



(43) International Publication Date
22 November 2012 (22.11.2012)

- (51) International Patent Classification:
C12N 15/85 (2006.01) *C12N 5/074* (2010.01)
C12N 5/0735 (2010.01)
- (21) International Application Number:
PCT/GB2012/051075
- (22) International Filing Date:
15 May 2012 (15.05.2012)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
1108116.3 16 May 2011 (16.05.2011) GB
- (71) Applicant (for all designated States except US): **ISIS INNOVATION LIMITED** [GB/GB]; Ewert House, Ewert Place, Summertown, Oxford, Oxfordshire OX2 7SG (GB).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): **MONACO, Zoia Larin** [GB/GB]; c/o The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, Oxfordshire OX3 7BN (GB). **MORALLI, Daniela** [IT/GB]; c/o The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, Oxfordshire OX3 7BN (GB). **WADE-MARTINS, Richard** [GB/GB]; c/o Department of Physiology Anatomy and Genetics, Le Gros Clark Building, South Parks Road, Oxford, Oxfordshire OX1 3QX (GB).
- (74) Agent: **BARKER BRETTELL LLP**; 100 Hagley Road, Edgbaston, Birmingham, West Midlands B16 8QQ (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, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, 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, 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))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: METHODS FOR PROVIDING HUMAN CELLS COMPRISING A HUMAN ARTIFICIAL CHROMOSOME

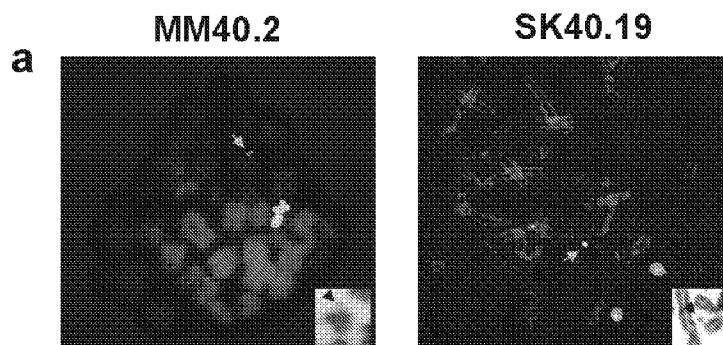


Fig. 5

(57) Abstract: A method for providing a eukaryotic cell comprising a human artificial chromosome (HAC), the method comprising the step of: transducing the eukaryotic cell with at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two viral vectors, wherein the at least one α -satellite DNA and at least one introduced DNA sequence are able to form one or more HACs inside the cell.



METHODS FOR PROVIDING HUMAN CELLS COMPRISING A HUMAN ARTIFICIAL CHROMOSOME

This invention relates to a method for providing a eukaryotic cell comprising a
5 human artificial chromosome (HAC). In particular, it relates to a method for
transducing a eukaryotic cell with DNA contained in one or more HSV-1
amplicon vectors to form a HAC inside the eukaryotic cell.

For the purposes of the present invention, all references as cited herein are
10 incorporated by reference in their entirety.

Human artificial chromosomes (HACs) are autonomous molecules that behave as
normal chromosomes in human cells. Together with the endogenous
chromosomes, HACs segregate during cell division and are maintained in the host
15 cell. *De novo* HACs are generated by introducing defined sequences such as α -
satellite (alphoid) DNA and specific introduced DNA sequences into eukaryotic
cells. α -satellite DNA, containing higher-order repeat sequences, and a
centromere protein B binding sequence (CENP-B box) are requirements for a
functional centromere in a HAC.

20

Known methods for the delivery of large pieces of DNA are inefficient, often
resulting in DNA shearing and degradation, which is a major obstacle in
developing a HAC expression system in different cell types.

25 Most gene expression studies in hESc utilize lentiviral, adenoviral and adeno-
associated (AAV) viral vectors for gene delivery. However, lentiviral vectors
integrate randomly at multiple sites within the host genome leading to insertional
mutagenesis, and although adenoviral vectors remain episomal, silencing post
transduction may occur. Another disadvantage is that the capacity of AAV and
30 lentiviral vectors is limited to approximately 5 and 10 kb of DNA respectively.

