTTATGGTA ATCGTGCGAG AGGGCGCAGG
1381 GACTTCCTTT GTCCCAAATC TGTGCGGAGC CGAAATCTGG GAGGCGCCGC CGCACCCCCT
1441 CTAGCGGGCG CGGGGCGAAG CGGTGCGGCG CCGGCAGGAA GGAAATGGGC GGGGAGGGCC
1501 TTCGTGCGTC GCCGCGCCGC CGTCCCCTTC TCCCTCTCCA GCCTCGGGGC TGTCCGCGGG
1561 GGGACGGGGC AGGGCGGGGT TCGGCTTCTG GCGTGTGACC GGCGGCTCTA GAGCCTCTGC
1621 TAACCATGTT CATGCCTTCT TCTTTTTCCT ACAGCTCCTG GGCAACGTGC TGGTTATTGT
1681 GCTGTCTCAT CATTTTGGCA AAGAATTCCT CGAGCATGTG GTCTGAGTTA AAAATCAGGA
1741 GCAACGACGG AGGTGAAGGA CCAGAGGACG CCAACGACCC CCGGGGAAAG GGGGTGCAAC
1801 ACATCCATAT CCAGCCATCT CTACCTGTTT ATGGACAGAG GGTTAGGGAT GGTGATAGGG
1861 GCAAACGTGA CTCGTACTGG TCTACTTCTC CTAGTGGTAG CACCACAAAA CCAGCATCAG
1921 GTTGGGAGAG GTCAAGTAAA GCCGACACAT GGTTGCTGAT TCTCTCATTC ACCCAGTGGG
1981 CTTTGTCAAT TGCCACAGTG ATCATCTGTA TCATAATTTC TGCTAGACAA GGGTATAGTA
2041 TGAAAGAGTA CTCAATGACT GTAGAGGCAT TGAACATGAG CAGCAGGGAG GTGAAAGAGT
2101 CACTTACCAG TCTAATAAGG CAAGAGGTTA TAGCAAGGGC TGTCAACATT CAGAGCTCTG
2161 TGCAAACCGG AATCCCAGTC TTGTTGAACA AAAACAGCAG GGATGTCATC CAGATGATTG
2221 ATAAGTCGTG CAGCAGACAA GAGCTCACTC AGCACTGTGA GAGTACGATC GCAGTCCACC
2281 ATGCCGATGG AATTGCCCCA CTTGAGCCAC ATAGTTTCTG GAGATGCCCT GTCGGAGAAC
2341 CGTATCTTAG CTCAGATCCT GAAATCTCAT TGCTGCCTGG TCCGAGCTTG TTATCTGGTT

2401 CTACAACGAT CTCTGGATGT GTTAGGCTCC CTTCACTCTC AATTGGCGAG
GCAATCTATG
2461 CCTATTCATC AAATCTCATT ACACAAGGTT GTGCTGACAT AGGGAAATCA
TATCAGGTCC
2521 TGCAGCTAGG GTACATATCA CTCAATTCAG ATATGTTCCC TGATCTTAAC
CCCGTAGTGT
2581 CCCACACTTA TGACATCAAC GACAATCGGA AATCATGCTC TGTGGTGGCA
ACCGGGACTA
2641 GGGGTTATCA GCTTTGCTCC ATGCCGACTG TAGACGAAAG AACCGACTAC
TCTAGTGATG
2701 GTATTGAGGA TCTGGTCCTT GATGTCCTGG ATCTCAAAGG GAGAACTAAG
TCTCACCGGT
2761 ATCGCAACAG CGAGGTAGAT CTTGATCACC CGTTCTCTGC ACTATACCCC
AGTGTAGGCA
2821 ACGGCATTGC AACAGAAGGC TCATTGATAT TTCTTGGGTA TGGTGGACTA
ACCACCCCTC
2881 TGCAGGGTGA TACAAAATGT AGGACCCAAG GATGCCAACA GGTGTCGCAA
GACACATGCA
2941 ATGAGGCTCT GAAAATTACA TGGCTAGGAG GGAAACAGGT GGTCAGCGTG
ATCATCCAGG
3001 TCAATGACTA TCTCTCAGAG AGGCCAAAGA TAAGAGTCAC AACCATTCCA
ATCACTCAAA
3061 ACTATCTCGG GGCGGAAGGT AGATTATTAA AATTGGGTGA TCGGGTGTAC
ATCTATACAA
3121 GATCATCAGG CTGGCACTCT CAACTGCAGA TAGGAGTACT TGATGTCAGC
CACCCTTTGA
3181 CTATCAACTG GACACCTCAT GAAGCCTTGT CTAGACCAGG AAATAAAGAG
TGCAATTGGT
3241 ACAATAAGTG TCCGAAGGAA TGCATATCAG GCGTATACAC TGATGCTTAT
CCATTGTCCC
3301 CTGATGCAGC TAACGTCGCT ACCGTCACGC TATATGCCAA TACATCGCGT
GTCAACCCAA
3361 CAATCATGTA TTCTAACACT ACTAACATTA TAAATATGTT AAGGATAAAG
GATGTTCAAT
3421 TAGAGGCTGC ATATACCACG ACATCGTGTA TCACGCATTT TGGTAAAGGC
TACTGCTTTC
3481 ACATCATCGA GATCAATCAG AAGAGCCTGA ATACCTTACA GCCGATGCTC
TTTAAGACTA
3541 GCATCCCTAA ATTATGCAAG GCCGAGTCTT AAGCGGCCGC GCATGCGAAT
TCACTCCTCA
3601 GGTGCAGGCT GCCTATCAGA AGGTGGTGGC TGGTGTGGCC AATGCCCTGG
CTCACAAATA
3661 CCACTGAGAT CTTTTTCCCT CTGCCAAAAA TTATGGGGAC ATCATGAAGC
CCCTTGAGCA
3721 TCTGACTTCT GGCTAATAAA GGAAATTTAT TTTCATTGCA ATAGTGTGTT
GGAATTTTTT
3781 GTGTCTCTCA CTCGGAAGGA CATATGGGAG GGCAAATCAT TTAAAACATC
AGAATGAGTA
3841 TTTGGTTTAG AGTTTGGCAA CATATGCCCA TATGCTGGCT GCCATGAACA
AAGGTTGGCT
3901 ATAAAGAGGT CATCAGTATA TGAAACAGCC CCCTGCTGTC TATTCCTTAT
TCCATAGAAA
3961 AGCCTTGACT TGAGGTTAGA TTTTTTTTAT ATTTTGTTTT GTGTTATTTT
TTTCTTTAAC
4021 ATCCCTAAAA TTTTCCTTAC ATGTTTTACT AGCCAGATTT TTCCTCCTCT
CCTGACTACT
4081 CCCAGTCATA GCTGTCCCTC TTCTCTTATG GAGATCCCTC GACCTGCAGC
CCAAGCTTGG

4141 CGTAATCATG GTCATAGCTG TTTCCTGTGT GAAATTGTTA TCCGCTCACA
ATTCCACACA
4201 ACATACGAGC CGGAAGCATA AAGTGTAAAG CCTGGGGTGC CTAATGAGTG
AGCTAACTCA
4261 CATTAATTGC GTTGCGCTCA CTGCCCGCTT TCCAGTCGGG AAACCTGTCG
TGCCAGCGGA
4321 TCCGCATCTC AATTAGTCAG CAACCATAGT CCCGCCCCTA ACTCCGCCCA
TCCCGCCCCT
4381 AACTCCGCCC AGTTCCGCCC ATTCTCCGCC CCATGGCTGA CTAATTTTTT
TTATTTATGC
4441 AGAGGCCGAG GCCGCCTCGG CCTCTGAGCT ATTCCAGAAG TAGTGAGGAG
GCTTTTTTGG
4501 AGGCCTAGGC TTTTGCAAAA AGCTAACTTG TTTATTGCAG CTTATAATGG
TTACAAATAA
4561 AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTT CACTGCATTC
TAGTTGTGGT
4621 TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGTCCGC TTCCTCGCTC
ACTGACTCGC
4681 TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG
GTAATACGGT
4741 TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC
CAGCAAAAGG
4801 CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC
CCCCCTGACG
4861 AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA
CTATAAAGAT
4921 ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC
CTGCCGCTTA
4981 CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAT
AGCTCACGCT
5041 GTAGGTATCT CAGTTCGGTG TAGGTCGTTC GCTCCAAGCT GGGCTGTGTG
CACGAACCCC
5101 CCGTTCAGCC CGACCGCTGC GCCTTATCCG GTAACTATCG TCTTGAGTCC
AACCCGGTAA
5161 GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA
GCGAGGTATG
5221 TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT
AGAAGAACAG
5281 TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT
GGTAGCTCTT
5341 GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG
CAGCAGATTA
5401 CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG
TCTGACGCTC
5461 AGTGGAACGA AAACTCACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA
AGGATCTTCA
5521 CCTAGATCCT TTTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA
TATGAGTAAA
5581 CTTGGTCTGA CAGTTAGAAA AACTCATCGA GCATCAAATG AAACTGCAAT
TTATTCATAT
5641 CAGGATTATC AATACCATAT TTTTGAAAAA GCCGTTTCTG TAATGAAGGA
GAAAACTCAC
5701 CGAGGCAGTT CCATAGGATG GCAAGATCCT GGTATCGGTC TGCGATTCCG
ACTCGTCCAA
5761 CATCAATACA ACCTATTAAT TTCCCCTCGT CAAAAATAAG GTTATCAAGT
GAGAAATCAC
5821 CATGAGTGAC GACTGAATCC GGTGAGAATG GCAACAGCTT ATGCATTTCT
TTCCAGACTT

5881 GTTCAACAGG CCAGCCATTA CGCTCGTCAT CAAAATCACT CGCATCAACC
AAACCGTTAT
5941 TCATTCGTGA TTGCGCCTGA GCGAGACGAA ATACGCGATC GCTGTTAAAA
GGACAATTAC
6001 AAACAGGAAT CGAATGCAAC CGGCGCAGGA ACACTGCCAG CGCATCAACA
ATATTTTCAC
6061 CTGAATCAGG ATATTCTTCT AATACCTGGA ATGCTGTTTT TCCGGGGATC
GCAGTGGTGA
6121 GTAACCATGC ATCATCAGGA GTACGGATAA AATGCTTGAT GGTCGGAAGA
GGCATAAATT
6181 CCGTCAGCCA GTTTAGTCTG ACCATCTCAT CTGTAACATC ATTGGCAACG
CTACCTTTGC
6241 CATGTTTCAG AAACAACTCT GGCGCATCGG GCTTCCCATA CAATCGATAG
ATTGTCGCAC
6301 CTGATTGCCC GACATTATCG CGAGCCCATT TATACCCATA TAAATCAGCA
TCCATGTTGG
6361 AATTTAATCG CGGCCTAGAG CAAGACGTTT CCCGTTGAAT ATGGCTCATA
ACACCCCTTG
6421 TATTACTGTT TATGTAAGCA GACAGTTTTA TTGTTCATGA TGATATATTT
TTATCTTGTG
6481 CAATGTAACA TCAGAGATTT TGAGACACAA CAATTGGTCG AC

SEQ ID NO: 6

1 AGATCTGTTA CATAACTTAT GGTAAATGGC CTGCCTGGCT GACTGCCCAA
TGACCCCTGC
61 CCAATGATGT CAATAATGAT GTATGTTCCC ATGTAATGCC AATAGGGACT
TTCCATTGAT
121 GTCAATGGGT GGAGTATTTA TGGTAACTGC CCACTTGGCA GTACATCAAG
TGTATCATAT
181 GCCAAGTATG CCCCCTATTG ATGTCAATGA TGGTAAATGG CCTGCCTGGC
ATTATGCCCA
241 GTACATGACC TTATGGGACT TTCCTACTTG GCAGTACATC TATGTATTAG
TCATTGCTAT
301 TACCATGGGA ATTCACTAGT GGAGAAGAGC ATGCTTGAGG GCTGAGTGCC
CCTCAGTGGG
361 CAGAGAGCAC ATGGCCCACA GTCCCTGAGA AGTTGGGGGG AGGGGTGGGC
AATTGAACTG
421 GTGCCTAGAG AAGGTGGGGC TTGGGTAAAC TGGGAAAGTG ATGTGGTGTA
CTGGCTCCAC
481 CTTTTTCCCC AGGGTGGGGG AGAACCATAT ATAAGTGCAG TAGTCTCTGT
GAACATTCAA
541 GCTTCTGCCT TCTCCCTCCT GTGAGTTTGC TAGC

SEQ ID NO: 7

1 GCTAGCCACC ATGCAGAGAA GCCCTCTGGA GAAGGCCTCT GTGGTGAGCA
AGCTGTTCTT
61 CAGCTGGACC AGGCCCATCC TGAGGAAGGG CTACAGGCAG AGACTGGAGC
TGTCTGACAT
121 CTACCAGATC CCCTCTGTGG ACTCTGCTGA CAACCTGTCT GAGAAGCTGG
AGAGGGAGTG
181 GGATAGAGAG CTGGCCAGCA AGAAGAACCC CAAGCTGATC AATGCCCTGA
GGAGATGCTT
241 CTTCTGGAGA TTCATGTTCT ATGGCATCTT CCTGTACCTG GGGGAAGTGA
CCAAGGCTGT

301 GCAGCCTCTG CTGCTGGGCA GAATCATTGC CAGCTATGAC CCTGACAACA AGGAGGAGAG
361 GAGCATTGCC ATCTACCTGG GCATTGGCCT GTGCCTGCTG TTCATTGTGA GGACCCTGCT
421 GCTGCACCCT GCCATCTTTG GCCTGCACCA CATTGGCATG CAGATGAGGA TTGCCATGTT
481 CAGCCTGATC TACAAGAAAA CCCTGAAGCT GTCCAGCAGA GTGCTGGACA AGATCAGCAT
541 TGGCCAGCTG GTGAGCCTGC TGAGCAACAA CCTGAACAAG TTTGATGAGG GCCTGGCCCT
601 GGCCCACTTT GTGTGGATTG CCCCTCTGCA GGTGGCCCTG CTGATGGGCC TGATTTGGGA
661 GCTGCTGCAG GCCTCTGCCT TTTGTGGCCT GGGCTTCCTG ATTGTGCTGG CCCTGTTTCA
721 GGCTGGCCTG GGCAGGATGA TGATGAAGTA CAGGGACCAG AGGGCAGGCA AGATCAGTGA
781 GAGGCTGGTG ATCACCTCTG AGATGATTGA GAACATCCAG TCTGTGAAGG CCTACTGTTG
841 GGAGGAAGCT ATGGAGAAGA TGATTGAAAA CCTGAGGCAG ACAGAGCTGA AGCTGACCAG
901 GAAGGCTGCC TATGTGAGAT ACTTCAACAG CTCTGCCTTC TTCTTCTCTG GCTTCTTTGT
961 GGTGTTCCTG TCTGTGCTGC CCTATGCCCT GATCAAGGGG ATCATCCTGA GAAAGATTTT
1021 CACCACCATC AGCTTCTGCA TTGTGCTGAG GATGGCTGTG ACCAGACAGT TCCCCTGGGC
1081 TGTGCAGACC TGGTATGACA GCCTGGGGGC CATCAACAAG ATCCAGGACT TCCTGCAGAA
1141 GCAGGAGTAC AAGACCCTGG AGTACAACCT GACCACCACA GAAGTGGTGA TGGAGAATGT
1201 GACAGCCTTC TGGGAGGAGG GCTTTGGGGA GCTGTTTGAG AAGGCCAAGC AGAACAACAA
1261 CAACAGAAAG ACCAGCAATG GGGATGACTC CCTGTTCTTC TCCAACTTCT CCCTGCTGGG
1321 CACACCTGTG CTGAAGGACA TCAACTTCAA GATTGAGAGG GGGCAGCTGC TGGCTGTGGC
1381 TGGATCTACA GGGGCTGGCA AGACCAGCCT GCTGATGATG ATCATGGGGG AGCTGGAGCC
1441 TTCTGAGGGC AAGATCAAGC ACTCTGGCAG GATCAGCTTT TGCAGCCAGT TCAGCTGGAT
1501 CATGCCTGGC ACCATCAAGG AGAACATCAT CTTTGGAGTG AGCTATGATG AGTACAGATA
1561 CAGGAGTGTG ATCAAGGCCT GCCAGCTGGA GGAGGACATC AGCAAGTTTG CTGAGAAGGA
1621 CAACATTGTG CTGGGGGAGG GAGGCATTAC ACTGTCTGGG GGCCAGAGAG CCAGAATCAG
1681 CCTGGCCAGG GCTGTGTACA AGGATGCTGA CCTGTACCTG CTGGACTCCC CCTTTGGCTA
1741 CCTGGATGTG CTGACAGAGA AGGAGATTTT TGAGAGCTGT GTGTGCAAGC TGATGGCCAA
1801 CAAGACCAGA ATCCTGGTGA CCAGCAAGAT GGAGCACCTG AAGAAGGCTG ACAAGATCCT
1861 GATCCTGCAT GAGGGCAGCA GCTACTTCTA TGGGACCTTC TCTGAGCTGC AGAACCTGCA
1921 GCCTGACTTC AGCTCTAAGC TGATGGGCTG TGACAGCTTT GACCAGTTCT CTGCTGAGAG
1981 GAGGAACAGC ATCCTGACAG AGACCCTGCA CAGATTCAGC CTGGAGGGAG ATGCCCCTGT

2041 GAGCTGGACA GAGACCAAGA AGCAGAGCTT CAAGCAGACA GGGGAGTTTG GGGAGAAGAG
2101 GAAGAACTCC ATCCTGAACC CCATCAACAG CATCAGGAAG TTCAGCATTG TGCAGAAAAC
2161 CCCCCTGCAG ATGAATGGCA TTGAGGAAGA TTCTGATGAG CCCCTGGAGA GGAGACTGAG
2221 CCTGGTGCCT GATTCTGAGC AGGGAGAGGC CATCCTGCCT AGGATCTCTG TGATCAGCAC
2281 AGGCCCTACA CTGCAGGCCA GAAGGAGGCA GTCTGTGCTG AACCTGATGA CCCACTCTGT
2341 GAACCAGGGC CAGAACATCC ACAGGAAAAC CACAGCCTCC ACCAGGAAAG TGAGCCTGGC
2401 CCCTCAGGCC AATCTGACAG AGCTGGACAT CTACAGCAGG AGGCTGTCTC AGGAGACAGG
2461 CCTGGAGATT TCTGAGGAGA TCAATGAGGA GGACCTGAAA GAGTGCTTCT TTGATGACAT
2521 GGAGAGCATC CCTGCTGTGA CCACCTGGAA CACCTACCTG AGATACATCA CAGTGCACAA
2581 GAGCCTGATC TTTGTGCTGA TCTGGTGCCT GGTGATCTTC CTGGCTGAAG TGGCTGCCTC
2641 TCTGGTGGTG CTGTGGCTGC TGGGAAACAC CCCACTGCAG GACAAGGGCA ACAGCACCCA
2701 CAGCAGGAAC AACAGCTATG CTGTGATCAT CACCTCCACC TCCAGCTACT ATGTGTTCTA
2761 CATCTATGTG GGAGTGGCTG ATACCCTGCT GGCTATGGGC TTCTTTAGAG GCCTGCCCCT
2821 GGTGCACACA CTGATCACAG TGAGCAAGAT CCTCCACCAC AAGATGCTGC ACTCTGTGCT
2881 GCAGGCTCCT ATGAGCACCC TGAATACCCT GAAGGCTGGG GGCATCCTGA ACAGATTCTC
2941 CAAGGATATT GCCATCCTGG ATGACCTGCT GCCTCTCACC ATCTTTGACT TCATCCAGCT
3001 GCTGCTGATT GTGATTGGGG CCATTGCTGT GGTGGCAGTG CTGCAGCCCT ACATCTTTGT
3061 GGCCACAGTG CCTGTGATTG TGGCCTTCAT CATGCTGAGG GCCTACTTTC TGCAGACCTC
3121 CCAGCAGCTG AAGCAGCTGG AGTCTGAGGG CAGAAGCCCC ATCTTCACCC ACCTGGTGAC
3181 AAGCCTGAAG GGCCTGTGGA CCCTGAGAGC CTTTGGCAGG CAGCCCTACT TTGAGACCCT
3241 GTTCCACAAG GCCCTGAACC TGCACACAGC CAACTGGTTC CTCTACCTGT CCACCCTGAG
3301 ATGGTTCCAG ATGAGAATTG AGATGATCTT TGTCATCTTC TTCATTGCTG TGACCTTCAT
3361 CAGCATTCTG ACCACAGGAG AGGGAGAGGG CAGAGTGGGC ATTATCCTGA CCCTGGCCAT
3421 GAACATCATG AGCACACTGC AGTGGGCAGT GAACAGCAGC ATTGATGTGG ACAGCCTGAT
3481 GAGGAGTGTG AGCAGAGTGT TCAAGTTCAT TGATATGCCC ACAGAGGGCA AGCCTACCAA
3541 GAGCACCAAG CCCTACAAGA ATGGCCAGCT GAGCAAAGTG ATGATCATTG AGAACAGCCA
3601 TGTGAAGAAG GATGATATCT GGCCCAGTGG AGGCCAGATG ACAGTGAAGG ACCTGACAGC
3661 CAAGTACACA GAGGGGGGCA ATGCTATCCT GGAGAACATC TCCTTCAGCA TCTCCCCTGG
3721 CCAGAGAGTG GGACTGCTGG GAAGAACAGG CTCTGGCAAG TCTACCCTGC TGTCTGCCTT

3781 CCTGAGGCTG CTGAACACAG AGGGAGAGAT CCAGATTGAT GGAGTGTCCT
GGGACAGCAT
3841 CACACTGCAG CAGTGGAGGA AGGCCTTTGG TGTGATCCCC CAGAAAGTGT
TCATCTTCAG
3901 TGGCACCTTC AGGAAGAACC TGGACCCCTA TGAGCAGTGG TCTGACCAGG
AGATTTGGAA
3961 AGTGGCTGAT GAAGTGGGCC TGAGAAGTGT GATTGAGCAG TTCCCTGGCA
AGCTGGACTT
4021 TGTCCTGGTG GATGGGGGCT GTGTGCTGAG CCATGGCCAC AAGCAGCTGA
TGTGCCTGGC
4081 CAGATCAGTG CTGAGCAAGG CCAAGATCCT GCTGCTGGAT GAGCCTTCTG
CCCACCTGGA
4141 TCCTGTGACC TACCAGATCA TCAGGAGGAC CCTCAAGCAG GCCTTTGCTG
ACTGCACAGT
4201 CATCCTGTGT GAGCACAGGA TTGAGGCCAT GCTGGAGTGC CAGCAGTTCC
TGGTGATTGA
4261 GGAGAACAAA GTGAGGCAGT ATGACAGCAT CCAGAAGCTG CTGAATGAGA
GGAGCCTGTT
4321 CAGGCAGGCC ATCAGCCCCT CTGATAGAGT GAAGCTGTTC CCCCACAGGA
ACAGCTCCAA
4381 GTGCAAGAGC AAGCCCCAGA TTGCTGCCCT GAAGGAGGAG ACAGAGGAGG
AAGTGCAGGA
4441 CACCAGGCTG TGAGGGCCC

SEQ ID NO: 8

1 GGGCCCAATC AACCTCTGGA TTACAAAATT TGTGAAAGAT TGACTGGTAT
TCTTAACTAT
61 GTTGCTCCTT TTACGCTATG TGGATACGCT GCTTTAATGC CTTTGTATCA
TGCTATTGCT
121 TCCCGTATGG CTTTCATTTT CTCCTCCTTG TATAAATCCT GGTTGCTGTC
TCTTTATGAG
181 GAGTTGTGGC CCGTTGTCAG GCAACGTGGC GTGGTGTGCA CTGTGTTTGC
TGACGCAACC
241 CCCACTGGTT GGGGCATTGC CACCACCTGT CAGCTCCTTT CCGGGACTTT
CGCTTTCCCC
301 CTCCCTATTG CCACGGCGGA ACTCATCGCC GCCTGCCTTG CCCGCTGCTG
GACAGGGGCT
361 CGGCTGTTGG GCACTGACAA TTCCGTGGTG TTGTCGGGGA AATCATCGTC
CTTTCCTTGG
421 CTGCTCGCCT GTGTTGCCAC CTGGATTCTG CGCGGGACGT CCTTCTGCTA
CGTCCCTTCG
481 GCCCTCAATC CAGCGGACCT TCCTTCCCGC GGCCTGCTGC CGGCTCTGCG
GCCTCTTCCG
541 CGTCTTCGCC TTCGCCCTCA GACGAGTCGG ATCTCCCTTT GGGCCGCCTC
CCCGCAAGCT

SEQ ID NO: 9

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT
AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT
ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT
AGTAATCAAT

181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
TTACGGTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA
TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT
ATTTACGGTA
361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC
CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAC
GGGACTTTCC
481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
GGTTTTGGCA
541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC
TCCACCCCAT
601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA
AATGTCGTAA
661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG
GTGGGAGGTC
721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC
AGCTTGAGCC
781 TGGGTGTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT
CCTTGGCTTA
841 GAAAGCTAAT AAACTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCT
CATTGACGCC
901 TCACTCTCTT GAACGGGAAT CTTCCTTACT GGGTTCTCTC TCTGACCCAG
GCGAGAGAAA
961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA
CAGCTGAGAA
1021 GGCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC
TTGGTGAGTA
1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG
AGGCCGTAGC
1141 CGTAACTACT CTTGGGCAAG TAGGGCAGGC GGTGGGTACG CAATGGGGGC
GGCTACCTCA
1201 GCACTAAATA GGAGACAATT AGACCAATTT GAGAAAATAC GACTTCGCCC
GAACGGAAAG
1261 AAAAAGTACC AAATTAAACA TTTAATATGG GCAGGCAAGG AGATGGAGCG
CTTCGGCCTC
1321 CATGAGAGGT TGTTGGAGAC AGAGGAGGGG TGTAAAAGAA TCATAGAAGT
CCTCTACCCC
1381 CTAGAACCAA CAGGATCGGA GGGCTTAAAA AGTCTGTTCA ATCTTGTGTG
CGTGCTATAT
1441 TGCTTGCACA AGGAACAGAA AGTGAAAGAC ACAGAGGAAG CAGTAGCAAC
AGTAAGACAA
1501 CACTGCCATC TAGTGGAAAA AGAAAAAAGT GCAACAGAGA CATCTAGTGG
ACAAAAGAAA
1561 AATGACAAGG GAATAGCAGC GCCACCTGGT GGCAGTCAGA ATTTTCCAGC
GCAACAACAA
1621 GGAAATGCCT GGGTACATGT ACCCTTGTCA CCGCGCACCT TAAATGCGTG
GGTAAAAGCA
1681 GTAGAGGAGA AAAAATTTGG AGCAGAAATA GTACCCATTT TTTTGTTTCA
AGCCCTATCG
1741 AATTCCCGTT TGTGCTAGGG TTCTTAGGCT TCTTGGGGGC TGCTGGAACT
GCAATGGGAG
1801 CAGCGGCGAC AGCCCTGACG GTCCAGTCTC AGCATTTGCT TGCTGGGATA
CTGCAGCAGC
1861 AGAAGAATCT GCTGGCGGCT GTGGAGGCTC AACAGCAGAT GTTGAAGCTG
ACCATTTGGG

1921 GTGTTAAAAA CCTCAATGCC CGCGTCACAG CCCTTGAGAA GTACCTAGAG
GATCAGGCAC
1981 GACTAAACTC CTGGGGGTGC GCATGGAAAC AAGTATGTCA TACCACAGTG
GAGTGGCCCT
2041 GGACAAATCG GACTCCGGAT TGGCAAAATA TGACTTGGTT GGAGTGGGAA
AGACAAATAG
2101 CTGATTTGGA AAGCAACATT ACGAGACAAT TAGTGAAGGC TAGAGAACAA
GAGGAAAAGA
2161 ATCTAGATGC CTATCAGAAG TTAACTAGTT GGTCAGATTT CTGGTCTTGG
TTCGATTTCT
2221 CAAAATGGCT TAACATTTTA AAAATGGGAT TTTTAGTAAT AGTAGGAATA
ATAGGGTTAA
2281 GATTACTTTA CACAGTATAT GGATGTATAG TGAGGGTTAG GCAGGGATAT
GTTCCTCTAT
2341 CTCCACAGAT CCATATCCGC GGCAATTTTA AAAGAAAGGG AGGAATAGGG
GGACAGACTT
2401 CAGCAGAGAG ACTAATTAAT ATAATAACAA CACAATTAGA AATACAACAT
TTACAAACCA
2461 AAATTCAAAA AATTTTAAAT TTTAGAGCCG CGGAGATCTG TTACATAACT
TATGGTAAAT
2521 GGCCTGCCTG GCTGACTGCC CAATGACCCC TGCCCAATGA TGTCAATAAT
GATGTATGTT
2581 CCCATGTAAT GCCAATAGGG ACTTTCCATT GATGTCAATG GGTGGAGTAT
TTATGGTAAC
2641 TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ATGCCCCCTA
TTGATGTCAA
2701 TGATGGTAAA TGGCCTGCCT GGCATTATGC CCAGTACATG ACCTTATGGG
ACTTTCCTAC
2761 TTGGCAGTAC ATCTATGTAT TAGTCATTGC TATTACCATG GGAATTCACT
AGTGGAGAAG
2821 AGCATGCTTG AGGGCTGAGT GCCCCTCAGT GGGCAGAGAG CACATGGCCC
ACAGTCCCTG
2881 AGAAGTTGGG GGGAGGGGTG GGCAATTGAA CTGGTGCCTA GAGAAGGTGG
GGCTTGGGTA
2941 AACTGGGAAA GTGATGTGGT GTACTGGCTC CACCTTTTTC CCCAGGGTGG
GGGAGAACCA
3001 TATATAAGTG CAGTAGTCTC TGTGAACATT CAAGCTTCTG CCTTCTCCCT
CCTGTGAGTT
3061 TGCTAGCCAC CATGCCCAGC TCTGTGTCCT GGGGCATTCT GCTGCTGGCT
GGCCTGTGCT
3121 GTCTGGTGCC TGTGTCCCTG GCTGAGGACC CTCAGGGGGA TGCTGCCCAG
AAAACAGACA
3181 CCTCCCACCA TGACCAGGAC CACCCCACCT TCAACAAGAT CACCCCCAAC
CTGGCAGAGT
3241 TTGCCTTCAG CCTGTACAGA CAGCTGGCCC ACCAGAGCAA CAGCACCAAC
ATCTTTTTCA
3301 GCCCTGTGTC CATTGCCACA GCCTTTGCCA TGCTGAGCCT GGGCACCAAG
GCTGACACCC
3361 ATGATGAGAT CCTGGAAGGC CTGAACTTCA ACCTGACAGA GATCCCTGAG
GCCCAGATCC
3421 ATGAGGGCTT CCAGGAACTG CTGAGAACCC TGAACCAGCC AGACAGCCAG
CTGCAGCTGA
3481 CAACAGGCAA TGGGCTGTTC CTGTCTGAGG GCCTGAAGCT GGTGGACAAG
TTTCTGGAAG
3541 ATGTGAAGAA GCTGTACCAC TCTGAGGCCT TCACAGTGAA CTTTGGGGAC
ACAGAAGAGG
3601 CCAAGAAACA GATCAATGAC TATGTGGAAA AGGGCACCCA GGGCAAGATT
GTGGACCTTG

3661 TGAAAGAGCT GGACAGGGAC ACTGTGTTTG CCCTTGTGAA CTACATCTTC
TTCAAGGGCA
3721 AGTGGGAGAG GCCCTTTGAA GTGAAGGACA CTGAGGAAGA GGACTTCCAT
GTGGACCAAG
3781 TGACCACAGT GAAGGTGCCA ATGATGAAGA GACTGGGGAT GTTCAATATC
CAGCACTGCA
3841 AGAAACTGAG CAGCTGGGTG CTGCTGATGA AGTACCTGGG CAATGCTACA
GCCATATTCT
3901 TTCTGCCTGA TGAGGGCAAG CTGCAGCACC TGGAAAATGA GCTGACCCAT
GACATCATCA
3961 CCAAATTTCT GGAAAATGAG GACAGAAGAT CTGCCAGCCT GCATCTGCCC
AAGCTGAGCA
4021 TCACAGGCAC ATATGACCTG AAGTCTGTGC TGGGACAGCT GGGAATCACC
AAGGTGTTCA
4081 GCAATGGGGC AGACCTGAGT GGAGTGACAG AGGAAGCCCC TCTGAAGCTG
TCCAAGGCTG
4141 TGCACAAGGC AGTGCTGACC ATTGATGAGA AGGGCACAGA GGCTGCTGGG
GCCATGTTTC
4201 TGGAAGCCAT CCCCATGTCC ATCCCCCCAG AAGTGAAGTT CAACAAGCCC
TTTGTGTTCC
4261 TGATGATTGA GCAGAACACC AAGAGCCCCC TGTTCATGGG CAAGGTTGTG
AACCCCACCC
4321 AGAAATGAGG GCCCAATCAA CCTCTGGATT ACAAAATTTG TGAAAGATTG
ACTGGTATTC
4381 TTAACTATGT TGCTCCTTTT ACGCTATGTG GATACGCTGC TTTAATGCCT
TTGTATCATG
4441 CTATTGCTTC CCGTATGGCT TTCATTTTCT CCTCCTTGTA TAAATCCTGG
TTGCTGTCTC
4501 TTTATGAGGA GTTGTGGCCC GTTGTCAGGC AACGTGGCGT GGTGTGCACT
GTGTTTGCTG
4561 ACGCAACCCC CACTGGTTGG GGCATTGCCA CCACCTGTCA GCTCCTTTCC
GGGACTTTCG
4621 CTTTCCCCCT CCCTATTGCC ACGGCGGAAC TCATCGCCGC CTGCCTTGCC
CGCTGCTGGA
4681 CAGGGGCTCG GCTGTTGGGC ACTGACAATT CCGTGGTGTT GTCGGGGAAA
TCATCGTCCT
4741 TTCCTTGGCT GCTCGCCTGT GTTGCCACCT GGATTCTGCG CGGGACGTCC
TTCTGCTACG
4801 TCCCTTCGGC CCTCAATCCA GCGGACCTTC CTTCCCGCGG CCTGCTGCCG
GCTCTGCGGC
4861 CTCTTCCGCG TCTTCGCCTT CGCCCTCAGA CGAGTCGGAT CTCCCTTTGG
GCCGCCTCCC
4921 CGCAAGCTTC GCACTTTTTA AAAGAAAAGG GAGGACTGGA TGGGATTTAT
TACTCCGATA
4981 GGACGCTGGC TTGTAACTCA GTCTCTTACT AGGAGACCAG CTTGAGCCTG
GGTGTTCGCT
5041 GGTTAGCCTA ACCTGGTTGG CCACCAGGGG TAAGGACTCC TTGGCTTAGA
AAGCTAATAA
5101 ACTTGCCTGC ATTAGAGCTC TTACGCGTCC CGGGCTCGAG ATCCGCATCT
CAATTAGTCA
5161 GCAACCATAG TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC
CAGTTCCGCC
5221 CATTCTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA
GGCCGCCTCG
5281 GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG
CTTTTGCAAA
5341 AAGCTAACTT GTTTATTGCA GCTTATAATG GTTACAAATA AAGCAATAGC
ATCACAAATT

5401 TCACAAATAA AGCATTTTTT TCACTGCATT CTAGTTGTGG TTTGTCCAAA CTCATCAATG
5461 TATCTTATCA TGTCTGTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG TCGTTCGGCT
5521 GCGGCGAGCG GTATCAGCTC ACTCAAAGGC GGTAATACGG TTATCCACAG AATCAGGGGA
5581 TAACGCAGGA AAGAACATGT GAGCAAAAGG CCAGCAAAAG GCCAGGAACC GTAAAAAGGC
5641 CGCGTTGCTG GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG
5701 CTCAAGTCAG AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
5761 AAGCTCCCTC GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCTT
5821 TCTCCCTTCG GGAAGCGTGG CGCTTTCTCA TAGCTCACGC TGTAGGTATC TCAGTTCGGT
5881 GTAGGTCGTT CGCTCCAAGC TGGGCTGTGT GCACGAACCC CCCGTTCAGC CCGACCGCTG
5941 CGCCTTATCC GGTAACTATC GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT
6001 GGCAGCAGCC ACTGGTAACA GGATTAGCAG AGCGAGGTAT GTAGGCGGTG CTACAGAGTT
6061 CTTGAAGTGG TGGCCTAACT ACGGCTACAC TAGAAGAACA GTATTTGGTA TCTGCGCTCT
6121 GCTGAAGCCA GTTACCTTCG GAAAAAGAGT TGGTAGCTCT TGATCCGGCA AACAAACCAC
6181 CGCTGGTAGC GGTGGTTTTT TTGTTTGCAA GCAGCAGATT ACGCGCAGAA AAAAAGGATC
6241 TCAAGAAGAT CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG
6301 TTAAGGGATT TTGGTCATGA GATTATCAAA AAGGATCTTC ACCTAGATCC TTTTAAATTA
6361 AAAATGAAGT TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTAGAA
6421 AAACTCATCG AGCATCAAAT GAAACTGCAA TTTATTCATA TCAGGATTAT CAATACCATA
6481 TTTTTGAAAA AGCCGTTTCT GTAATGAAGG AGAAAACTCA CCGAGGCAGT TCCATAGGAT
6541 GGCAAGATCC TGGTATCGGT CTGCGATTCC GACTCGTCCA ACATCAATAC AACCTATTAA
6601 TTTCCCCTCG TCAAAAATAA GGTTATCAAG TGAGAAATCA CCATGAGTGA CGACTGAATC
6661 CGGTGAGAAT GGCAACAGCT TATGCATTTC TTTCCAGACT TGTTCAACAG GCCAGCCATT
6721 ACGCTCGTCA TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTCGTG ATTGCGCCTG
6781 AGCGAGACGA AATACGCGAT CGCTGTTAAA AGGACAATTA CAAACAGGAA TCGAATGCAA
6841 CCGGCGCAGG AACACTGCCA GCGCATCAAC AATATTTTCA CCTGAATCAG GATATTCTTC
6901 TAATACCTGG AATGCTGTTT TTCCGGGGAT CGCAGTGGTG AGTAACCATG CATCATCAGG
6961 AGTACGGATA AAATGCTTGA TGGTCGGAAG AGGCATAAAT TCCGTCAGCC AGTTTAGTCT
7021 GACCATCTCA TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTTCA GAAACAACTC
7081 TGGCGCATCG GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC

7141 GCGAGCCCAT TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC
GCGGCCTAGA
7201 GCAAGACGTT TCCCGTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT
TTATGTAAGC
7261 AGACAGTTTT ATTGTTCATG ATGATATATT TTTATCTTGT GCAATGTAAC
ATCAGAGATT
7321 TTGAGACACA ACAATTGGTC GACGGATCC

SEQ ID NO: 10

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT
AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT
ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT
AGTAATCAAT
181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
TTACGGTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA
TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT
ATTTACGGTA
361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC
CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAC
GGGACTTTCC
481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
GGTTTTGGCA
541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC
TCCACCCCAT
601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA
AATGTCGTAA
661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG
GTGGGAGGTC
721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC
AGCTTGAGCC
781 TGGGTGTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT
CCTTGGCTTA
841 GAAAGCTAAT AAACTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCT
CATTGACGCC
901 TCACTCTCTT GAACGGGAAT CTTCCTTACT GGGTTCTCTC TCTGACCCAG
GCGAGAGAAA
961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA
CAGCTGAGAA
1021 GGCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC
TTGGTGAGTA
1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG
AGGCCGTAGC
1141 CGTAACTACT CTTGGGCAAG TAGGGCAGGC GGTGGGTACG CAATGGGGGC
GGCTACCTCA
1201 GCACTAAATA GGAGACAATT AGACCAATTT GAGAAAATAC GACTTCGCCC
GAACGGAAAG
1261 AAAAAGTACC AAATTAAACA TTTAATATGG GCAGGCAAGG AGATGGAGCG
CTTCGGCCTC
1321 CATGAGAGGT TGTTGGAGAC AGAGGAGGGG TGTAAAAGAA TCATAGAAGT
CCTCTACCCC

1381 CTAGAACCAA CAGGATCGGA GGGCTTAAAA AGTCTGTTCA ATCTTGTGTG CGTGCTATAT
1441 TGCTTGCACA AGGAACAGAA AGTGAAAGAC ACAGAGGAAG CAGTAGCAAC AGTAAGACAA
1501 CACTGCCATC TAGTGGAAAA AGAAAAAAGT GCAACAGAGA CATCTAGTGG ACAAAAGAAA
1561 AATGACAAGG GAATAGCAGC GCCACCTGGT GGCAGTCAGA ATTTTCCAGC GCAACAACAA
1621 GGAAATGCCT GGGTACATGT ACCCTTGTCA CCGCGCACCT TAAATGCGTG GGTAAAAGCA
1681 GTAGAGGAGA AAAAATTTGG AGCAGAAATA GTACCCATTT TTTTGTTTCA AGCCCTATCG
1741 AATTCCCGTT TGTGCTAGGG TTCTTAGGCT TCTTGGGGGC TGCTGGAACT GCAATGGGAG
1801 CAGCGGCGAC AGCCCTGACG GTCCAGTCTC AGCATTTGCT TGCTGGGATA CTGCAGCAGC
1861 AGAAGAATCT GCTGGCGGCT GTGGAGGCTC AACAGCAGAT GTTGAAGCTG ACCATTTGGG
1921 GTGTTAAAAA CCTCAATGCC CGCGTCACAG CCCTTGAGAA GTACCTAGAG GATCAGGCAC
1981 GACTAAACTC CTGGGGGTGC GCATGGAAAC AAGTATGTCA TACCACAGTG GAGTGGCCCT
2041 GGACAAATCG GACTCCGGAT TGGCAAAATA TGACTTGGTT GGAGTGGGAA AGACAAATAG
2101 CTGATTTGGA AAGCAACATT ACGAGACAAT TAGTGAAGGC TAGAGAACAA GAGGAAAAGA
2161 ATCTAGATGC CTATCAGAAG TTAACTAGTT GGTCAGATTT CTGGTCTTGG TTCGATTTCT
2221 CAAAATGGCT TAACATTTTA AAAATGGGAT TTTTAGTAAT AGTAGGAATA ATAGGGTTAA
2281 GATTACTTTA CACAGTATAT GGATGTATAG TGAGGGTTAG GCAGGGATAT GTTCCTCTAT
2341 CTCCACAGAT CCATATCCGC GGCAATTTTA AAAGAAAGGG AGGAATAGGG GGACAGACTT
2401 CAGCAGAGAG ACTAATTAAT ATAATAACAA CACAATTAGA AATACAACAT TTACAAACCA
2461 AAATTCAAAA AATTTTAAAT TTTAGAGCCG CGGAGATCTG TTACATAACT TATGGTAAAT
2521 GGCCTGCCTG GCTGACTGCC CAATGACCCC TGCCCAATGA TGTCAATAAT GATGTATGTT
2581 CCCATGTAAT GCCAATAGGG ACTTTCCATT GATGTCAATG GGTGGAGTAT TTATGGTAAC
2641 TGCCCACTTG GCAGTACATC AAGTGTATCA TATGCCAAGT ATGCCCCCTA TTGATGTCAA
2701 TGATGGTAAA TGGCCTGCCT GGCATTATGC CCAGTACATG ACCTTATGGG ACTTTCCTAC
2761 TTGGCAGTAC ATCTATGTAT TAGTCATTGC TATTACCATG GGAATTCACT AGTGGAGAAG
2821 AGCATGCTTG AGGGCTGAGT GCCCCTCAGT GGGCAGAGAG CACATGGCCC ACAGTCCCTG
2881 AGAAGTTGGG GGGAGGGGTG GGCAATTGAA CTGGTGCCTA GAGAAGGTGG GGCTTGGGTA
2941 AACTGGGAAA GTGATGTGGT GTACTGGCTC CACCTTTTTC CCCAGGGTGG GGGAGAACCA
3001 TATATAAGTG CAGTAGTCTC TGTGAACATT CAAGCTTCTG CCTTCTCCCT CCTGTGAGTT
3061 TGCTAGCCAC CATGGGAGTG AAGGTGCTGT TTGCCCTGAT CTGCATTGCT GTGGCTGAGG