Previous studies introduced input HAC DNA (404 kb) containing the
hypoxanthine-guanine phosphoribosyl transferase (HPRT) gene into human HPRT
deficient HT1080 fibrosarcoma cells, and the HAC generated successfully

complemented the deficiency. The input HAC DNA was delivered by standard transfection, but it was found that this method was inefficient for routine large DNA transfer and did not work in other cell types.

- 5 The present invention provides a new approach to delivering large DNA sequences into cells to form a HAC. This delivery approach is based on the use of a viral vector system, preferably a herpes simplex virus type 1 (HSV-1) amplicon vector system. The DNA may be delivered to the cell using two or more viral vectors which may be selected from vectors based on any virus, for
10 example, an HSV virus, a retrovirus, an adeno-associated virus (AAV), an adenovirus, a vaccinia virus or a plant virus.

In one embodiment the two viral vectors may both be HSV-1 amplicon vectors. HSV-1 amplicon vectors differ from HSV-1 vectors. HSV-1 vectors are HSV-1
15 viruses, which have been deleted of essential genes, but are still formed mostly of the viral genome. By contrast, in the HSV-1 amplicon system, the vector may be, for example, a BAC or PAC input HAC DNA that only contains two short viral sequences, the origin of replication OriS, and the packaging signal Pac. HSV-1 amplicon vectors are high-capacity vectors that can hold up to 150 kb of
20 additional DNA sequence and successfully deliver a large introduced DNA sequence (for example bacterial artificial chromosomes (BACs) or P1 artificial chromosomes (PACs)) intact into different cell types in the absence of contaminating viral genes. The advantage of the HSV-1 amplicon system is that HSV-1 amplicons have a high capacity for large DNA delivery (up to 150 kb) and
25 can efficiently introduce HAC input DNA into cells, for example immortalized cell lines, or into stem cells. The efficiency of DNA delivery into cells using the HSV-1 amplicon system is significantly greater than with other methods of chemical transfection (by a factor of 10^4).

- 30 Surprisingly the HSV-1 amplicon vector system can be used to transduce stem cells. It is possible to co-transduce two HSV-1 amplicon vectors into the same cell, both in differentiated cells and stem cells. While it is known that a cell where HSV-1 has established latency can be superinfected by a second, different strain of HSV, it has generally been thought that the simultaneous infection by

two identical HSV-1 virions is not possible. This is because, since HSV-1 directly delivers its DNA to the nucleus, even if two transgenes could enter the cell via different amplicons, they would not be able to interact, and so would probably integrate, or form HACs independently of one another. Surprisingly, the two HSV-1 amplicon vectors delivered to cells were able to recombine and form a single HAC. This new approach opens up many possibilities for forming HACs inside cells because large amounts of DNA can be introduced into the cells efficiently and the DNA can form into HACs that are stably contained inside the cells.

10

According to a first aspect the invention provides a method for providing a eukaryotic cell comprising a human artificial chromosome (HAC), the method comprising the step of:

transducing the eukaryotic cell with at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two viral vectors, preferably at least two HSV-1 amplicon vectors, wherein the α -satellite DNA and the at least one introduced DNA sequence are able to form one or more HACs inside the cell.

An α -satellite DNA is needed for each HAC that is formed. As well as the α -satellite DNA the HAC may comprise one or more introduced DNA sequences. All of the DNA that is required to form a HAC may be comprised on at least two viral vectors, preferably at least two HSV-1 amplicon vectors. This means that the total DNA that is required to form a HAC is distributed between the two or more viral vectors, preferably two or more HSV-1 amplicon vectors. The two or more viral vectors preferably two or more HSV-1 amplicon vectors, may also comprise other DNA in addition to the DNA required to form a HAC. The DNA that is required to form a HAC may be comprised on or distributed between the two or more HSV-1 amplicon vectors in any convenient way, for example, the α -satellite DNA may be on one of the vectors, while the one or more introduced DNA sequences may be on one or more additional HSV-1 amplicon vectors. Alternatively, the α -satellite DNA may be on one of the HSV-1 amplicon vectors with some of the introduced DNA and the remainder of the introduced DNA may be on one or more additional HSV-1 amplicon vectors. In another embodiment

the α -satellite DNA and the introduced DNA may each be divided between the one or more HSV-1 amplicon vectors.