3121 CCAAGCCCAC AGAGAACAAT GAGGACTTCA ACATTGTGGC TGTGGCCAGC
AACTTTGCCA
3181 CCACAGACCT GGATGCTGAC AGGGGCAAGC TGCCTGGCAA GAAGCTGCCC
CTGGAAGTCC
3241 TGAAAGAGAT GGAAGCCAAT GCCAGGAAGG CTGGCTGCAC AAGAGGCTGT
CTGATCTGCC
3301 TGAGCCACAT CAAGTGCACC CCCAAGATGA AGAAGTTCAT CCCTGGCAGG
TGCCACACCT
3361 ATGAAGGGGA CAAAGAGTCT GCCCAGGGGG GAATTGGAGA GGCCATTGTG
GACATCCCTG
3421 AGATCCCTGG CTTCAAGGAC CTGGAACCCA TGGAACAGTT CATTGCCCAG
GTGGACCTGT
3481 GTGTGGACTG CACTACAGGC TGTCTCAAGG GCCTGGCCAA TGTGCAGTGC
TCTGACCTGC
3541 TGAAGAAGTG GCTGCCCCAG AGATGTGCCA CCTTTGCCAG CAAGATCCAG
GGCCAGGTGG
3601 ACAAGATCAA GGGAGCTGGG GGAGATTGAT GAGGGCCCAA TCAACCTCTG
GATTACAAAA
3661 TTTGTGAAAG ATTGACTGGT ATTCTTAACT ATGTTGCTCC TTTTACGCTA
TGTGGATACG
3721 CTGCTTTAAT GCCTTTGTAT CATGCTATTG CTTCCCGTAT GGCTTTCATT
TTCTCCTCCT
3781 TGTATAAATC CTGGTTGCTG TCTCTTTATG AGGAGTTGTG GCCCGTTGTC
AGGCAACGTG
3841 GCGTGGTGTG CACTGTGTTT GCTGACGCAA CCCCCACTGG TTGGGGCATT
GCCACCACCT
3901 GTCAGCTCCT TTCCGGGACT TTCGCTTTCC CCCTCCCTAT TGCCACGGCG
GAACTCATCG
3961 CCGCCTGCCT TGCCCGCTGC TGGACAGGGG CTCGGCTGTT GGGCACTGAC
AATTCCGTGG
4021 TGTTGTCGGG GAAATCATCG TCCTTTCCTT GGCTGCTCGC CTGTGTTGCC
ACCTGGATTC
4081 TGCGCGGGAC GTCCTTCTGC TACGTCCCTT CGGCCCTCAA TCCAGCGGAC
CTTCCTTCCC
4141 GCGGCCTGCT GCCGGCTCTG CGGCCTCTTC CGCGTCTTCG CCTTCGCCCT
CAGACGAGTC
4201 GGATCTCCCT TTGGGCCGCC TCCCCGCAAG CTTCGCACTT TTTAAAAGAA
AAGGGAGGAC
4261 TGGATGGGAT TTATTACTCC GATAGGACGC TGGCTTGTAA CTCAGTCTCT
TACTAGGAGA
4321 CCAGCTTGAG CCTGGGTGTT CGCTGGTTAG CCTAACCTGG TTGGCCACCA
GGGGTAAGGA
4381 CTCCTTGGCT TAGAAAGCTA ATAAACTTGC CTGCATTAGA GCTCTTACGC
GTCCCGGGCT
4441 CGAGATCCGC ATCTCAATTA GTCAGCAACC ATAGTCCCGC CCCTAACTCC
GCCCATCCCG
4501 CCCCTAACTC CGCCCAGTTC CGCCCATTCT CCGCCCCATG GCTGACTAAT
TTTTTTTATT
4561 TATGCAGAGG CCGAGGCCGC CTCGGCCTCT GAGCTATTCC AGAAGTAGTG
AGGAGGCTTT
4621 TTTGGAGGCC TAGGCTTTTG CAAAAAGCTA ACTTGTTTAT TGCAGCTTAT
AATGGTTACA
4681 AATAAAGCAA TAGCATCACA AATTTCACAA ATAAAGCATT TTTTTCACTG
CATTCTAGTT
4741 GTGGTTTGTC CAAACTCATC AATGTATCTT ATCATGTCTG TCCGCTTCCT
CGCTCACTGA
4801 CTCGCTGCGC TCGGTCGTTC GGCTGCGGCG AGCGGTATCA GCTCACTCAA
AGGCGGTAAT

4861 ACGGTTATCC ACAGAATCAG GGGATAACGC AGGAAAGAAC ATGTGAGCAA
AAGGCCAGCA
4921 AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT TTCCATAGGC
TCCGCCCCCC
4981 TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG CGAAACCCGA
CAGGACTATA
5041 AAGATACCAG GCGTTTCCCC CTGGAAGCTC CCTCGTGCGC TCTCCTGTTC
CGACCCTGCC
5101 GCTTACCGGA TACCTGTCCG CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT
CTCATAGCTC
5161 ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC AAGCTGGGCT
GTGTGCACGA
5221 ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC TATCGTCTTG
AGTCCAACCC
5281 GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT AACAGGATTA
GCAGAGCGAG
5341 GTATGTAGGC GGTGCTACAG AGTTCTTGAA GTGGTGGCCT AACTACGGCT
ACACTAGAAG
5401 AACAGTATTT GGTATCTGCG CTCTGCTGAA GCCAGTTACC TTCGGAAAAA
GAGTTGGTAG
5461 CTCTTGATCC GGCAAACAAA CCACCGCTGG TAGCGGTGGT TTTTTTGTTT
GCAAGCAGCA
5521 GATTACGCGC AGAAAAAAAG GATCTCAAGA AGATCCTTTG ATCTTTTCTA
CGGGGTCTGA
5581 CGCTCAGTGG AACGAAAACT CACGTTAAGG GATTTTGGTC ATGAGATTAT
CAAAAAGGAT
5641 CTTCACCTAG ATCCTTTTAA ATTAAAAATG AAGTTTTAAA TCAATCTAAA
GTATATATGA
5701 GTAAACTTGG TCTGACAGTT AGAAAAACTC ATCGAGCATC AAATGAAACT
GCAATTTATT
5761 CATATCAGGA TTATCAATAC CATATTTTTG AAAAAGCCGT TTCTGTAATG
AAGGAGAAAA
5821 CTCACCGAGG CAGTTCCATA GGATGGCAAG ATCCTGGTAT CGGTCTGCGA
TTCCGACTCG
5881 TCCAACATCA ATACAACCTA TTAATTTCCC CTCGTCAAAA ATAAGGTTAT
CAAGTGAGAA
5941 ATCACCATGA GTGACGACTG AATCCGGTGA GAATGGCAAC AGCTTATGCA
TTTCTTTCCA
6001 GACTTGTTCA ACAGGCCAGC CATTACGCTC GTCATCAAAA TCACTCGCAT
CAACCAAACC
6061 GTTATTCATT CGTGATTGCG CCTGAGCGAG ACGAAATACG CGATCGCTGT
TAAAAGGACA
6121 ATTACAAACA GGAATCGAAT GCAACCGGCG CAGGAACACT GCCAGCGCAT
CAACAATATT
6181 TTCACCTGAA TCAGGATATT CTTCTAATAC CTGGAATGCT GTTTTTCCGG
GGATCGCAGT
6241 GGTGAGTAAC CATGCATCAT CAGGAGTACG GATAAAATGC TTGATGGTCG
GAAGAGGCAT
6301 AAATTCCGTC AGCCAGTTTA GTCTGACCAT CTCATCTGTA ACATCATTGG
CAACGCTACC
6361 TTTGCCATGT TTCAGAAACA ACTCTGGCGC ATCGGGCTTC CCATACAATC
GATAGATTGT
6421 CGCACCTGAT TGCCCGACAT TATCGCGAGC CCATTTATAC CCATATAAAT
CAGCATCCAT
6481 GTTGGAATTT AATCGCGGCC TAGAGCAAGA CGTTTCCCGT TGAATATGGC
TCATAACACC
6541 CCTTGTATTA CTGTTTATGT AAGCAGACAG TTTTATTGTT CATGATGATA
TATTTTTATC

6601 TTGTGCAATG TAACATCAGA GATTTTGAGA CACAACAATT GGTCGACGGA TCC

SEQ ID NO: 11

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT AGTAATCAAT
181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA
361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAC GGGACTTTCC
481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA
541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT
601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA
661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG GTGGGAGGTC
721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC AGCTTGAGCC
781 TGGGTGTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT CCTTGGCTTA
841 GAAAGCTAAT AAACTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCT CATTGACGCC
901 TCACTCTCTT GAACGGGAAT CTTCCTTACT GGGTTCTCTC TCTGACCCAG GCGAGAGAAA
961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA CAGCTGAGAA
1021 GGCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC TTGGTGAGTA
1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG AGGCCGTAGC
1141 CGTAACTACT CTGGGCAAGT AGGGCAGGCG GTGGGTACGC AATGGGGGCG GCTACCTCAG
1201 CACTAAATAG GAGACAATTA GACCAATTTG AGAAAATACG ACTTCGCCCG AACGGAAAGA
1261 AAAAGTACCA AATTAAACAT TTAATATGGG CAGGCAAGGA GATGGAGCGC TTCGGCCTCC
1321 ATGAGAGGTT GTTGGAGACA GAGGAGGGGT GTAAAAGAAT CATAGAAGTC CTCTACCCCC
1381 TAGAACCAAC AGGATCGGAG GGCTTAAAAA GTCTGTTCAA TCTTGTGTGC GTGCTATATT
1441 GCTTGCACAA GGAACAGAAA GTGAAAGACA CAGAGGAAGC AGTAGCAACA GTAAGACAAC
1501 ACTGCCATCT AGTGGAAAAA GAAAAAAGTG CAACAGAGAC ATCTAGTGGA CAAAAGAAAA

1561 ATGACAAGGG AATAGCAGCG CCACCTGGTG GCAGTCAGAA TTTTCCAGCG CAACAACAAG
1621 GAAATGCCTG GGTACATGTA CCCTTGTCAC CGCGCACCTT AAATGCGTGG GTAAAAGCAG
1681 TAGAGGAGAA AAAATTTGGA GCAGAAATAG TACCCATGTT TCAAGCCCTA TCGAATTCCC
1741 GTTTGTGCTA GGGTTCTTAG GCTTCTTGGG GGCTGCTGGA ACTGCAATGG GAGCAGCGGC
1801 GACAGCCCTG ACGGTCCAGT CTCAGCATTT GCTTGCTGGG ATACTGCAGC AGCAGAAGAA
1861 TCTGCTGGCG GCTGTGGAGG CTCAACAGCA GATGTTGAAG CTGACCATTT GGGGTGTTAA
1921 AAACCTCAAT GCCCGCGTCA CAGCCCTTGA GAAGTACCTA GAGGATCAGG CACGACTAAA
1981 CTCCTGGGGG TGCGCATGGA AACAAGTATG TCATACCACA GTGGAGTGGC CCTGGACAAA
2041 TCGGACTCCG GATTGGCAAA ATATGACTTG GTTGGAGTGG GAAAGACAAA TAGCTGATTT
2101 GGAAAGCAAC ATTACGAGAC AATTAGTGAA GGCTAGAGAA CAAGAGGAAA AGAATCTAGA
2161 TGCCTATCAG AAGTTAACTA GTTGGTCAGA TTTCTGGTCT TGGTTCGATT TCTCAAAATG
2221 GCTTAACATT TTAAAAATGG GATTTTTAGT AATAGTAGGA ATAATAGGGT TAAGATTACT
2281 TTACACAGTA TATGGATGTA TAGTGAGGGT TAGGCAGGGA TATGTTCCTC TATCTCCACA
2341 GATCCATATC CGCGGCAATT TTAAAAGAAA GGGAGGAATA GGGGGACAGA CTTCAGCAGA
2401 GAGACTAATT AATATAATAA CAACACAATT AGAAATACAA CATTTACAAA CCAAAATTCA
2461 AAAAATTTTA AATTTTAGAG CCGCGGAGAT CTCAATATTG GCCATTAGCC ATATTATTCA
2521 TTGGTTATAT AGCATAAATC AATATTGGCT ATTGGCCATT GCATACGTTG TATCTATATC
2581 ATAATATGTA CATTTATATT GGCTCATGTC CAATATGACC GCCATGTTGG CATTGATTAT
2641 TGACTAGTTA TTAATAGTAA TCAATTACGG GGTCATTAGT TCATAGCCCA TATATGGAGT
2701 TCCGCGTTAC ATAACTTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCCGCC
2761 CATTGACGTC AATAATGACG TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC
2821 GTCAATGGGT GGAGTATTTA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATCATA
2881 TGCCAAGTCC GCCCCCTATT GACGTCAATG ACGGTAAATG GCCCGCCTGG CATTATGCCC
2941 AGTACATGAC CTTACGGGAC TTTCCTACTT GGCAGTACAT CTACGTATTA GTCATCGCTA
3001 TTACCATGGT GATGCGGTTT TGGCAGTACA CCAATGGGCG TGGATAGCGG TTTGACTCAC
3061 GGGGATTTCC AAGTCTCCAC CCCATTGACG TCAATGGGAG TTTGTTTTGG CACCAAAATC
3121 AACGGGACTT TCCAAAATGT CGTAATAACC CCGCCCCGTT GACGCAAATG GGCGGTAGGC
3181 GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTCGTTTAGT GAACCGTCAG ATCACTAGAA
3241 GCTTTATTGC GGTAGTTTAT CACAGTTAAA TTGCTAACGC AGTCAGTGCT TCTGACACAA

3301 CAGTCTCGAA CTTAAGCTGC AGAAGTTGGT CGTGAGGCAC TGGGCAGGCT AGCCACCAAT
3361 GCAGATTGAG CTGAGCACCT GCTTCTTCCT GTGCCTGCTG AGGTTCTGCT TCTCTGCCAC
3421 CAGGAGATAC TACCTGGGGG CTGTGGAGCT GAGCTGGGAC TACATGCAGT CTGACCTGGG
3481 GGAGCTGCCT GTGGATGCCA GGTTCCCCCC CAGAGTGCCC AAGAGCTTCC CCTTCAACAC
3541 CTCTGTGGTG TACAAGAAGA CCCTGTTTGT GGAGTTCACT GACCACCTGT TCAACATTGC
3601 CAAGCCCAGG CCCCCCTGGA TGGGCCTGCT GGGCCCCACC ATCCAGGCTG AGGTGTATGA
3661 CACTGTGGTG ATCACCCTGA AGAACATGGC CAGCCACCCT GTGAGCCTGC ATGCTGTGGG
3721 GGTGAGCTAC TGGAAGGCCT CTGAGGGGGC TGAGTATGAT GACCAGACCA GCCAGAGGGA
3781 GAAGGAGGAT GACAAGGTGT TCCCTGGGGG CAGCCACACC TATGTGTGGC AGGTGCTGAA
3841 GGAGAATGGC CCCATGGCCT CTGACCCCCT GTGCCTGACC TACAGCTACC TGAGCCATGT
3901 GGACCTGGTG AAGGACCTGA ACTCTGGCCT GATTGGGGCC CTGCTGGTGT GCAGGGAGGG
3961 CAGCCTGGCC AAGGAGAAGA CCCAGACCCT GCACAAGTTC ATCCTGCTGT TTGCTGTGTT
4021 TGATGAGGGC AAGAGCTGGC ACTCTGAAAC CAAGAACAGC CTGATGCAGG ACAGGGATGC
4081 TGCCTCTGCC AGGGCCTGGC CCAAGATGCA CACTGTGAAT GGCTATGTGA ACAGGAGCCT
4141 GCCTGGCCTG ATTGGCTGCC ACAGGAAGTC TGTGTACTGG CATGTGATTG GCATGGGCAC
4201 CACCCCTGAG GTGCACAGCA TCTTCCTGGA GGGCCACACC TTCCTGGTCA GGAACCACAG
4261 GCAGGCCAGC CTGGAGATCA GCCCCATCAC CTTCCTGACT GCCCAGACCC TGCTGATGGA
4321 CCTGGGCCAG TTCCTGCTGT TCTGCCACAT CAGCAGCCAC CAGCATGATG GCATGGAGGC
4381 CTATGTGAAG GTGGACAGCT GCCCTGAGGA GCCCCAGCTG AGGATGAAGA ACAATGAGGA
4441 GGCTGAGGAC TATGATGATG ACCTGACTGA CTCTGAGATG GATGTGGTGA GGTTTGATGA
4501 TGACAACAGC CCCAGCTTCA TCCAGATCAG GTCTGTGGCC AAGAAGCACC CCAAGACCTG
4561 GGTGCACTAC ATTGCTGCTG AGGAGGAGGA CTGGGACTAT GCCCCCCTGG TGCTGGCCCC
4621 TGATGACAGG AGCTACAAGA GCCAGTACCT GAACAATGGC CCCCAGAGGA TTGGCAGGAA
4681 GTACAAGAAG GTCAGGTTCA TGGCCTACAC TGATGAAACC TTCAAGACCA GGGAGGCCAT
4741 CCAGCATGAG TCTGGCATCC TGGGCCCCCT GCTGTATGGG GAGGTGGGGG ACACCCTGCT
4801 GATCATCTTC AAGAACCAGG CCAGCAGGCC CTACAACATC TACCCCCATG GCATCACTGA
4861 TGTGAGGCCC CTGTACAGCA GGAGGCTGCC CAAGGGGGTG AAGCACCTGA AGGACTTCCC
4921 CATCCTGCCT GGGGAGATCT TCAAGTACAA GTGGACTGTG ACTGTGGAGG ATGGCCCCAC
4981 CAAGTCTGAC CCCAGGTGCC TGACCAGATA CTACAGCAGC TTTGTGAACA TGGAGAGGGA

5041 CCTGGCCTCT GGCCTGATTG GCCCCCTGCT GATCTGCTAC AAGGAGTCTG
TGGACCAGAG
5101 GGGCAACCAG ATCATGTCTG ACAAGAGGAA TGTGATCCTG TTCTCTGTGT
TTGATGAGAA
5161 CAGGAGCTGG TACCTGACTG AGAACATCCA GAGGTTCCTG CCCAACCCTG
CTGGGGTGCA
5221 GCTGGAGGAC CCTGAGTTCC AGGCCAGCAA CATCATGCAC AGCATCAATG
GCTATGTGTT
5281 TGACAGCCTG CAGCTGTCTG TGTGCCTGCA TGAGGTGGCC TACTGGTACA
TCCTGAGCAT
5341 TGGGGCCCAG ACTGACTTCC TGTCTGTGTT CTTCTCTGGC TACACCTTCA
AGCACAAGAT
5401 GGTGTATGAG GACACCCTGA CCCTGTTCCC CTTCTCTGGG GAGACTGTGT
TCATGAGCAT
5461 GGAGAACCCT GGCCTGTGGA TTCTGGGCTG CCACAACTCT GACTTCAGGA
ACAGGGGCAT
5521 GACTGCCCTG CTGAAAGTCT CCAGCTGTGA CAAGAACACT GGGGACTACT
ATGAGGACAG
5581 CTATGAGGAC ATCTCTGCCT ACCTGCTGAG CAAGAACAAT GCCATTGAGC
CCAGGAGCTT
5641 CAGCCAGAAT GCCACTAATG TGTCTAACAA CAGCAACACC AGCAATGACA
GCAATGTGTC
5701 TCCCCCAGTG CTGAAGAGGC ACCAGAGGGA GATCACCAGG ACCACCCTGC
AGTCTGACCA
5761 GGAGGAGATT GACTATGATG ACACCATCTC TGTGGAGATG AAGAAGGAGG
ACTTTGACAT
5821 CTACGACGAG GACGAGAACC AGAGCCCCAG GAGCTTCCAG AAGAAGACCA
GGCACTACTT
5881 CATTGCTGCT GTGGAGAGGC TGTGGGACTA TGGCATGAGC AGCAGCCCCC
ATGTGCTGAG
5941 GAACAGGGCC CAGTCTGGCT CTGTGCCCCA GTTCAAGAAG GTGGTGTTCC
AGGAGTTCAC
6001 TGATGGCAGC TTCACCCAGC CCCTGTACAG AGGGGAGCTG AATGAGCACC
TGGGCCTGCT
6061 GGGCCCCTAC ATCAGGGCTG AGGTGGAGGA CAACATCATG GTGACCTTCA
GGAACCAGGC
6121 CAGCAGGCCC TACAGCTTCT ACAGCAGCCT GATCAGCTAT GAGGAGGACC
AGAGGCAGGG
6181 GGCTGAGCCC AGGAAGAACT TTGTGAAGCC CAATGAAACC AAGACCTACT
TCTGGAAGGT
6241 GCAGCACCAC ATGGCCCCCA CCAAGGATGA GTTTGACTGC AAGGCCTGGG
CCTACTTCTC
6301 TGATGTGGAC CTGGAGAAGG ATGTGCACTC TGGCCTGATT GGCCCCCTGC
TGGTGTGCCA
6361 CACCAACACC CTGAACCCTG CCCATGGCAG GCAGGTGACT GTGCAGGAGT
TTGCCCTGTT
6421 CTTCACCATC TTTGATGAAA CCAAGAGCTG GTACTTCACT GAGAACATGG
AGAGGAACTG
6481 CAGGGCCCCC TGCAACATCC AGATGGAGGA CCCCACCTTC AAGGAGAACT
ACAGGTTCCA
6541 TGCCATCAAT GGCTACATCA TGGACACCCT GCCTGGCCTG GTGATGGCCC
AGGACCAGAG
6601 GATCAGGTGG TACCTGCTGA GCATGGGCAG CAATGAGAAC ATCCACAGCA
TCCACTTCTC
6661 TGGCCATGTG TTCACTGTGA GGAAGAAGGA GGAGTACAAG ATGGCCCTGT
ACAACCTGTA
6721 CCCTGGGGTG TTTGAGACTG TGGAGATGCT GCCCAGCAAG GCTGGCATCT
GGAGGGTGGA

6781 GTGCCTGATT GGGGAGCACC TGCATGCTGG CATGAGCACC CTGTTCCTGG
TGTACAGCAA
6841 CAAGTGCCAG ACCCCCCTGG GCATGGCCTC TGGCCACATC AGGGACTTCC
AGATCACTGC
6901 CTCTGGCCAG TATGGCCAGT GGGCCCCCAA GCTGGCCAGG CTGCACTACT
CTGGCAGCAT
6961 CAATGCCTGG AGCACCAAGG AGCCCTTCAG CTGGATCAAG GTGGACCTGC
TGGCCCCCAT
7021 GATCATCCAT GGCATCAAGA CCCAGGGGGC CAGGCAGAAG TTCAGCAGCC
TGTACATCAG
7081 CCAGTTCATC ATCATGTACA GCCTGGATGG CAAGAAGTGG CAGACCTACA
GGGGCAACAG
7141 CACTGGCACC CTGATGGTGT TCTTTGGCAA TGTGGACAGC TCTGGCATCA
AGCACAACAT
7201 CTTCAACCCC CCCATCATTG CCAGATACAT CAGGCTGCAC CCCACCCACT
ACAGCATCAG
7261 GAGCACCCTG AGGATGGAGC TGATGGGCTG TGACCTGAAC AGCTGCAGCA
TGCCCCTGGG
7321 CATGGAGAGC AAGGCCATCT CTGATGCCCA GATCACTGCC AGCAGCTACT
TCACCAACAT
7381 GTTTGCCACC TGGAGCCCCA GCAAGGCCAG GCTGCACCTG CAGGGCAGGA
GCAATGCCTG
7441 GAGGCCCCAG GTCAACAACC CCAAGGAGTG GCTGCAGGTG GACTTCCAGA
AGACCATGAA
7501 GGTGACTGGG GTGACCACCC AGGGGGTGAA GAGCCTGCTG ACCAGCATGT
ATGTGAAGGA
7561 GTTCCTGATC AGCAGCAGCC AGGATGGCCA CCAGTGGACC CTGTTCTTCC
AGAATGGCAA
7621 GGTGAAGGTG TTCCAGGGCA ACCAGGACAG CTTCACCCCT GTGGTGAACA
GCCTGGACCC
7681 CCCCCTGCTG ACCAGATACC TGAGGATTCA CCCCCAGAGC TGGGTGCACC
AGATTGCCCT
7741 GAGGATGGAG GTGCTGGGCT GTGAGGCCCA GGACCTGTAC TGAGCGGCCG
CGGGCCCAAT
7801 CAACCTCTGG ATTACAAAAT TTGTGAAAGA TTGACTGGTA TTCTTAACTA
TGTTGCTCCT
7861 TTTACGCTAT GTGGATACGC TGCTTTAATG CCTTTGTATC ATGCTATTGC
TTCCCGTATG
7921 GCTTTCATTT TCTCCTCCTT GTATAAATCC TGGTTGCTGT CTCTTTATGA
GGAGTTGTGG
7981 CCCGTTGTCA GGCAACGTGG CGTGGTGTGC ACTGTGTTTG CTGACGCAAC
CCCCACTGGT
8041 TGGGGCATTG CCACCACCTG TCAGCTCCTT TCCGGGACTT TCGCTTTCCC
CCTCCCTATT
8101 GCCACGGCGG AACTCATCGC CGCCTGCCTT GCCCGCTGCT GGACAGGGGC
TCGGCTGTTG
8161 GGCACTGACA ATTCCGTGGT GTTGTCGGGG AAATCATCGT CCTTTCCTTG
GCTGCTCGCC
8221 TGTGTTGCCA CCTGGATTCT GCGCGGGACG TCCTTCTGCT ACGTCCCTTC
GGCCCTCAAT
8281 CCAGCGGACC TTCCTTCCCG CGGCCTGCTG CCGGCTCTGC GGCCTCTTCC
GCGTCTTCGC
8341 CTTCGCCCTC AGACGAGTCG GATCTCCCTT TGGGCCGCCT CCCCGCAAGC
TTCGCACTTT
8401 TTAAAAGAAA AGGGAGGACT GGATGGGATT TATTACTCCG ATAGGACGCT
GGCTTGTAAC
8461 TCAGTCTCTT ACTAGGAGAC CAGCTTGAGC CTGGGTGTTC GCTGGTTAGC
CTAACCTGGT

8521 TGGCCACCAG GGGTAAGGAC TCCTTGGCTT AGAAAGCTAA TAAACTTGCC
TGCATTAGAG
8581 CTCTTACGCG TCCCGGGCTC GAGATCCGCA TCTCAATTAG TCAGCAACCA
TAGTCCCGCC
8641 CCTAACTCCG CCCATCCCGC CCCTAACTCC GCCCAGTTCC GCCCATTCTC
CGCCCCATGG
8701 CTGACTAATT TTTTTTATTT ATGCAGAGGC CGAGGCCGCC TCGGCCTCTG
AGCTATTCCA
8761 GAAGTAGTGA GGAGGCTTTT TTGGAGGCCT AGGCTTTTGC AAAAAGCTAA
CTTGTTTATT
8821 GCAGCTTATA ATGGTTACAA ATAAAGCAAT AGCATCACAA ATTTCACAAA
TAAAGCATTT
8881 TTTTCACTGC ATTCTAGTTG TGGTTTGTCC AAACTCATCA ATGTATCTTA
TCATGTCTGT
8941 CCGCTTCCTC GCTCACTGAC TCGCTGCGCT CGGTCGTTCG GCTGCGGCGA
GCGGTATCAG
9001 CTCACTCAAA GGCGGTAATA CGGTTATCCA CAGAATCAGG GGATAACGCA
GGAAAGAACA
9061 TGTGAGCAAA AGGCCAGCAA AAGGCCAGGA ACCGTAAAAA GGCCGCGTTG
CTGGCGTTTT
9121 TCCATAGGCT CCGCCCCCCT GACGAGCATC ACAAAAATCG ACGCTCAAGT
CAGAGGTGGC
9181 GAAACCCGAC AGGACTATAA AGATACCAGG CGTTTCCCCC TGGAAGCTCC
CTCGTGCGCT
9241 CTCCTGTTCC GACCCTGCCG CTTACCGGAT ACCTGTCCGC CTTTCTCCCT
TCGGGAAGCG
9301 TGGCGCTTTC TCATAGCTCA CGCTGTAGGT ATCTCAGTTC GGTGTAGGTC
GTTCGCTCCA
9361 AGCTGGGCTG TGTGCACGAA CCCCCCGTTC AGCCCGACCG CTGCGCCTTA
TCCGGTAACT
9421 ATCGTCTTGA GTCCAACCCG GTAAGACACG ACTTATCGCC ACTGGCAGCA
GCCACTGGTA
9481 ACAGGATTAG CAGAGCGAGG TATGTAGGCG GTGCTACAGA GTTCTTGAAG
TGGTGGCCTA
9541 ACTACGGCTA CACTAGAAGA ACAGTATTTG GTATCTGCGC TCTGCTGAAG
CCAGTTACCT
9601 TCGGAAAAAG AGTTGGTAGC TCTTGATCCG GCAAACAAAC CACCGCTGGT
AGCGGTGGTT
9661 TTTTTGTTTG CAAGCAGCAG ATTACGCGCA GAAAAAAAGG ATCTCAAGAA
GATCCTTTGA
9721 TCTTTTCTAC GGGGTCTGAC GCTCAGTGGA ACGAAAACTC ACGTTAAGGG
ATTTTGGTCA
9781 TGAGATTATC AAAAAGGATC TTCACCTAGA TCCTTTTAAA TTAAAAATGA
AGTTTTAAAT
9841 CAATCTAAAG TATATATGAG TAAACTTGGT CTGACAGTTA GAAAAACTCA
TCGAGCATCA
9901 AATGAAACTG CAATTTATTC ATATCAGGAT TATCAATACC ATATTTTTGA
AAAAGCCGTT
9961 TCTGTAATGA AGGAGAAAAC TCACCGAGGC AGTTCCATAG GATGGCAAGA
TCCTGGTATC
10021 GGTCTGCGAT TCCGACTCGT CCAACATCAA TACAACCTAT TAATTTCCCC
TCGTCAAAAA
10081 TAAGGTTATC AAGTGAGAAA TCACCATGAG TGACGACTGA ATCCGGTGAG
AATGGCAACA
10141 GCTTATGCAT TTCTTTCCAG ACTTGTTCAA CAGGCCAGCC ATTACGCTCG
TCATCAAAAT
10201 CACTCGCATC AACCAAACCG TTATTCATTC GTGATTGCGC CTGAGCGAGA
CGAAATACGC

10261 GATCGCTGTT AAAAGGACAA TTACAAACAG GAATCGAATG CAACCGGCGC AGGAACACTG
10321 CCAGCGCATC AACAATATTT TCACCTGAAT CAGGATATTC TTCTAATACC TGGAATGCTG
10381 TTTTTCCGGG GATCGCAGTG GTGAGTAACC ATGCATCATC AGGAGTACGG ATAAAATGCT
10441 TGATGGTCGG AAGAGGCATA AATTCCGTCA GCCAGTTTAG TCTGACCATC TCATCTGTAA
10501 CATCATTGGC AACGCTACCT TTGCCATGTT TCAGAAACAA CTCTGGCGCA TCGGGCTTCC
10561 CATACAATCG ATAGATTGTC GCACCTGATT GCCCGACATT ATCGCGAGCC CATTTATACC
10621 CATATAAATC AGCATCCATG TTGGAATTTA ATCGCGGCCT AGAGCAAGAC GTTTCCCGTT
10681 GAATATGGCT CATAACACCC CTTGTATTAC TGTTTATGTA AGCAGACAGT TTTATTGTTC
10741 ATGATGATAT ATTTTTATCT TGTGCAATGT AACATCAGAG ATTTTGAGAC ACAACAATTG
10801 GTCGACGGAT CC

SEQ ID NO: 12

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT AGTAATCAAT
181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC TTACGGTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT ATTTACGGTA
361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAC GGGACTTTCC
481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC GGTTTTGGCA
541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC TCCACCCCAT
601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA AATGTCGTAA
661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG GTGGGAGGTC
721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC AGCTTGAGCC
781 TGGGTGTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT CCTTGGCTTA
841 GAAAGCTAAT AAACTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCT CATTGACGCC
901 TCACTCTCTT GAACGGGAAT CTTCCTTACT GGGTTCTCTC TCTGACCCAG GCGAGAGAAA
961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA CAGCTGAGAA

1021 GGCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC
TTGGTGAGTA
1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG
AGGCCGTAGC
1141 CGTAACTACT CTGGGCAAGT AGGGCAGGCG GTGGGTACGC AATGGGGGCG
GCTACCTCAG
1201 CACTAAATAG GAGACAATTA GACCAATTTG AGAAAATACG ACTTCGCCCG
AACGGAAAGA
1261 AAAAGTACCA AATTAAACAT TTAATATGGG CAGGCAAGGA GATGGAGCGC
TTCGGCCTCC
1321 ATGAGAGGTT GTTGGAGACA GAGGAGGGGT GTAAAAGAAT CATAGAAGTC
CTCTACCCCC
1381 TAGAACCAAC AGGATCGGAG GGCTTAAAAA GTCTGTTCAA TCTTGTGTGC
GTGCTATATT
1441 GCTTGCACAA GGAACAGAAA GTGAAAGACA CAGAGGAAGC AGTAGCAACA
GTAAGACAAC
1501 ACTGCCATCT AGTGGAAAAA GAAAAAAGTG CAACAGAGAC ATCTAGTGGA
CAAAAGAAAA
1561 ATGACAAGGG AATAGCAGCG CCACCTGGTG GCAGTCAGAA TTTTCCAGCG
CAACAACAAG
1621 GAAATGCCTG GGTACATGTA CCCTTGTCAC CGCGCACCTT AAATGCGTGG
GTAAAAGCAG
1681 TAGAGGAGAA AAAATTTGGA GCAGAAATAG TACCCATGTT TCAAGCCCTA
TCGAATTCCC
1741 GTTTGTGCTA GGGTTCTTAG GCTTCTTGGG GGCTGCTGGA ACTGCAATGG
GAGCAGCGGC
1801 GACAGCCCTG ACGGTCCAGT CTCAGCATTT GCTTGCTGGG ATACTGCAGC
AGCAGAAGAA
1861 TCTGCTGGCG GCTGTGGAGG CTCAACAGCA GATGTTGAAG CTGACCATTT
GGGGTGTTAA
1921 AAACCTCAAT GCCCGCGTCA CAGCCCTTGA GAAGTACCTA GAGGATCAGG
CACGACTAAA
1981 CTCCTGGGGG TGCGCATGGA AACAAGTATG TCATACCACA GTGGAGTGGC
CCTGGACAAA
2041 TCGGACTCCG GATTGGCAAA ATATGACTTG GTTGGAGTGG GAAAGACAAA
TAGCTGATTT
2101 GGAAAGCAAC ATTACGAGAC AATTAGTGAA GGCTAGAGAA CAAGAGGAAA
AGAATCTAGA
2161 TGCCTATCAG AAGTTAACTA GTTGGTCAGA TTTCTGGTCT TGGTTCGATT
TCTCAAAATG
2221 GCTTAACATT TTAAAAATGG GATTTTTAGT AATAGTAGGA ATAATAGGGT
TAAGATTACT
2281 TTACACAGTA TATGGATGTA TAGTGAGGGT TAGGCAGGGA TATGTTCCTC
TATCTCCACA
2341 GATCCATATC CGCGGCAATT TTAAAAGAAA GGGAGGAATA GGGGGACAGA
CTTCAGCAGA
2401 GAGACTAATT AATATAATAA CAACACAATT AGAAATACAA CATTTACAAA
CCAAAATTCA
2461 AAAAATTTTA AATTTTAGAG CCGCGGAGAT CTGTTACATA ACTTATGGTA
AATGGCCTGC
2521 CTGGCTGACT GCCCAATGAC CCCTGCCCAA TGATGTCAAT AATGATGTAT
GTTCCCATGT
2581 AATGCCAATA GGGACTTTCC ATTGATGTCA ATGGGTGGAG TATTTATGGT
AACTGCCCAC
2641 TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTATGCCCC CTATTGATGT
CAATGATGGT
2701 AAATGGCCTG CCTGGCATTA TGCCCAGTAC ATGACCTTAT GGGACTTTCC
TACTTGGCAG

2761 TACATCTATG TATTAGTCAT TGCTATTACC ATGGGAATTC ACTAGTGGAG AAGAGCATGC
2821 TTGAGGGCTG AGTGCCCCTC AGTGGGCAGA GAGCACATGG CCCACAGTCC CTGAGAAGTT
2881 GGGGGGAGGG GTGGGCAATT GAACTGGTGC CTAGAGAAGG TGGGGCTTGG GTAAACTGGG
2941 AAAGTGATGT GGTGTACTGG CTCCACCTTT TTCCCCAGGG TGGGGGAGAA CCATATATAA
3001 GTGCAGTAGT CTCTGTGAAC ATTCAAGCTT CTGCCTTCTC CCTCCTGTGA GTTTGCTAGC
3061 CACCAATGCA GATTGAGCTG AGCACCTGCT TCTTCCTGTG CCTGCTGAGG TTCTGCTTCT
3121 CTGCCACCAG GAGATACTAC CTGGGGGCTG TGGAGCTGAG CTGGGACTAC ATGCAGTCTG
3181 ACCTGGGGGA GCTGCCTGTG GATGCCAGGT TCCCCCCCAG AGTGCCCAAG AGCTTCCCCT
3241 TCAACACCTC TGTGGTGTAC AAGAAGACCC TGTTTGTGGA GTTCACTGAC CACCTGTTCA
3301 ACATTGCCAA GCCCAGGCCC CCCTGGATGG GCCTGCTGGG CCCCACCATC CAGGCTGAGG
3361 TGTATGACAC TGTGGTGATC ACCCTGAAGA ACATGGCCAG CCACCCTGTG AGCCTGCATG
3421 CTGTGGGGGT GAGCTACTGG AAGGCCTCTG AGGGGGCTGA GTATGATGAC CAGACCAGCC
3481 AGAGGGAGAA GGAGGATGAC AAGGTGTTCC CTGGGGGCAG CCACACCTAT GTGTGGCAGG
3541 TGCTGAAGGA GAATGGCCCC ATGGCCTCTG ACCCCCTGTG CCTGACCTAC AGCTACCTGA
3601 GCCATGTGGA CCTGGTGAAG GACCTGAACT CTGGCCTGAT TGGGGCCCTG CTGGTGTGCA
3661 GGGAGGGCAG CCTGGCCAAG GAGAAGACCC AGACCCTGCA CAAGTTCATC CTGCTGTTTG
3721 CTGTGTTTGA TGAGGGCAAG AGCTGGCACT CTGAAACCAA GAACAGCCTG ATGCAGGACA
3781 GGGATGCTGC CTCTGCCAGG GCCTGGCCCA AGATGCACAC TGTGAATGGC TATGTGAACA
3841 GGAGCCTGCC TGGCCTGATT GGCTGCCACA GGAAGTCTGT GTACTGGCAT GTGATTGGCA
3901 TGGGCACCAC CCCTGAGGTG CACAGCATCT TCCTGGAGGG CCACACCTTC CTGGTCAGGA
3961 ACCACAGGCA GGCCAGCCTG GAGATCAGCC CCATCACCTT CCTGACTGCC CAGACCCTGC
4021 TGATGGACCT GGGCCAGTTC CTGCTGTTCT GCCACATCAG CAGCCACCAG CATGATGGCA
4081 TGGAGGCCTA TGTGAAGGTG GACAGCTGCC CTGAGGAGCC CCAGCTGAGG ATGAAGAACA
4141 ATGAGGAGGC TGAGGACTAT GATGATGACC TGACTGACTC TGAGATGGAT GTGGTGAGGT
4201 TTGATGATGA CAACAGCCCC AGCTTCATCC AGATCAGGTC TGTGGCCAAG AAGCACCCCA
4261 AGACCTGGGT GCACTACATT GCTGCTGAGG AGGAGGACTG GGACTATGCC CCCCTGGTGC
4321 TGGCCCCTGA TGACAGGAGC TACAAGAGCC AGTACCTGAA CAATGGCCCC CAGAGGATTG
4381 GCAGGAAGTA CAAGAAGGTC AGGTTCATGG CCTACACTGA TGAAACCTTC AAGACCAGGG
4441 AGGCCATCCA GCATGAGTCT GGCATCCTGG GCCCCCTGCT GTATGGGGAG GTGGGGGACA

4501 CCCTGCTGAT CATCTTCAAG AACCAGGCCA GCAGGCCCTA CAACATCTAC
CCCCATGGCA
4561 TCACTGATGT GAGGCCCCTG TACAGCAGGA GGCTGCCCAA GGGGGTGAAG
CACCTGAAGG
4621 ACTTCCCCAT CCTGCCTGGG GAGATCTTCA AGTACAAGTG GACTGTGACT
GTGGAGGATG
4681 GCCCCACCAA GTCTGACCCC AGGTGCCTGA CCAGATACTA CAGCAGCTTT
GTGAACATGG
4741 AGAGGGACCT GGCCTCTGGC CTGATTGGCC CCCTGCTGAT CTGCTACAAG
GAGTCTGTGG
4801 ACCAGAGGGG CAACCAGATC ATGTCTGACA AGAGGAATGT GATCCTGTTC
TCTGTGTTTG
4861 ATGAGAACAG GAGCTGGTAC CTGACTGAGA ACATCCAGAG GTTCCTGCCC
AACCCTGCTG
4921 GGGTGCAGCT GGAGGACCCT GAGTTCCAGG CCAGCAACAT CATGCACAGC
ATCAATGGCT
4981 ATGTGTTTGA CAGCCTGCAG CTGTCTGTGT GCCTGCATGA GGTGGCCTAC
TGGTACATCC
5041 TGAGCATTGG GGCCCAGACT GACTTCCTGT CTGTGTTCTT CTCTGGCTAC
ACCTTCAAGC
5101 ACAAGATGGT GTATGAGGAC ACCCTGACCC TGTTCCCCTT CTCTGGGGAG
ACTGTGTTCA
5161 TGAGCATGGA GAACCCTGGC CTGTGGATTC TGGGCTGCCA CAACTCTGAC
TTCAGGAACA
5221 GGGGCATGAC TGCCCTGCTG AAAGTCTCCA GCTGTGACAA GAACACTGGG
GACTACTATG
5281 AGGACAGCTA TGAGGACATC TCTGCCTACC TGCTGAGCAA GAACAATGCC
ATTGAGCCCA
5341 GGAGCTTCAG CCAGAATGCC ACTAATGTGT CTAACAACAG CAACACCAGC
AATGACAGCA
5401 ATGTGTCTCC CCCAGTGCTG AAGAGGCACC AGAGGGAGAT CACCAGGACC
ACCCTGCAGT
5461 CTGACCAGGA GGAGATTGAC TATGATGACA CCATCTCTGT GGAGATGAAG
AAGGAGGACT
5521 TTGACATCTA CGACGAGGAC GAGAACCAGA GCCCCAGGAG CTTCCAGAAG
AAGACCAGGC
5581 ACTACTTCAT TGCTGCTGTG GAGAGGCTGT GGGACTATGG CATGAGCAGC
AGCCCCCATG
5641 TGCTGAGGAA CAGGGCCCAG TCTGGCTCTG TGCCCCAGTT CAAGAAGGTG
GTGTTCCAGG
5701 AGTTCACTGA TGGCAGCTTC ACCCAGCCCC TGTACAGAGG GGAGCTGAAT
GAGCACCTGG
5761 GCCTGCTGGG CCCCTACATC AGGGCTGAGG TGGAGGACAA CATCATGGTG
ACCTTCAGGA
5821 ACCAGGCCAG CAGGCCCTAC AGCTTCTACA GCAGCCTGAT CAGCTATGAG
GAGGACCAGA
5881 GGCAGGGGGC TGAGCCCAGG AAGAACTTTG TGAAGCCCAA TGAAACCAAG
ACCTACTTCT
5941 GGAAGGTGCA GCACCACATG GCCCCCACCA AGGATGAGTT TGACTGCAAG
GCCTGGGCCT
6001 ACTTCTCTGA TGTGGACCTG GAGAAGGATG TGCACTCTGG CCTGATTGGC
CCCCTGCTGG
6061 TGTGCCACAC CAACACCCTG AACCCTGCCC ATGGCAGGCA GGTGACTGTG
CAGGAGTTTG
6121 CCCTGTTCTT CACCATCTTT GATGAAACCA AGAGCTGGTA CTTCACTGAG
AACATGGAGA
6181 GGAACTGCAG GGCCCCCTGC AACATCCAGA TGGAGGACCC CACCTTCAAG
GAGAACTACA