- An introduced DNA sequence may be any piece of DNA (excluding the α -satellite DNA) that is included on an HSV-1 amplicon vector, which is introduced into a cell in order to become part of a HAC. The introduced DNA may include a coding and/or a non-coding piece of DNA, for example all or part of one or more genes, regulatory regions and/or exons. The introduced DNA may be, for example, chromosomal DNA, an artificial DNA, a cDNA, or mitochondrial DNA.
- The introduced DNA may be identical to or different from the DNA of the host cell that it is introduced into. In one embodiment the introduced DNA comprises a gene that can express a protein or a peptide of interest inside the cell once a HAC is formed. Alternatively the introduced DNA may encode oligos or regulatory RNA molecules or a molecule that prevents transcription of a target gene, or interferes with the translation of a target protein or a non coding RNA/DNA controlling cellular processes. Preferably, the introduced DNA is configured to be at least transcribed, and preferably translated when the HAC is formed.
- The minimum requirement for forming a HAC in a cell is an α -satellite DNA sequence, for example the sequence of the core alpha satellite DNA from human chromosome 17 shown in Figure 12. In addition a HAC may include one or more introduced DNA sequences that may be expressed, or at least transcribed, in the host cell. Preferably the HAC has α -satellite DNA which provides the HAC with its own centromere and thus the HAC is maintained as an autonomous molecule in the cell without selection. In one embodiment telomeric sequences may also be included in the HAC. Preferably, the HAC behaves as a normal chromosome in a eukaryotic cell and segregates during cell division along with the endogenous chromosomes of the host cell. Optionally a HAC may also comprise a reporter gene, such as GFP, and selectable gene that help detection of HACs in the cells.

The minimum amount of input α satellite DNA sufficient to generate a HAC in the target cells is 40kb. Typically the total DNA needed to form a HAC (the HAC input DNA) ranges between 50 and 200 kb. If the HAC input DNA is

delivered by HSV-1 amplicons, then the upper limit of the amount of HAC input DNA is about 150kb or multiple of about 150 kb if more than one amplicon is used to deliver it, because each HSV-1 amplicon vector can introduce about 150kb of DNA.

5

An advantage of forming a HAC in a cell is that it can contain and express large pieces of DNA. The large capacity of a HAC allows it to express whole genes. An important factor in the expression of some genes is that endogenous promoters or controlling regions may be located several kilobases upstream or downstream of the gene of interest. The position of the promoters or controlling regions with respect to the gene is important for overall gene regulation. Hence the larger the region that can be incorporated into the HAC, the higher the chance of obtaining physiological levels of expression.

10

15 A HAC is maintained for a long time in the cells with or without selection. In the presence of selection a HAC may be maintained in the cell for more than four years. In the absence of selection a HAC is stable for longer than 9 months. The HAC replicates and segregates into the daughter cells when the cells divide.

20 It has always been difficult to get large pieces of DNA into cells. Current methods of introducing DNA into cells include, use of calcium phosphate, electroporation and viral and non-viral vectors. Each of these methods shares the problem that only small pieces of DNA can be put into cells without the DNA being damaged. Large DNA fragments have very low transfection efficiencies or break up when entering cells using current methods. This has made it very difficult in the past to get sufficiently large pieces of DNA into cells to form a HAC.

25

If the DNA used to form a HAC is delivered by chemical or physical transfection, in theory there is no limit to the size of DNA. However, in practice, the larger the vector, the lower the delivery efficiency (for example an 8 kb vector can be delivered to HT1080 cells by lipofection up to 10^5 times more efficiently than a 400 kb vector). Furthermore, assembling a single vector containing at least 40 kb

30

of alpha satellite and more than 200kb of other DNA is technically difficult and time consuming.

In order to use a HAC for gene therapy the efficiency of delivery of existing methods needs to be improved.

The DNA may be delivered to the cell using two or more viral vectors which may be selected from vectors based on any virus, for example an HSV virus, a retrovirus, an adeno-associated virus (aav), an adenovirus, a vaccinia virus or a plant virus. In one embodiment the two or more viral vectors may be based on different viruses. In another embodiment the two or more vectors may be based on the same virus. The advantage of delivering the DNA using viral vectors is that they deliver DNA efficiently to the cell. Two or more viral vectors may be introduced into the cell simultaneously and therefore more DNA can be introduced into the cell to form a HAC.

Using the HSV-1 amplicon delivery system, a single HSV-1 amplicon vector can deliver up to 150kb of DNA. Using more than one HSV-1 amplicon vector introducing different pieces of DNA simultaneously, multiples of 150kb can be delivered with high efficiency to any cell type. Using standard delivery methods, DNA as large as 400kb can be delivered but with very low efficiency so that HAC formation has previously only been achieved in HT1080 fibrosarcoma cells.

In the present invention two or more HSV-1 amplicon vectors, preferably HSV-1 amplicon vectors, that comprise all of the DNA necessary to form a HAC are used to transduce the cells with the DNA needed to form a HAC and the HAC forms inside the transduced cell.

The requirements for the HSV-1 amplicon system are the HSV-1 origin of replication OriS and the packaging signal pac, in addition to α -satellite DNA required for HAC formation, for example the core α -satellite DNA from human chromosome 17 shown in figure 12. The DNA required to form a HAC, including the introduced DNA and at least 40 kb of α -satellite DNA, may be retrofitted in *E. coli* by loxP-Cre recombination with a smaller plasmid carrying the HSV-1

origin of replication, OriS, and the HSV-1 packaging signal, Pac. The vector DNA may be extracted from the *E. coli* bacteria by alkaline lysis, and lipofected using commercial reagents into the packaging cell line Vero 2-2, along with the fHSV Δ pac Δ 270+ vector and the pEBHICP27 constructs. fHSV Δ pac Δ 270+ vector and pEBHICP27 provide, in trans, the necessary proteins and enzymes to replicate and package the input HAC DNA which contains the OriS and Pac sequences. The input HAC DNA is thus packaged into HSV-1 capsids (these infectious particles are now termed amplicons), which are released following Vero 2-2 cell lysis by ultrasonic disruption. Following centrifugation to concentrate the amplicons, they are applied to the target cells in a small volume of medium for 24 hours. After this period, excess amplicons may be removed by washing the cells with medium, and selection is applied to recover stably transduced clones.

In one embodiment the cells are transduced with one α -satellite DNA sequence and one HAC is formed inside the cell. One α -satellite DNA is needed for each HAC, but any number of introduced DNA sequences may be on the HAC.

The HAC may comprise one or more than one introduced DNA sequence, for example one or more than one gene. The total DNA which forms the HAC (the input HAC DNA), including the α -satellite DNA and the one or more introduced DNA sequences, may be divided between the two or more HSV-1 amplicon vectors. In one embodiment the α -satellite DNA may be on one of the HSV-1 amplicon vectors and the one or more introduced DNA sequences may be on another one or more HSV-1 amplicon vectors. In another embodiment the α -satellite DNA sequence may be divided between the two or more HSV-1 amplicon vectors and the one or more introduced DNA sequences may be divided between the two or more HSV-1 amplicon vectors. All or part of the α -satellite DNA sequence may be on the same HSV-1 amplicon vector as one or more introduced DNA sequences.

In one embodiment an α -satellite DNA sequence may be on one HSV-1 amplicon vector, while one or more introduced DNA sequences may be on another HSV-1 amplicon vector. In another embodiment three, four, five, six or more than six

HSV-1 amplicon vectors may be transduced into the cells and may be introduced simultaneously into one cell. One of the HSV-1 amplicon vectors may contain an α -satellite DNA sequence and the other co-transduced HSV-1 amplicon vectors may comprise one or more introduced DNA sequences. These sequences may
5 combine to form a HAC inside the cell. Preferably one α -satellite DNA sequence is required for each HAC that will be generated.

In one embodiment the at least one α -satellite DNA sequence and the at least one introduced DNA sequence may be divided between three or more, four or more,
10 five or more, six or more, seven or more, eight or more, nine or more or ten or more HSV-1 amplicon vectors.

Each HSV-1 amplicon vector can deliver up to about 150 kb of introduced DNA. Therefore, if a large number of HSV-1 amplicon vectors are used in the present
15 invention more introduced DNA can be delivered into the cell. The introduced DNA from the large number of HSV-1 amplicon vectors may join together in the cell to form one or more HAC. This is advantageous because a HAC with a larger introduced sequence or a larger number of introduced sequences may be formed in the cell.