6241 GGTTCCATGC CATCAATGGC TACATCATGG ACACCCTGCC TGGCCTGGTG ATGGCCCAGG
6301 ACCAGAGGAT CAGGTGGTAC CTGCTGAGCA TGGGCAGCAA TGAGAACATC CACAGCATCC
6361 ACTTCTCTGG CCATGTGTTC ACTGTGAGGA AGAAGGAGGA GTACAAGATG GCCCTGTACA
6421 ACCTGTACCC TGGGGTGTTT GAGACTGTGG AGATGCTGCC CAGCAAGGCT GGCATCTGGA
6481 GGGTGGAGTG CCTGATTGGG GAGCACCTGC ATGCTGGCAT GAGCACCCTG TTCCTGGTGT
6541 ACAGCAACAA GTGCCAGACC CCCCTGGGCA TGGCCTCTGG CCACATCAGG GACTTCCAGA
6601 TCACTGCCTC TGGCCAGTAT GGCCAGTGGG CCCCCAAGCT GGCCAGGCTG CACTACTCTG
6661 GCAGCATCAA TGCCTGGAGC ACCAAGGAGC CCTTCAGCTG GATCAAGGTG GACCTGCTGG
6721 CCCCCATGAT CATCCATGGC ATCAAGACCC AGGGGGCCAG GCAGAAGTTC AGCAGCCTGT
6781 ACATCAGCCA GTTCATCATC ATGTACAGCC TGGATGGCAA GAAGTGGCAG ACCTACAGGG
6841 GCAACAGCAC TGGCACCCTG ATGGTGTTCT TTGGCAATGT GGACAGCTCT GGCATCAAGC
6901 ACAACATCTT CAACCCCCCC ATCATTGCCA GATACATCAG GCTGCACCCC ACCCACTACA
6961 GCATCAGGAG CACCCTGAGG ATGGAGCTGA TGGGCTGTGA CCTGAACAGC TGCAGCATGC
7021 CCCTGGGCAT GGAGAGCAAG GCCATCTCTG ATGCCCAGAT CACTGCCAGC AGCTACTTCA
7081 CCAACATGTT TGCCACCTGG AGCCCCAGCA AGGCCAGGCT GCACCTGCAG GGCAGGAGCA
7141 ATGCCTGGAG GCCCCAGGTC AACAACCCCA AGGAGTGGCT GCAGGTGGAC TTCCAGAAGA
7201 CCATGAAGGT GACTGGGGTG ACCACCCAGG GGGTGAAGAG CCTGCTGACC AGCATGTATG
7261 TGAAGGAGTT CCTGATCAGC AGCAGCCAGG ATGGCCACCA GTGGACCCTG TTCTTCCAGA
7321 ATGGCAAGGT GAAGGTGTTC CAGGGCAACC AGGACAGCTT CACCCCTGTG GTGAACAGCC
7381 TGGACCCCCC CCTGCTGACC AGATACCTGA GGATTCACCC CCAGAGCTGG GTGCACCAGA
7441 TTGCCCTGAG GATGGAGGTG CTGGGCTGTG AGGCCCAGGA CCTGTACTGA GCGGCCGCGG
7501 GCCCAATCAA CCTCTGGATT ACAAAATTTG TGAAAGATTG ACTGGTATTC TTAACTATGT
7561 TGCTCCTTTT ACGCTATGTG GATACGCTGC TTTAATGCCT TTGTATCATG CTATTGCTTC
7621 CCGTATGGCT TTCATTTTCT CCTCCTTGTA TAAATCCTGG TTGCTGTCTC TTTATGAGGA
7681 GTTGTGGCCC GTTGTCAGGC AACGTGGCGT GGTGTGCACT GTGTTTGCTG ACGCAACCCC
7741 CACTGGTTGG GGCATTGCCA CCACCTGTCA GCTCCTTTCC GGGACTTTCG CTTTCCCCCT
7801 CCCTATTGCC ACGGCGGAAC TCATCGCCGC CTGCCTTGCC CGCTGCTGGA CAGGGGCTCG
7861 GCTGTTGGGC ACTGACAATT CCGTGGTGTT GTCGGGGAAA TCATCGTCCT TTCCTTGGCT
7921 GCTCGCCTGT GTTGCCACCT GGATTCTGCG CGGGACGTCC TTCTGCTACG TCCCTTCGGC

7981 CCTCAATCCA GCGGACCTTC CTTCCCGCGG CCTGCTGCCG GCTCTGCGGC
CTCTTCCGCG
8041 TCTTCGCCTT CGCCCTCAGA CGAGTCGGAT CTCCCTTTGG GCCGCCTCCC
CGCAAGCTTC
8101 GCACTTTTTA AAAGAAAAGG GAGGACTGGA TGGGATTTAT TACTCCGATA
GGACGCTGGC
8161 TTGTAACTCA GTCTCTTACT AGGAGACCAG CTTGAGCCTG GGTGTTCGCT
GGTTAGCCTA
8221 ACCTGGTTGG CCACCAGGGG TAAGGACTCC TTGGCTTAGA AAGCTAATAA
ACTTGCCTGC
8281 ATTAGAGCTC TTACGCGTCC CGGGCTCGAG ATCCGCATCT CAATTAGTCA
GCAACCATAG
8341 TCCCGCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC CAGTTCCGCC
CATTCTCCGC
8401 CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA GGCCGCCTCG
GCCTCTGAGC
8461 TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG CTTTTGCAAA
AAGCTAACTT
8521 GTTTATTGCA GCTTATAATG GTTACAAATA AAGCAATAGC ATCACAAATT
TCACAAATAA
8581 AGCATTTTTT TCACTGCATT CTAGTTGTGG TTTGTCCAAA CTCATCAATG
TATCTTATCA
8641 TGTCTGTCCG CTTCCTCGCT CACTGACTCG CTGCGCTCGG TCGTTCGGCT
GCGGCGAGCG
8701 GTATCAGCTC ACTCAAAGGC GGTAATACGG TTATCCACAG AATCAGGGGA
TAACGCAGGA
8761 AAGAACATGT GAGCAAAAGG CCAGCAAAAG GCCAGGAACC GTAAAAAGGC
CGCGTTGCTG
8821 GCGTTTTTCC ATAGGCTCCG CCCCCCTGAC GAGCATCACA AAAATCGACG
CTCAAGTCAG
8881 AGGTGGCGAA ACCCGACAGG ACTATAAAGA TACCAGGCGT TTCCCCCTGG
AAGCTCCCTC
8941 GTGCGCTCTC CTGTTCCGAC CCTGCCGCTT ACCGGATACC TGTCCGCCTT
TCTCCCTTCG
9001 GGAAGCGTGG CGCTTTCTCA TAGCTCACGC TGTAGGTATC TCAGTTCGGT
GTAGGTCGTT
9061 CGCTCCAAGC TGGGCTGTGT GCACGAACCC CCCGTTCAGC CCGACCGCTG
CGCCTTATCC
9121 GGTAACTATC GTCTTGAGTC CAACCCGGTA AGACACGACT TATCGCCACT
GGCAGCAGCC
9181 ACTGGTAACA GGATTAGCAG AGCGAGGTAT GTAGGCGGTG CTACAGAGTT
CTTGAAGTGG
9241 TGGCCTAACT ACGGCTACAC TAGAAGAACA GTATTTGGTA TCTGCGCTCT
GCTGAAGCCA
9301 GTTACCTTCG GAAAAAGAGT TGGTAGCTCT TGATCCGGCA AACAAACCAC
CGCTGGTAGC
9361 GGTGGTTTTT TTGTTTGCAA GCAGCAGATT ACGCGCAGAA AAAAAGGATC
TCAAGAAGAT
9421 CCTTTGATCT TTTCTACGGG GTCTGACGCT CAGTGGAACG AAAACTCACG
TTAAGGGATT
9481 TTGGTCATGA GATTATCAAA AAGGATCTTC ACCTAGATCC TTTTAAATTA
AAAATGAAGT
9541 TTTAAATCAA TCTAAAGTAT ATATGAGTAA ACTTGGTCTG ACAGTTAGAA
AAACTCATCG
9601 AGCATCAAAT GAAACTGCAA TTTATTCATA TCAGGATTAT CAATACCATA
TTTTTGAAAA
9661 AGCCGTTTCT GTAATGAAGG AGAAAACTCA CCGAGGCAGT TCCATAGGAT
GGCAAGATCC

9721 TGGTATCGGT CTGCGATTCC GACTCGTCCA ACATCAATAC AACCTATTAA
TTTCCCCTCG
9781 TCAAAAATAA GGTTATCAAG TGAGAAATCA CCATGAGTGA CGACTGAATC
CGGTGAGAAT
9841 GGCAACAGCT TATGCATTTC TTTCCAGACT TGTTCAACAG GCCAGCCATT
ACGCTCGTCA
9901 TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTCGTG ATTGCGCCTG
AGCGAGACGA
9961 AATACGCGAT CGCTGTTAAA AGGACAATTA CAAACAGGAA TCGAATGCAA
CCGGCGCAGG
10021 AACACTGCCA GCGCATCAAC AATATTTTCA CCTGAATCAG GATATTCTTC
TAATACCTGG
10081 AATGCTGTTT TTCCGGGGAT CGCAGTGGTG AGTAACCATG CATCATCAGG
AGTACGGATA
10141 AAATGCTTGA TGGTCGGAAG AGGCATAAAT TCCGTCAGCC AGTTTAGTCT
GACCATCTCA
10201 TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTTCA GAAACAACTC
TGGCGCATCG
10261 GGCTTCCCAT ACAATCGATA GATTGTCGCA CCTGATTGCC CGACATTATC
GCGAGCCCAT
10321 TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC GCGGCCTAGA
GCAAGACGTT
10381 TCCCGTTGAA TATGGCTCAT AACACCCCTT GTATTACTGT TTATGTAAGC
AGACAGTTTT
10441 ATTGTTCATG ATGATATATT TTTATCTTGT GCAATGTAAC ATCAGAGATT
TTGAGACACA
10501 ACAATTGGTC GACGGATCC

SEQ ID NO: 13

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT
AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT
ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT
AGTAATCAAT
181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
TTACGGTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA
TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT
ATTTACGGTA
361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC
CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAC
GGGACTTTCC
481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
GGTTTTGGCA
541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC
TCCACCCCAT
601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA
AATGTCGTAA
661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG
GTGGGAGGTC
721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC
AGCTTGAGCC

781 TGGGTGTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT
CCTTGGCTTA
841 GAAAGCTAAT AAACTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCT
CATTGACGCC
901 TCACTCTCTT GAACGGGAAT CTTCCTTACT GGGTTCTCTC TCTGACCCAG
GCGAGAGAAA
961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA
CAGCTGAGAA
1021 GGCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC
TTGGTGAGTA
1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG
AGGCCGTAGC
1141 CGTAACTACT CTGGGCAAGT AGGGCAGGCG GTGGGTACGC AATGGGGGCG
GCTACCTCAG
1201 CACTAAATAG GAGACAATTA GACCAATTTG AGAAAATACG ACTTCGCCCG
AACGGAAAGA
1261 AAAAGTACCA AATTAAACAT TTAATATGGG CAGGCAAGGA GATGGAGCGC
TTCGGCCTCC
1321 ATGAGAGGTT GTTGGAGACA GAGGAGGGGT GTAAAAGAAT CATAGAAGTC
CTCTACCCCC
1381 TAGAACCAAC AGGATCGGAG GGCTTAAAAA GTCTGTTCAA TCTTGTGTGC
GTGCTATATT
1441 GCTTGCACAA GGAACAGAAA GTGAAAGACA CAGAGGAAGC AGTAGCAACA
GTAAGACAAC
1501 ACTGCCATCT AGTGGAAAAA GAAAAAAGTG CAACAGAGAC ATCTAGTGGA
CAAAAGAAAA
1561 ATGACAAGGG AATAGCAGCG CCACCTGGTG GCAGTCAGAA TTTTCCAGCG
CAACAACAAG
1621 GAAATGCCTG GGTACATGTA CCCTTGTCAC CGCGCACCTT AAATGCGTGG
GTAAAAGCAG
1681 TAGAGGAGAA AAAATTTGGA GCAGAAATAG TACCCATGTT TCAAGCCCTA
TCGAATTCCC
1741 GTTTGTGCTA GGGTTCTTAG GCTTCTTGGG GGCTGCTGGA ACTGCAATGG
GAGCAGCGGC
1801 GACAGCCCTG ACGGTCCAGT CTCAGCATTT GCTTGCTGGG ATACTGCAGC
AGCAGAAGAA
1861 TCTGCTGGCG GCTGTGGAGG CTCAACAGCA GATGTTGAAG CTGACCATTT
GGGGTGTTAA
1921 AAACCTCAAT GCCCGCGTCA CAGCCCTTGA GAAGTACCTA GAGGATCAGG
CACGACTAAA
1981 CTCCTGGGGG TGCGCATGGA AACAAGTATG TCATACCACA GTGGAGTGGC
CCTGGACAAA
2041 TCGGACTCCG GATTGGCAAA ATATGACTTG GTTGGAGTGG GAAAGACAAA
TAGCTGATTT
2101 GGAAAGCAAC ATTACGAGAC AATTAGTGAA GGCTAGAGAA CAAGAGGAAA
AGAATCTAGA
2161 TGCCTATCAG AAGTTAACTA GTTGGTCAGA TTTCTGGTCT TGGTTCGATT
TCTCAAAATG
2221 GCTTAACATT TTAAAAATGG GATTTTTAGT AATAGTAGGA ATAATAGGGT
TAAGATTACT
2281 TTACACAGTA TATGGATGTA TAGTGAGGGT TAGGCAGGGA TATGTTCCTC
TATCTCCACA
2341 GATCCATATC CGCGGCAATT TTAAAAGAAA GGGAGGAATA GGGGGACAGA
CTTCAGCAGA
2401 GAGACTAATT AATATAATAA CAACACAATT AGAAATACAA CATTTACAAA
CCAAAATTCA
2461 AAAAATTTTA AATTTTAGAG CCGCGGAGAT CTCAATATTG GCCATTAGCC
ATATTATTCA

2521 TTGGTTATAT AGCATAAATC AATATTGGCT ATTGGCCATT GCATACGTTG TATCTATATC
2581 ATAATATGTA CATTTATATT GGCTCATGTC CAATATGACC GCCATGTTGG CATTGATTAT
2641 TGACTAGTTA TTAATAGTAA TCAATTACGG GGTCATTAGT TCATAGCCCA TATATGGAGT
2701 TCCGCGTTAC ATAACTTACG GTAAATGGCC CGCCTGGCTG ACCGCCCAAC GACCCCCGCC
2761 CATTGACGTC AATAATGACG TATGTTCCCA TAGTAACGCC AATAGGGACT TTCCATTGAC
2821 GTCAATGGGT GGAGTATTTA CGGTAAACTG CCCACTTGGC AGTACATCAA GTGTATCATA
2881 TGCCAAGTCC GCCCCCTATT GACGTCAATG ACGGTAAATG GCCCGCCTGG CATTATGCCC
2941 AGTACATGAC CTTACGGGAC TTTCCTACTT GGCAGTACAT CTACGTATTA GTCATCGCTA
3001 TTACCATGGT GATGCGGTTT TGGCAGTACA CCAATGGGCG TGGATAGCGG TTTGACTCAC
3061 GGGGATTTCC AAGTCTCCAC CCCATTGACG TCAATGGGAG TTTGTTTTGG CACCAAAATC
3121 AACGGGACTT TCCAAAATGT CGTAATAACC CCGCCCCGTT GACGCAAATG GGCGGTAGGC
3181 GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTCGTTTAGT GAACCGTCAG ATCACTAGAA
3241 GCTTTATTGC GGTAGTTTAT CACAGTTAAA TTGCTAACGC AGTCAGTGCT TCTGACACAA
3301 CAGTCTCGAA CTTAAGCTGC AGAAGTTGGT CGTGAGGCAC TGGGCAGGCT AGCCACCAAT
3361 GCAGATTGAG CTGAGCACCT GCTTCTTCCT GTGCCTGCTG AGGTTCTGCT TCTCTGCCAC
3421 CAGGAGATAC TACCTGGGGG CTGTGGAGCT GAGCTGGGAC TACATGCAGT CTGACCTGGG
3481 GGAGCTGCCT GTGGATGCCA GGTTCCCCCC CAGAGTGCCC AAGAGCTTCC CCTTCAACAC
3541 CTCTGTGGTG TACAAGAAGA CCCTGTTTGT GGAGTTCACT GACCACCTGT TCAACATTGC
3601 CAAGCCCAGG CCCCCCTGGA TGGGCCTGCT GGGCCCCACC ATCCAGGCTG AGGTGTATGA
3661 CACTGTGGTG ATCACCCTGA AGAACATGGC CAGCCACCCT GTGAGCCTGC ATGCTGTGGG
3721 GGTGAGCTAC TGGAAGGCCT CTGAGGGGGC TGAGTATGAT GACCAGACCA GCCAGAGGGA
3781 GAAGGAGGAT GACAAGGTGT TCCCTGGGGG CAGCCACACC TATGTGTGGC AGGTGCTGAA
3841 GGAGAATGGC CCCATGGCCT CTGACCCCCT GTGCCTGACC TACAGCTACC TGAGCCATGT
3901 GGACCTGGTG AAGGACCTGA ACTCTGGCCT GATTGGGGCC CTGCTGGTGT GCAGGGAGGG
3961 CAGCCTGGCC AAGGAGAAGA CCCAGACCCT GCACAAGTTC ATCCTGCTGT TTGCTGTGTT
4021 TGATGAGGGC AAGAGCTGGC ACTCTGAAAC CAAGAACAGC CTGATGCAGG ACAGGGATGC
4081 TGCCTCTGCC AGGGCCTGGC CCAAGATGCA CACTGTGAAT GGCTATGTGA ACAGGAGCCT
4141 GCCTGGCCTG ATTGGCTGCC ACAGGAAGTC TGTGTACTGG CATGTGATTG GCATGGGCAC
4201 CACCCCTGAG GTGCACAGCA TCTTCCTGGA GGGCCACACC TTCCTGGTCA GGAACCACAG

4261 GCAGGCCAGC CTGGAGATCA GCCCCATCAC CTTCCTGACT GCCCAGACCC
TGCTGATGGA
4321 CCTGGGCCAG TTCCTGCTGT TCTGCCACAT CAGCAGCCAC CAGCATGATG
GCATGGAGGC
4381 CTATGTGAAG GTGGACAGCT GCCCTGAGGA GCCCCAGCTG AGGATGAAGA
ACAATGAGGA
4441 GGCTGAGGAC TATGATGATG ACCTGACTGA CTCTGAGATG GATGTGGTGA
GGTTTGATGA
4501 TGACAACAGC CCCAGCTTCA TCCAGATCAG GTCTGTGGCC AAGAAGCACC
CCAAGACCTG
4561 GGTGCACTAC ATTGCTGCTG AGGAGGAGGA CTGGGACTAT GCCCCCCTGG
TGCTGGCCCC
4621 TGATGACAGG AGCTACAAGA GCCAGTACCT GAACAATGGC CCCCAGAGGA
TTGGCAGGAA
4681 GTACAAGAAG GTCAGGTTCA TGGCCTACAC TGATGAAACC TTCAAGACCA
GGGAGGCCAT
4741 CCAGCATGAG TCTGGCATCC TGGGCCCCCT GCTGTATGGG GAGGTGGGGG
ACACCCTGCT
4801 GATCATCTTC AAGAACCAGG CCAGCAGGCC CTACAACATC TACCCCCATG
GCATCACTGA
4861 TGTGAGGCCC CTGTACAGCA GGAGGCTGCC CAAGGGGGTG AAGCACCTGA
AGGACTTCCC
4921 CATCCTGCCT GGGGAGATCT TCAAGTACAA GTGGACTGTG ACTGTGGAGG
ATGGCCCCAC
4981 CAAGTCTGAC CCCAGGTGCC TGACCAGATA CTACAGCAGC TTTGTGAACA
TGGAGAGGGA
5041 CCTGGCCTCT GGCCTGATTG GCCCCCTGCT GATCTGCTAC AAGGAGTCTG
TGGACCAGAG
5101 GGGCAACCAG ATCATGTCTG ACAAGAGGAA TGTGATCCTG TTCTCTGTGT
TTGATGAGAA
5161 CAGGAGCTGG TACCTGACTG AGAACATCCA GAGGTTCCTG CCCAACCCTG
CTGGGGTGCA
5221 GCTGGAGGAC CCTGAGTTCC AGGCCAGCAA CATCATGCAC AGCATCAATG
GCTATGTGTT
5281 TGACAGCCTG CAGCTGTCTG TGTGCCTGCA TGAGGTGGCC TACTGGTACA
TCCTGAGCAT
5341 TGGGGCCCAG ACTGACTTCC TGTCTGTGTT CTTCTCTGGC TACACCTTCA
AGCACAAGAT
5401 GGTGTATGAG GACACCCTGA CCCTGTTCCC CTTCTCTGGG GAGACTGTGT
TCATGAGCAT
5461 GGAGAACCCT GGCCTGTGGA TTCTGGGCTG CCACAACTCT GACTTCAGGA
ACAGGGGCAT
5521 GACTGCCCTG CTGAAAGTCT CCAGCTGTGA CAAGAACACT GGGGACTACT
ATGAGGACAG
5581 CTATGAGGAC ATCTCTGCCT ACCTGCTGAG CAAGAACAAT GCCATTGAGC
CCAGGAGCTT
5641 CAGCCAGAAC AGCAGGCACC CCAGCACCAG GCAGAAGCAG TTCAATGCCA
CCACCATCCC
5701 TGAGAATGAC ATAGAGAAGA CAGACCCATG GTTTGCCCAC CGGACCCCCA
TGCCCAAGAT
5761 CCAGAATGTG AGCAGCTCTG ACCTGCTGAT GCTGCTGAGG CAGAGCCCCA
CCCCCCATGG
5821 CCTGAGCCTG TCTGACCTGC AGGAGGCCAA GTATGAAACC TTCTCTGATG
ACCCCAGCCC
5881 TGGGGCCATT GACAGCAACA ACAGCCTGTC TGAGATGACC CACTTCAGGC
CCCAGCTGCA
5941 CCACTCTGGG GACATGGTGT TCACCCCTGA GTCTGGCCTG CAGCTGAGGC
TGAATGAGAA

6001 GCTGGGCACC ACTGCTGCCA CTGAGCTGAA GAAGCTGGAC TTCAAAGTCT CCAGCACCAG
6061 CAACAACCTG ATCAGCACCA TCCCCTCTGA CAACCTGGCT GCTGGCACTG ACAACACCAG
6121 CAGCCTGGGC CCCCCCAGCA TGCCTGTGCA CTATGACAGC CAGCTGGACA CCACCCTGTT
6181 TGGCAAGAAG AGCAGCCCCC TGACTGAGTC TGGGGGCCCC CTGAGCCTGT CTGAGGAGAA
6241 CAATGACAGC AAGCTGCTGG AGTCTGGCCT GATGAACAGC CAGGAGAGCA GCTGGGGCAA
6301 GAATGTGAGC AGCAGGGAGA TCACCAGGAC CACCCTGCAG TCTGACCAGG AGGAGATTGA
6361 CTATGATGAC ACCATCTCTG TGGAGATGAA GAAGGAGGAC TTTGACATCT ACGACGAGGA
6421 CGAGAACCAG AGCCCCAGGA GCTTCCAGAA GAAGACCAGG CACTACTTCA TTGCTGCTGT
6481 GGAGAGGCTG TGGGACTATG GCATGAGCAG CAGCCCCCAT GTGCTGAGGA ACAGGGCCCA
6541 GTCTGGCTCT GTGCCCCAGT TCAAGAAGGT GGTGTTCCAG GAGTTCACTG ATGGCAGCTT
6601 CACCCAGCCC CTGTACAGAG GGGAGCTGAA TGAGCACCTG GGCCTGCTGG GCCCCTACAT
6661 CAGGGCTGAG GTGGAGGACA ACATCATGGT GACCTTCAGG AACCAGGCCA GCAGGCCCTA
6721 CAGCTTCTAC AGCAGCCTGA TCAGCTATGA GGAGGACCAG AGGCAGGGGG CTGAGCCCAG
6781 GAAGAACTTT GTGAAGCCCA ATGAAACCAA GACCTACTTC TGGAAGGTGC AGCACCACAT
6841 GGCCCCCACC AAGGATGAGT TTGACTGCAA GGCCTGGGCC TACTTCTCTG ATGTGGACCT
6901 GGAGAAGGAT GTGCACTCTG GCCTGATTGG CCCCCTGCTG GTGTGCCACA CCAACACCCT
6961 GAACCCTGCC CATGGCAGGC AGGTGACTGT GCAGGAGTTT GCCCTGTTCT TCACCATCTT
7021 TGATGAAACC AAGAGCTGGT ACTTCACTGA GAACATGGAG AGGAACTGCA GGGCCCCCTG
7081 CAACATCCAG ATGGAGGACC CCACCTTCAA GGAGAACTAC AGGTTCCATG CCATCAATGG
7141 CTACATCATG GACACCCTGC CTGGCCTGGT GATGGCCCAG GACCAGAGGA TCAGGTGGTA
7201 CCTGCTGAGC ATGGGCAGCA ATGAGAACAT CCACAGCATC CACTTCTCTG GCCATGTGTT
7261 CACTGTGAGG AAGAAGGAGG AGTACAAGAT GGCCCTGTAC AACCTGTACC CTGGGGTGTT
7321 TGAGACTGTG GAGATGCTGC CCAGCAAGGC TGGCATCTGG AGGGTGGAGT GCCTGATTGG
7381 GGAGCACCTG CATGCTGGCA TGAGCACCCT GTTCCTGGTG TACAGCAACA AGTGCCAGAC
7441 CCCCCTGGGC ATGGCCTCTG GCCACATCAG GGACTTCCAG ATCACTGCCT CTGGCCAGTA
7501 TGGCCAGTGG GCCCCCAAGC TGGCCAGGCT GCACTACTCT GGCAGCATCA ATGCCTGGAG
7561 CACCAAGGAG CCCTTCAGCT GGATCAAGGT GGACCTGCTG GCCCCCATGA TCATCCATGG
7621 CATCAAGACC CAGGGGGCCA GGCAGAAGTT CAGCAGCCTG TACATCAGCC AGTTCATCAT
7681 CATGTACAGC CTGGATGGCA AGAAGTGGCA GACCTACAGG GGCAACAGCA CTGGCACCCT

7741 GATGGTGTTC TTTGGCAATG TGGACAGCTC TGGCATCAAG CACAACATCT
TCAACCCCCC
7801 CATCATTGCC AGATACATCA GGCTGCACCC CACCCACTAC AGCATCAGGA
GCACCCTGAG
7861 GATGGAGCTG ATGGGCTGTG ACCTGAACAG CTGCAGCATG CCCCTGGGCA
TGGAGAGCAA
7921 GGCCATCTCT GATGCCCAGA TCACTGCCAG CAGCTACTTC ACCAACATGT
TTGCCACCTG
7981 GAGCCCCAGC AAGGCCAGGC TGCACCTGCA GGGCAGGAGC AATGCCTGGA
GGCCCCAGGT
8041 CAACAACCCC AAGGAGTGGC TGCAGGTGGA CTTCCAGAAG ACCATGAAGG
TGACTGGGGT
8101 GACCACCCAG GGGGTGAAGA GCCTGCTGAC CAGCATGTAT GTGAAGGAGT
TCCTGATCAG
8161 CAGCAGCCAG GATGGCCACC AGTGGACCCT GTTCTTCCAG AATGGCAAGG
TGAAGGTGTT
8221 CCAGGGCAAC CAGGACAGCT TCACCCCTGT GGTGAACAGC CTGGACCCCC
CCCTGCTGAC
8281 CAGATACCTG AGGATTCACC CCCAGAGCTG GGTGCACCAG ATTGCCCTGA
GGATGGAGGT
8341 GCTGGGCTGT GAGGCCCAGG ACCTGTACTG AGCGGCCGCG GGCCCAATCA
ACCTCTGGAT
8401 TACAAAATTT GTGAAAGATT GACTGGTATT CTTAACTATG TTGCTCCTTT
TACGCTATGT
8461 GGATACGCTG CTTTAATGCC TTTGTATCAT GCTATTGCTT CCCGTATGGC
TTTCATTTTC
8521 TCCTCCTTGT ATAAATCCTG GTTGCTGTCT CTTTATGAGG AGTTGTGGCC
CGTTGTCAGG
8581 CAACGTGGCG TGGTGTGCAC TGTGTTTGCT GACGCAACCC CCACTGGTTG
GGGCATTGCC
8641 ACCACCTGTC AGCTCCTTTC CGGGACTTTC GCTTTCCCCC TCCCTATTGC
CACGGCGGAA
8701 CTCATCGCCG CCTGCCTTGC CCGCTGCTGG ACAGGGGCTC GGCTGTTGGG
CACTGACAAT
8761 TCCGTGGTGT TGTCGGGGAA ATCATCGTCC TTTCCTTGGC TGCTCGCCTG
TGTTGCCACC
8821 TGGATTCTGC GCGGGACGTC CTTCTGCTAC GTCCCTTCGG CCCTCAATCC
AGCGGACCTT
8881 CCTTCCCGCG GCCTGCTGCC GGCTCTGCGG CCTCTTCCGC GTCTTCGCCT
TCGCCCTCAG
8941 ACGAGTCGGA TCTCCCTTTG GGCCGCCTCC CCGCAAGCTT CGCACTTTTT
AAAAGAAAAG
9001 GGAGGACTGG ATGGGATTTA TTACTCCGAT AGGACGCTGG CTTGTAACTC
AGTCTCTTAC
9061 TAGGAGACCA GCTTGAGCCT GGGTGTTCGC TGGTTAGCCT AACCTGGTTG
GCCACCAGGG
9121 GTAAGGACTC CTTGGCTTAG AAAGCTAATA AACTTGCCTG CATTAGAGCT
CTTACGCGTC
9181 CCGGGCTCGA GATCCGCATC TCAATTAGTC AGCAACCATA GTCCCGCCCC
TAACTCCGCC
9241 CATCCCGCCC CTAACTCCGC CCAGTTCCGC CCATTCTCCG CCCCATGGCT
GACTAATTTT
9301 TTTTATTTAT GCAGAGGCCG AGGCCGCCTC GGCCTCTGAG CTATTCCAGA
AGTAGTGAGG
9361 AGGCTTTTTT GGAGGCCTAG GCTTTTGCAA AAAGCTAACT TGTTTATTGC
AGCTTATAAT
9421 GGTTACAAAT AAAGCAATAG CATCACAAAT TTCACAAATA AAGCATTTTT
TTCACTGCAT

9481 TCTAGTTGTG GTTTGTCCAA ACTCATCAAT GTATCTTATC ATGTCTGTCC
GCTTCCTCGC
9541 TCACTGACTC GCTGCGCTCG GTCGTTCGGC TGCGGCGAGC GGTATCAGCT
CACTCAAAGG
9601 CGGTAATACG GTTATCCACA GAATCAGGGG ATAACGCAGG AAAGAACATG
TGAGCAAAAG
9661 GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG CCGCGTTGCT GGCGTTTTTC
CATAGGCTCC
9721 GCCCCCCTGA CGAGCATCAC AAAAATCGAC GCTCAAGTCA GAGGTGGCGA
AACCCGACAG
9781 GACTATAAAG ATACCAGGCG TTTCCCCCTG GAAGCTCCCT CGTGCGCTCT
CCTGTTCCGA
9841 CCCTGCCGCT TACCGGATAC CTGTCCGCCT TTCTCCCTTC GGGAAGCGTG
GCGCTTTCTC
9901 ATAGCTCACG CTGTAGGTAT CTCAGTTCGG TGTAGGTCGT TCGCTCCAAG
CTGGGCTGTG
9961 TGCACGAACC CCCCGTTCAG CCCGACCGCT GCGCCTTATC CGGTAACTAT
CGTCTTGAGT
10021 CCAACCCGGT AAGACACGAC TTATCGCCAC TGGCAGCAGC CACTGGTAAC
AGGATTAGCA
10081 GAGCGAGGTA TGTAGGCGGT GCTACAGAGT TCTTGAAGTG GTGGCCTAAC
TACGGCTACA
10141 CTAGAAGAAC AGTATTTGGT ATCTGCGCTC TGCTGAAGCC AGTTACCTTC
GGAAAAAGAG
10201 TTGGTAGCTC TTGATCCGGC AAACAAACCA CCGCTGGTAG CGGTGGTTTT
TTTGTTTGCA
10261 AGCAGCAGAT TACGCGCAGA AAAAAAGGAT CTCAAGAAGA TCCTTTGATC
TTTTCTACGG
10321 GGTCTGACGC TCAGTGGAAC GAAAACTCAC GTTAAGGGAT TTTGGTCATG
AGATTATCAA
10381 AAAGGATCTT CACCTAGATC CTTTTAAATT AAAAATGAAG TTTTAAATCA
ATCTAAAGTA
10441 TATATGAGTA AACTTGGTCT GACAGTTAGA AAAACTCATC GAGCATCAAA
TGAAACTGCA
10501 ATTTATTCAT ATCAGGATTA TCAATACCAT ATTTTTGAAA AAGCCGTTTC
TGTAATGAAG
10561 GAGAAAACTC ACCGAGGCAG TTCCATAGGA TGGCAAGATC CTGGTATCGG
TCTGCGATTC
10621 CGACTCGTCC AACATCAATA CAACCTATTA ATTTCCCCTC GTCAAAAATA
AGGTTATCAA
10681 GTGAGAAATC ACCATGAGTG ACGACTGAAT CCGGTGAGAA TGGCAACAGC
TTATGCATTT
10741 CTTTCCAGAC TTGTTCAACA GGCCAGCCAT TACGCTCGTC ATCAAAATCA
CTCGCATCAA
10801 CCAAACCGTT ATTCATTCGT GATTGCGCCT GAGCGAGACG AAATACGCGA
TCGCTGTTAA
10861 AAGGACAATT ACAAACAGGA ATCGAATGCA ACCGGCGCAG GAACACTGCC
AGCGCATCAA
10921 CAATATTTTC ACCTGAATCA GGATATTCTT CTAATACCTG GAATGCTGTT
TTTCCGGGGA
10981 TCGCAGTGGT GAGTAACCAT GCATCATCAG GAGTACGGAT AAAATGCTTG
ATGGTCGGAA
11041 GAGGCATAAA TTCCGTCAGC CAGTTTAGTC TGACCATCTC ATCTGTAACA
TCATTGGCAA
11101 CGCTACCTTT GCCATGTTTC AGAAACAACT CTGGCGCATC GGGCTTCCCA
TACAATCGAT
11161 AGATTGTCGC ACCTGATTGC CCGACATTAT CGCGAGCCCA TTTATACCCA
TATAAATCAG

11221 CATCCATGTT GGAATTTAAT CGCGGCCTAG AGCAAGACGT TTCCCGTTGA
ATATGGCTCA
11281 TAACACCCCT TGTATTACTG TTTATGTAAG CAGACAGTTT TATTGTTCAT
GATGATATAT
11341 TTTTATCTTG TGCAATGTAA CATCAGAGAT TTTGAGACAC AACAATTGGT
CGACGGATCC

SEQ ID NO: 14

1 GGTACCTCAA TATTGGCCAT TAGCCATATT ATTCATTGGT TATATAGCAT
AAATCAATAT
61 TGGCTATTGG CCATTGCATA CGTTGTATCT ATATCATAAT ATGTACATTT
ATATTGGCTC
121 ATGTCCAATA TGACCGCCAT GTTGGCATTG ATTATTGACT AGTTATTAAT
AGTAATCAAT
181 TACGGGGTCA TTAGTTCATA GCCCATATAT GGAGTTCCGC GTTACATAAC
TTACGGTAAA
241 TGGCCCGCCT GGCTGACCGC CCAACGACCC CCGCCCATTG ACGTCAATAA
TGACGTATGT
301 TCCCATAGTA ACGCCAATAG GGACTTTCCA TTGACGTCAA TGGGTGGAGT
ATTTACGGTA
361 AACTGCCCAC TTGGCAGTAC ATCAAGTGTA TCATATGCCA AGTCCGCCCC
CTATTGACGT
421 CAATGACGGT AAATGGCCCG CCTGGCATTA TGCCCAGTAC ATGACCTTAC
GGGACTTTCC
481 TACTTGGCAG TACATCTACG TATTAGTCAT CGCTATTACC ATGGTGATGC
GGTTTTGGCA
541 GTACACCAAT GGGCGTGGAT AGCGGTTTGA CTCACGGGGA TTTCCAAGTC
TCCACCCCAT
601 TGACGTCAAT GGGAGTTTGT TTTGGCACCA AAATCAACGG GACTTTCCAA
AATGTCGTAA
661 CAACTGCGAT CGCCCGCCCC GTTGACGCAA ATGGGCGGTA GGCGTGTACG
GTGGGAGGTC
721 TATATAAGCA GAGCTCGCTG GCTTGTAACT CAGTCTCTTA CTAGGAGACC
AGCTTGAGCC
781 TGGGTGTTCG CTGGTTAGCC TAACCTGGTT GGCCACCAGG GGTAAGGACT
CCTTGGCTTA
841 GAAAGCTAAT AAACTTGCCT GCATTAGAGC TTATCTGAGT CAAGTGTCCT
CATTGACGCC
901 TCACTCTCTT GAACGGGAAT CTTCCTTACT GGGTTCTCTC TCTGACCCAG
GCGAGAGAAA
961 CTCCAGCAGT GGCGCCCGAA CAGGGACTTG AGTGAGAGTG TAGGCACGTA
CAGCTGAGAA
1021 GGCGTCGGAC GCGAAGGAAG CGCGGGGTGC GACGCGACCA AGAAGGAGAC
TTGGTGAGTA
1081 GGCTTCTCGA GTGCCGGGAA AAAGCTCGAG CCTAGTTAGA GGACTAGGAG
AGGCCGTAGC
1141 CGTAACTACT CTTGGGCAAG TAGGGCAGGC GGTGGGTACG CAATGGGGGC
GGCTACCTCA
1201 GCACTAAATA GGAGACAATT AGACCAATTT GAGAAAATAC GACTTCGCCC
GAACGGAAAG
1261 AAAAAGTACC AAATTAAACA TTTAATATGG GCAGGCAAGG AGATGGAGCG
CTTCGGCCTC
1321 CATGAGAGGT TGTTGGAGAC AGAGGAGGGG TGTAAAAGAA TCATAGAAGT
CCTCTACCCC
1381 CTAGAACCAA CAGGATCGGA GGGCTTAAAA AGTCTGTTCA ATCTTGTGTG
CGTGCTATAT

1441 TGCTTGCACA AGGAACAGAA AGTGAAAGAC ACAGAGGAAG CAGTAGCAAC AGTAAGACAA
1501 CACTGCCATC TAGTGGAAAA AGAAAAAAGT GCAACAGAGA CATCTAGTGG ACAAAAGAAA
1561 AATGACAAGG GAATAGCAGC GCCACCTGGT GGCAGTCAGA ATTTTCCAGC GCAACAACAA
1621 GGAAATGCCT GGGTACATGT ACCCTTGTCA CCGCGCACCT TAAATGCGTG GGTAAAAGCA
1681 GTAGAGGAGA AAAAATTTGG AGCAGAAATA GTACCCATGT TTCAAGCCCT ATCGAATTCC
1741 CGTTTGTGCT AGGGTTCTTA GGCTTCTTGG GGGCTGCTGG AACTGCAATG GGAGCAGCGG
1801 CGACAGCCCT GACGGTCCAG TCTCAGCATT TGCTTGCTGG GATACTGCAG CAGCAGAAGA
1861 ATCTGCTGGC GGCTGTGGAG GCTCAACAGC AGATGTTGAA GCTGACCATT TGGGGTGTTA
1921 AAAACCTCAA TGCCCGCGTC ACAGCCCTTG AGAAGTACCT AGAGGATCAG GCACGACTAA
1981 ACTCCTGGGG GTGCGCATGG AAACAAGTAT GTCATACCAC AGTGGAGTGG CCCTGGACAA
2041 ATCGGACTCC GGATTGGCAA AATATGACTT GGTTGGAGTG GGAAAGACAA ATAGCTGATT
2101 TGGAAAGCAA CATTACGAGA CAATTAGTGA AGGCTAGAGA ACAAGAGGAA AAGAATCTAG
2161 ATGCCTATCA GAAGTTAACT AGTTGGTCAG ATTTCTGGTC TTGGTTCGAT TTCTCAAAAT
2221 GGCTTAACAT TTTAAAAATG GGATTTTTAG TAATAGTAGG AATAATAGGG TTAAGATTAC
2281 TTTACACAGT ATATGGATGT ATAGTGAGGG TTAGGCAGGG ATATGTTCCT CTATCTCCAC
2341 AGATCCATAT CCGCGGCAAT TTTAAAAGAA AGGGAGGAAT AGGGGGACAG ACTTCAGCAG
2401 AGAGACTAAT TAATATAATA ACAACACAAT TAGAAATACA ACATTTACAA ACCAAAATTC
2461 AAAAAATTTT AAATTTTAGA GCCGCGGAGA TCTGTTACAT AACTTATGGT AAATGGCCTG
2521 CCTGGCTGAC TGCCCAATGA CCCCTGCCCA ATGATGTCAA TAATGATGTA TGTTCCCATG
2581 TAATGCCAAT AGGGACTTTC CATTGATGTC AATGGGTGGA GTATTTATGG TAACTGCCCA
2641 CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTATGCCC CCTATTGATG TCAATGATGG
2701 TAAATGGCCT GCCTGGCATT ATGCCCAGTA CATGACCTTA TGGGACTTTC CTACTTGGCA
2761 GTACATCTAT GTATTAGTCA TTGCTATTAC CATGGGAATT CACTAGTGGA GAAGAGCATG
2821 CTTGAGGGCT GAGTGCCCCT CAGTGGGCAG AGAGCACATG GCCCACAGTC CCTGAGAAGT
2881 TGGGGGGAGG GGTGGGCAAT TGAACTGGTG CCTAGAGAAG GTGGGGCTTG GGTAAACTGG
2941 GAAAGTGATG TGGTGTACTG GCTCCACCTT TTTCCCCAGG GTGGGGGAGA ACCATATATA
3001 AGTGCAGTAG TCTCTGTGAA CATTCAAGCT TCTGCCTTCT CCCTCCTGTG AGTTTGCTAG
3061 CCACCAATGC AGATTGAGCT GAGCACCTGC TTCTTCCTGT GCCTGCTGAG GTTCTGCTTC
3121 TCTGCCACCA GGAGATACTA CCTGGGGGCT GTGGAGCTGA GCTGGGACTA CATGCAGTCT