20 A HAC may be used to express a gene that is lacking in particular cells. They are therefore useful in the treatment of genetic diseases, for example in gene therapy, where a particular gene is not expressed or a particular protein is expressed in cells in a form that does not have the normal activity. Furthermore, HACs can
25 be used as model chromosomes to analyze carcinogenesis and tumour progression; to characterize the effect of compounds and treatments on chromosome structure, protein composition, replication, segregation and behaviour; to study biochemical or enzymatic pathways; to introduce extra copies of a gene to analyse the effects of increased copy number and overexpression of
30 genes of certain biochemical or cellular assays; to introduce tagged copies of different proteins to analyse the assembly or structure of a protein complex; as vectors for vaccinia production and as vectors for induced pluripotent stem cell (iPS) generation.

Other advantages of the use of HACs is that they behave as normal chromosomes in the cells, and assemble both euchromatin and heterochromatin, thus maintaining correct structure and behaviour. Due to their similarity to endogenous chromosomes they respond in a physiological way to cellular/extracellular stimuli and controlling systems. Furthermore, they occupy in the nucleus a specific position, thus ensuring that they are exposed to the correct proteins and/or modifying enzymes.

The cell that is transduced with one or more HSV-1 amplicon vectors may be any type of eukaryotic cell for example a cell derived from a human and mammalian established or immortalized cell lines; embryonic stem cells; adult stem cells; induced pluripotent stem cells (iPS). In one embodiment the cell may be induced to have the characteristics of a stem cell, for example it may be an induced pluripotent stem cell. The cell may be made into an induced pluripotent stem cell before transduction with the HSV-1 amplicon vectors according to the present invention. Alternatively, the cell may be induced to become an induced pluripotent stem cell by including the necessary genes on one or more of the HSV-1 amplicon vectors introduced into the cell in the present invention. All or combinations of the following genes: Oct4, Sox-2, KLF4, c-myc, Nanog and Lin28, SV40 large T antigen, telomerase (hTERT) may be required to induce a cell to become an induced pluripotent stem cell. This is advantageous because a differentiated cell from a particular patient may be transduced according to the present invention, so that it includes a HAC and also is induced to form a pluripotent stem cell at the same time. The cell may then be caused to differentiate into the required cell type and be reintroduced into the patient. This allows cells comprising a HAC to be produced that are immunologically compatible with the patient.

In another embodiment one or more, preferably two or more, HSV-1 amplicon vectors according to the present invention may be transduced into a stem cell. The stem cell may be a human pluripotent stem cell or a human induced pluripotent stem cell. This is advantageous because the stem cell can be induced to differentiate into the required cell type.

In one embodiment the eukaryotic cell transduced with the one or more HSV-1 amplicon vectors is a mammalian cell, preferably a human cell. The mammalian cell may be a stem cell or an induced pluripotent stem cell. The stem cell may be a pluripotent or totipotent stem cell.

5

Preferably the eukaryotic cell is an adult stem cell or a cell derived from an adult, preferably a stem cell derived from brain, bone marrow, peripheral blood, blood vessels, skeletal muscle, skin, teeth, heart, gut, liver, ovarian epithelium, or testis. Suitably the eukaryotic cell is a stem cell derived from umbilical cord blood or menstrual blood. Preferably the eukaryotic cell is a hematopoietic stem cell or a mesenchymal stem cell.

In one embodiment the eukaryotic cell may be a cell from a human or mammalian established or immortalized cell line. The eukaryotic cell may be an embryonic stem cell, an adult stem cell or an induced pluripotent stem cell (iPS). The embryonic stem cell may be a human embryonic stem cell. The embryonic stem cell may be derived from an embryo or from an embryonic stem cell line. Preferably the embryonic stem cell is derived from a commercially available embryonic stem cell line. Preferably the stem cell is not derived from a human embryo. Preferably the stem cell is a human stem cell that is not a human embryonic stem cell.

Human embryonic stem cells (hESc) are an important tool in clinical and basic research. Due to their high replicative lifespan and ability to differentiate into the three different germ layers, they are used for human developmental biology, tissue regeneration, transplant therapies and drug discovery studies. Safe and efficient in vitro genetic manipulation of hESc is an essential step in realizing their full potential for clinical applications.

It is advantageous from the point of view of gene therapy to transform a human stem cell or induced pluripotent stem cell with the HSV-1 amplicon to produce a HAC in the cell because the cell can be differentiated to form the required cell type. The cell may be obtained from a subject and, once the HAC is in place in the cell, the cell may be reintroduced into the subject where inserted genes may

be expressed from the HAC in the cells. Alternatively, the HAC may be introduced in the cells *in vivo*.

It is possible to co-transduce various cell-types, including stem cell lines, with at least two different HSV-1 amplicon vectors. This is advantageous because it allows a larger amount of DNA to be introduced into a cell at one time. The cells can be transduced with a number of HSV-1 amplicon vectors each carrying part of the DNA needed to form a HAC. Once all of the vectors are inside the cell the DNA in the different HSV-1 amplicon vectors can combine to form a HAC.

10

The transduction of cells may be done *in vitro* and the cells introduced into a patient once a HAC has been formed. Cells with the correctly formed HAC can be selected *in vitro* before introducing the selected cells into a patient.

15 According to another aspect the present invention provides a method for providing a human cell comprising a human artificial chromosome (HAC) comprising the steps of:

providing a human pluripotent stem cell or an induced human pluripotent stem cell;

20 transducing the human stem cell with at least two different HSV-1 amplicon vectors wherein one HSV-1 amplicon vector comprises an α -satellite DNA sequence and the other HSV-1 amplicon vector comprises an introduced DNA sequence, and wherein the α -satellite DNA and at least one introduced DNA sequence are able to form a HAC inside the cell.

25

The human stem cell may be transduced with a further one or more HSV-1 amplicon vectors which comprise introduced DNA sequence and form part of the HAC inside the cell.

30 In a further aspect the present invention provides a composition comprising two or more HSV-1 amplicon vectors as defined in any one of the aspects of the invention for use in medicine.

Preferably the composition comprises one or more α -satellite DNA sequences and one or more introduced DNA sequences comprised of two or more, preferably three, four, five, six, seven, eight, nine or ten or more HSV-1 amplicon vectors. The composition may also comprise appropriate solvents, diluents, excipients or
5 carriers.

The composition may comprise two or more HSV-1 amplicon vectors as defined in any aspect of the invention for use in the treatment of diseases preventable or treatable by introducing introduced DNA sequences into cells. For example, the
10 compositions and methods of the present invention may be used for the treatment of diseases due to the absence of large genes or genomic regions (for example Duchenne muscular dystrophy; hemophilia; DNA repair diseases) and for the treatment of diseases due to microdeletions or uniparental disomies (for example Prader-Willi syndrome, Angelman syndrome, Beckwith-Wiedemann syndrome).
15 The compositions and methods of the present invention may be used to increase the expression of certain genes in cells that have regulatory mutations or variants which decrease normal gene expression. Furthermore HACs may be useful to efficiently deliver large interfering RNAs, to prevent the expression of large dominant genes.

20 In a still further aspect the present invention provides a eukaryotic cell as defined in any one of the preceding aspects for use in medicine. For example the present invention provides a eukaryotic cell, preferably a mammalian cell, more preferably a human cell, such as a stem cell, for example a human pluripotent
25 stem cell, an induced pluripotent stem cell, an embryonic stem cell, a human adult stem cell or a cell derived from an adult, preferably a stem cell derived from brain, bone marrow, peripheral blood, blood vessels, skeletal muscle, skin, teeth, heart, gut, liver, ovarian epithelium, or testis, suitably a stem cell derived from, umbilical cord blood or menstrual blood or a hematopoietic stem cell, or a
30 mesenchymal stem cell comprising a HAC for use in medicine.

In another aspect the present invention provides a eukaryotic cell provided by the method according to any one of the preceding aspects for use in medicine,

preferably for use in the treatment of diseases preventable or treatable by introducing introduced DNA sequences into cells.

In a yet further aspect the present invention provides a method of treating
5 diseases treatable or preventable by introducing introduced DNA sequences into eukaryotic cells, the method comprising:

transducing the eukaryotic cell with at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two HSV-1 amplicon vectors, wherein the at least one α -satellite DNA and at least one introduced DNA
10 sequence are able to form one or more HACs inside the cell.

The method may further comprise the step of administering the transduced eukaryotic cells to a subject, such as a human. More preferably a method of treating diseases treatable or preventable by introducing introduced DNA
15 sequences into cells, the method comprising administering at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two HSV-1 amplicon vectors to the person, wherein the α -satellite DNA and at least one introduced DNA sequence are able to form one or more HACs inside the cells of the subject.

20

In a further aspect the present invention provides a kit comprising at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two HSV-1 amplicon vectors, wherein the α -satellite DNA and the at least one introduced DNA sequence are able to form one or more HACs inside a eukaryotic
25 cell.