3181 GACCTGGGGG AGCTGCCTGT GGATGCCAGG TTCCCCCCCA GAGTGCCCAA
GAGCTTCCCC
3241 TTCAACACCT CTGTGGTGTA CAAGAAGACC CTGTTTGTGG AGTTCACTGA
CCACCTGTTC
3301 AACATTGCCA AGCCCAGGCC CCCCTGGATG GGCCTGCTGG GCCCCACCAT
CCAGGCTGAG
3361 GTGTATGACA CTGTGGTGAT CACCCTGAAG AACATGGCCA GCCACCCTGT
GAGCCTGCAT
3421 GCTGTGGGGG TGAGCTACTG GAAGGCCTCT GAGGGGGCTG AGTATGATGA
CCAGACCAGC
3481 CAGAGGGAGA AGGAGGATGA CAAGGTGTTC CCTGGGGGCA GCCACACCTA
TGTGTGGCAG
3541 GTGCTGAAGG AGAATGGCCC CATGGCCTCT GACCCCCTGT GCCTGACCTA
CAGCTACCTG
3601 AGCCATGTGG ACCTGGTGAA GGACCTGAAC TCTGGCCTGA TTGGGGCCCT
GCTGGTGTGC
3661 AGGGAGGGCA GCCTGGCCAA GGAGAAGACC CAGACCCTGC ACAAGTTCAT
CCTGCTGTTT
3721 GCTGTGTTTG ATGAGGGCAA GAGCTGGCAC TCTGAAACCA AGAACAGCCT
GATGCAGGAC
3781 AGGGATGCTG CCTCTGCCAG GGCCTGGCCC AAGATGCACA CTGTGAATGG
CTATGTGAAC
3841 AGGAGCCTGC CTGGCCTGAT TGGCTGCCAC AGGAAGTCTG TGTACTGGCA
TGTGATTGGC
3901 ATGGGCACCA CCCCTGAGGT GCACAGCATC TTCCTGGAGG GCCACACCTT
CCTGGTCAGG
3961 AACCACAGGC AGGCCAGCCT GGAGATCAGC CCCATCACCT TCCTGACTGC
CCAGACCCTG
4021 CTGATGGACC TGGGCCAGTT CCTGCTGTTC TGCCACATCA GCAGCCACCA
GCATGATGGC
4081 ATGGAGGCCT ATGTGAAGGT GGACAGCTGC CCTGAGGAGC CCCAGCTGAG
GATGAAGAAC
4141 AATGAGGAGG CTGAGGACTA TGATGATGAC CTGACTGACT CTGAGATGGA
TGTGGTGAGG
4201 TTTGATGATG ACAACAGCCC CAGCTTCATC CAGATCAGGT CTGTGGCCAA
GAAGCACCCC
4261 AAGACCTGGG TGCACTACAT TGCTGCTGAG GAGGAGGACT GGGACTATGC
CCCCCTGGTG
4321 CTGGCCCCTG ATGACAGGAG CTACAAGAGC CAGTACCTGA ACAATGGCCC
CCAGAGGATT
4381 GGCAGGAAGT ACAAGAAGGT CAGGTTCATG GCCTACACTG ATGAAACCTT
CAAGACCAGG
4441 GAGGCCATCC AGCATGAGTC TGGCATCCTG GGCCCCCTGC TGTATGGGGA
GGTGGGGGAC
4501 ACCCTGCTGA TCATCTTCAA GAACCAGGCC AGCAGGCCCT ACAACATCTA
CCCCCATGGC
4561 ATCACTGATG TGAGGCCCCT GTACAGCAGG AGGCTGCCCA AGGGGGTGAA
GCACCTGAAG
4621 GACTTCCCCA TCCTGCCTGG GGAGATCTTC AAGTACAAGT GGACTGTGAC
TGTGGAGGAT
4681 GGCCCCACCA AGTCTGACCC CAGGTGCCTG ACCAGATACT ACAGCAGCTT
TGTGAACATG
4741 GAGAGGGACC TGGCCTCTGG CCTGATTGGC CCCCTGCTGA TCTGCTACAA
GGAGTCTGTG
4801 GACCAGAGGG GCAACCAGAT CATGTCTGAC AAGAGGAATG TGATCCTGTT
CTCTGTGTTT
4861 GATGAGAACA GGAGCTGGTA CCTGACTGAG AACATCCAGA GGTTCCTGCC
CAACCCTGCT

4921 GGGGTGCAGC TGGAGGACCC TGAGTTCCAG GCCAGCAACA TCATGCACAG
CATCAATGGC
4981 TATGTGTTTG ACAGCCTGCA GCTGTCTGTG TGCCTGCATG AGGTGGCCTA
CTGGTACATC
5041 CTGAGCATTG GGGCCCAGAC TGACTTCCTG TCTGTGTTCT TCTCTGGCTA
CACCTTCAAG
5101 CACAAGATGG TGTATGAGGA CACCCTGACC CTGTTCCCCT TCTCTGGGGA
GACTGTGTTC
5161 ATGAGCATGG AGAACCCTGG CCTGTGGATT CTGGGCTGCC ACAACTCTGA
CTTCAGGAAC
5221 AGGGGCATGA CTGCCCTGCT GAAAGTCTCC AGCTGTGACA AGAACACTGG
GGACTACTAT
5281 GAGGACAGCT ATGAGGACAT CTCTGCCTAC CTGCTGAGCA AGAACAATGC
CATTGAGCCC
5341 AGGAGCTTCA GCCAGAACAG CAGGCACCCC AGCACCAGGC AGAAGCAGTT
CAATGCCACC
5401 ACCATCCCTG AGAATGACAT AGAGAAGACA GACCCATGGT TTGCCCACCG
GACCCCCATG
5461 CCCAAGATCC AGAATGTGAG CAGCTCTGAC CTGCTGATGC TGCTGAGGCA
GAGCCCCACC
5521 CCCCATGGCC TGAGCCTGTC TGACCTGCAG GAGGCCAAGT ATGAAACCTT
CTCTGATGAC
5581 CCCAGCCCTG GGGCCATTGA CAGCAACAAC AGCCTGTCTG AGATGACCCA
CTTCAGGCCC
5641 CAGCTGCACC ACTCTGGGGA CATGGTGTTC ACCCCTGAGT CTGGCCTGCA
GCTGAGGCTG
5701 AATGAGAAGC TGGGCACCAC TGCTGCCACT GAGCTGAAGA AGCTGGACTT
CAAAGTCTCC
5761 AGCACCAGCA ACAACCTGAT CAGCACCATC CCCTCTGACA ACCTGGCTGC
TGGCACTGAC
5821 AACACCAGCA GCCTGGGCCC CCCCAGCATG CCTGTGCACT ATGACAGCCA
GCTGGACACC
5881 ACCCTGTTTG GCAAGAAGAG CAGCCCCCTG ACTGAGTCTG GGGGCCCCCT
GAGCCTGTCT
5941 GAGGAGAACA ATGACAGCAA GCTGCTGGAG TCTGGCCTGA TGAACAGCCA
GGAGAGCAGC
6001 TGGGGCAAGA ATGTGAGCAG CAGGGAGATC ACCAGGACCA CCCTGCAGTC
TGACCAGGAG
6061 GAGATTGACT ATGATGACAC CATCTCTGTG GAGATGAAGA AGGAGGACTT
TGACATCTAC
6121 GACGAGGACG AGAACCAGAG CCCCAGGAGC TTCCAGAAGA AGACCAGGCA
CTACTTCATT
6181 GCTGCTGTGG AGAGGCTGTG GGACTATGGC ATGAGCAGCA GCCCCCATGT
GCTGAGGAAC
6241 AGGGCCCAGT CTGGCTCTGT GCCCCAGTTC AAGAAGGTGG TGTTCCAGGA
GTTCACTGAT
6301 GGCAGCTTCA CCCAGCCCCT GTACAGAGGG GAGCTGAATG AGCACCTGGG
CCTGCTGGGC
6361 CCCTACATCA GGGCTGAGGT GGAGGACAAC ATCATGGTGA CCTTCAGGAA
CCAGGCCAGC
6421 AGGCCCTACA GCTTCTACAG CAGCCTGATC AGCTATGAGG AGGACCAGAG
GCAGGGGGCT
6481 GAGCCCAGGA AGAACTTTGT GAAGCCCAAT GAAACCAAGA CCTACTTCTG
GAAGGTGCAG
6541 CACCACATGG CCCCCACCAA GGATGAGTTT GACTGCAAGG CCTGGGCCTA
CTTCTCTGAT
6601 GTGGACCTGG AGAAGGATGT GCACTCTGGC CTGATTGGCC CCCTGCTGGT
GTGCCACACC

6661 AACACCCTGA ACCCTGCCCA TGGCAGGCAG GTGACTGTGC AGGAGTTTGC CCTGTTCTTC
6721 ACCATCTTTG ATGAAACCAA GAGCTGGTAC TTCACTGAGA ACATGGAGAG GAACTGCAGG
6781 GCCCCCTGCA ACATCCAGAT GGAGGACCCC ACCTTCAAGG AGAACTACAG GTTCCATGCC
6841 ATCAATGGCT ACATCATGGA CACCCTGCCT GGCCTGGTGA TGGCCCAGGA CCAGAGGATC
6901 AGGTGGTACC TGCTGAGCAT GGGCAGCAAT GAGAACATCC ACAGCATCCA CTTCTCTGGC
6961 CATGTGTTCA CTGTGAGGAA GAAGGAGGAG TACAAGATGG CCCTGTACAA CCTGTACCCT
7021 GGGGTGTTTG AGACTGTGGA GATGCTGCCC AGCAAGGCTG GCATCTGGAG GGTGGAGTGC
7081 CTGATTGGGG AGCACCTGCA TGCTGGCATG AGCACCCTGT TCCTGGTGTA CAGCAACAAG
7141 TGCCAGACCC CCCTGGGCAT GGCCTCTGGC CACATCAGGG ACTTCCAGAT CACTGCCTCT
7201 GGCCAGTATG GCCAGTGGGC CCCCAAGCTG GCCAGGCTGC ACTACTCTGG CAGCATCAAT
7261 GCCTGGAGCA CCAAGGAGCC CTTCAGCTGG ATCAAGGTGG ACCTGCTGGC CCCCATGATC
7321 ATCCATGGCA TCAAGACCCA GGGGGCCAGG CAGAAGTTCA GCAGCCTGTA CATCAGCCAG
7381 TTCATCATCA TGTACAGCCT GGATGGCAAG AAGTGGCAGA CCTACAGGGG CAACAGCACT
7441 GGCACCCTGA TGGTGTTCTT TGGCAATGTG GACAGCTCTG GCATCAAGCA CAACATCTTC
7501 AACCCCCCCA TCATTGCCAG ATACATCAGG CTGCACCCCA CCCACTACAG CATCAGGAGC
7561 ACCCTGAGGA TGGAGCTGAT GGGCTGTGAC CTGAACAGCT GCAGCATGCC CCTGGGCATG
7621 GAGAGCAAGG CCATCTCTGA TGCCCAGATC ACTGCCAGCA GCTACTTCAC CAACATGTTT
7681 GCCACCTGGA GCCCCAGCAA GGCCAGGCTG CACCTGCAGG GCAGGAGCAA TGCCTGGAGG
7741 CCCCAGGTCA ACAACCCCAA GGAGTGGCTG CAGGTGGACT TCCAGAAGAC CATGAAGGTG
7801 ACTGGGGTGA CCACCCAGGG GGTGAAGAGC CTGCTGACCA GCATGTATGT GAAGGAGTTC
7861 CTGATCAGCA GCAGCCAGGA TGGCCACCAG TGGACCCTGT TCTTCCAGAA TGGCAAGGTG
7921 AAGGTGTTCC AGGGCAACCA GGACAGCTTC ACCCCTGTGG TGAACAGCCT GGACCCCCCC
7981 CTGCTGACCA GATACCTGAG GATTCACCCC CAGAGCTGGG TGCACCAGAT TGCCCTGAGG
8041 ATGGAGGTGC TGGGCTGTGA GGCCCAGGAC CTGTACTGAG CGGCCGCGGG CCCAATCAAC
8101 CTCTGGATTA CAAAATTTGT GAAAGATTGA CTGGTATTCT TAACTATGTT GCTCCTTTTA
8161 CGCTATGTGG ATACGCTGCT TTAATGCCTT TGTATCATGC TATTGCTTCC CGTATGGCTT
8221 TCATTTTCTC CTCCTTGTAT AAATCCTGGT TGCTGTCTCT TTATGAGGAG TTGTGGCCCG
8281 TTGTCAGGCA ACGTGGCGTG GTGTGCACTG TGTTTGCTGA CGCAACCCCC ACTGGTTGGG
8341 GCATTGCCAC CACCTGTCAG CTCCTTTCCG GGACTTTCGC TTTCCCCCTC CCTATTGCCA

8401 CGGCGGAACT CATCGCCGCC TGCCTTGCCC GCTGCTGGAC AGGGGCTCGG
CTGTTGGGCA
8461 CTGACAATTC CGTGGTGTTG TCGGGGAAAT CATCGTCCTT TCCTTGGCTG
CTCGCCTGTG
8521 TTGCCACCTG GATTCTGCGC GGGACGTCCT TCTGCTACGT CCCTTCGGCC
CTCAATCCAG
8581 CGGACCTTCC TTCCCGCGGC CTGCTGCCGG CTCTGCGGCC TCTTCCGCGT
CTTCGCCTTC
8641 GCCCTCAGAC GAGTCGGATC TCCCTTTGGG CCGCCTCCCC GCAAGCTTCG
CACTTTTTAA
8701 AAGAAAAGGG AGGACTGGAT GGGATTTATT ACTCCGATAG GACGCTGGCT
TGTAACTCAG
8761 TCTCTTACTA GGAGACCAGC TTGAGCCTGG GTGTTCGCTG GTTAGCCTAA
CCTGGTTGGC
8821 CACCAGGGGT AAGGACTCCT TGGCTTAGAA AGCTAATAAA CTTGCCTGCA
TTAGAGCTCT
8881 TACGCGTCCC GGGCTCGAGA TCCGCATCTC AATTAGTCAG CAACCATAGT
CCCGCCCCTA
8941 ACTCCGCCCA TCCCGCCCCT AACTCCGCCC AGTTCCGCCC ATTCTCCGCC
CCATGGCTGA
9001 CTAATTTTTT TTATTTATGC AGAGGCCGAG GCCGCCTCGG CCTCTGAGCT
ATTCCAGAAG
9061 TAGTGAGGAG GCTTTTTTGG AGGCCTAGGC TTTTGCAAAA AGCTAACTTG
TTTATTGCAG
9121 CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA
GCATTTTTTT
9181 CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT
GTCTGTCCGC
9241 TTCCTCGCTC ACTGACTCGC TGCGCTCGGT CGTTCGGCTG CGGCGAGCGG
TATCAGCTCA
9301 CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA
AGAACATGTG
9361 AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC GCGTTGCTGG
CGTTTTTCCA
9421 TAGGCTCCGC CCCCCTGACG AGCATCACAA AAATCGACGC TCAAGTCAGA
GGTGGCGAAA
9481 CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG
TGCGCTCTCC
9541 TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG
GAAGCGTGGC
9601 GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTTC
GCTCCAAGCT
9661 GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC GCCTTATCCG
GTAACTATCG
9721 TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG GCAGCAGCCA
CTGGTAACAG
9781 GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT
GGCCTAACTA
9841 CGGCTACACT AGAAGAACAG TATTTGGTAT CTGCGCTCTG CTGAAGCCAG
TTACCTTCGG
9901 AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAACCACC GCTGGTAGCG
GTGGTTTTTT
9961 TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT CAAGAAGATC
CTTTGATCTT
10021 TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT TAAGGGATTT
TGGTCATGAG
10081 ATTATCAAAA AGGATCTTCA CCTAGATCCT TTTAAATTAA AAATGAAGTT
TTAAATCAAT

10141 CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTAGAAA AACTCATCGA
GCATCAAATG
10201 AAACTGCAAT TTATTCATAT CAGGATTATC AATACCATAT TTTTGAAAAA
GCCGTTTCTG
10261 TAATGAAGGA GAAAACTCAC CGAGGCAGTT CCATAGGATG GCAAGATCCT
GGTATCGGTC
10321 TGCGATTCCG ACTCGTCCAA CATCAATACA ACCTATTAAT TTCCCCTCGT
CAAAAATAAG
10381 GTTATCAAGT GAGAAATCAC CATGAGTGAC GACTGAATCC GGTGAGAATG
GCAACAGCTT
10441 ATGCATTTCT TTCCAGACTT GTTCAACAGG CCAGCCATTA CGCTCGTCAT
CAAAATCACT
10501 CGCATCAACC AAACCGTTAT TCATTCGTGA TTGCGCCTGA GCGAGACGAA
ATACGCGATC
10561 GCTGTTAAAA GGACAATTAC AAACAGGAAT CGAATGCAAC CGGCGCAGGA
ACACTGCCAG
10621 CGCATCAACA ATATTTTCAC CTGAATCAGG ATATTCTTCT AATACCTGGA
ATGCTGTTTT
10681 TCCGGGGATC GCAGTGGTGA GTAACCATGC ATCATCAGGA GTACGGATAA
AATGCTTGAT
10741 GGTCGGAAGA GGCATAAATT CCGTCAGCCA GTTTAGTCTG ACCATCTCAT
CTGTAACATC
10801 ATTGGCAACG CTACCTTTGC CATGTTTCAG AAACAACTCT GGCGCATCGG
GCTTCCCATA
10861 CAATCGATAG ATTGTCGCAC CTGATTGCCC GACATTATCG CGAGCCCATT
TATACCCATA
10921 TAAATCAGCA TCCATGTTGG AATTTAATCG CGGCCTAGAG CAAGACGTTT
CCCGTTGAAT
10981 ATGGCTCATA ACACCCCTTG TATTACTGTT TATGTAAGCA GACAGTTTTA
TTGTTCATGA
11041 TGATATATTT TTATCTTGTG CAATGTAACA TCAGAGATTT TGAGACACAA
CAATTGGTCG
11101 ACGGATCC

SEQ ID NO: 15

ATGCCCAGCTCTGTGTCCTGGGGCATTCTGCTGCTGGCTGGCCTGTGCTGTCTGGTGCCTGTGTCCCTGG
CTGAGGACCCTCAGGGGGATGCTGCCCAGAAAACAGACACCTCCCACCATGACCAGGACCACCCCACCTT
CAACAAGATCACCCCCAACCTGGCAGAGTTTGCCTTCAGCCTGTACAGACAGCTGGCCCACCAGAGCAAC
AGCACCAACATCTTTTTCAGCCCTGTGTCCATTGCCACAGCCTTTGCCATGCTGAGCCTGGGCACCAAGG
CTGACACCCATGATGAGATCCTGGAAGGCCTGAACTTCAACCTGACAGAGATCCCTGAGGCCCAGATCCA
TGAGGGCTTCCAGGAACTGCTGAGAACCCTGAACCAGCCAGACAGCCAGCTGCAGCTGACAACAGGCAAT
GGGCTGTTCCTGTCTGAGGGCCTGAAGCTGGTGGACAAGTTTCTGGAAGATGTGAAGAAGCTGTACCACT
CTGAGGCCTTCACAGTGAACTTTGGGGACACAGAAGAGGCCAAGAAACAGATCAATGACTATGTGGAAAA
GGGCACCCAGGGCAAGATTGTGGACCTTGTCAAAGAGCTGGACAGGGACACTGTGTTTGCCCTTGTGAAC
TACATCTTCTTCAAGGGCAAGTGGGAGAGGCCCTTTGAAGTGAAGGACACTGAGGAAGAGGACTTCCATG
TGGACCAAGTGACCACAGTGAAGGTGCCAATGATGAAGAGACTGGGGATGTTCAATATCCAGCACTGCAA
GAAACTGAGCAGCTGGGTGCTGCTGATGAAGTACCTGGGCAATGCTACAGCCATATTCTTTCTGCCTGAT
GAGGGCAAGCTGCAGCACCTGGAAAATGAGCTGACCCATGACATCATCACCAAATTTCTGGAAAATGAGG
ACAGAAGATCTGCCAGCCTGCATCTGCCCAAGCTGAGCATCACAGGCACATATGACCTGAAGTCTGTGCT
GGGACAGCTGGGAATCACCAAGGTGTTCAGCAATGGGGCAGACCTGAGTGGAGTGACAGAGGAAGCCCCT
CTGAAGCTGTCCAAGGCTGTGCACAAGGCAGTGCTGACCATTGATGAGAAGGGCACAGAGGCTGCTGGGG
CCATGTTTCTGGAAGCCATCCCCATGTCCATCCCCCCAGAAGTGAAGTTCAACAAGCCCTTTGTGTTCCT
GATGATTGAGCAGAACACCAAGAGCCCCCTGTTCATGGGCAAGGTTGTGAACCCCACCCAGAAATGA

SEQ ID NO: 16

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGAT
ACTACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGGGGAGCTGCCTGTGGATGC
CAGGTTCCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCTGTGGTGTACAAGAAGACCCTGTTT
GTGGAGTTCACTGACCACCTGTTCAACATTGCCAAGCCCAGGCCCCCCTGGATGGGCCTGCTGGGCCCCA
CCATCCAGGCTGAGGTGTATGACACTGTGGTGATCACCCTGAAGAACATGGCCAGCCACCCTGTGAGCCT
GCATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGACCAGCCAGAGG
GAGAAGGAGGATGACAAGGTGTTCCCTGGGGGCAGCCACACCTATGTGTGGCAGGTGCTGAAGGAGAATG
GCCCCATGGCCTCTGACCCCCTGTGCCTGACCTACAGCTACCTGAGCCATGTGGACCTGGTGAAGGACCT
GAACTCTGGCCTGATTGGGGCCCTGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACC
CTGCACAAGTTCATCCTGCTGTTTGCTGTGTTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACA
GCCTGATGCAGGACAGGGATGCTGCCTCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGT
GAACAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTGTGTACTGGCATGTGATTGGCATGGGC
ACCACCCCTGAGGTGCACAGCATCTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGGCAGGCCA
GCCTGGAGATCAGCCCCATCACCTTCCTGACTGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCT
GTTCTGCCACATCAGCAGCCACCAGCATGATGGCATGGAGGCCTATGTGAAGGTGGACAGCTGCCCTGAG
GAGCCCCAGCTGAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGA
TGGATGTGGTGAGGTTTGATGATGACAACAGCCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAGCA
CCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCCCTGGTGCTGGCC
CCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGAGGATTGGCAGGAAGTACAAGA
AGGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATCCAGCATGAGTCTGGCAT
CCTGGGCCCCCTGCTGTATGGGGAGGTGGGGGACACCCTGCTGATCATCTTCAAGAACCAGGCCAGCAGG
CCCTACAACATCTACCCCCATGGCATCACTGATGTGAGGCCCCTGTACAGCAGGAGGCTGCCCAAGGGGG
TGAAGCACCTGAAGGACTTCCCCATCCTGCCTGGGGAGATCTTCAAGTACAAGTGGACTGTGACTGTGGA
GGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGATACTACAGCAGCTTTGTGAACATGGAGAGG
GACCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAAGGAGTCTGTGGACCAGAGGGGCAACC
AGATCATGTCTGACAAGAGGAATGTGATCCTGTTCTCTGTGTTTGATGAGAACAGGAGCTGGTACCTGAC
TGAGAACATCCAGAGGTTCCTGCCCAACCCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGC
AACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCTGCAGCTGTCTGTGTGCCTGCATGAGGTGG
CCTACTGGTACATCCTGAGCATTGGGGCCCAGACTGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTT
CAAGCACAAGATGGTGTATGAGGACACCCTGACCCTGTTCCCCTTCTCTGGGGAGACTGTGTTCATGAGC
ATGGAGAACCCTGGCCTGTGGATTCTGGGCTGCCACAACTCTGACTTCAGGAACAGGGGCATGACTGCCC
TGCTGAAAGTCTCCAGCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGC
CTACCTGCTGAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAACAGCAGGCACCCCAGCACC
AGGCAGAAGCAGTTCAATGCCACCACCATCCCTGAGAATGACATAGAGAAGACAGACCCATGGTTTGCCC
ACCGGACCCCCATGCCCAAGATCCAGAATGTGAGCAGCTCTGACCTGCTGATGCTGCTGAGGCAGAGCCC
CACCCCCCATGGCCTGAGCCTGTCTGACCTGCAGGAGGCCAAGTATGAAACCTTCTCTGATGACCCCAGC
CCTGGGGCCATTGACAGCAACAACAGCCTGTCTGAGATGACCCACTTCAGGCCCCAGCTGCACCACTCTG
GGGACATGGTGTTCACCCCTGAGTCTGGCCTGCAGCTGAGGCTGAATGAGAAGCTGGGCACCACTGCTGC
CACTGAGCTGAAGAAGCTGGACTTCAAAGTCTCCAGCACCAGCAACAACCTGATCAGCACCATCCCCTCT
GACAACCTGGCTGCTGGCACTGACAACACCAGCAGCCTGGGCCCCCCCAGCATGCCTGTGCACTATGACA
GCCAGCTGGACACCACCCTGTTTGGCAAGAAGAGCAGCCCCCTGACTGAGTCTGGGGGCCCCCTGAGCCT
GTCTGAGGAGAACAATGACAGCAAGCTGCTGGAGTCTGGCCTGATGAACAGCCAGGAGAGCAGCTGGGGC
AAGAATGTGAGCAGCAGGGAGATCACCAGGACCACCCTGCAGTCTGACCAGGAGGAGATTGACTATGATG
ACACCATCTCTGTGGAGATGAAGAAGGAGGACTTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAG
GAGCTTCCAGAAGAAGACCAGGCACTACTTCATTGCTGCTGTGGAGAGGCTGTGGGACTATGGCATGAGC
AGCAGCCCCCATGTGCTGAGGAACAGGGCCCAGTCTGGCTCTGTGCCCCAGTTCAAGAAGGTGGTGTTCC
AGGAGTTCACTGATGGCAGCTTCACCCAGCCCCTGTACAGAGGGGAGCTGAATGAGCACCTGGGCCTGCT
GGGCCCCTACATCAGGGCTGAGGTGGAGGACAACATCATGGTGACCTTCAGGAACCAGGCCAGCAGGCCC
TACAGCTTCTACAGCAGCCTGATCAGCTATGAGGAGGACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACT
TTGTGAAGCCCAATGAAACCAAGACCTACTTCTGGAAGGTGCAGCACCACATGGCCCCCACCAAGGATGA
GTTTGACTGCAAGGCCTGGGCCTACTTCTCTGATGTGGACCTGGAGAAGGATGTGCACTCTGGCCTGATT
GGCCCCCTGCTGGTGTGCCACACCAACACCCTGAACCCTGCCCATGGCAGGCAGGTGACTGTGCAGGAGT
TTGCCCTGTTCTTCACCATCTTTGATGAAACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAACTG
CAGGGCCCCCTGCAACATCCAGATGGAGGACCCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAAT
GGCTACATCATGGACACCCTGCCTGGCCTGGTGATGGCCCAGGACCAGAGGATCAGGTGGTACCTGCTGA
GCATGGGCAGCAATGAGAACATCCACAGCATCCACTTCTCTGGCCATGTGTTCACTGTGAGGAAGAAGGA
GGAGTACAAGATGGCCCTGTACAACCTGTACCCTGGGGTGTTTGAGACTGTGGAGATGCTGCCCAGCAAG
GCTGGCATCTGGAGGGTGGAGTGCCTGATTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGG

TGTACAGCAACAAGTGCCAGACCCCCCTGGGCATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGC
CTCTGGCCAGTATGGCCAGTGGGCCCCCAAGCTGGCCAGGCTGCACTACTCTGGCAGCATCAATGCCTGG
AGCACCAAGGAGCCCTTCAGCTGGATCAAGGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAGA
CCCAGGGGGCCAGGCAGAAGTTCAGCAGCCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGATGG
CAAGAAGTGGCAGACCTACAGGGGCAACAGCACTGGCACCCTGATGGTGTTCTTTGGCAATGTGGACAGC
TCTGGCATCAAGCACAACATCTTCAACCCCCCCATCATTGCCAGATACATCAGGCTGCACCCCACCCACT
ACAGCATCAGGAGCACCCTGAGGATGGAGCTGATGGGCTGTGACCTGAACAGCTGCAGCATGCCCCTGGG
CATGGAGAGCAAGGCCATCTCTGATGCCCACATCACTGCCAGCAGCTACTTCACCAACATGTTTGCCACC
TGGAGCCCCAGCAAGGCCAGGCTGCACCTGCAGGGCAGGAGCAATGCCTGGAGGCCCCAGGTCAACAACC
CCAAGGAGTGGCTGCAGGTGGACTTCCAGAAGACCATGAAGGTGACTGGGGTGACCACCCAGGGGGTGAA
GAGCCTGCTGACCAGCATGTATGTGAAGGAGTTCCTGATCAGCAGCAGCCAGGATGGCCACCAGTGGACC
CTGTTCTTCCAGAATGGCAAGGTGAAGGTGTTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACA
GCCTGGACCCCCCCCTGCTGACCAGATACCTGAGGATTCACCCCCAGAGCTGGGTGCACCAGATTGCCCT
GAGGATGGAGGTGCTGGGCTGTGAGGCCCAGGACCTGTACTGA

SEQ ID NO: 17

CCGCGGAGATCTCAATATTGGCCATTAGCCATATTATTCATTGGTTATATAGCATAAATCAATATTGGCT
ATTGGCCATTGCATACGTTGTATCTATATCATAATATGTACATTTATATTGGCTCATGTCCAATATGACC
GCCATGTTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCC
CATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGT
GGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTCCGCCCCCTATT
GACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTACGGGACTTTCCTACTT
GGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACACCAATGGGCG
TGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGG
CACCAAAATCAACGGGACTTTCCAAAATGTCGTAATAACCCCGCCCCGTTGACGCAAATGGGCGGTAGGC
GTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCACTAGAAGCTTTATTGC
GGTAGTTTATCACAGTTAAATTGCTAACGCAGTCAGTGCTTCTGACACAACAGTCTCGAACTTAAGCTGC
AGAAGTTGGTCGTGAGGCACTGGGCAGGCTAGC

SEQ ID NO: 18

TCGAGATGTGGTCTGAGTTAAAAATCAGGAGCAACGACGGAGGTGAAGGACCAGACGCCAACGACCC

SEQ ID NO: 19

CCGGGGGTCGTTGGCGTCTGGTCCTTCACCTCCGTCGTTGCTCCTGATTTTTAACTCAGACCACATC

SEQ ID NO: 20

CCGGGGAAAGGGGGTGCAACACATCCATATCCAGCCATCTCTACCTGTTTATGGACA

SEQ ID NO: 21

ACCCTCTGTCCATAAACAGGTAGAGATGGCTGGATATGGATGTGTTGCACCCCTTTCC

SEQ ID NO: 22

GGGTTAGGTGGTTGCTGATTCTCTCATTCACCCAGTGGG

SEQ ID NO: 23

GATCCCCACTGGGTGAATGAGAGAATCAGCAACCACCTA

SEQ ID NO: 24

GAGACTCGAGATGTGGTCTGAGTTAAAAATCAGG

SEQ ID NO: 25

AGAGGTAGACCAGTACGAGTCACGTTTGCCCCTATCACCATCCCTAACCCTCTGTCATAAAC

SEQ ID NO: 26

TACGGGTCGAGACACAGGACCCCGTAAGACGACGACCGACCGGACACGACAGACCACGGACACAGGGACC
GACTCCTGGGAGTCCCCCTACGACGGGTCTTTTGTCTGTGGAGGGTGGTACTGGTCCTGGTGGGGTGGAA
GTTGTTCTAGTGGGGGTTGGACCGTCTCAAACGGAAGTCGGACATGTCTGTCGACCGGGTGGTCTCGTTG
TCGTGGTTGTAGAAAAAGTCGGGACACAGGTAACGGTGTCGGAAACGGTACGACTCGGACCCGTGGTTCC
GACTGTGGGTACTACTCTAGGACCTTCCGGACTTGAAGTTGGACTGTCTCTAGGGACTCCGGGTCTAGGT
ACTCCCGAAGGTCCTTGACGACTCTTGGGACTTGGTCGGTCTGTCGGTCGACGTCGACTGTTGTCCGTTA
CCCGACAAGGACAGACTCCCGGACTTCGACCACCTGTTCAAAGACCTTCTACACTTCTTCGACATGGTGA
GACTCCGGAAGTGTCACTTGAAACCCCTGTGTCTTCTCCGGTTCTTTGTCTAGTTACTGATACACCTTTT
CCCGTGGGTCCCGTTCTAACACCTGGAACACTTTCTCGACCTGTCCCTGTGACACAAACGGGAACACTTG
ATGTAGAAGAAGTTCCCGTTCACCCTCTCCGGGAAACTTCACTTCCTGTGACTCCTTCTCCTGAAGGTAC
ACCTGGTTCACTGGTGTCACTTCCACGGTTACTACTTCTCTGACCCCTACAAGTTATAGGTCGTGACGTT
CTTTGACTCGTCGACCCACGACGACTACTTCATGGACCCGTTACGATGTCGGTATAAGAAAGACGGACTA
CTCCCGTTCGACGTCGTGGACCTTTTACTCGACTGGGTACTGTAGTAGTGGTTTAAAGACCTTTTACTCC
TGTCTTCTAGACGGTCGGACGTAGACGGGTTCGACTCGTAGTGTCCGTGTATACTGGACTTCAGACACGA
CCCTGTCGACCCTTAGTGGTTCCACAAGTCGTTACCCCGTCTGGACTCACCTCACTGTCTCCTTCGGGGA
GACTTCGACAGGTTCCGACACGTGTTCCGTCACGACTGGTAACTACTCTTCCCGTGTCTCCGACGACCCC
GGTACAAAGACCTTCGGTAGGGGTACAGGTAGGGGGGTCTTCACTTCAAGTTGTTCGGGAAACACAAGGA
CTACTAACTCGTCTTGTGGTTCTCGGGGGACAAGTACCCGTTCCAACACTTGGGGTGGGTCTTTACT

SEQ ID NO: 27

AEDPQGDAAQKTDTSHHDQDHPTFAEDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLYRQLAHQSN
STNIFFSPVSIATAFAMLSLGTKADTHDEILEGLNFNLTEIPEAQIHEGFQELLRTLNQPDSQLQLTTGNG
LFLSEGLKLVDKFLEDVKKLYHSEAFTVNFGDTEEAKKQINDYVEKGTQGKIVDLVKELDRDTVFALVNYI
FFKGKWERPFEVKDTEEEDFHVDQVTTVKVPMMKRLGMFNIQHCKKLSSWVLLMKYLGNATAIFFLPDEGK
LQHLENELTHDIITKFLENEDRRSASLHLPKLSITGTYDLKSVLGQLGITKVFSNGADLSGVTEEAPLKLS
KAVHKAVLTIDEKGTEAAGAMFLEAIPMSIPPEVKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK

SEQ ID NO: 28

TACGTCTAACTCGACTCGTGGACGAAGAAGGACACGGACGACTCCAAGACGAAGAGACGGTGGTCCTCTA
TGATGGACCCCCGACACCTCGACTCGACCCTGATGTACGTCAGACTGGACCCCCTCGACGGACACCTACG
GTCCAAGGGGGGGTCTCACGGGTTCTCGAAGGGGAAGTTGTGGAGACACCACATGTTCTTCTGGGACAAA
CACCTCAAGTGACTGGTGGACAAGTTGTAACGGTTCGGGTCCGGGGGGACCTACCCGGACGACCCGGGGT
GGTAGGTCCGACTCCACATACTGTGACACCACTAGTGGGACTTCTTGTACCGGTCGGTGGGACACTCGGA
CGTACGACACCCCCACTCGATGACCTTCCGGAGACTCCCCCGACTCATACTACTGGTCTGGTCGGTCTCC
CTCTTCCTCCTACTGTTCCACAAGGGACCCCCGTCGGTGTGGATACACACCGTCCACGACTTCCTCTTAC
CGGGGTACCGGAGACTGGGGGACACGGACTGGATGTCGATGGACTCGGTACACCTGGACCACTTCCTGGA
CTTGAGACCGGACTAACCCCGGGACGACCACACGTCCCTCCCGTCGGACCGGTTCCTCTTCTGGGTCTGG
GACGTGTTCAAGTAGGACGACAAACGACACAAACTACTCCCGTTCTCGACCGTGAGACTTTGGTTCTTGT
CGGACTACGTCCTGTCCCTACGACGGAGACGGTCCCGGACCGGGTTCTACGTGTGACACTTACCGATACA
CTTGTCCTCGGACGGACCGGACTAACCGACGGTGTCCTTCAGACACATGACCGTACACTAACCGTACCCG
TGGTGGGGACTCCACGTGTCGTAGAAGGACCTCCCGGTGTGGAAGGACCAGTCCTTGGTGTCCGTCCGGT
CGGACCTCTAGTCGGGGTAGTGGAAGGACTGACGGGTCTGGGACGACTACCTGGACCCGGTCAAGGACGA
CAAGACGGTGTAGTCGTCGGTGGTCGTACTACCGTACCTCCGGATACACTTCCACCTGTCGACGGGACTC
CTCGGGGTCGACTCCTACTTCTTGTTACTCCTCCGACTCCTGATACTACTACTGGACTGACTGAGACTCT
ACCTACACCACTCCAAACTACTACTGTTGTCGGGGTCGAAGTAGGTCTAGTCCAGACACCGGTTCTTCGT

GGGGTTCTGGACCCACGTGATGTAACGACGACTCCTCCTCCTGACCCTGATACGGGGGGACCACGACCGG
GGACTACTGTCCTCGATGTTCTCGGTCATGGACTTGTTACCGGGGGTCTCCTAACCGTCCTTCATGTTCT
TCCAGTCCAAGTACCGGATGTGACTACTTTGGAAGTTCTGGTCCCTCCGGTAGGTCGTACTCAGACCGTA
GGACCCGGGGGACGACATACCCCTCCACCCCCTGTGGGACGACTAGTAGAAGTTCTTGGTCCGGTCGTCC
GGGATGTTGTAGATGGGGGTACCGTAGTGACTACACTCCGGGGACATGTCGTCCTCCGACGGGTTCCCCC
ACTTCGTGGACTTCCTGAAGGGGTAGGACGGACCCCTCTAGAAGTTCATGTTCACCTGACACTGACACCT
CCTACCGGGGTGGTTCAGACTGGGGTCCACGGACTGGTCTATGATGTCGTCGAAACACTTGTACCTCTCC
CTGGACCGGAGACCGGACTAACCGGGGGACCACTAGACGATGTTCCTCAGACACCTGGTCTCCCCGTTGG
TCTAGTACAGACTGTTCTCCTTACACTAGGACAAGAGACACAAACTACTCTTGTCCTCGACCATGGACTG
ACTCTTGTAGGTCTCCAAGGACGGGTTGGGACGACCCCACGTCGACCTCCTGGGACTCAAGGTCCGGTCG
TTGTAGTACGTGTCGTAGTTACCGATACACAAACTGTCGGACGTCGACAGACACACGGACGTACTCCACC
GGATGACCATGTAGGACTCGTAACCCCGGGTCTGACTGAAGGACAGACACAAGAAGAGACCGATGTGGAA
GTTCGTGTTCTACCACATACTCCTGTGGGACTGGGACAAGGGGAAGAGACCCCTCTGACACAAGTACTCG
TACCTCTTGGGACCGGACACCTAAGACCCGACGGTGTTGAGACTGAAGTCCTTGTCCCCGTACTGACGGG
ACGACTTTCAGAGGTCGACACTGTTCTTGTGACCCCTGATGATACTCCTGTCGATACTCCTGTAGAGACG
GATGGACGACTCGTTCTTGTTACGGTAACTCGGGTCCTCGAAGTCGGTCTTGTCGTCCGTGGGGTCGTGG
TCCGTCTTCGTCAAGTTACGGTGGTGGTAGGGACTCTTACTGTATCTCTTCTGTCTGGGTACCAAACGGG
TGGCCTGGGGGTACGGGTTCTAGGTCTTACACTCGTCGAGACTGGACGACTACGACGACTCCGTCTCGGG
GTGGGGGGTACCGGACTCGGACAGACTGGACGTCCTCCGGTTCATACTTTGGAAGAGACTACTGGGGTCG
GGACCCCGGTAACTGTCGTTGTTGTCGGACAGACTCTACTGGGTGAAGTCCGGGGTCGACGTGGTGAGAC
CCCTGTACCACAAGTGGGGACTCAGACCGGACGTCGACTCCGACTTACTCTTCGACCCGTGGTGACGACG
GTGACTCGACTTCTTCGACCTGAAGTTTCAGAGGTCGTGGTCGTTGTTGGACTAGTCGTGGTAGGGGAGA
CTGTTGGACCGACGACCGTGACTGTTGTGGTCGTCGGACCCGGGGGGGTCGTACGGACACGTGATACTGT
CGGTCGACCTGTGGTGGGACAAACCGTTCTTCTCGTCGGGGGACTGACTCAGACCCCCGGGGGACTCGGA
CAGACTCCTCTTGTTACTGTCGTTCGACGACCTCAGACCGGACTACTTGTCGGTCCTCTCGTCGACCCCG
TTCTTACACTCGTCGTCCCTCTAGTGGTCCTGGTGGGACGTCAGACTGGTCCTCCTCTAACTGATACTAC
TGTGGTAGAGACACCTCTACTTCTTCCTCCTGAAACTGTAGATGCTGCTCCTGCTCTTGGTCTCGGGGTC
CTCGAAGGTCTTCTTCTGGTCCGTGATGAAGTAACGACGACACCTCTCCGACACCCTGATACCGTACTCG
TCGTCGGGGGTACACGACTCCTTGTCCCGGGTCAGACCGAGACACGGGGTCAAGTTCTTCCACCACAAGG
TCCTCAAGTGACTACCGTCGAAGTGGGTCGGGGACATGTCTCCCCTCGACTTACTCGTGGACCCGGACGA
CCCGGGGATGTAGTCCCGACTCCACCTCCTGTTGTAGTACCACTGGAAGTCCTTGGTCCGGTCGTCCGGG
ATGTCGAAGATGTCGTCGGACTAGTCGATACTCCTCCTGGTCTCCGTCCCCCGACTCGGGTCCTTCTTGA
AACACTTCGGGTTACTTTGGTTCTGGATGAAGACCTTCCACGTCGTGGTGTACCGGGGGTGGTTCCTACT
CAAACTGACGTTCCGGACCCGGATGAAGAGACTACACCTGGACCTCTTCCTACACGTGAGACCGGACTAA
CCGGGGGACGACCACACGGTGTGGTTGTGGGACTTGGGACGGGTACCGTCCGTCCACTGACACGTCCTCA
AACGGGACAAGAAGTGGTAGAAACTACTTTGGTTCTCGACCATGAAGTGACTCTTGTACCTCTCCTTGAC
GTCCCGGGGGACGTTGTAGGTCTACCTCCTGGGGTGGAAGTTCCTCTTGATGTCCAAGGTACGGTAGTTA
CCGATGTAGTACCTGTGGGACGGACCGGACCACTACCGGGTCCTGGTCTCCTAGTCCACCATGGACGACT
CGTACCCGTCGTTACTCTTGTAGGTGTCGTAGGTGAAGAGACCGGTACACAAGTGACACTCCTTCTTCCT
CCTCATGTTCTACCGGGACATGTTGGACATGGGACCCCACAAACTCTGACACCTCTACGACGGGTCGTTC
CGACCGTAGACCTCCCACCTCACGGACTAACCCCTCGTGGACGTACGACCGTACTCGTGGGACAAGGACC
ACATGTCGTTGTTCACGGTCTGGGGGGACCCGTACCGGAGACCGGTGTAGTCCCTGAAGGTCTAGTGACG
GAGACCGGTCATACCGGTCACCCGGGGGTTCGACCGGTCCGACGTGATGAGACCGTCGTAGTTACGGACC
TCGTGGTTCCTCGGGAAGTCGACCTAGTTCCACCTGGACGACCGGGGGTACTAGTAGGTACCGTAGTTCT
GGGTCCCCCGGTCCGTCTTCAAGTCGTCGGACATGTAGTCGGTCAAGTAGTAGTACATGTCGGACCTACC
GTTCTTCACCGTCTGGATGTCCCCGTTGTCGTGACCGTGGGACTACCACAAGAAACCGTTACACCTGTCG
AGACCGTAGTTCGTGTTGTAGAAGTTGGGGCGGTAGTAACGGTCTATGTAGTCCGACGTGGGGTGGGTGA
TGTCGTAGTCCTCGTGGGACTCCTACCTCGACTACCCGACACTGGACTTGTCGACGTCGTACGGGGACCC
GTACCTCTCGTTCCGGTAGAGACTACGGGTCTAGTGACGGTCGTCGATGAAGTGGTTGTACAAACGGTGG
ACCTCGGGGTCGTTCCGGTCCGACGTGGACGTCCCGTCCTCGTTACGGACCTCCGGGGTCCAGTTGTTGG
GGTTCCTCACCGACGTCCACCTGAAGGTCTTCTGGTACTTCCACTGACCCCACTGGTGGGTCCCCCACTT
CTCGGACGACTGGTCGTACATACACTTCCTCAAGGACTAGTCGTCGTCGGTCCTACCGGTGGTCACCTGG
GACAAGAAGGTCTTACCGTTCCACTTCCACAAGGTCCCGTTGGTCCTGTCGAAGTGGGGACACCACTTGT
CGGACCTGGGGGGGGACGACTGGTCTATGGACTCCTAAGTGGGGGTCTCGACCCACGTGGTCTAACGGGA
CTCCTACCTCCACGACCCGACACTCCGGGTCCTGGACATGACT