The skilled person will appreciate that all preferred features of the invention described with reference to only some aspects of the invention can be applied to all aspects of the invention.

30

There now follows by way of example only a detailed description of the present invention with reference to the accompanying drawings, in which;

Figure 1 shows a diagram of the assembly of HAC vectors into HSV-1 particles;

Figure 2 shows a diagram of the alternative outcomes of double infection;

Figure 3 shows co-transduction of two vectors, one expressing the RFP gene (red) and one the GFP gene (green). Cells co-infected by both amplicon types are yellow (arrows).

Figure 4 shows (a): a schematic representation of vectors pHGNeo4 and p α 40 not drawn to scale. (b): GFP expression in HUES-2, HUES-10 and HT1080 cells 24 hours after transduction with pHGNeo4 (left) and p α 40 (right) amplicons.

Figure 5 shows a HAC analysis. (a): FISH analysis with a 17 α DNA probe (green), and a vector probe (red). The chromosomes are counterstained in DAPI, blue. The HAC are identified by yellow arrows. The insets show DAPI staining only (HAC, red arrows), in black and white. (b): Fibre FISH on HAC clones, with 17 α DNA (green) and vector DNA (red) probes. C: ImmunoFISH with anti-CENP C antibody (green), and 17 α HAC probe (red)

Figure 6 shows a pluripotency analysis. (a): RT-PCR with primers specific for pluripotency genes at time-points zero, 30, 60 and 90 days in culture (0, 30, 60, 90) in the presence (on) or absence (off) of selection. (RT-) control PCR performed in the absence of the reverse transcriptase, using GAPDH primers, (-) no template control. (b): Immuno-staining with pluripotency markers antibodies (red). The cells are counterstained with DAPI, blue. (c): RT-PCR analysis of germinal layer transcripts on cDNA from in vitro differentiated MM40.2 and SK40.19 cells, with primers specific for pluripotency (Oct4, Sox2, Nanog), endoderm (α FP, HFN3 α , α 1AT), mesoderm (GATA-2), epidermis (CK-14, CK-5, HSK), and neuronal markers (Pax6). U: undifferentiated cells, D: differentiated cells. (RT-): control PCR performed in the absence of the reverse transcriptase, using GAPDH primers. (-) no template control. (d): Neuronal differentiated cells stained with anti- β III tubulin antibody (red).

Figure 7 shows a teratoma assay. (a): Haematoxylin/eosin stained MM40.2-teratoma derived sections. (b): immunostaining with antibodies for ectoderm

(β III-tubulin, green), mesoderm (α -actinin, red) and endoderm α -feto protein, red) markers.

Figure 8 shows (a): Schematic representation of the p α 60 and p α 100 vectors, not drawn to scale, and GFP expression in HUES-2, and HT1080 cells 24 hours after transduction with p α 60 (left) and p α 100 (right) amplicons. (b): FISH on metaphase spreads from HUES-2 HAC clone MM60.1 and MM100.3. In green, a 17 α DNA probe, which also recognized the endogenous chromosomes 17; in red, a vector probe, which is HAC specific. The chromosomes are counterstained in DAPI, blue. The HAC are identified by yellow arrows. The small insets show the HAC DAPI staining only (red arrows), as black and white images.

Figure 9 shows FISH analysis on neuronal differentiated cells from MM40.2 and SK40.19, with a HAC specific probe (green signal) to identify HAC (yellow arrows) in cells stained with anti- β III tubulin antibody (red). The nuclei are counterstained in DAPI (blue).

Figure 10 shows real time qPCR analysis of HAC frequency in MM40.2 differentiated cells, using HAC specific primers.

20

Figure 11 shows heterogeneity of GFP expression in MM40.2 EB derived cells. Left, bright-field, right, GFP.

Figure 12 shows the sequence of the core alpha satellite DNA from human chromosome 17 (Seq ID No: 1).

25

Figure 13 shows the backbone sequence used in the vectors p α 40 and p α 100. The backbone sequence is all of the sequence of the vector excluding the α -satellite DNA sequence. The backbone sequence is the same for vectors p α 40 and p α 100 but p α 40 has 40 kb of α -satellite DNA while p α 100 has 100 kb of α -satellite DNA (Seq ID No: 2).

30

Figure 14 shows the backbone sequence of the vector p α 60. The backbone sequence is all of the sequence of the vector excluding the α -satellite DNA sequence. p α 60 has 60 kb of α -satellite DNA (Seq ID No: 4).