SEQ ID NO: 29

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYKKTLFV
EFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQREK
EDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQTLHK
FILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMGTTPE
VHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSHQHDGMEAYVKVDSCPEEPQLR
MKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLAPDDRSY
KSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLLIIFKNQASRPYNIYPH
GITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSSFVNMERDLASGLIG
PLLICYKESVDQRGNQIMSDKRNVILFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQASNIMHSINGY
VFDSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMSMENPGLWILG
CHNSDFRNRGMTALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNSRHPSTRQKQFNATTIP
ENDIEKTDPWFAHRTPMPKIQNVSSSDLLMLLRQSPTPHGLSLSDLQEAKYETFSDDPSPGAIDSNNSLSE
MTHFRPQLHHSGDMVFTPESGLQLRLNEKLGTTAATELKKLDFKVSSTSNNLISTIPSDNLAAGTDNTSSL
GPPSMPVHYDSQLDTTLFGKKSSPLTESGGPLSLSEENNDSKLLESGLMNSQESSWGKNVSSREITRTTLQ
SDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKKTRHYFIAAVERLWDYGMSSSPHVLRNRAQSGSV
PQFKKVVFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVEDNIMVTFRNQASRPYSFYSSLISYEEDQRQ
GAEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKAWAYFSDVDLEKDVHSGLIGPLLVCHTNTLNPAHGR
QVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQMEDPTFKENYRFHAINGYIMDTLPGLVMAQDQRI
RWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNLYPGVFETVEMLPSKAGIWRVECLIGEHLHAGMS
TLFLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAPKLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIH
GIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGNSTGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHP
THYSIRSTLRMELMGCDLNSCSMPLGMESKAISDAQITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVN
NPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVKEFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVN
SLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEAQDLY

SEQ ID NO: 30

ATGCAGATTGAGCTGAGCACCTGCTTCTTCCTGTGCCTGCTGAGGTTCTGCTTCTCTGCCACCAGGAGAT
ACTACCTGGGGGCTGTGGAGCTGAGCTGGGACTACATGCAGTCTGACCTGGGGGGAGCTGCCTGTGGATGC
CAGGTTCCCCCCCAGAGTGCCCAAGAGCTTCCCCTTCAACACCTCTGTGGTGTACAAGAAGACCCTGTTT
GTGGAGTTCACTGACCACCTGTTCAACATTGCCAAGCCCAGGCCCCCCTGGATGGGCCTGCTGGGCCCCA
CCATCCAGGCTGAGGTGTATGACACTGTGGTGATCACCCTGAAGAACATGGCCAGCCACCCTGTGAGCCT
GCATGCTGTGGGGGTGAGCTACTGGAAGGCCTCTGAGGGGGCTGAGTATGATGACCAGACCAGCCAGAGG
GAGAAGGAGGATGACAAGGTGTTCCCTGGGGGCAGCCACACCTATGTGTGGCAGGTGCTGAAGGAGAATG
GCCCCATGGCCTCTGACCCCCTGTGCCTGACCTACAGCTACCTGAGCCATGTGGACCTGGTGAAGGACCT
GAACTCTGGCCTGATTGGGGCCCTGCTGGTGTGCAGGGAGGGCAGCCTGGCCAAGGAGAAGACCCAGACC
CTGCACAAGTTCATCCTGCTGTTTGCTGTGTTTGATGAGGGCAAGAGCTGGCACTCTGAAACCAAGAACA
GCCTGATGCAGGACAGGGATGCTGCCTCTGCCAGGGCCTGGCCCAAGATGCACACTGTGAATGGCTATGT
GAACAGGAGCCTGCCTGGCCTGATTGGCTGCCACAGGAAGTCTGTGTACTGGCATGTGATTGGCATGGGC
ACCACCCCTGAGGTGCACAGCATCTTCCTGGAGGGCCACACCTTCCTGGTCAGGAACCACAGGCAGGCCA
GCCTGGAGATCAGCCCCATCACCTTCCTGACTGCCCAGACCCTGCTGATGGACCTGGGCCAGTTCCTGCT
GTTCTGCCACATCAGCAGCCACCAGCATGATGGCATGGAGGCCTATGTGAAGGTGGACAGCTGCCCTGAG
GAGCCCCAGCTGAGGATGAAGAACAATGAGGAGGCTGAGGACTATGATGATGACCTGACTGACTCTGAGA
TGGATGTGGTGAGGTTTGATGATGACAACACCCCAGCTTCATCCAGATCAGGTCTGTGGCCAAGAAGCA
CCCCAAGACCTGGGTGCACTACATTGCTGCTGAGGAGGAGGACTGGGACTATGCCCCCCTGGTGCTGGCC
CCTGATGACAGGAGCTACAAGAGCCAGTACCTGAACAATGGCCCCCAGAGGATTGGCAGGAAGTACAAGA
AGGTCAGGTTCATGGCCTACACTGATGAAACCTTCAAGACCAGGGAGGCCATCCAGCATGAGTCTGGCAT
CCTGGGCCCCCTGCTGTATGGGGAGGTGGGGGACACCCTGCTGATCATCTTCAAGAACCAGGCCAGCAGG
CCCTACAACATCTACCCCCATGGCATCACTGATGTGAGGCCCCTGTACAGCAGGAGGCTGCCCAAGGGGG
TGAAGCACCTGAAGGACTTCCCCATCCTGCCTGGGGAGATCTTCAAGTACAAGTGGACTGTGACTGTGGA
GGATGGCCCCACCAAGTCTGACCCCAGGTGCCTGACCAGATACTACAGCAGCTTTGTGAACATGGAGAGG
GACCTGGCCTCTGGCCTGATTGGCCCCCTGCTGATCTGCTACAAGGAGTCTGTGGACCAGAGGGGCAACC
AGATCATGTCTGACAAGAGGAATGTGATCCTGTTCTCTGTGTTTGATGAGAACAGGAGCTGGTACCTGAC
TGAGAACATCCAGAGGTTCCTGCCCAACCCTGCTGGGGTGCAGCTGGAGGACCCTGAGTTCCAGGCCAGC
AACATCATGCACAGCATCAATGGCTATGTGTTTGACAGCCTGCAGCTGTCTGTGTGCCTGCATGAGGTGG
CCTACTGGTACATCCTGAGCATTGGGGCCCAGACTGACTTCCTGTCTGTGTTCTTCTCTGGCTACACCTT
CAAGCACAAGATGGTGTATGAGGACACCCTGACCCTGTTCCCCTTCTCTGGGGAGACTGTGTTCATGAGC
ATGGAGAACCCTGGCCTGTGGATTCTGGGCTGCCACAACTCTGACTTCAGGAACAGGGGCATGACTGCCC

TGCTGAAAGTCTCCAGCTGTGACAAGAACACTGGGGACTACTATGAGGACAGCTATGAGGACATCTCTGC
CTACCTGCTGAGCAAGAACAATGCCATTGAGCCCAGGAGCTTCAGCCAGAATGCCACTAATGTGTCTAAC
AACAGCAACACCAGCAATGACAGCAATGTGTCTCCCCCAGTGCTGAAGAGGCACCAGAGGGAGATCACCA
GGACCACCCTGCAGTCTGACCAGGAGGAGATTGACTATGATGACACCATCTCTGTGGAGATGAAGAAGGA
GGACTTTGACATCTACGACGAGGACGAGAACCAGAGCCCCAGGAGCTTCCAGAAGAAGACCAGGCACTAC
TTCATTGCTGCTGTGGAGAGGCTGTGGGACTATGGCATGAGCAGCAGCCCCCATGTGCTGAGGAACAGGG
CCCAGTCTGGCTCTGTGCCCCAGTTCAAGAAGGTGGTGTTCCAGGAGTTCACTGATGGCAGCTTCACCCA
GCCCCTGTACAGAGGGGAGCTGAATGAGCACCTGGGCCTGCTGGGCCCCTACATCAGGGCTGAGGTGGAG
GACAACATCATGGTGACCTTCAGGAACCAGGCCAGCAGGCCCTACAGCTTCTACAGCAGCCTGATCAGCT
ATGAGGAGGACCAGAGGCAGGGGGCTGAGCCCAGGAAGAACTTTGTGAAGCCCAATGAAACCAAGACCTA
CTTCTGGAAGGTGCAGCACCACATGGCCCCCACCAAGGATGAGTTTGACTGCAAGGCCTGGGCCTACTTC
TCTGATGTGGACCTGGAGAAGGATGTGCACTCTGGCCTGATTGGCCCCCTGCTGGTGTGCCACACCAACA
CCCTGAACCCTGCCCATGGCAGGCAGGTGACTGTGCAGGAGTTTGCCCTGTTCTTCACCATCTTTGATGA
AACCAAGAGCTGGTACTTCACTGAGAACATGGAGAGGAACTGCAGGGCCCCCTGCAACATCCAGATGGAG
GACCCCACCTTCAAGGAGAACTACAGGTTCCATGCCATCAATGGCTACATCATGGACACCCTGCCTGGCC
TGGTGATGGCCCAGGACCAGAGGATCAGGTCGTACCTGCTGAGCATGGGCAGCAATGAGAACATCCACAG
CATCCACTTCTCTGGCCATGTGTTCACTGTGAGGAAGAAGGAGGAGTACAAGATGGCCCTGTACAACCTG
TACCCTGGGGTGTTTGAGACTGTGGAGATGCTGCCCAGCAAGGCTGGCATCTGGAGGGTGGAGTGCCTGA
TTGGGGAGCACCTGCATGCTGGCATGAGCACCCTGTTCCTGGTGTACAGCAACAAGTGCCAGACCCCCCT
GGGCATGGCCTCTGGCCACATCAGGGACTTCCAGATCACTGCCTCTGGCCAGTATGGCCAGTGGGCCCCC
AAGCTGGCCAGGCTGCACTACTCTGGCAGCATCAATGCCTGGAGCACCAAGGAGCCCTTCAGCTGGATCA
AGGTGGACCTGCTGGCCCCCATGATCATCCATGGCATCAAGACCCAGGGGGCCAGGCAGAAGTTCAGCAG
CCTGTACATCAGCCAGTTCATCATCATGTACAGCCTGGATGGCAAGAAGTGGCAGACCTACAGGGGCAAC
AGCACTGGCACCCTGATGGTGTTCTTTGGCAATGTGGACAGCTCTGGCATCAAGCACAACATCTTCAACC
CCCCCATCATTGCCAGATACATCAGGCTGCACCCCACCCACTACAGCATCAGGAGCACCCTGAGGATGGA
GCTGATGGGCTGTGACCTGAACAGCTGCAGCATGCCCCTGGGCATGGAGAGCAAGGCCATCTCTGATGCC
CAGATCACTGCCAGCAGCTACTTCACCAACATGTTTGCCACCTGGAGCCCCAGCAAGGCCAGGCTGCACC
TGCAGGGCAGGAGCAATGCCTGGAGGCCCCAGGTCAACAACCCCAAGGAGTGGCTGCAGGTGGACTTCCA
GAAGACCATGAAGGTGACTGGGGTGACCACCCAGGGGGTGAAGAGCCTGCTGACCAGCATGTATGTGAAG
GAGTTCCTGATCAGCAGCAGCCAGGATGGCCACCAGTGGACCCTGTTCTTCCAGAATGGCAAGGTGAAGG
TGTTCCAGGGCAACCAGGACAGCTTCACCCCTGTGGTGAACAGCCTGGACCCCCCCCTGCTGACCAGATA
CCTGAGGATTCACCCCCAGAGCTGGGTGCACCAGATTGCCCTGAGGATGGAGGTGCTGGGCTGTGAGGCC
CAGGACCTGTACTGA

SEQ ID NO: 31

TACGTCTAACTCGACTCGTGGACGAAGAAGCACACGGACGACTCCAAGACGAAGAGACGGTGGTCCTCTA
TGATGGACCCCCGACACCTCGACTCGACCCTGATGTACGTCAGACTGGACCCCCTCGACGGACACCTACG
GTCCAAGGGGGGGTCTCACGGGTTCTCGAAGGGGAAGTTGTGGAGACACCACATGTTCTTCTGGGACAAA
CACCTCAAGTGACTGGTGGACAAGTTGTAACGGTTCGGGTCCGGGGGGACCTACCCGGACGACCCGGGGT
GGTAGGTCCGACTCCACATACTGTGACACCACTAGTGGGACTTCTTGTACCGGTCGGTGGGACACTCGGA
CGTACGACACCCCCACTCGATGACCTTCCGGAGACTCCCCCGACTCATACTACTGGTCTGGTCGGTCTCC
CTCTTCCTCCTACTGTTCCACAAGGGACCCCCGTCGGTGTGGATACACACCGTCCACGACTTCCTCTTAC
CGGGGTACCGGAGACTGGGGGACACGGACTGGATGTCGATGGACTCGGTACACCTGGACCACTTCCTGGA
CTTGAGACCGGACTAACCCCGGGACGACCACACGTCCCTCCCGTCGGACCGGTTCCTCTTCTGGGTCTGG
GACGTGTTCAAGTAGGACGACAAACGACACAAACTACTCCCGTTCTCGACCGTGAGACTTTGGTTCTTGT
CGGACTACGTCCTGTCCCTACGACGGAGACGGTCCCGGACCGGGTTCTACGTGTGACACTTACCGATACA
CTTGTCCTCGGACGGACCGGACTAACCGACGGTGTCCTTCAGACACATGACCGTACACTAACCGTACCCG
TGGTGGGGACTCCACGTGTCGTAGAAGGACCTCCCGGTGTGGAAGGACCAGTCCTTGGTGTCCGTCCGGT
CGGACCTCTAGTCGGGGTAGTGGAAGGACTGACGGGTCTGGGACGACTACCTGGACCCGGTCAAGGACGA
CAAGACGGTGTAGTCGTCGGTGGTCGTACTACCGTACCTCCGGATACACTTCCACCTGTCGACGGGACTC
CTCGGGGTCGACTCCTACTTCTTGTTACTCCTCCGACTCCTGATACTACTACTGGACTGACTGAGACTCT
ACCTACACCACTCCAAACTACTACTGTTGTCGGGGTCGAAGTAGGTCTAGTCCAGACACCGGTTCTTCGT
GGGGTTCTGGACCCACGTGATGTAACGACGACTCCTCCTCCTGACCCTGATACGGGGGGACCACGACCGG
GGACTACTGTCCTCGATGTTCTCGGTCATGGACTTGTTACCGGGGGTCTCCTAACCGTCCTTCATGTTCT
TCCAGTCCAAGTACCGGATGTGACTACTTTGGAAGTTCTGGTCCCTCCGGTAGGTCGTACTCAGACCGTA
GGACCCGGGGGACGACATACCCCTCCACCCCCTGTGGGACGACTAGTAGAAGTTCTTGGTCCGGTCGTCC
GGGATGTTGTAGATGGGGGTACCGTAGTGACTACACTCCGGGGACATGTCGTCCTCCGACGGGTTCCCCC

ACTTCGTGGACTTCCTGAAGGGGTAGGACGGACCCCTCTAGAAGTTCATGTTCACCTGACACTGACACCT
CCTACCGGGGTGGTTCAGACTGGGGTCCACGGACTGGTCTATGATGTCGTCGAAACACTTGTACCTCTCC
CTGGACCGGAGACCGGACTAACCGGGGGACGACTAGACGATGTTCCTCAGACACCTGGTCTCCCCGTTGG
TCTAGTACAGACTGTTCTCCTTACACTAGGACAAGAGACACAAACTACTCTTGTCCTCGACCATGGACTG
ACTCTTGTAGGTCTCCAAGGACGGGTTGGGACGACCCCACGTCGACCTCCTGGGACTCAAGGTCCGGTCG
TTGTAGTACGTGTCGTAGTTACCGATACACAAACTGTCGGACGTCGACAGACACACGGACGTACTCCACC
GGATGACCATGTAGGACTCGTAACCCCGGGTCTGACTGAAGGACAGACACAAGAAGAGACCGATGTGGAA
GTTCGTGTTCTACCACATACTCCTGTGGGACTGGGACAAGGGGAAGAGACCCCTCTGACACAAGTACTCG
TACCTCTTGGGACCGGACACCTAAGACCCGACGGTGTTGAGACTGAAGTCCTTGTCCCCGTACTGACGGG
ACGACTTTCAGAGGTCGACACTGTTCTTGTGACCCCTGATGATACTCCTGTCGATACTCCTGTAGAGACG
GATGGACGACTCGTTCTTGTTACGGTAACTCGGGTCCTCGAAGTCGGTCTTACGGTGATTACACAGATTG
TTGTCGTTGTGGTCGTTACTGTCGTTACACAGAGGGGGTCACGACTTCTCCGTGGTCTCCCTCTAGTGGT
CCTGGTGGGACGTCAGACTGGTCCTCCTCTAACTGATACTACTGTGGTAGAGACACCTCTACTTCTTCCT
CCTGAAACTGTAGATGCTGCTCCTGCTCTTGGTCTCGGGGTCCTCGAAGGTCTTCTTCTGGTCCGTGATG
AAGTAACGACGACACCTCTCCGACACCCTGATACCGTACTCGTCGTCGGGGGTACACGACTCCTTGTCCC
GGGTCAGACCGAGACACGGGGTCAAGTTCTTCCACCACAAGGTCCTCAAGTGACTACCGTCGAAGTGGGT
CGGGGACATGTCTCCCCTCGACTTACTCGTGGACCCGGACGACCCGGGGATGTAGTCCCGACTCCACCTC
CTGTTGTAGTACCACTGGAAGTCCTTGGTCCGGTCGTCCGGGATGTCGAAGATGTCGTCGGACTAGTCGA
TACTCCTCCTGGTCTCCGTCCCCCGACTCGGGTCCTTCTTGAAACACTTCGGGTTACTTTGGTTCTGGAT
GAAGACCTTCCACGTCGTGGTGTACCGGGGGTGGTTCCTACTCAAACTGACGTTCCGGACCCGGATGAAG
AGACTACACCTGGACCTCTTCCTACACGTGAGACCGGACTAACCGGGGGACGACCACACGGTGTGGTTGT
GGGACTTGGGACGGGTACCGTCCGTCCACTGACACGTCCTCAAACGGGACAAGAAGTGGTAGAAACTACT
TTGGTTCTCGACCATGAAGTGACTCTTGTACCTCTCCTTGACGTCCCGGGGGACGTTGTAGGTCTACCTC
CTGGGGTGGAAGTTCCTCTTGATGTCCAAGGTACGGTAGTTACCGATGTAGTACCTGTGGGACGGACCGG
ACCACTACCGGGTCCTGGTCTCCTAGTCCACCATGGACGACTCGTACCCGTCGTTACTCTTGTAGGTGTC
GTAGGTGAAGAGACCGGTACACAAGTGACACTCCTTCTTCCTCCTCATGTTCTACCGGGACATGTTGGAC
ATGGGACCCCACAAACTCTGACACCTCTACGACGGGTCGTTCCGACCGTAGACCTCCCACCTCACGGACT
AACCCCTCGTGGACGTACGACCGTACTCGTGGGACAAGGACCACATGTCGTTGTTCACGGTCTGGGGGGA
CCCGTACCGGAGACCGGTGTAGTCCCTGAAGGTCTAGTGACGGAGACCGGTCATACCGGTCACCCGGGGG
TTCGACCGGTCCGACGTGATGAGACCGTCGTAGTTACGGACCTCGTGGTTCCTCGGGAAGTCGACCTAGT
TCCACCTGGACGACCGGGGGTACTAGTAGGTACCGTAGTTCTGGGTCCCCCGGTCCGTCTTCAAGTCGTC
GGACATGTAGTCGGTCAAGTAGTAGTACATGTCGGACCTACCGTTCTTCACCGTCTGGATGTCCCCGTTG
TCGTGACCGTGGGACTACCACAAGAAACCGTTACACCTGTCGAGACCGTAGTTCGTGTTGTAGAAGTTGG
GGGGGTAGTAACGGTCTATGTAGTCCGACGTGGGGTGGGTGATGTCGTAGTCCTCGTGGGACTCCTACCT
CGACTACCCGACACTGGACTTGTCGACGTCGTACGGGGACCCGTACCTCTCGTTCCGGTAGAGACTACGG
GTCTAGTGACGGTCGTCGATGAAGTGGTTGTACAAACGGTGGACCTCGGGGTCGTTCCGGTCCGACGTGG
ACGTCCCGTCCTCGTTACGGACCTCCGGGGTCCAGTTGTTGGGGTTCCTCACCGACGTCCACCTGAAGGT
CTTCTGGTACTTCCACTGACCCCACTGGTGGGTCCCCCACTTCTCGGACGACTGGTCGTACATACACTTC
CTCAAGGACTAGTCGTCGTCGGTCCTACCGGTGGTCACCTGGGACAAGAAGGTCTTACCGTTCCACTTCC
ACAAGGTCCCGTTGGTCCTGTCGAAGTGGGGACACCACTTGTCGGACCTGGGGGGGGACGACTGGTCTAT
GGACTCCTAAGTGGGGGTCTCGACCCACGTGGTCTAACGGGACTCCTACCTCCACGACCCGACACTCCGG
GTCCTGGACATGACT

SEQ ID NO: 32

MQIELSTCFFLCLLRFCFSATRRYYLGAVELSWDYMQSDLGELPVDARFPPRVPKSFPFNTSVVYKKTLF
VEFTDHLFNIAKPRPPWMGLLGPTIQAEVYDTVVITLKNMASHPVSLHAVGVSYWKASEGAEYDDQTSQR
EKEDDKVFPGGSHTYVWQVLKENGPMASDPLCLTYSYLSHVDLVKDLNSGLIGALLVCREGSLAKEKTQT
LHKFILLFAVFDEGKSWHSETKNSLMQDRDAASARAWPKMHTVNGYVNRSLPGLIGCHRKSVYWHVIGMG
TTPEVHSIFLEGHTFLVRNHRQASLEISPITFLTAQTLLMDLGQFLLFCHISSHQHDGMEAYVKVDSCPE
EPQLRMKNNEEAEDYDDDLTDSEMDVVRFDDDNSPSFIQIRSVAKKHPKTWVHYIAAEEEDWDYAPLVLA
PDDRSYKSQYLNNGPQRIGRKYKKVRFMAYTDETFKTREAIQHESGILGPLLYGEVGDTLLIIFKNQASR
PYNIYPHGITDVRPLYSRRLPKGVKHLKDFPILPGEIFKYKWTVTVEDGPTKSDPRCLTRYYSSFVNMER
DLASGLIGPLLICYKESVDQRGNQIMSDKRNVILFSVFDENRSWYLTENIQRFLPNPAGVQLEDPEFQAS
NIMHSINGYVFDSLQLSVCLHEVAYWYILSIGAQTDFLSVFFSGYTFKHKMVYEDTLTLFPFSGETVFMS
MENPGLWILGCHNSDFRNRGMTALLKVSSCDKNTGDYYEDSYEDISAYLLSKNNAIEPRSFSQNATNVSN
NSNTSNDSNVSPPVLKRHQREITRTTLQSDQEEIDYDDTISVEMKKEDFDIYDEDENQSPRSFQKKTRHY
FIAAVERLWDYGMSSSPHVLRNRAQSGSVPQFKKVVFQEFTDGSFTQPLYRGELNEHLGLLGPYIRAEVE

DNIMVTFRNQASRPYSFYSSLISYEEDQRQGAEPRKNFVKPNETKTYFWKVQHHMAPTKDEFDCKAWAYF
SDVDLEKDVHSGLIGPLLVCHTNTLNPAHGRQVTVQEFALFFTIFDETKSWYFTENMERNCRAPCNIQME
DPTFKENYRFHAINGYIMDTLPGLVMAQDQRIRWYLLSMGSNENIHSIHFSGHVFTVRKKEEYKMALYNL
YPGVFETVEMLPSKAGIWRVECLIGEHLHAGMSTLFLVYSNKCQTPLGMASGHIRDFQITASGQYGQWAP
KLARLHYSGSINAWSTKEPFSWIKVDLLAPMIIHGIKTQGARQKFSSLYISQFIIMYSLDGKKWQTYRGN
STGTLMVFFGNVDSSGIKHNIFNPPIIARYIRLHPTHYSIRSTLRMELMGCDLNSCSMPLGMESKAISDA
QITASSYFTNMFATWSPSKARLHLQGRSNAWRPQVNNPKEWLQVDFQKTMKVTGVTTQGVKSLLTSMYVK
EFLISSSQDGHQWTLFFQNGKVKVFQGNQDSFTPVVNSLDPPLLTRYLRIHPQSWVHQIALRMEVLGCEA
QDLY

SEQ ID NO: 33

GGCGCCTCTAGAGTTATAACCGGTAATCGGTATAATAAGTAACCAATATATCGTATTTAGTTATAACCGA
TAACCGGTAACGTATGCAACATAGATATAGTATTATACATGTAAATATAACCGAGTACAGGTTATACTGG
CGGTACAACCGTAACTAATAACTGATCAATAATTATCATTAGTTAATGCCCCAGTAATCAAGTATCGGGT
ATATACCTCAAGGCGCAATGTATTGAATGCCATTTACCGGGCGGACCGACTGGCGGGTTGCTGGGGGCGG
GTAACTGCAGTTATTACTGCATACAAGGGTATCATTGCGGTTATCCCTGAAAGGTAACTGCAGTTACCCA
CCTCATAAATGCCATTTGACGGGTGAACCGTCATGTAGTTCACATAGTATACGGTTCAGGCGGGGGATAA
CTGCAGTTACTGCCATTTACCGGGCGGACCGTAATACGGGTCATGTACTGGAATGCCCTGAAAGGATGAA
CCGTCATGTAGATGCATAATCAGTAGCGATAATGGTACCACTACGCCAAAACCGTCATGTGGTTACCCGC
ACCTATCGCCAAACTGAGTGCCCCTAAAGGTTCAGAGGTGGGGTAACTGCAGTTACCCTCAAACAAAACC
GTGGTTTTAGTTGCCCTGAAAGGTTTTACAGCATTATTGGGGCGGGGCAACTGCGTTTACCCGCCATCCG
CACATGCCACCCTCCAGATATATTCGTCTCGAGCAAATCACTTGGCAGTCTAGTGATCTTCGAAATAACG
CCATCAAATAGTGTCAATTTAACGATTGCGTCAGTCACGAAGACTGTGTTGTCAGAGCTTGAATTCGACG
TCTTCAACCAGCACTCCGTGACCCGTCCGATCG

CLAIMS

1. A lentiviral vector pseudotyped with hemagglutinin-neuraminidase (HN) and fusion (F) proteins from a respiratory paramyxovirus, wherein said lentiviral vector comprises a hybrid human CMV enhancer/EF1a (hCEF) promoter and a transgene, and wherein the lentiviral vector lacks an(y) intron positioned between said promoter and said transgene.

2. The lentiviral vector according to Claim 1, wherein the lentiviral vector is selected from the group consisting of a Human immunodeficiency virus (HIV) vector, a Simian immunodeficiency virus (SIV) vector, a Feline immunodeficiency virus (FIV) vector, an Equine infectious anaemia virus (EIAV) vector, and a Visna/maedi virus vector.

3. The lentiviral vector according to Claim 1 or 2, wherein lentiviral vector is a SIV vector.

4. The lentiviral vector according to any one of Claims 1 to 3, wherein the respiratory paramyxovirus is a Sendai virus.

5. The lentiviral vector according to any one of Claims 1 to 4, wherein the transgene is selected from:

(a) a transgene which encodes a secreted therapeutic protein, optionally Alpha-1 Antitrypsin (*A1AT*), Factor VIII, Surfactant Protein B (*SFTPB*), Factor VII, Factor IX, Factor X, Factor XI, von Willebrand Factor, Granulocyte-Macrophage Colony-Stimulating Factor (*GM-CSF*) or a monoclonal antibody against an infectious agent; or

(b) *CFTR, DNAH5, DNAH11, DNAI1, or DNAI2*.

6. The lentiviral vector according to any one of Claims 1 to 5, wherein:

(a) the transgene encodes CFTR;

(b) the transgene encodes A1AT; or

(c) the transgene encodes FVIII.

Date Reçue/Date Received 2021-06-23

7. A method of producing lentivirus as defined in any one of Claims 1 to 6, said method comprising the following steps:

(a) growing cells in suspension;

(b) transfecting the cells with one or more plasmids;

(c) adding a nuclease;

(d) harvesting the lentivirus;

(e) adding trypsin; and

(f) purification.

8. The method according to Claim 7, wherein steps (a) to (f) are carried out sequentially.

9. The method according to Claim 7 or Claim 8, wherein the cells are HEK293T or 293T/17 cells.

10. The method according to any one of Claims 7 to 9, wherein the addition of the nuclease is at the pre-harvest stage.

11. The method according to any one of Claims 7 to 10, wherein the addition of trypsin is at the post-harvest stage.

12. The method according to any one of Claims 7 to 11, wherein the purification step comprises a chromatography step.

13. Use of the lentiviral vector according to any one of Claims 1 to 6 for the treatment of a disease.

14. Use of the lentiviral vector according to any one of Claims 1 to 6 in the manufacture of a medicament for the treatment of a disease.

15. The use according to Claim 13 or 14, wherein the disease is a lung disease selected from: cystic fibrosis (CF); Primary Ciliary Dyskinesia (PCD); Surfactant Protein B (SP-B) deficiency; Alpha 1-antitrypsin Deficiency (A1AD); Pulmonary Alveolar Proteinosis (PAP); and Chronic obstructive pulmonary disease (COPD).

Date Reçue/Date Received 2021-06-23

16. The use according to Claim 15, wherein the disease is CF.

17. The use according to Claim 13 or 14, wherein the disease is A1AT deficiency.

18. The use according to Claim 13 or 14, wherein the disease is Haemophilia.

19. The use according to Claim 18, wherein the transgene is Factor VIII.

20. A host cell comprising the vector according to any one of Claims 1 to 6.

21. A composition comprising a vector according to any one of Claims 1 to 6 and a pharmaceutically-acceptable carrier.

Date Reçue/Date Received 2021-06-23

Figure 1:

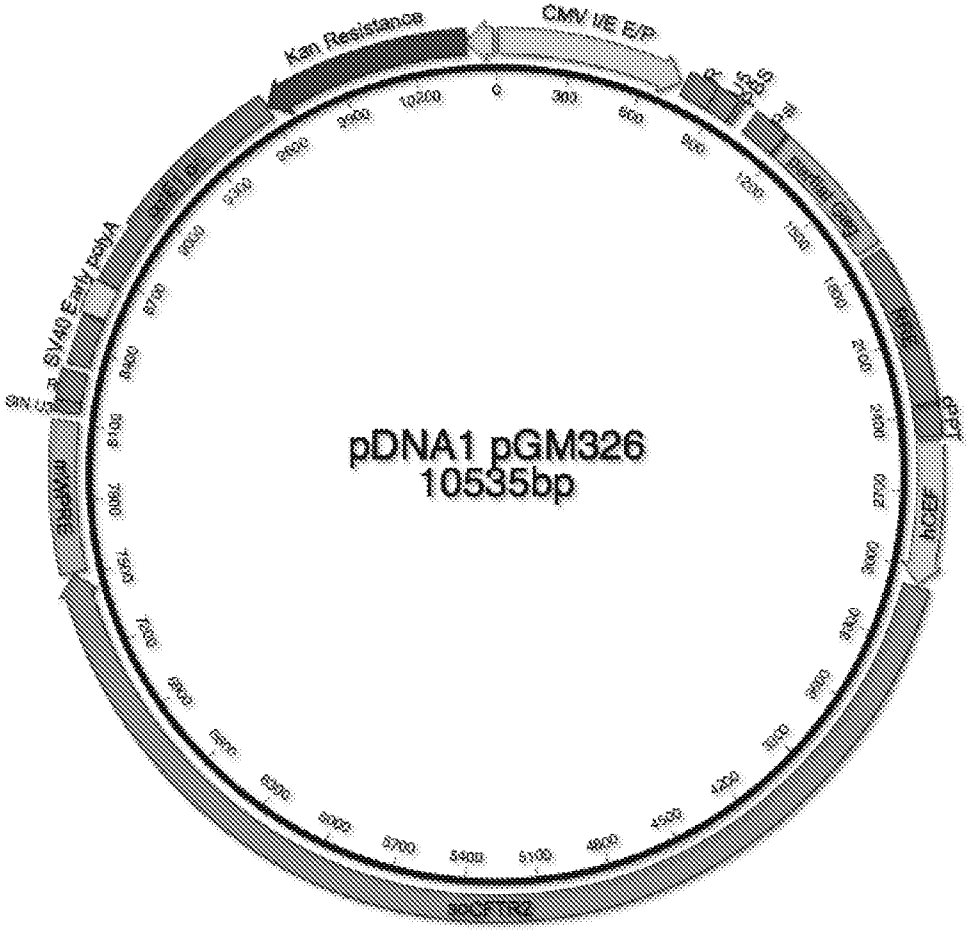


Figure 1A

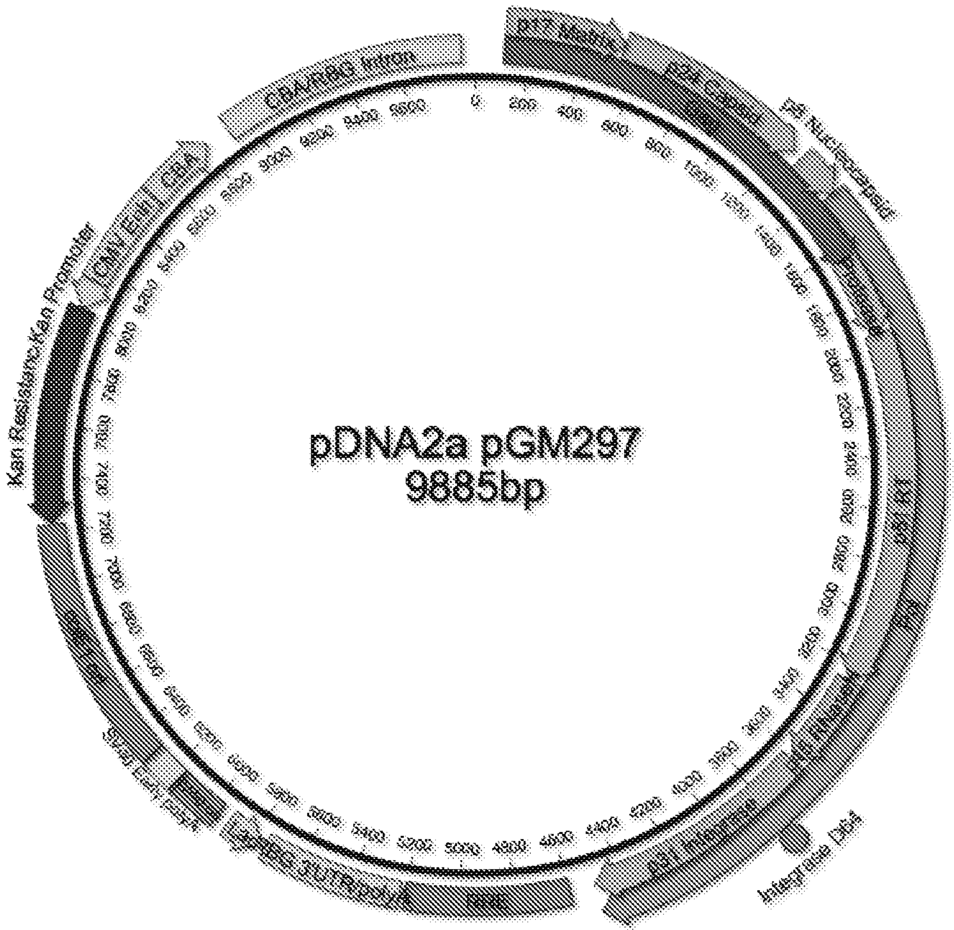

pDNA2a pGM297
9885bp
p17 Matrix
p24 Capsid
p6 Nucleocapsid
CBA/RBG intron
CBA
Kan Resistance Kan Promoter
p51 RT
p31 Integrase
Integrase D64
RBG 3'UTR/polyA

Figure 1B

Figure 1C

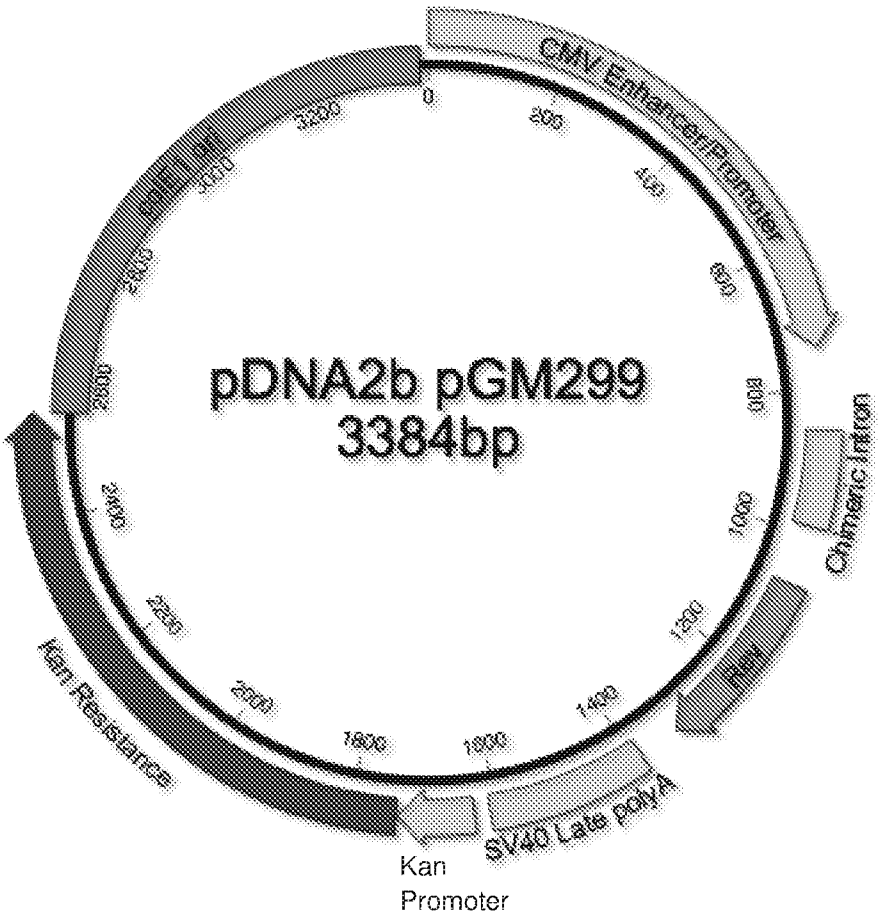
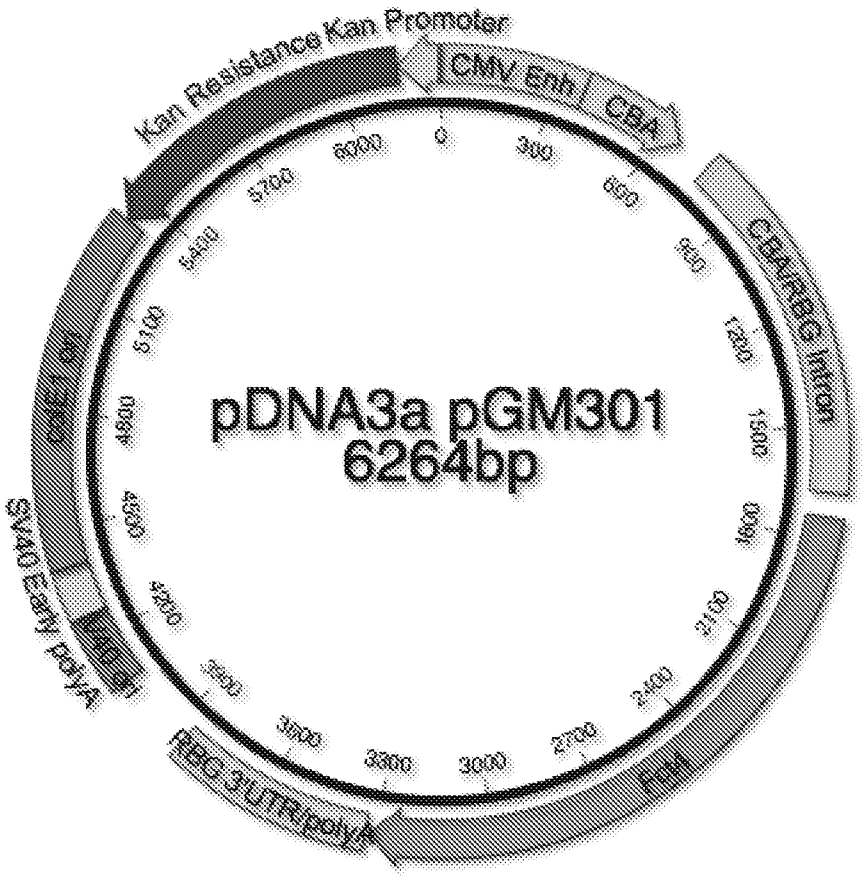

pDNA2b pGM299
3384bp
CMV Enhancer/IE promoter
Chimeric Intron
SV40 Late polyA
Kan
Promoter
Kan Resistance
0
200
400
600
800
1000
1200
1400
1600
1800
2000
2200
2400
2600
2800
3000
3200

Figure 1D


pDNA3a pGM301
6264bp
Kan Resistance
Kan Promoter
CMV Enh
CBA
CBA/RBG Intron
hGH polyA
pUC ori
Sv40 Early polyA
RBG 3'UTR polyA
0
300
650
900
1200
1500
1900
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800
5100
5400
5700
6000

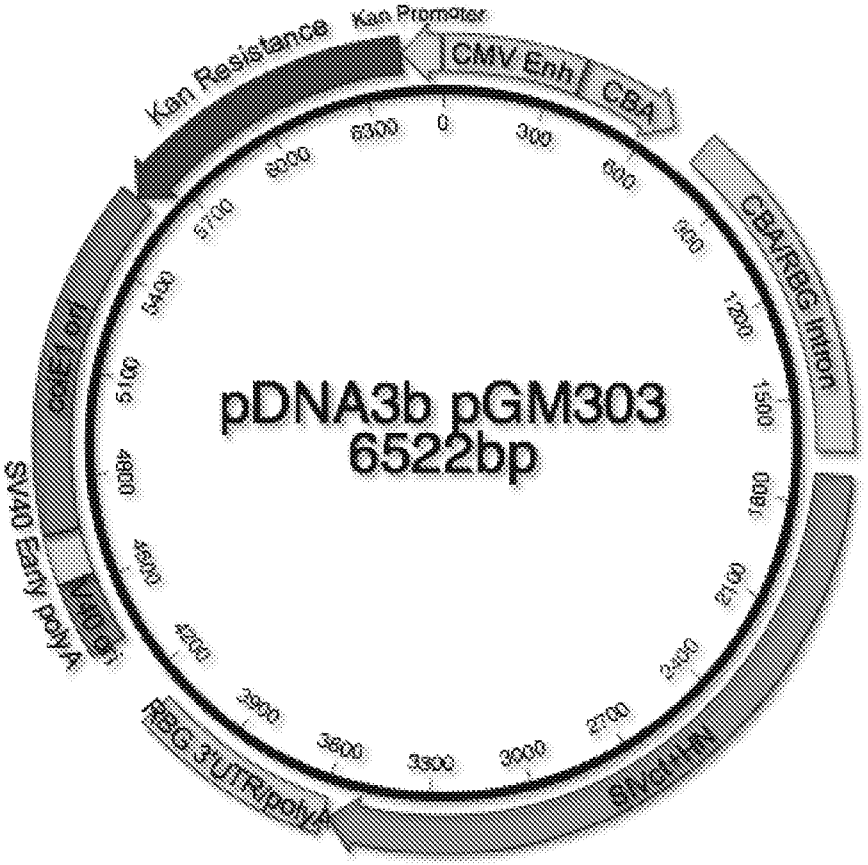

pDNA3b pGM303
6522bp
Kan Resistance
Kan Promoter
CMV Enh
CBA
CBA/CBE Intron
SV40 Early polyA
F1 Ori
RBG 3'UTR polyA
pUC Ori
0
300
600
900
1200
1500
1800
2100
2400
2700
3000
3300
3600
3900
4200
4500
4800
5100
5400
5700
6000
6300

Figure 1E

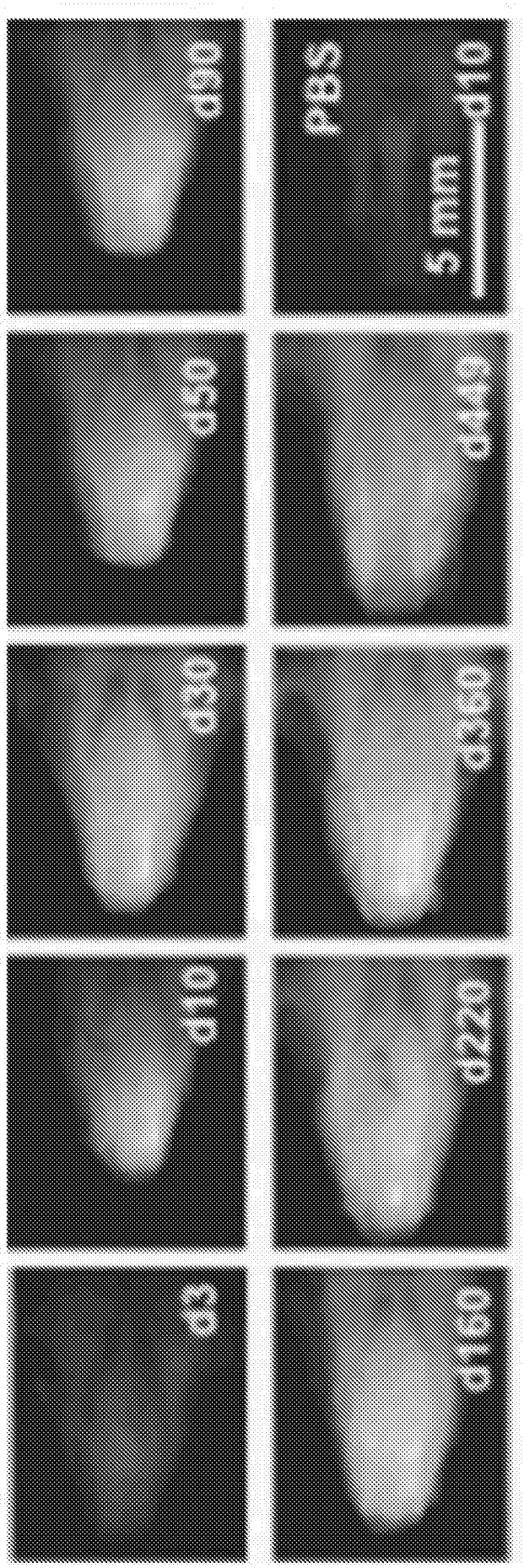

0dp
0hp
0d30p
0h30p
0d60p
0h60p
0d90p
0h90p
0d120p
0h120p
0d150p
0h150p
0d180p
0h180p
5 mm

Figure 2

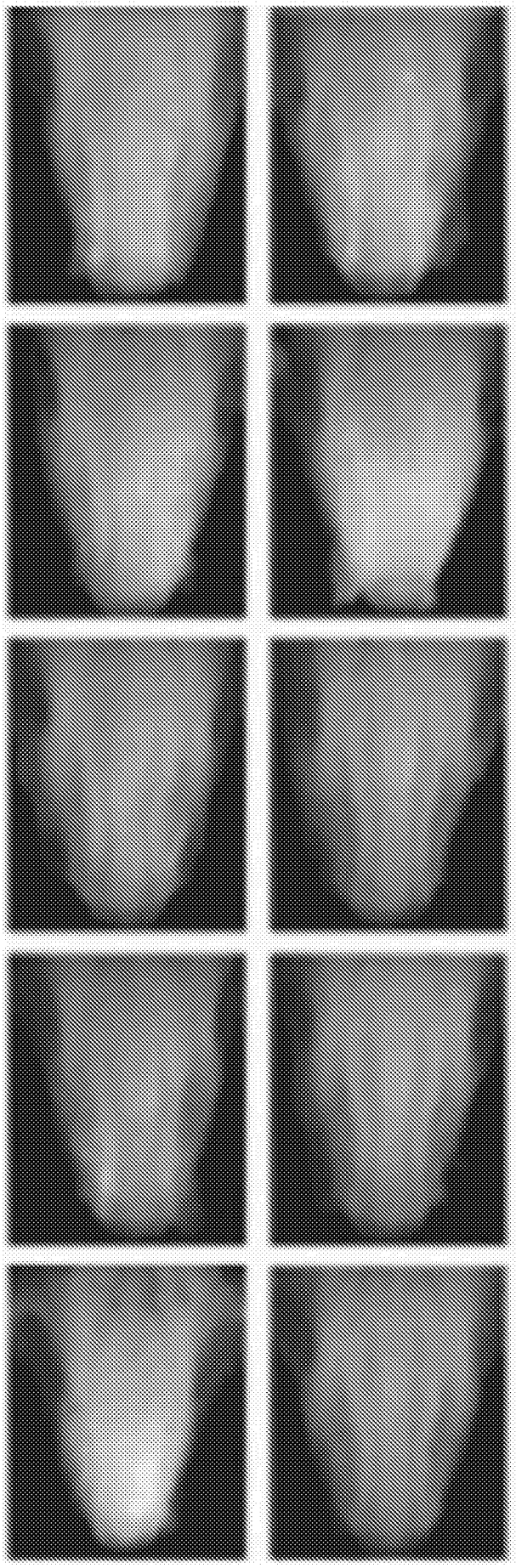


Figure 3

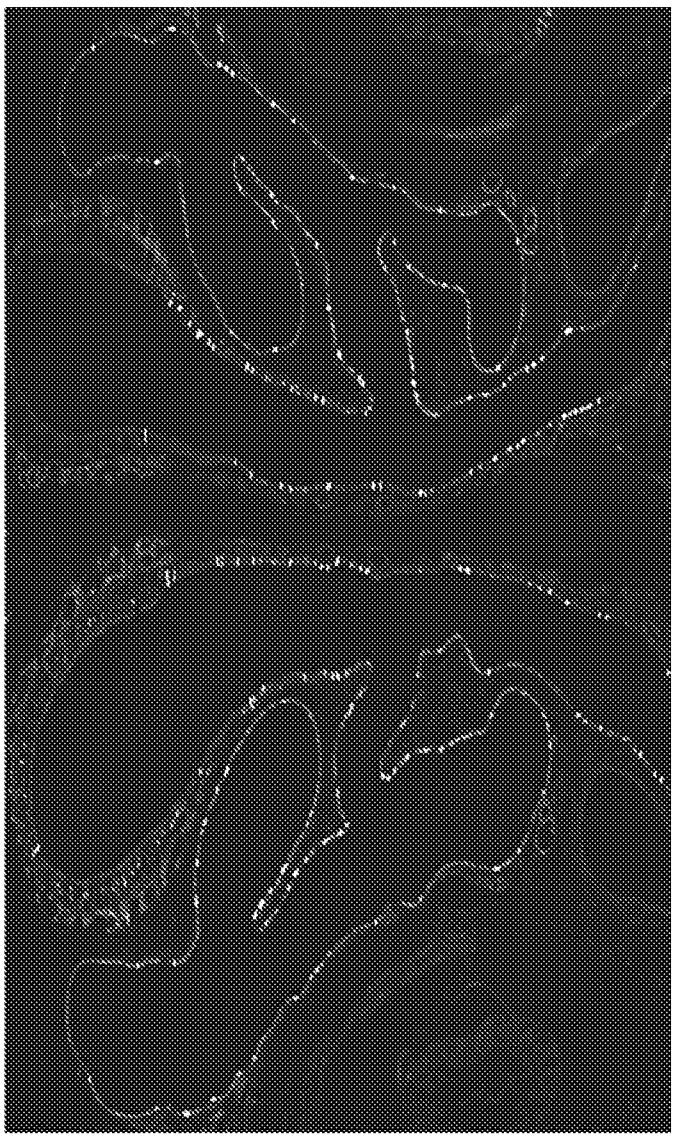


Figure 4

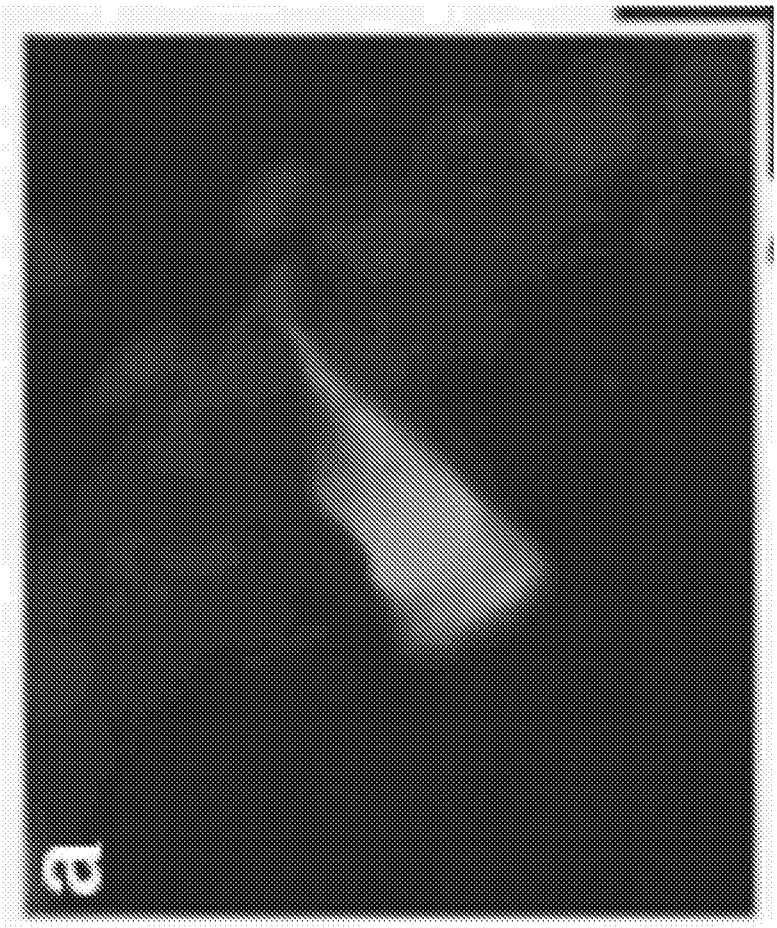


Figure 5

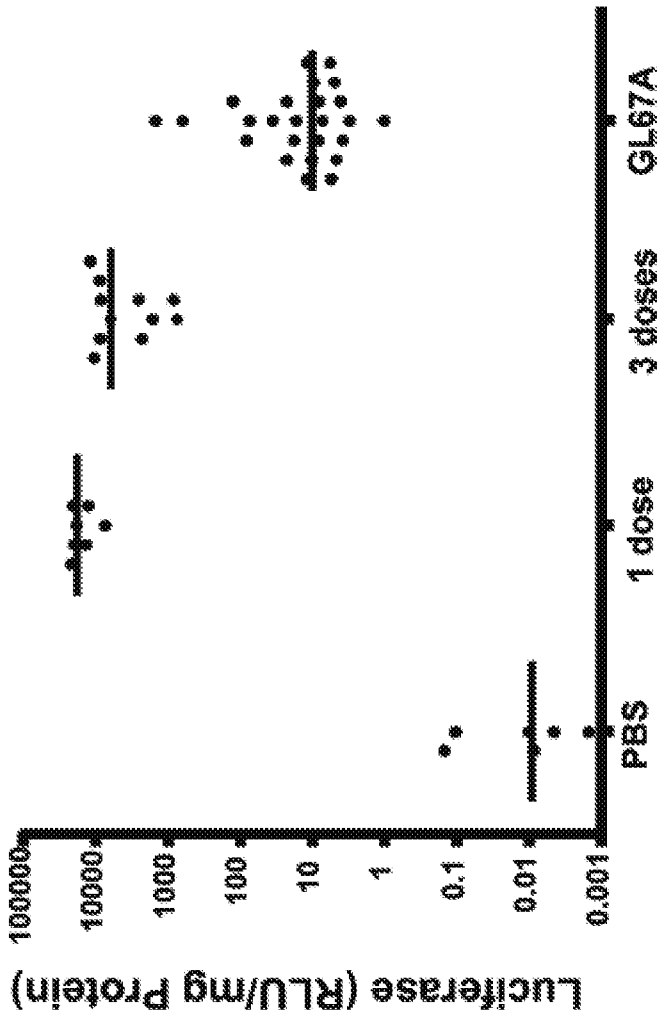

Luciferase (RLU/mg Protein)
100000
10000
1000
100
10
1
0.1
0.01
0.001
PBS
1 dose
3 doses
GL67A

Figure 6

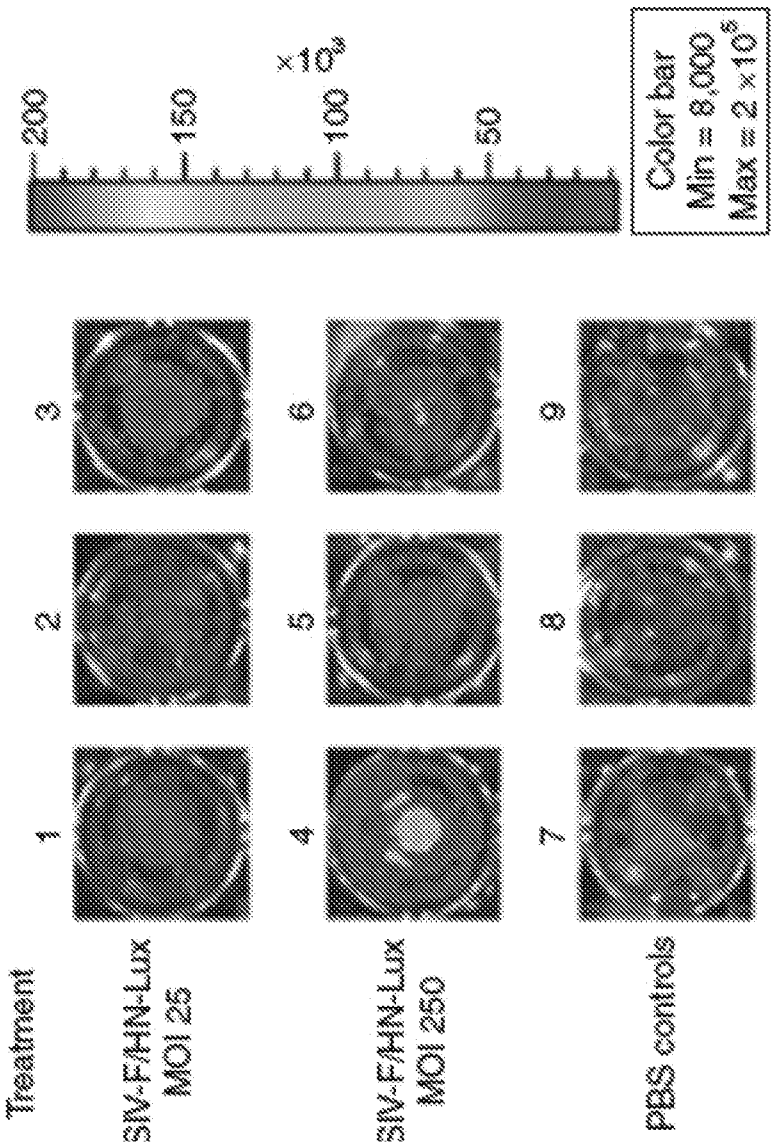

Color bar
Min = 8,000
Max = 2 x10^6
x10^3
200
150
100
50
Treatment
1
2
3
4
5
6
7
8
9
SIV-F/HN-Lux MOI 25
SIV-F/HN-Lux MOI 250
PBS controls

Figure 7

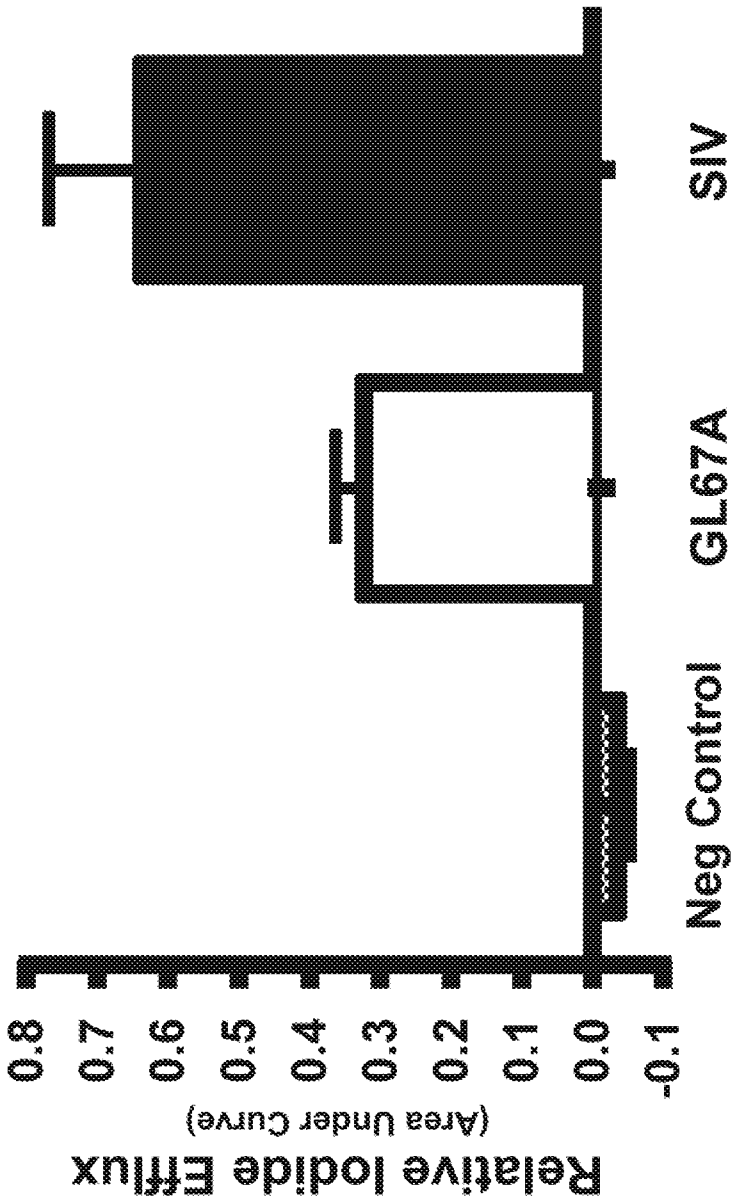

Relative Iodide Efflux
(Area Under Curve)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
Neg Control
GL67A
SIV

Figure 8

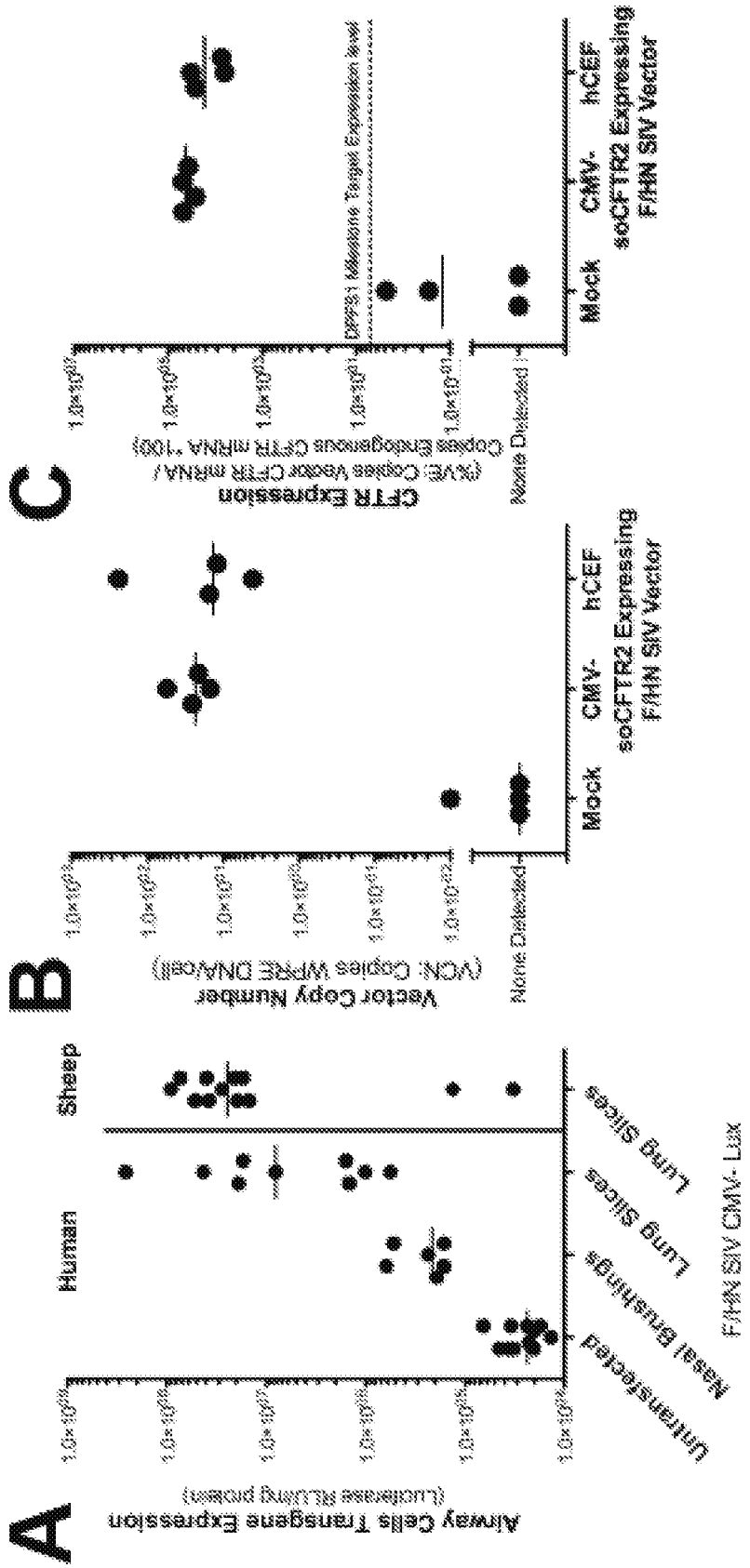

A
Airway Cells Transgene Expression
(Luciferase RLU/mg protein)
Human
Sheep
Untransfected
Nasal Brushings
Lung Slices
Lung Slices
F/HN SIV CMV- Lux
B
Vector Copy Number
(VCN: Copies WPRE DNA/cell)
None Detected
Mock
CMV-hCEF soCFTR2 Expressing F/HN SIV Vector
C
CFTR Expression
(RVE: Copies Vector CFTR mRNA/
Copies Endogenous CFTR mRNA *100)
None Detected
DF/3611 Mutation Target Expression level
Mock
CMV-hCEF soCFTR2 Expressing F/HN SIV Vector

Figure 9

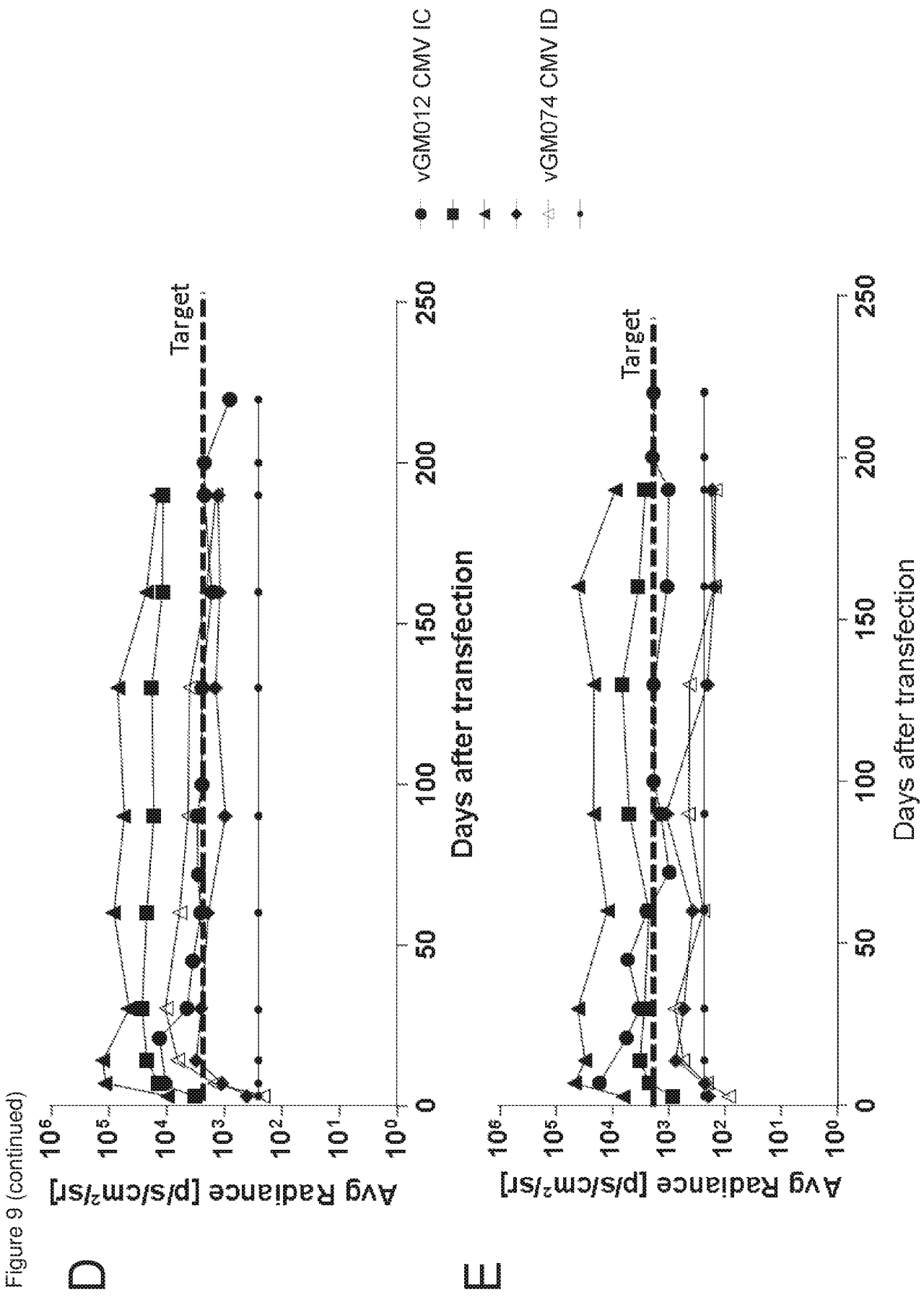


Figure 9 (continued)

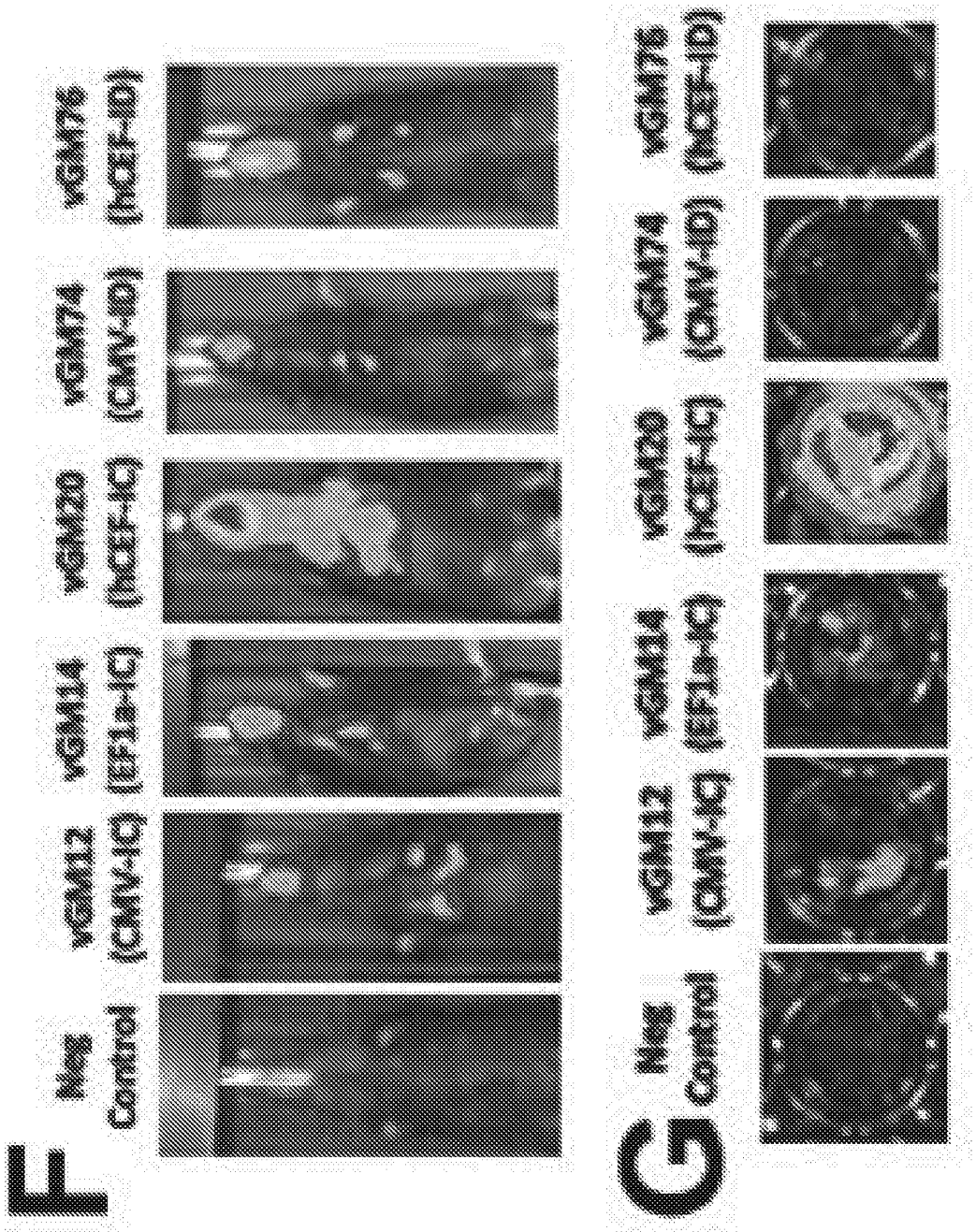


Figure 9 (continued)

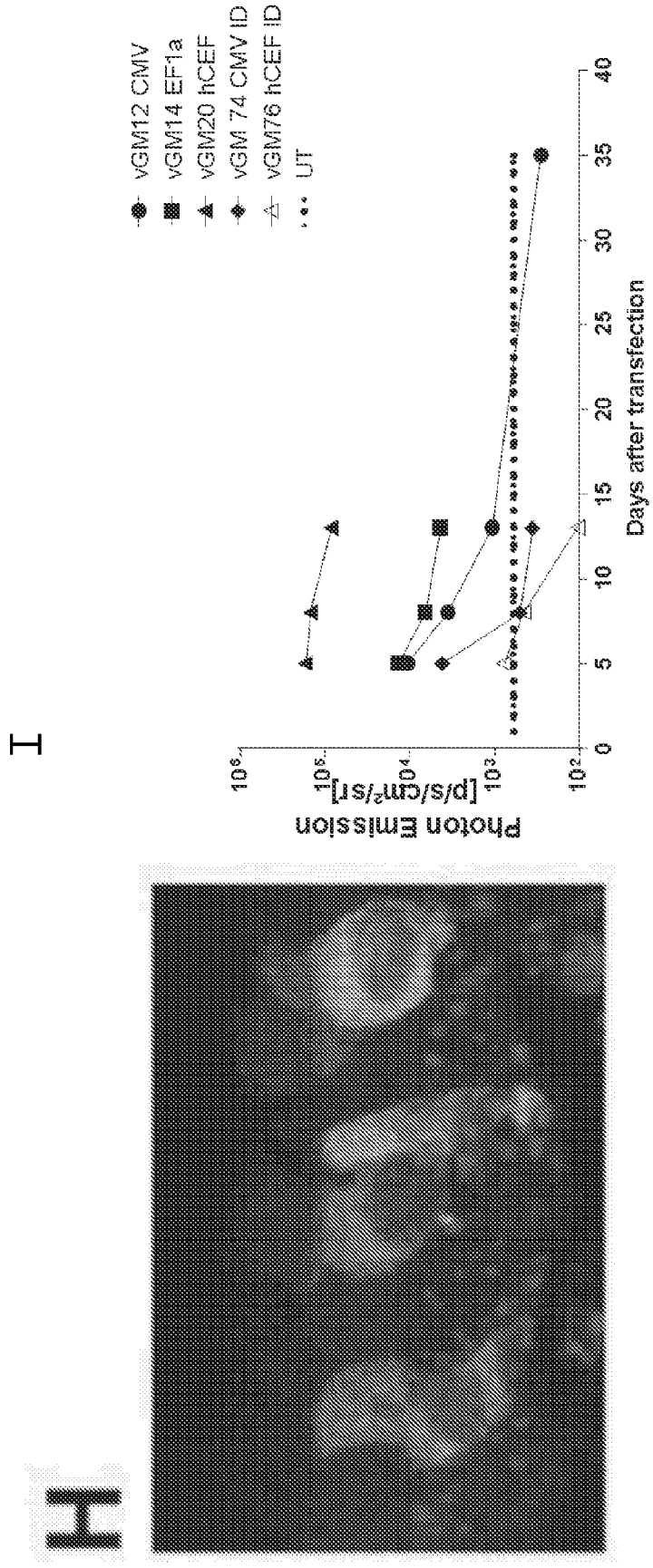


Figure 9 (continued)

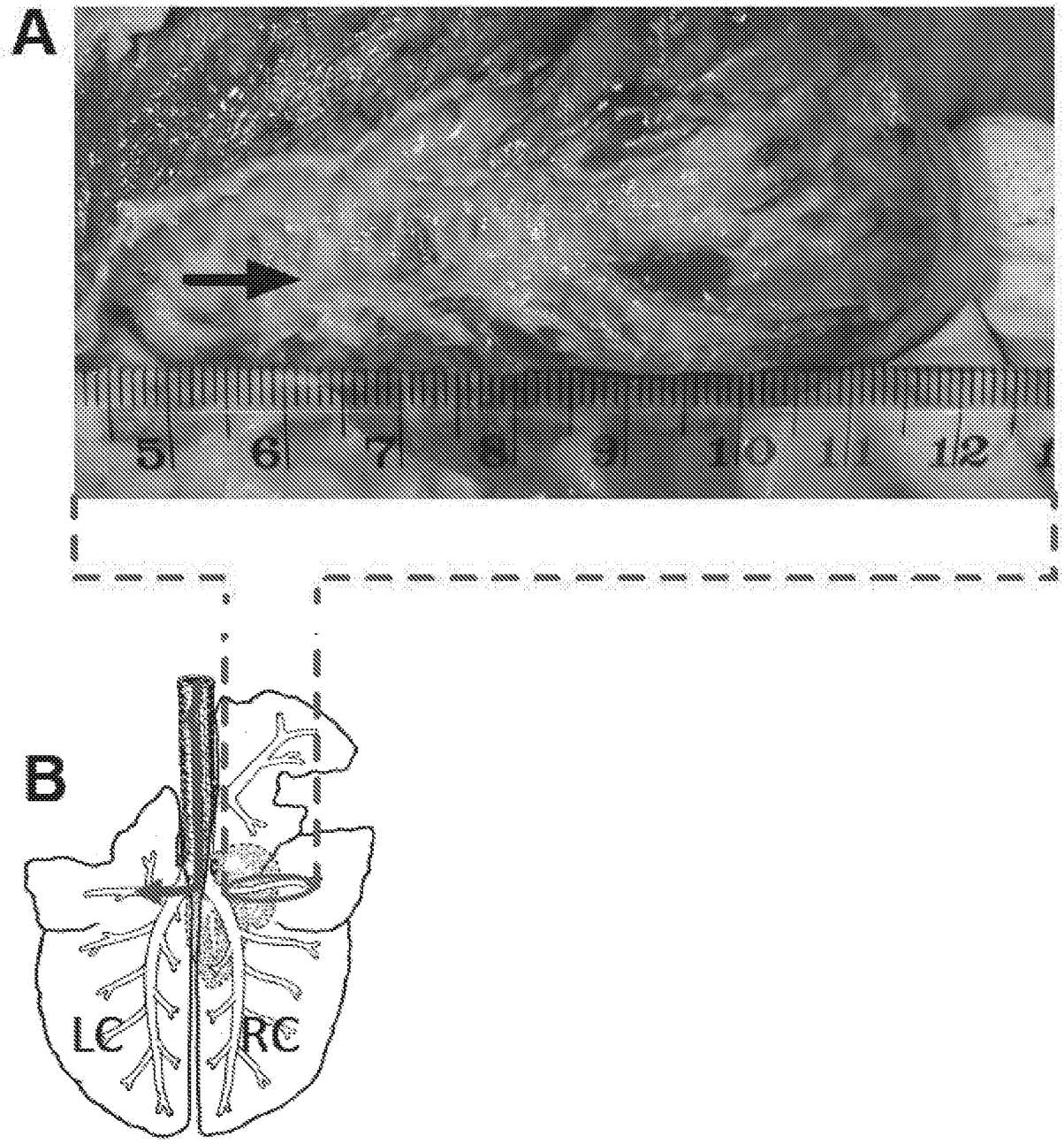
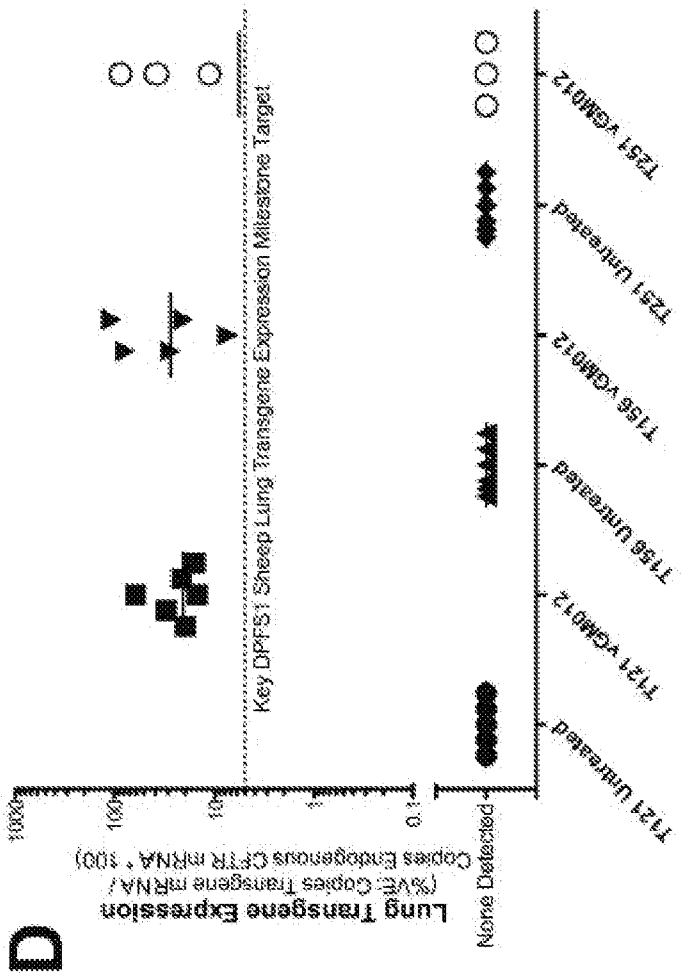


Figure 10



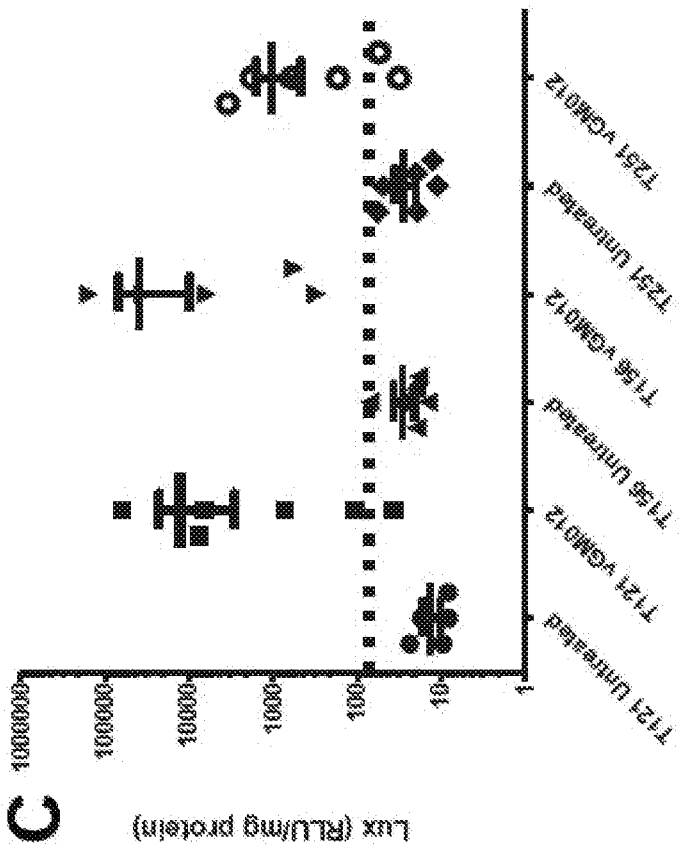
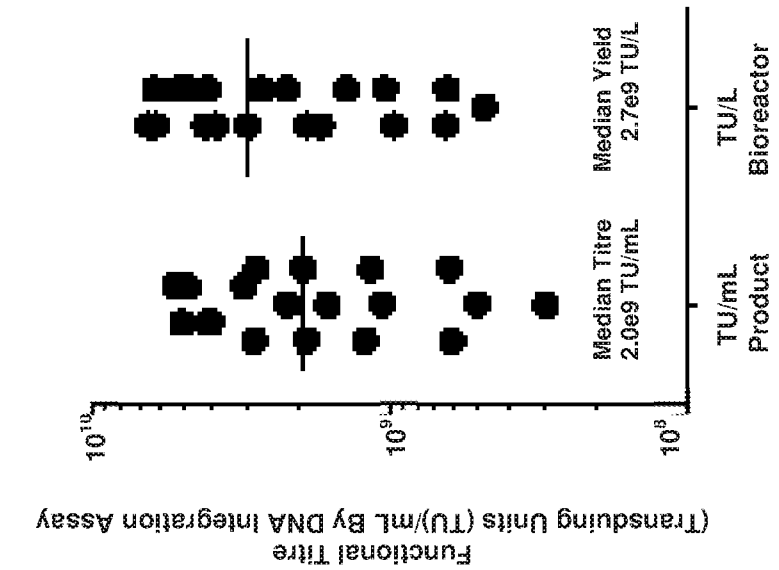


Figure 10 (continued)


B
Functional Titre
(Transducing Units (TU)/mL By DNA Integration Assay
10^{10}
10^{9}
10^{8}
Median Titre
2.0e9 TU/mL
Median Yield
2.7e9 TU/L
TU/mL
Product
TU/L
Bioreactor

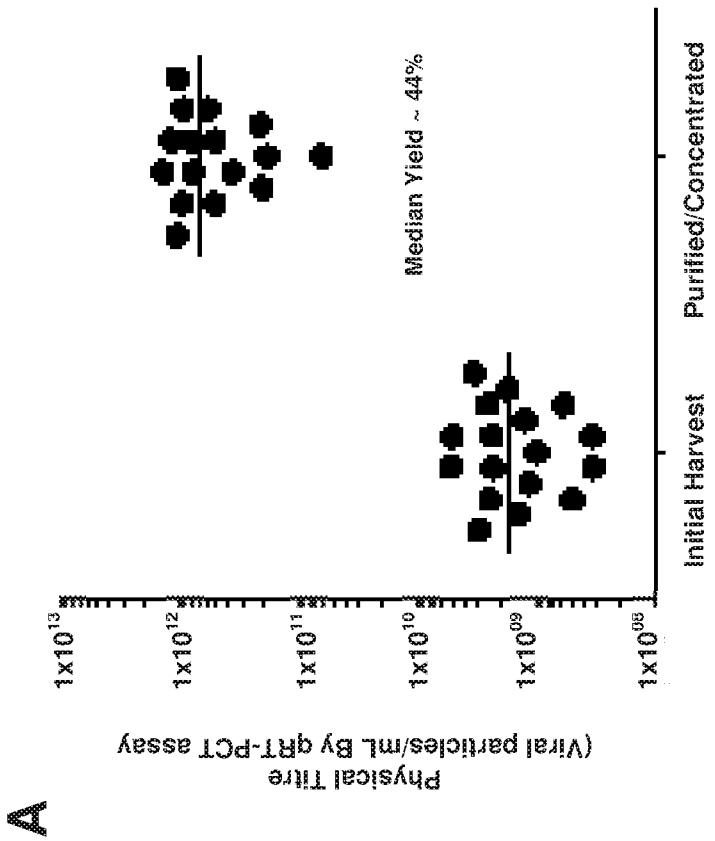

A
Physical Titre
(Viral particles/mL By qRT-PCT assay
$1x10^{13}$
$1x10^{12}$
$1x10^{11}$
$1x10^{10}$
$1x10^{09}$
$1x10^{08}$
Median Yield ~ 44%
Initial Harvest
Purified/Concentrated

Figure 11

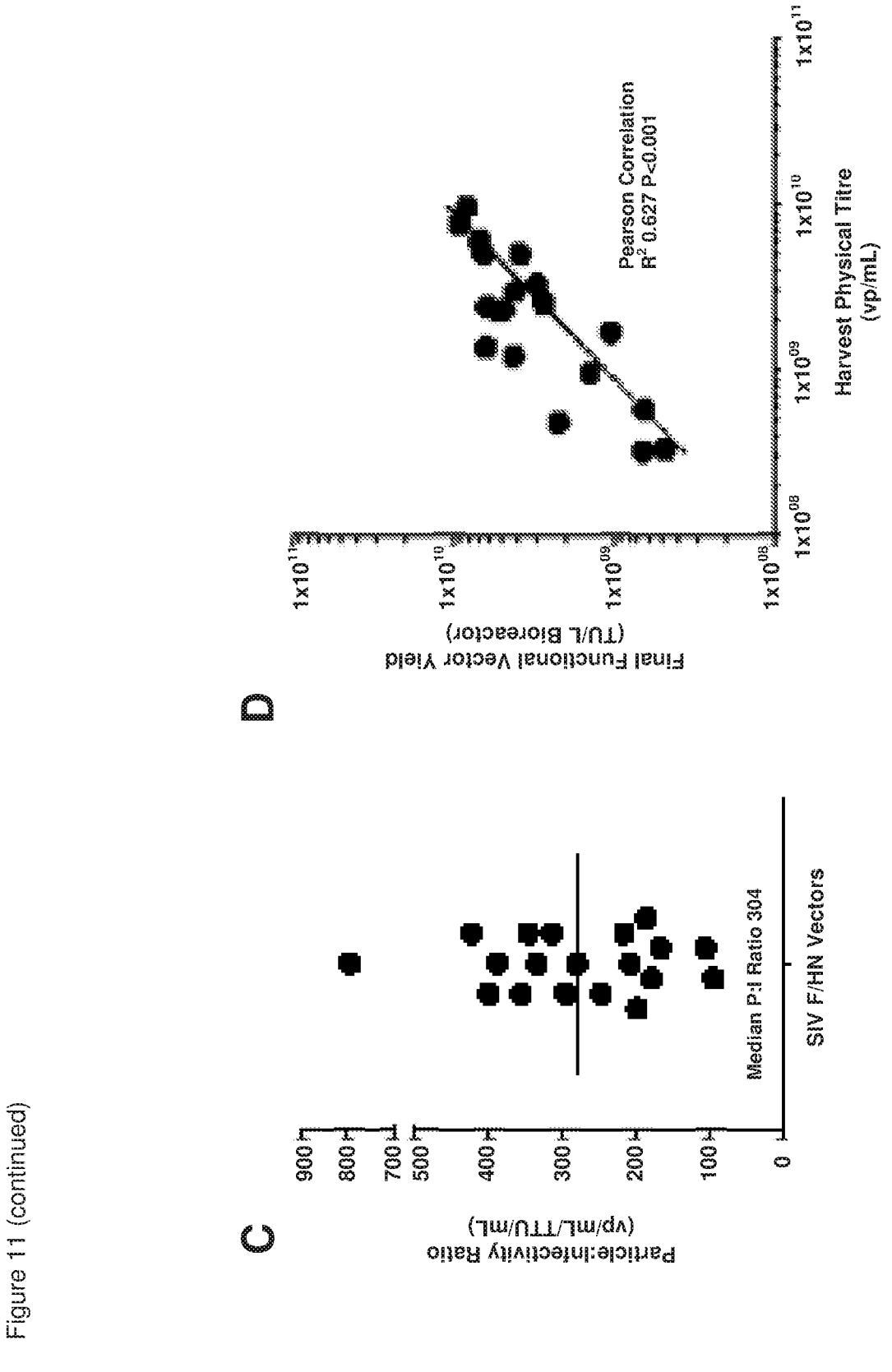


Figure 11 (continued)

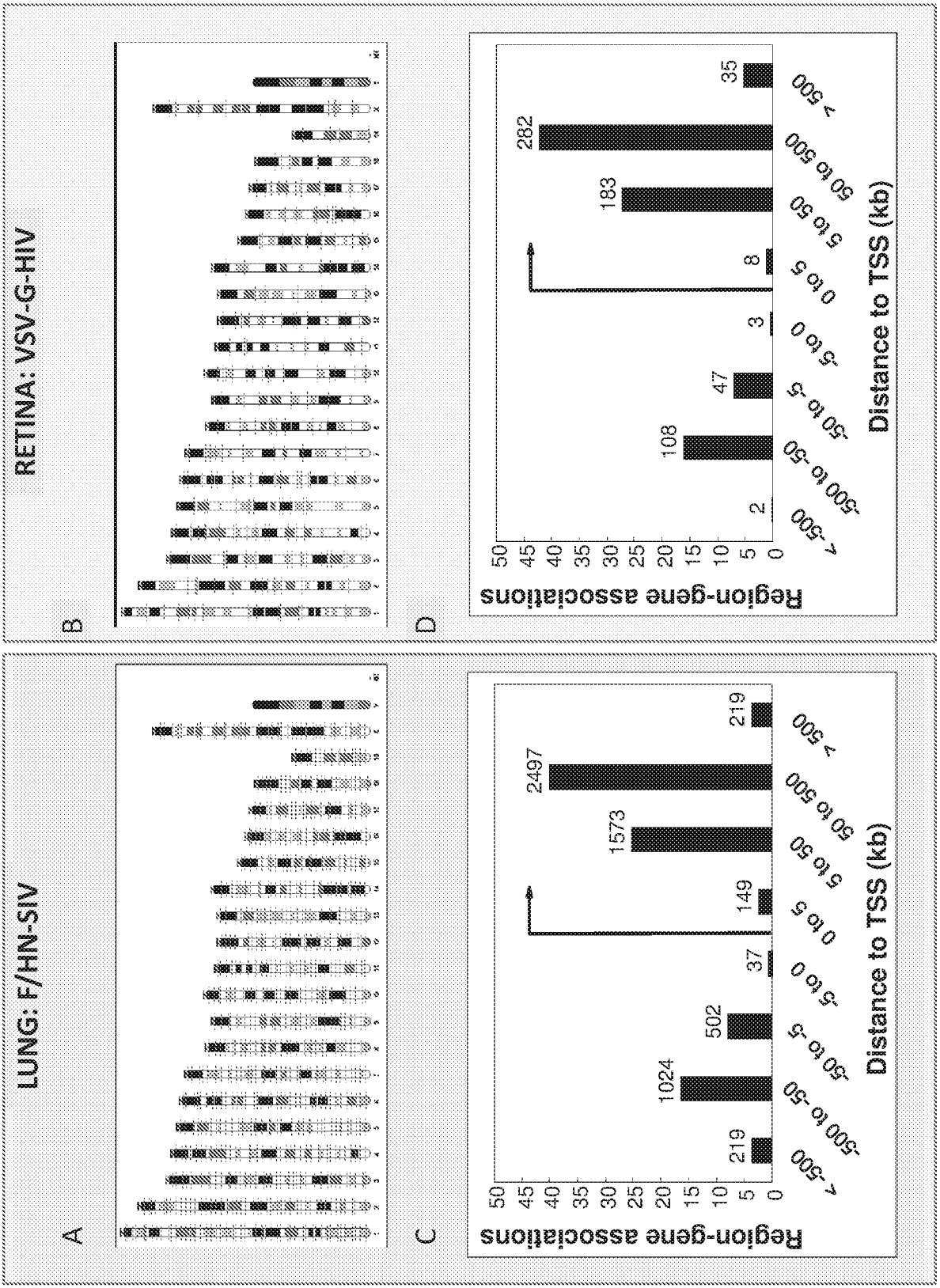
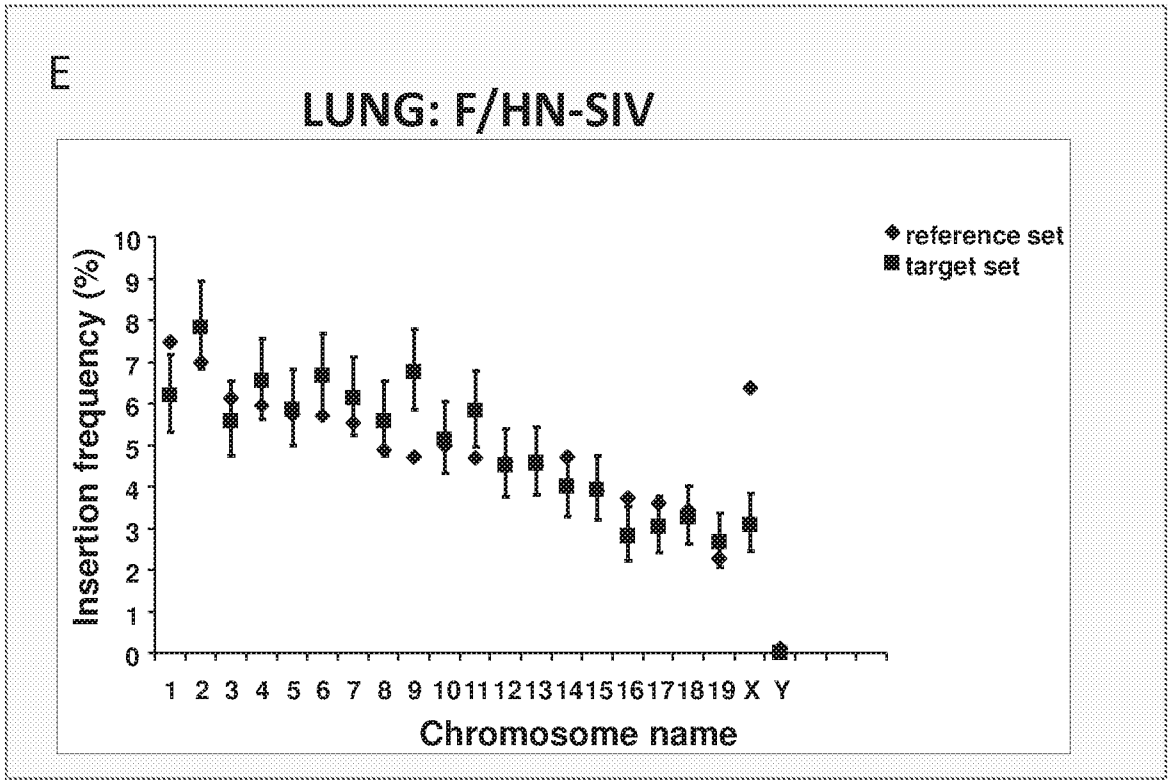
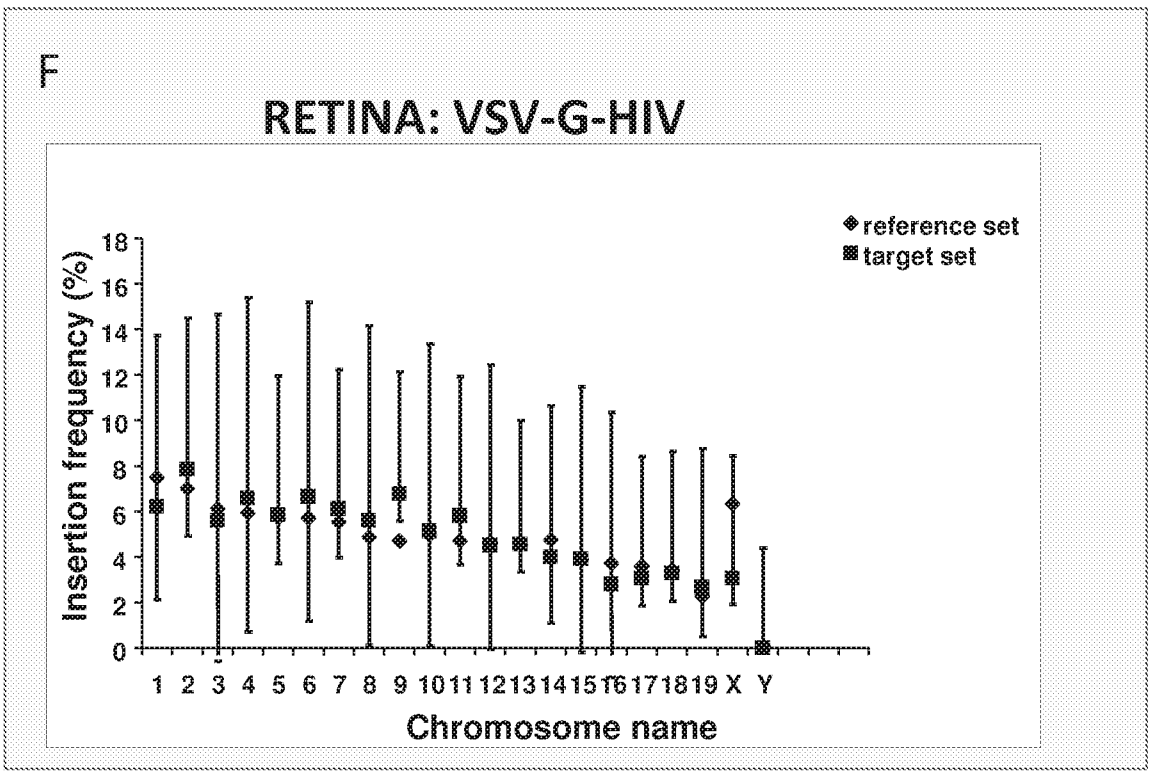

A
LUNG: F/HN-SIV
B
RETINA: VSV-G-HIV
C
Region-gene associations
Distance to TSS (kb)
219
1024
502
37
149
1573
2497
219
D
Region-gene associations
Distance to TSS (kb)
2
108
47
3
8
183
282
35
> 500
500 to 50
50 to 5
5 to 0
0 to 5
5 to 50
50 to 500
> 500

Figure 12

Figure 12 (continued)





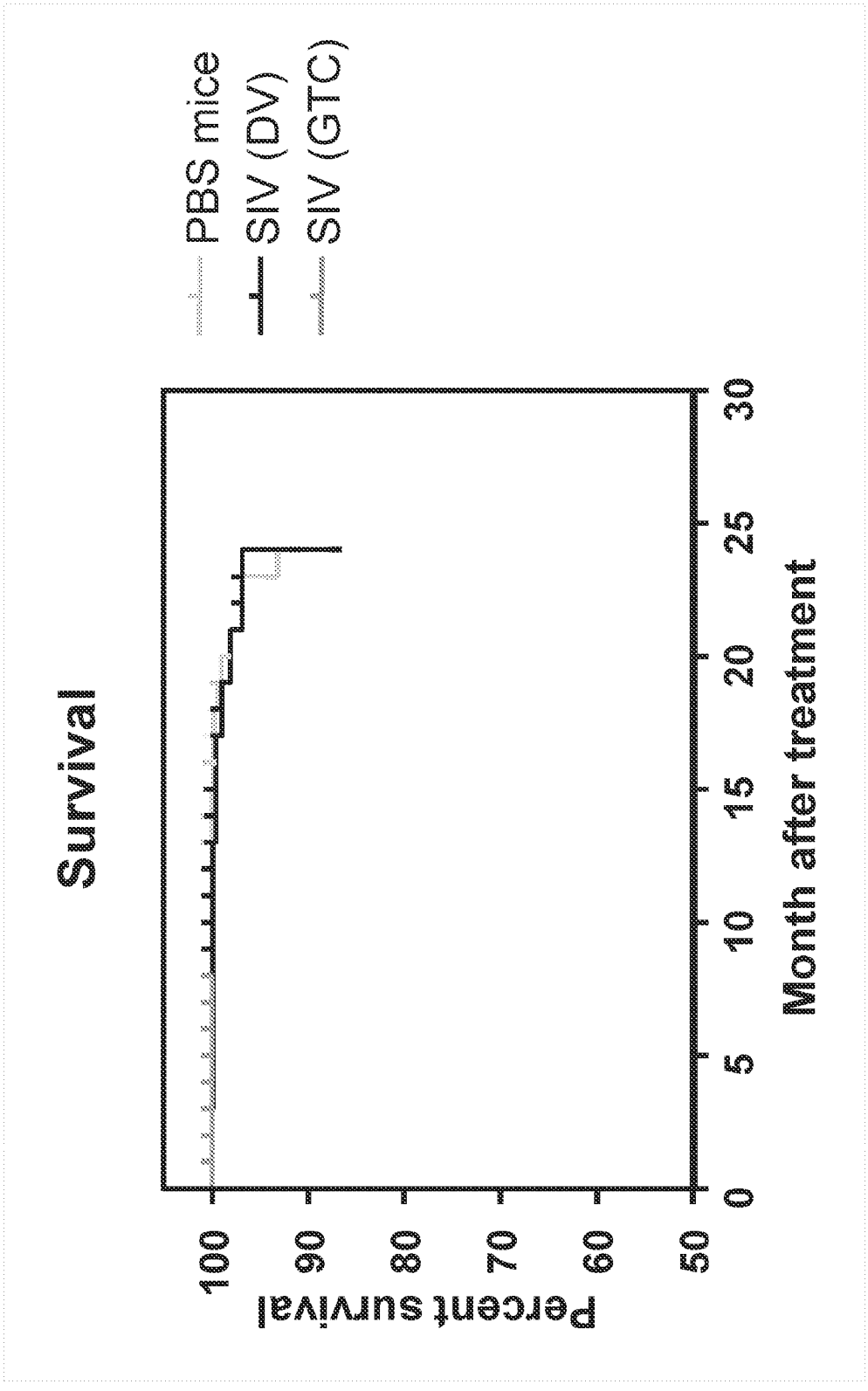


Figure 12 (continued)

G

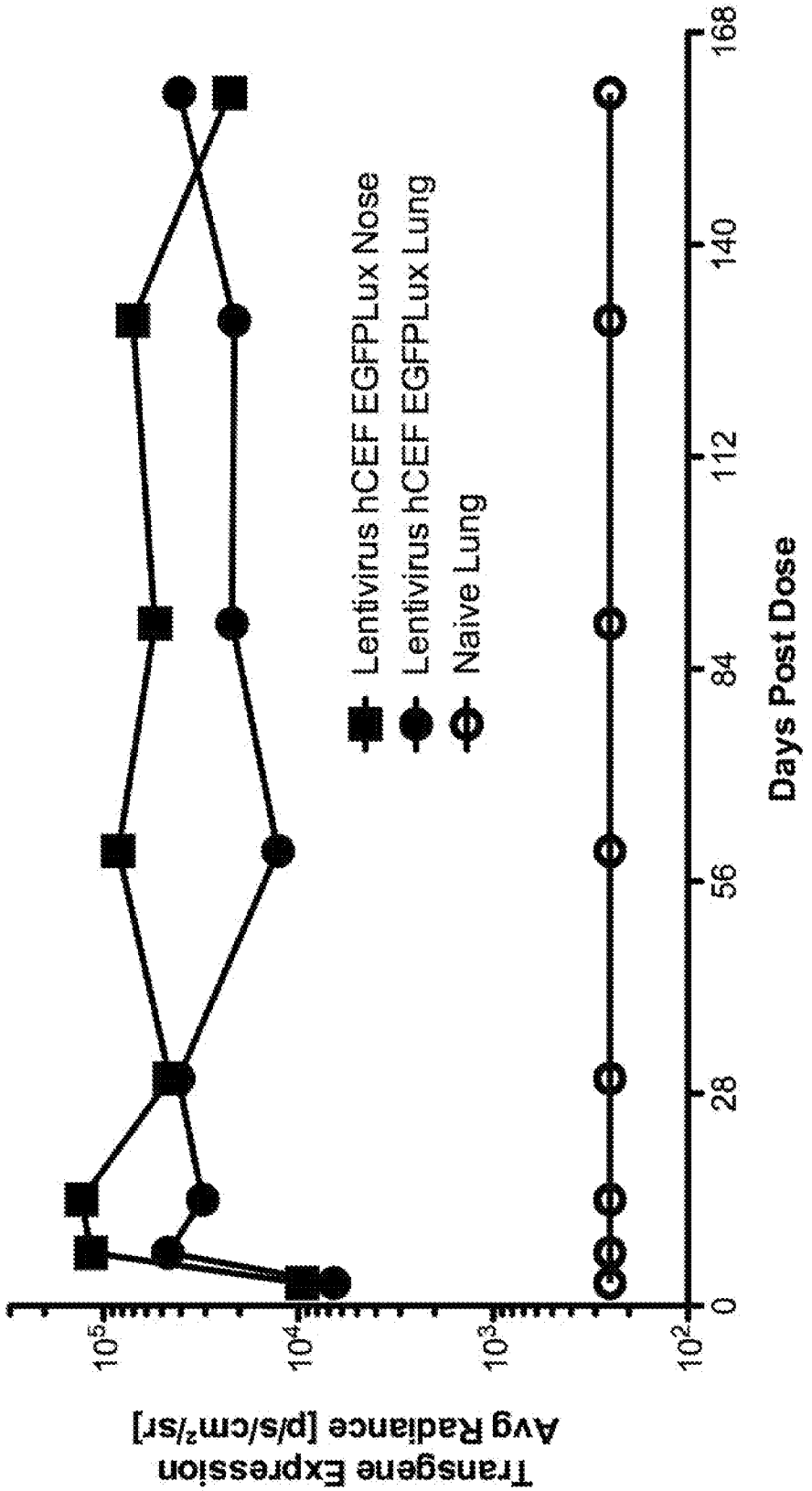
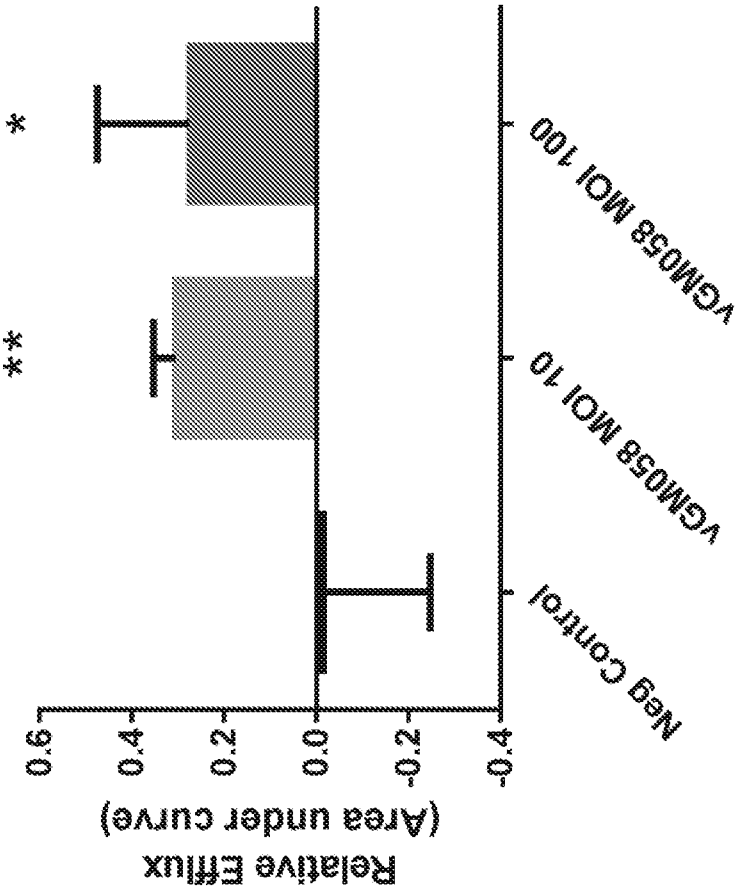

Transgene Expression
Avg Radiance [p/s/cm²/sr]
10⁵
10⁴
10³
10²
0
28
56
84
112
140
168
Days Post Dose
Lentivirus hCEF EGFPLux Nose
Lentivirus hCEF EGFPLux Lung
Naive Lung

Figure 13


Relative Efflux
(Area under curve)
0.6
0.4
0.2
0.0
-0.2
-0.4
Neg Control
VGM058 MOI 10
VGM058 MOI 100
**
*

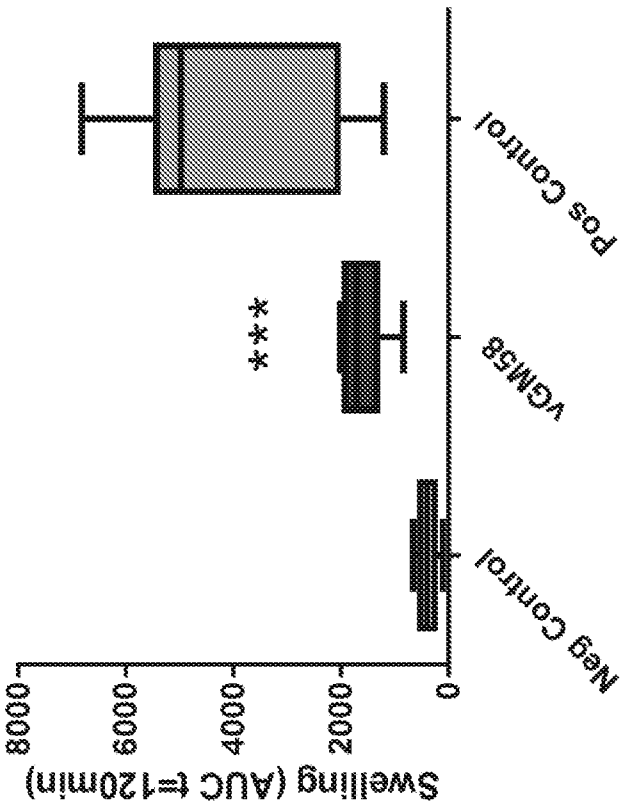
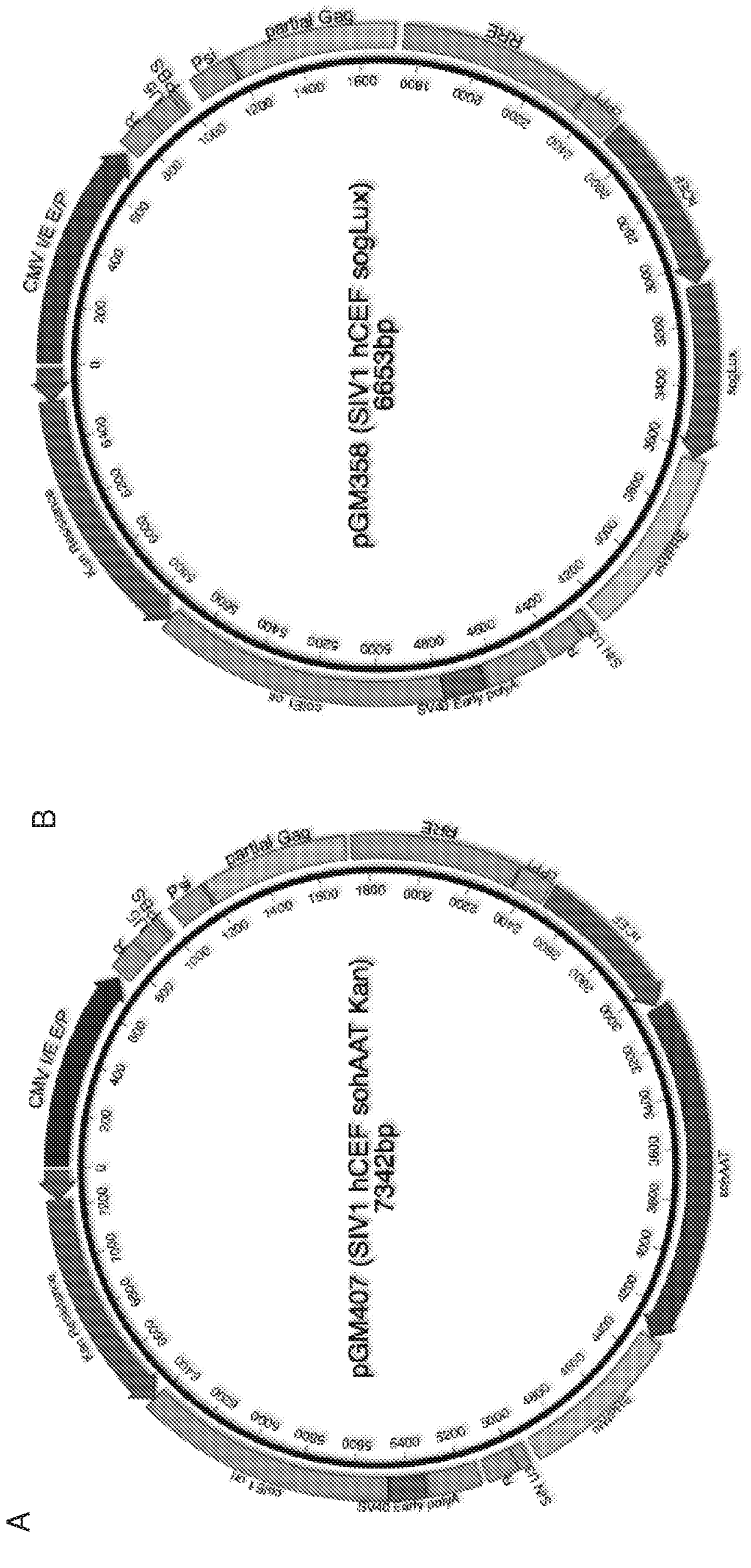

Swelling (AUC t=120min)
8000
6000
4000
2000
0
Neg Control
VGM58
Pos Control
***

Figure 14

Figure 15


A
pGM407 (SIV1 hCEF sohAAT Kan)
7342bp
CMV I/E E/P
partial Gag
Psi
RRE
SV40 Early polyA
B
pGM358 (SIV1 hCEF sogLux)
6653bp
CMV I/E E/P
partial Gag
Psi
RRE
SV40 Early polyA

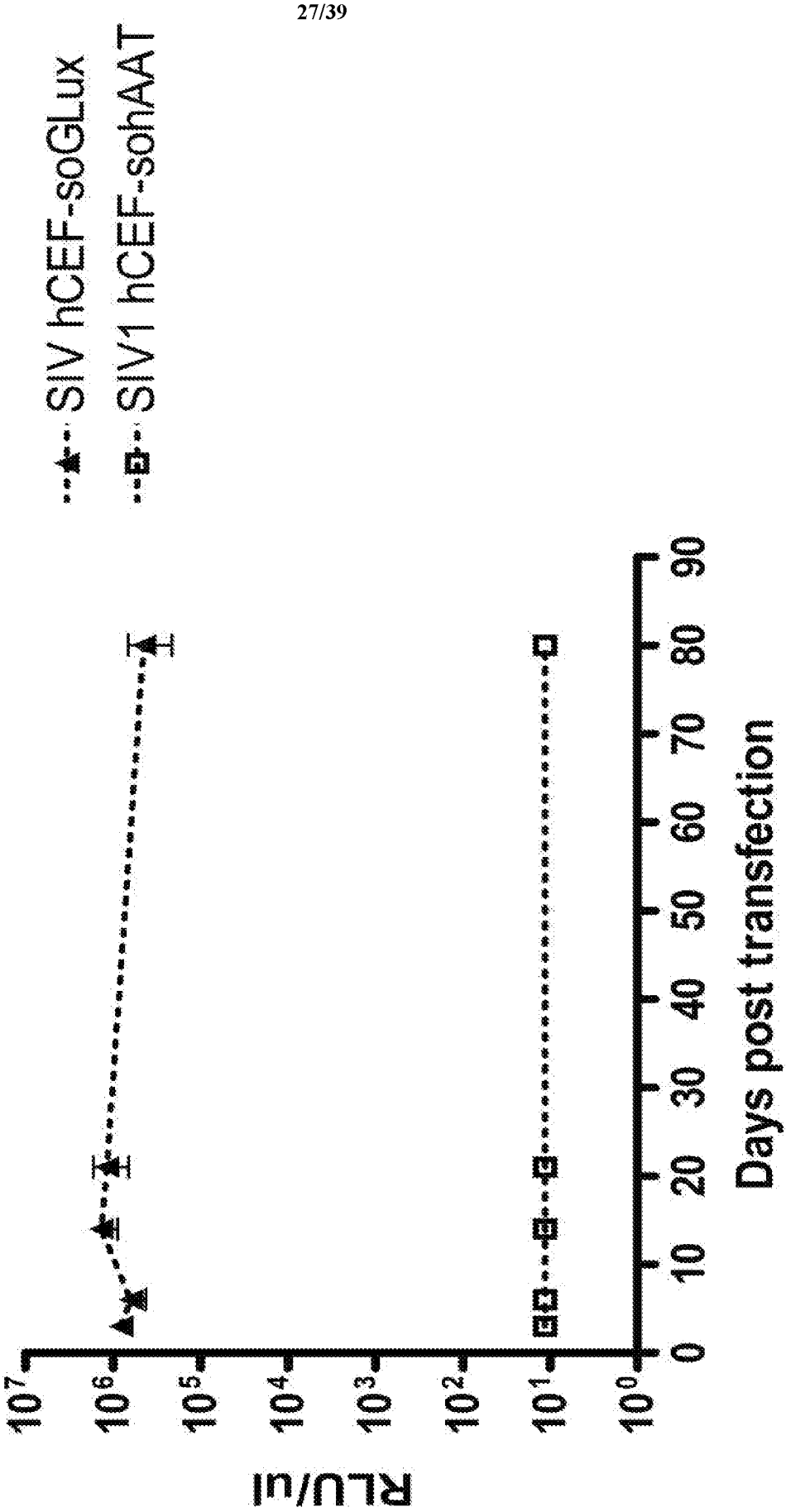
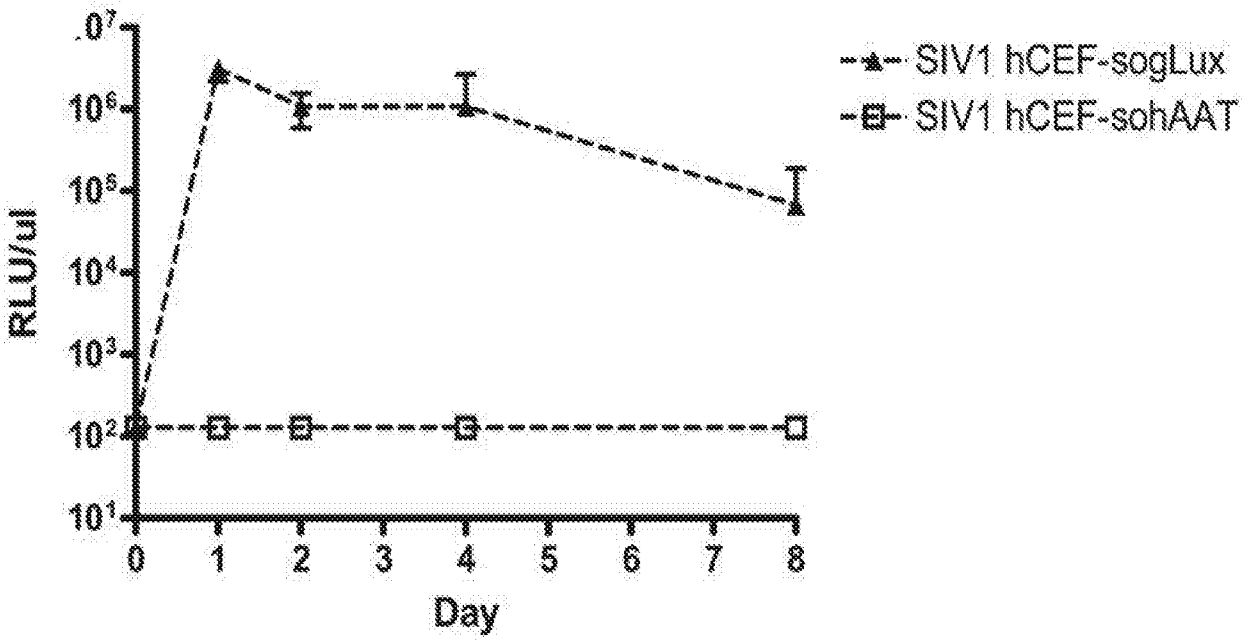

SIV hCEF-soGLux
SIV1 hCEF-sohAAT
RLU/ul
10^7
10^6
10^5
10^4
10^3
10^2
10^1
10^0
0 10 20 30 40 50 60 70 80 90
Days post transfection

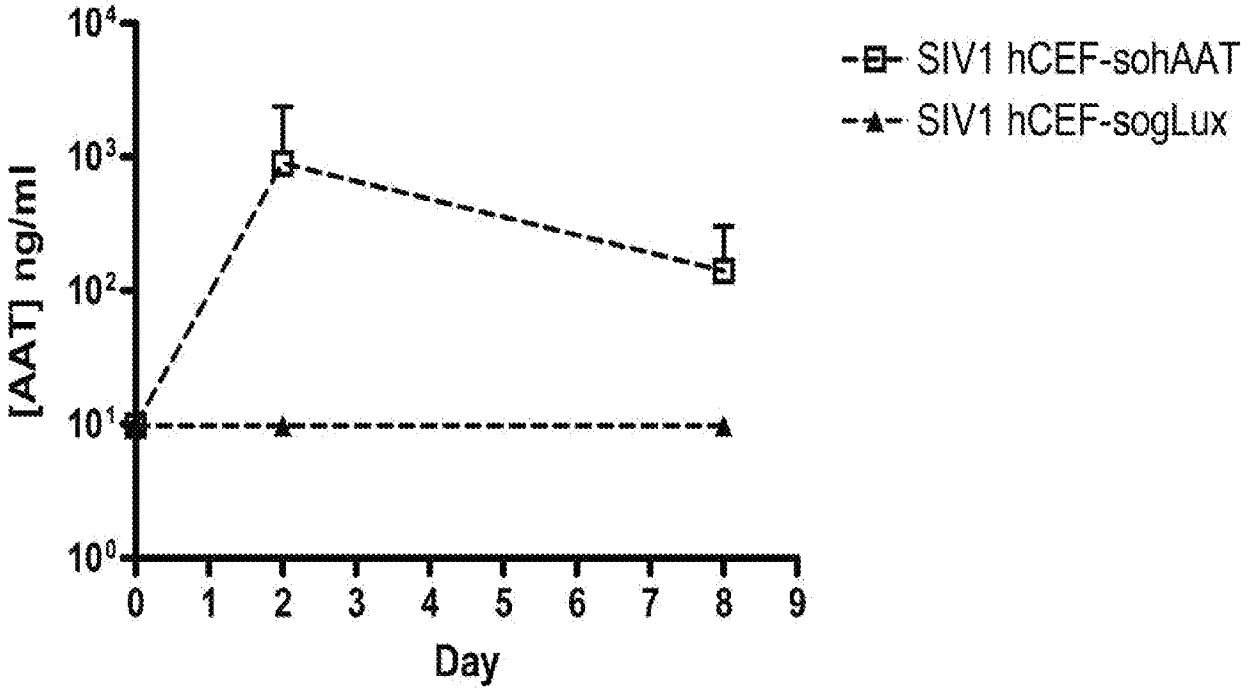
Figure 16

Figure 17

A


RLU/ul
10^7
10^6
10^5
10^4
10^3
10^2
10^1
0 1 2 3 4 5 6 7 8
Day
SIV1 hCEF-sogLux
SIV1 hCEF-sohAAT

B


[AAT] ng/ml
10^4
10^3
10^2
10^1
10^0
0 1 2 3 4 5 6 7 8 9
Day
SIV1 hCEF-sohAAT
SIV1 hCEF-sogLux

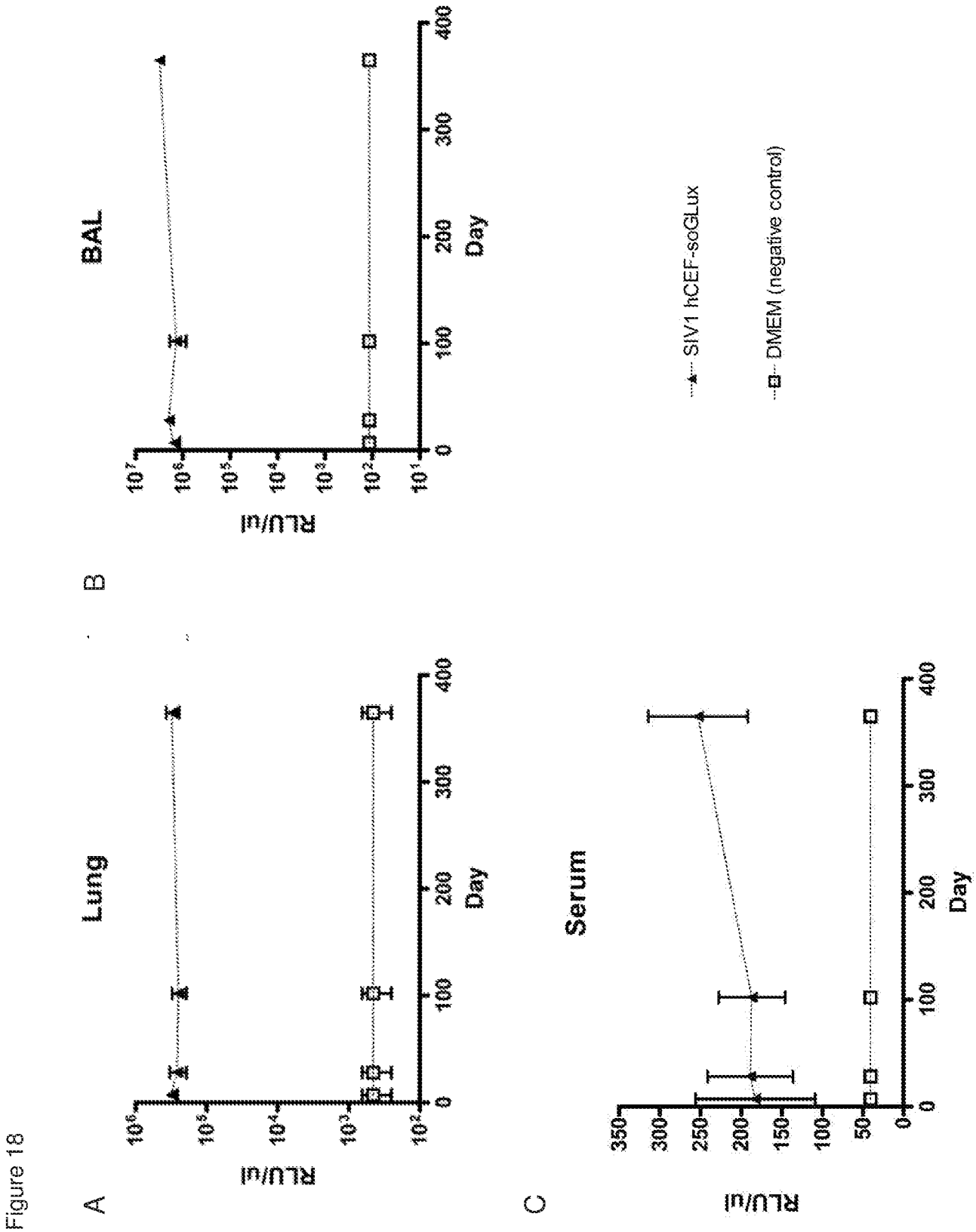
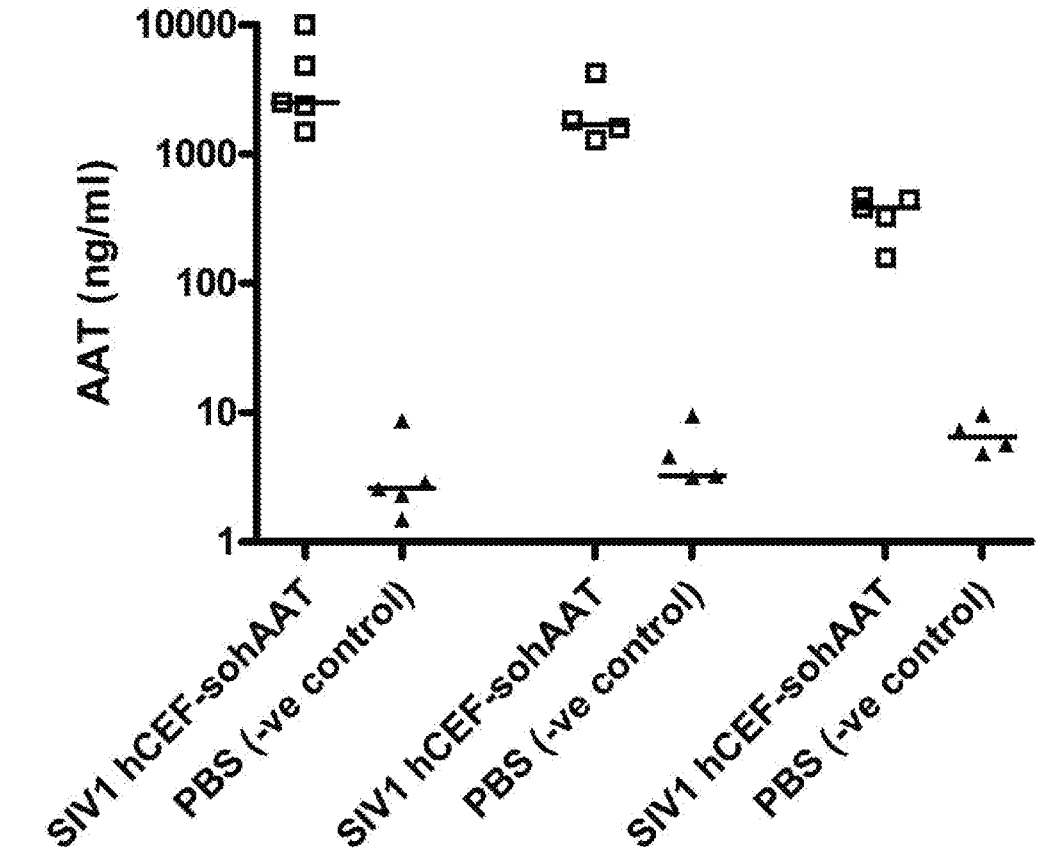

A
Lung
RLU/ul
Day
B
BAL
RLU/ul
Day
SIV1 hCEF-soGLux
DMEM (negative control)
C
Serum
RLU/ul
Day

Figure 18

Figure 19


10000
1000
100
10
1
AAT (ng/ml)
SIV1 hCEF-sohAAT
PBS (-ve control)
SIV1 hCEF-sohAAT
PBS (-ve control)
SIV1 hCEF-sohAAT
PBS (-ve control)
Lung
BAL
Serum

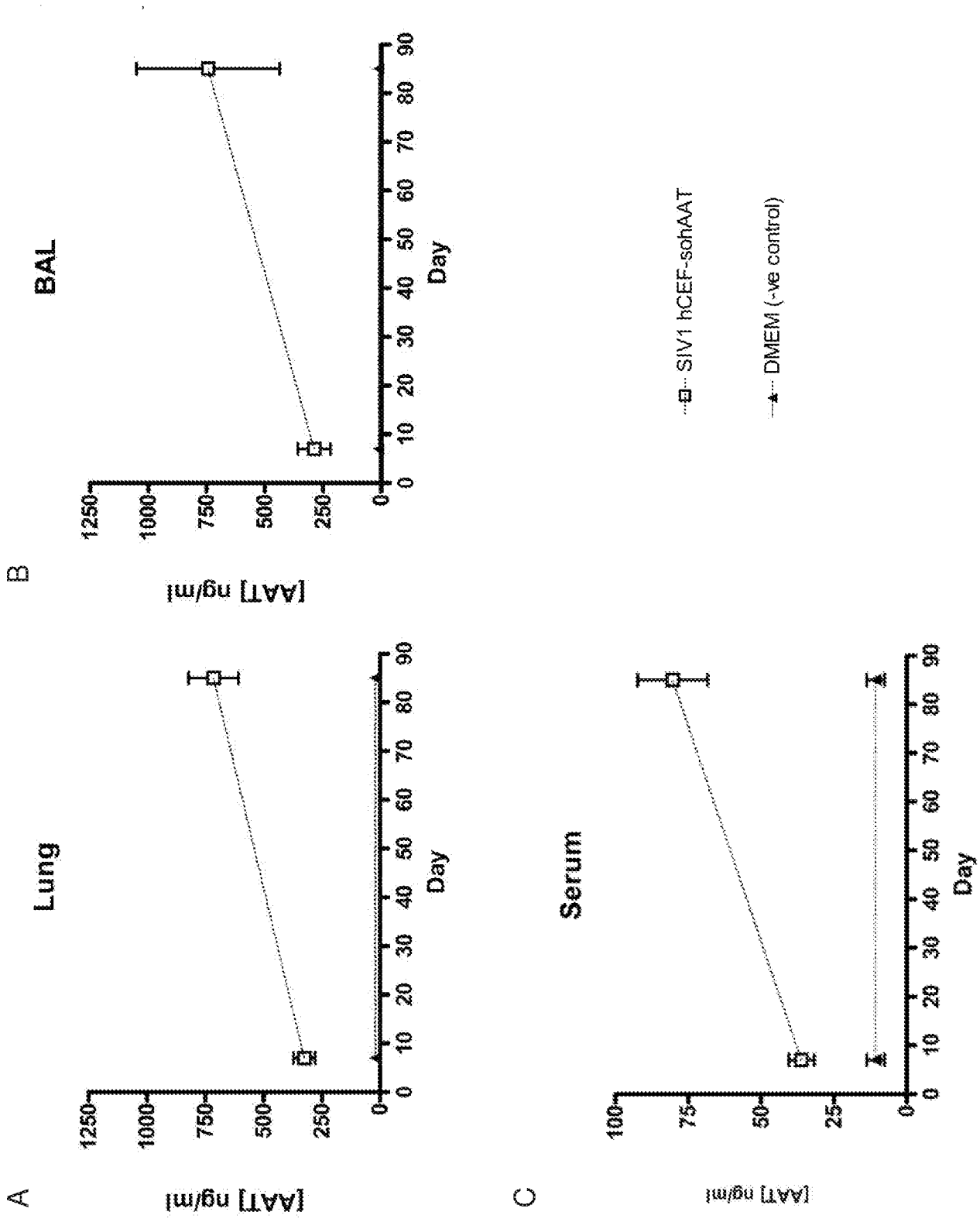
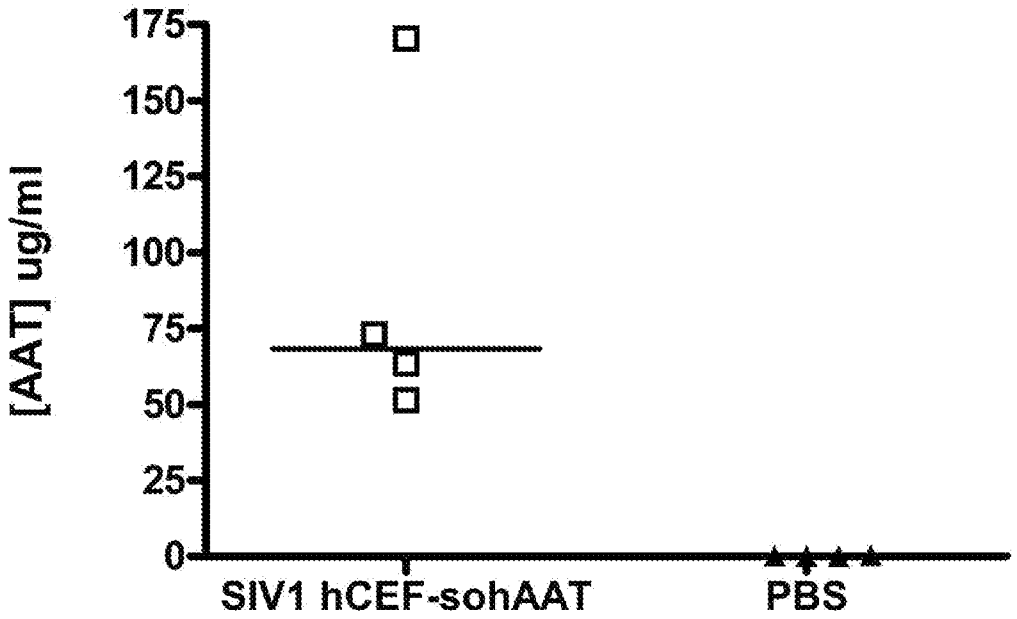

A
Lung
[AAT] ng/ml
Day
B
BAL
[AAT] ng/ml
Day
SIV1 hCEF-sohAAT
DMEM (-ve control)
C
Serum
[AAT] ng/ml
Day

Figure 20

Figure 21



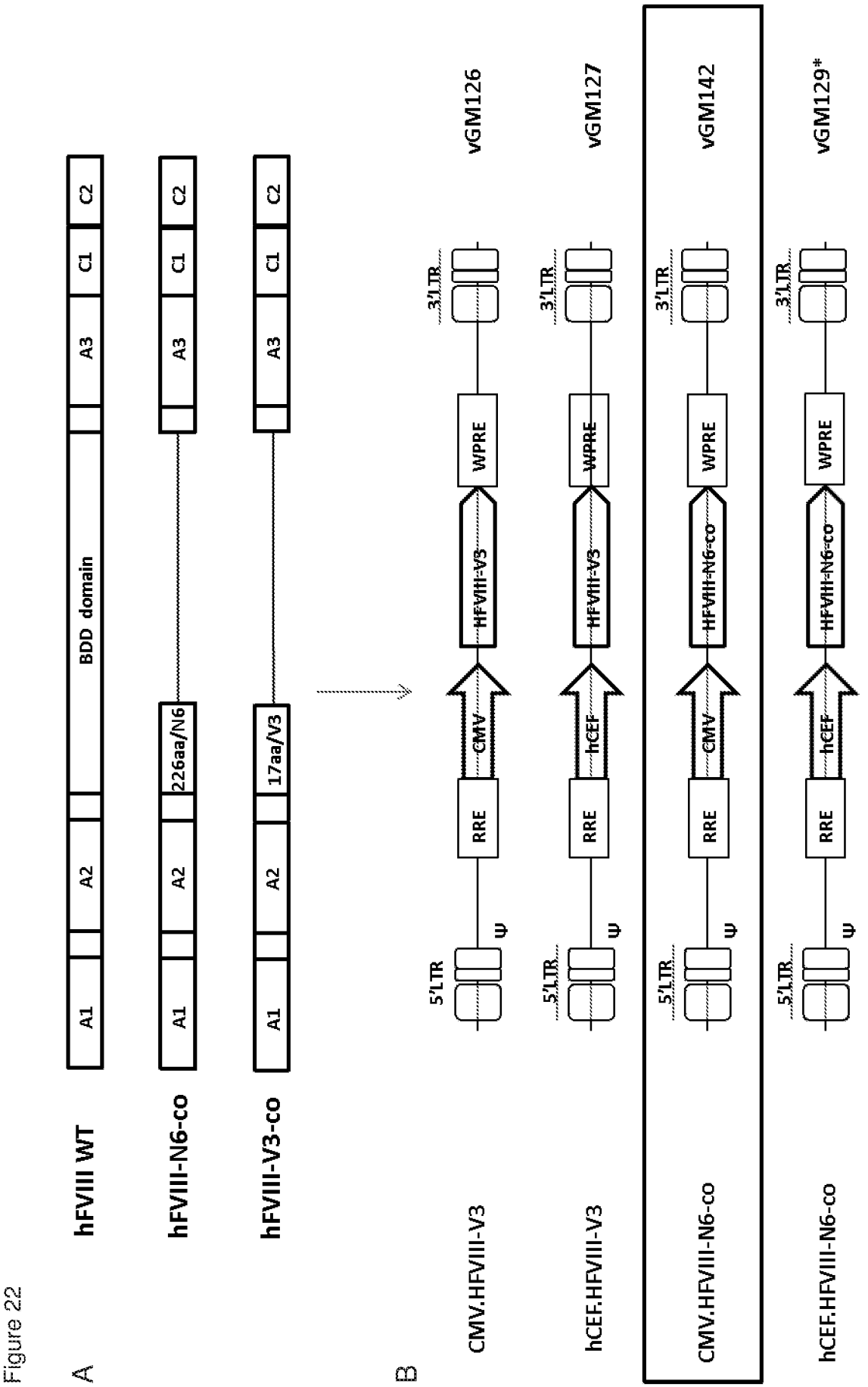

A
hFVIII WT
hFVIII-N6-co
hFVIII-V3-co
A1
A2
BDD domain
A3
C1
C2
226aa/N6
17aa/V3
B
CMV.HFVIII-V3
hCEF.HFVIII-V3
CMV.HFVIII-N6-co
hCEF.HFVIII-N6-co
vGM126
vGM127
vGM142
vGM129*
5'LTR
Ψ
RRE
CMV
hCEF
HFVIII-V3
HFVIII-N6-co
WPRE
3'LTR

Figure 22

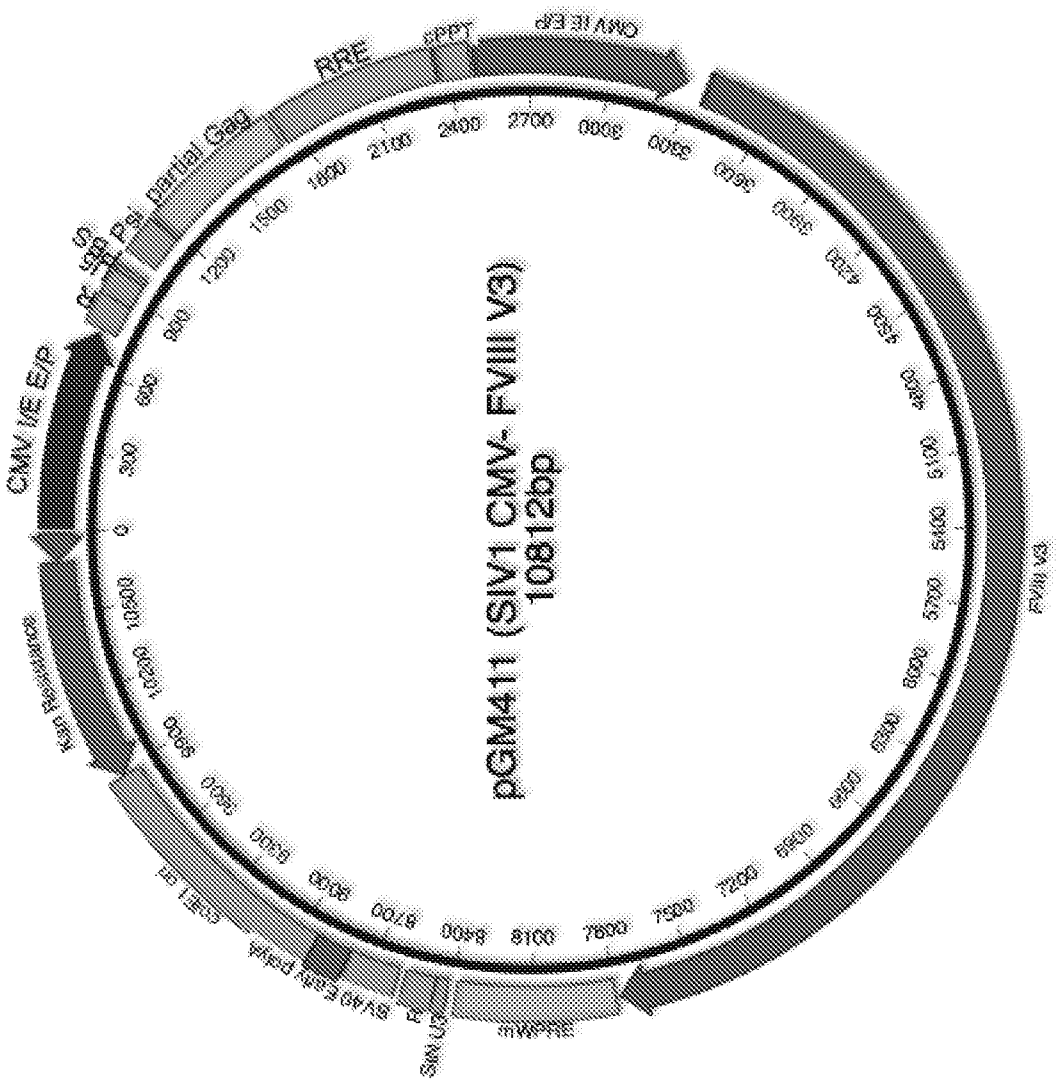
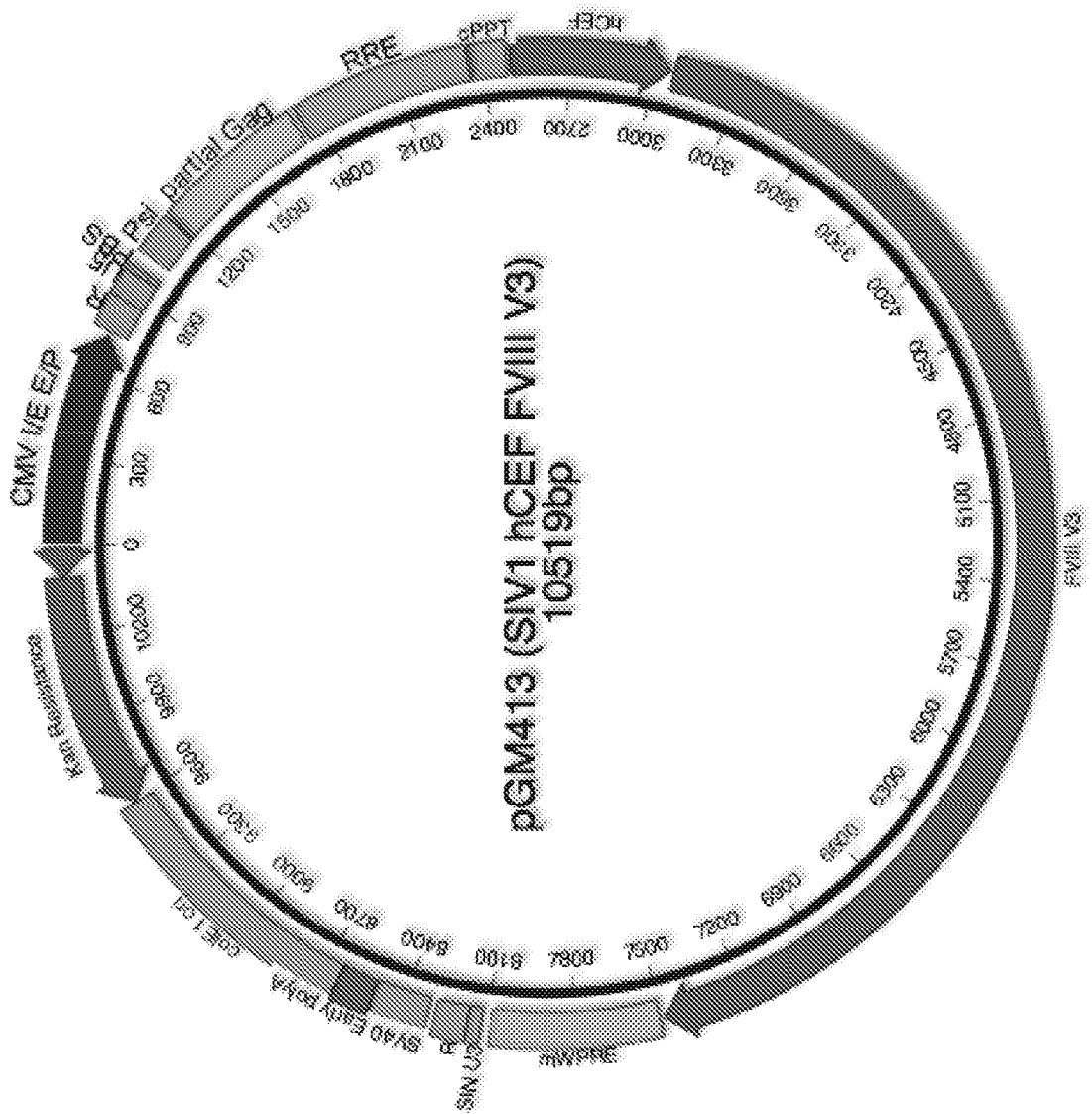


Figure 22 (continued)

C



D

Figure 22 (continued)

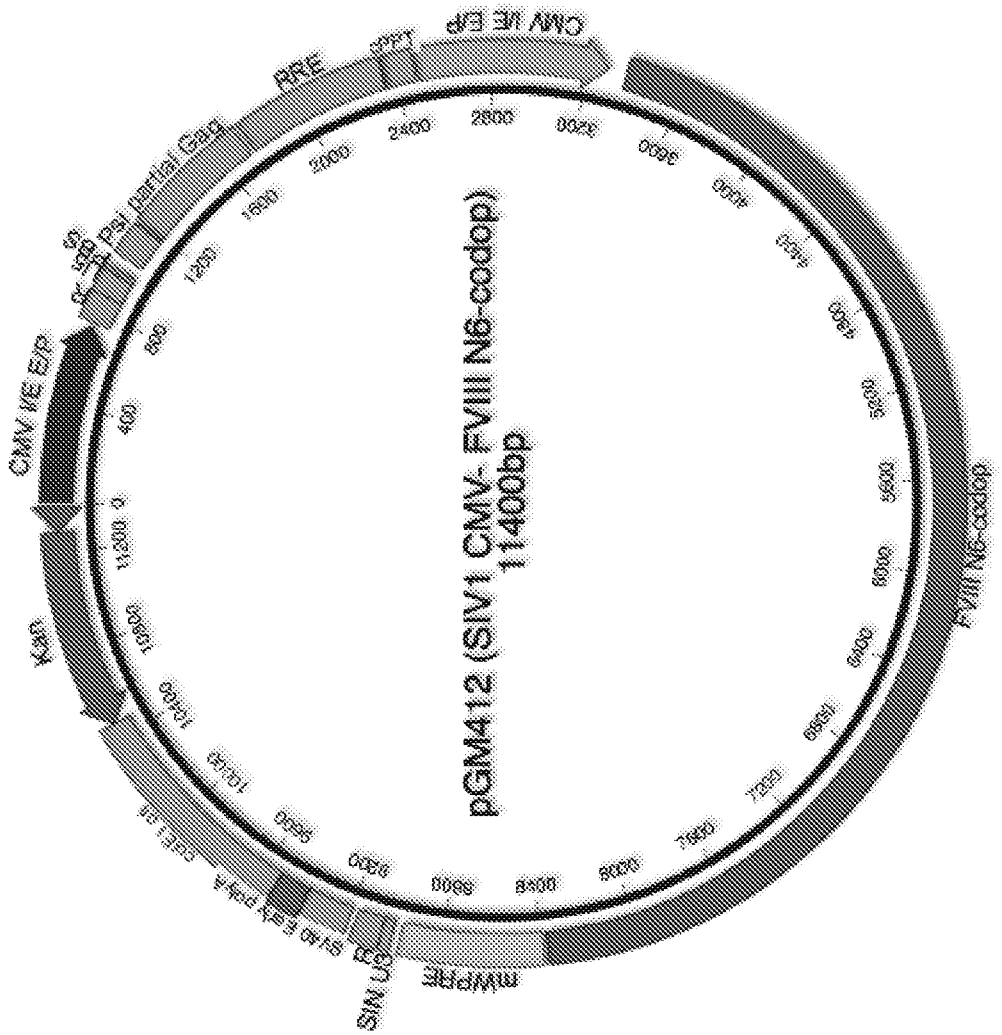


Figure 22 (continued)

E

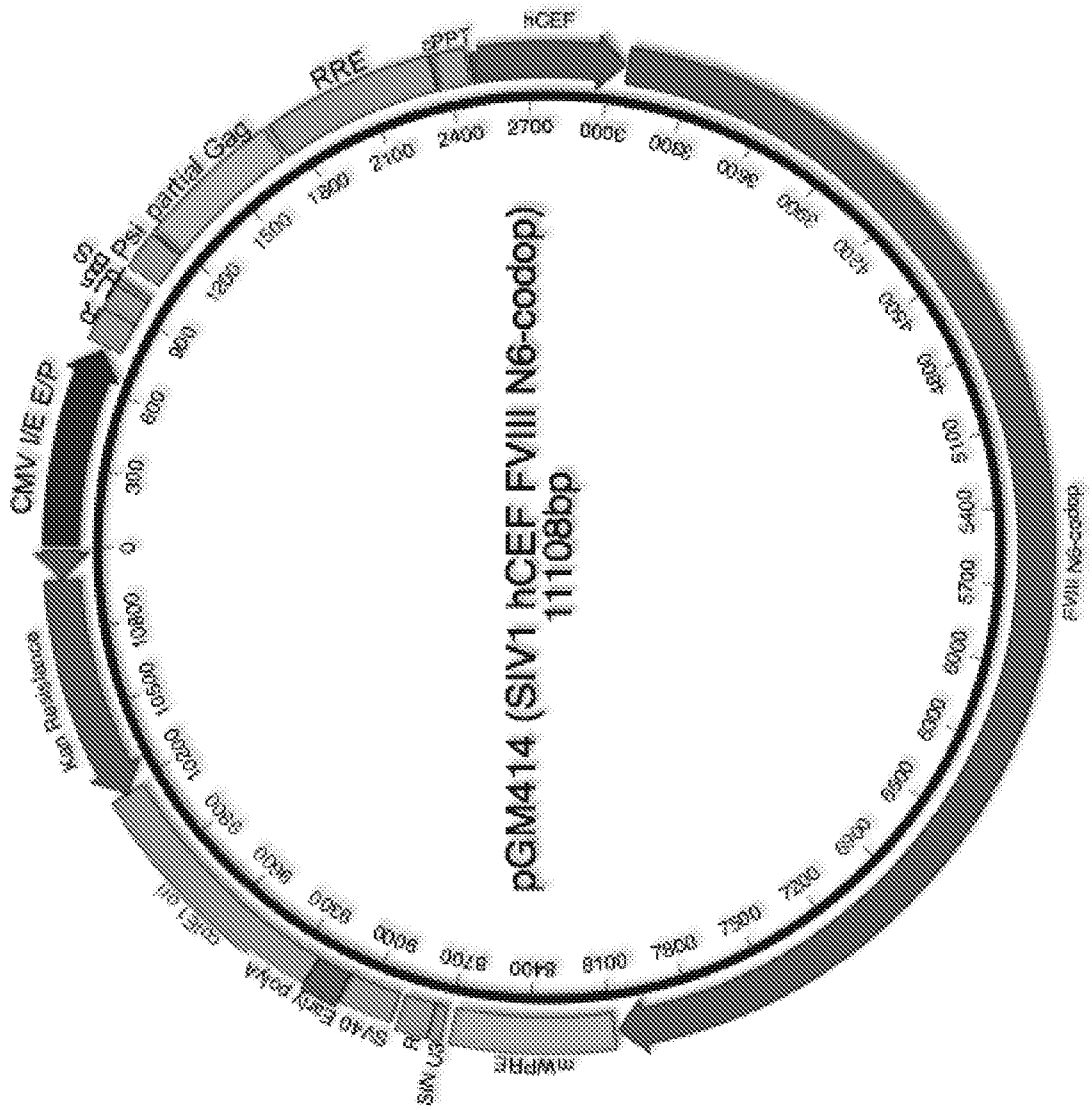


Figure 22 (continued)

F

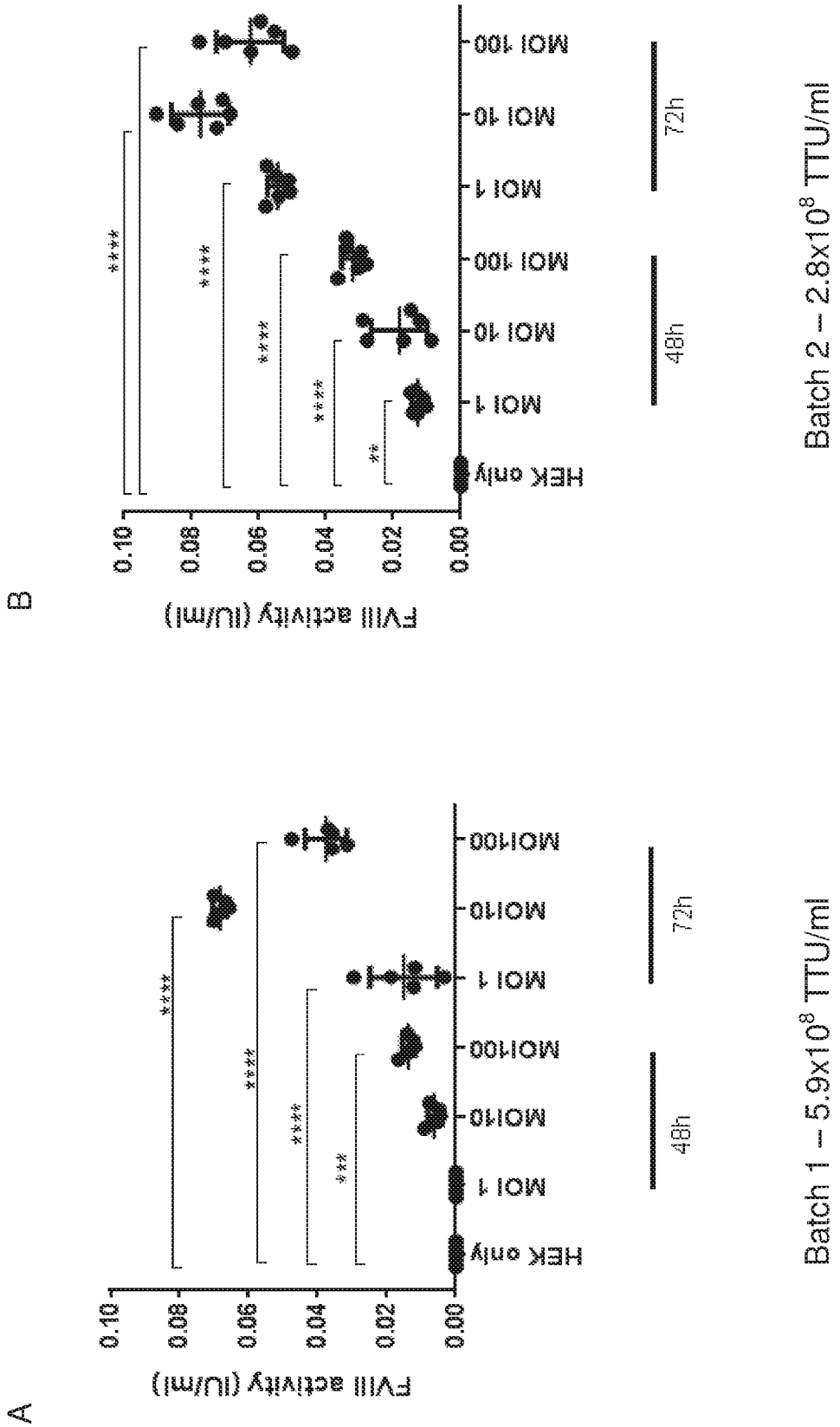

A
FVIII activity (IU/ml)
0.10
0.08
0.06
0.04
0.02
0.00
HEK only
MOI 1
MOI10
MOI100
MOI 1
MOI10
MOI100
****
***
****
****
48h
72h
Batch 1 – 5.9x10^8 TTU/ml
B
FVIII activity (IU/ml)
0.10
0.08
0.06
0.04
0.02
0.00
HEK only
MOI 1
MOI 10
MOI 100
MOI 1
MOI 10
MOI 100
**
****
****
****
****
48h
72h
Batch 2 – 2.8x10^8 TTU/ml

Figure 23

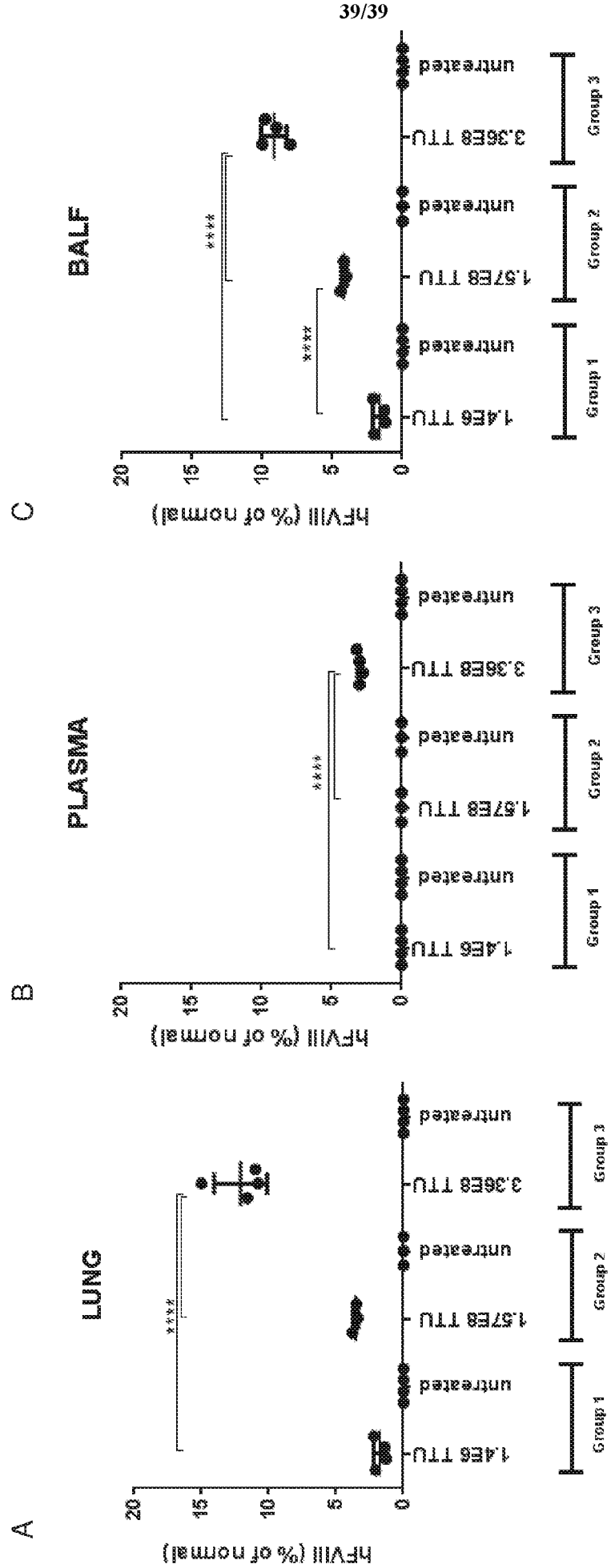
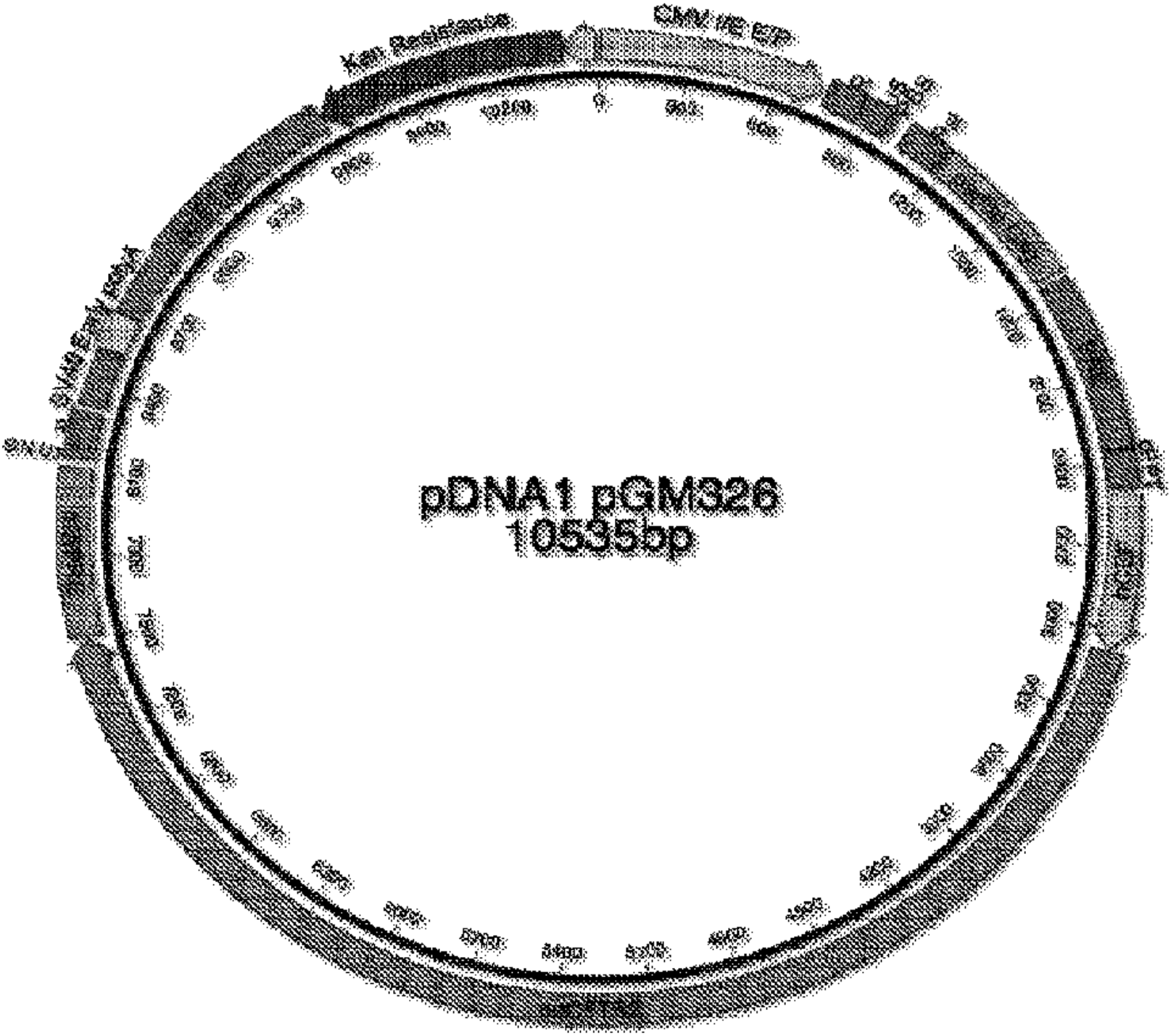

A
LUNG
hFVIII (% of normal)
20
15
10
5
0
****
1.4E6 TTU
untreated
1.57E8 TTU
untreated
3.36E8 TTU
Group 1
Group 2
Group 3
B
PLASMA
hFVIII (% of normal)
20
15
10
5
0
****
1.4E6 TTU
untreated
1.57E8 TTU
untreated
3.36E8 TTU
Group 1
Group 2
Group 3
C
BALF
hFVIII (% of normal)
20
15
10
5
0
****
****
1.4E6 TTU
untreated
1.57E8 TTU
untreated
3.36E8 TTU
Group 1
Group 2
Group 3

Figure 24


pDNA1 pGM326
10535bp