- 5 **Figure 15** shows the sequence of the vector pHG-Neo4 (Seq ID No: 4).

Figure 16 illustrates the efficiency of transduction in hESc, iPSc and HT1080, compared to the control cell line G16-9.

- 10 **Figure 17** shows the Dual transduction in iPS cells. Left panel: cells expressing GFP (green). Mid panel: cells expressing RFP (red). Right panel: merged picture. Cells expressing both GFP and RFP are circled. If the GFP and RFP are expressed at the same level, the cells appear yellow (bottom circle). If the GFP fluorescent protein is expressed more than RFP, the cells will appear mostly green (top
15 circle).

HAC formation by HSV-1 amplicon dual transduction

- The HSV-1 amplicon system is highly efficient at delivering DNA and forming a HAC in a variety of cell types. However, the size of the insert DNA is limited to
20 150 kb by the packaging capacity of the virion. Thus depending on the size of the initial vector, multiple copies of the HAC construct will be assembled as a linear molecule up to the size of 150kb, and packaged into the viral capsid as shown in Figure 1.

- 25 It is essential to include α -satellite DNA (alphoid) in the HAC vector to seed the formation of a new centromere on the nascent HAC. The α -satellite DNA may be about 40 kb in length and this reduces the remaining capacity for incorporating therapeutic genes and their regulatory regions. To overcome the problem in the present invention cells are simultaneously co-infected with two different HSV-1
30 vectors, one containing at least the α -satellite DNA and the other containing at least introduced DNA sequence, this may be a gene. The two vectors when transduced into a host cell can generate a single HAC containing both HSV-1 constructs, and thus carrying the introduced DNA sequence, gene of interest and the α -satellite DNA. A series of experiments were set up, as outlined in Figure 2,

in each case using a vector carrying α -satellite DNA and the red fluorescent protein gene, RFP, as a marker, and a second vector carrying the green fluorescent protein gene, GFP, and either a selectable marker gene or the HPRT minigene.

- 5 Several independent experiments were conducted, using the HSV-1 amplicon at different multiplicities of infection (MOI), to transduce human HT1080 cells. In all of the experiments, the two independent amplicons were able to co-infect the same cell, with various efficiencies, as estimated by counting the number of cells expressing GFP+RFP, compared to those expressing either GFP or RFP only
- 10 (corresponding to cells transduced by one or the other vector) (examples are shown in Figure 3, and reported in Table 1). The efficiency of co-transduction was slightly improved by subjecting the cells to centrifugation at low speed (750g for 45 minutes) (Table 1).
- 15 **Table 1:** examples of transduction efficiency in two different HSV-1 delivery experiments

Experiment 1	Vector 1 (GFP)/ Vector 2 (RFP)	
	MOI 1	MOI 5
Vector 1 GFP pHGHPR	11 %	15 %
Vector 2 RFP 227bsdRFP	4 %	13.3 %
Vector 1+vector 2 GFP+RFP	0.39 %	4.6 %
Experiment 2		
	MOI 1	MOI 5
Vector 1 GFP	18 %	23%
Vector 2 RFP	13 %	6 %
Vector 1+vector 2	2 %	10 %

GFP+RFP		
---------	--	--

Stable clones were recovered in each double transduction experiment, and the presence of a HAC was verified by fluorescence in situ hybridization (FISH) with
5 specific probes on chromosome metaphase spreads.

In the most recent set of experiments, the majority of the recovered clones (19 out of 20) contained a HAC, in percentages ranging from 20% to 70%. In all cases, the HAC was composed of both vectors.

10

In summary, the results show that when two HSV-1 amplicons are used to deliver different HAC constructs to cells, the exogenous DNAs recombine to generate a single HAC. This demonstrates that co-infection can at least double the capacity of the artificial chromosome vectors.

15

Transduction in Human Embryonic Stem Cells.

In this study, input HAC DNA vectors (ranging from 55 to 115kb) containing the marker gene for GFP were delivered into HUES-2 and HUES-10 human embryonic stem cells (hESc) lines following transduction with HSV-1 amplicons.
20 Mitotically stable, gene expressing, functional HACs were generated. The HACs were present in up to 70% of the hESc, and gene expression was maintained in the absence of selection over a period of 60 days and following cell differentiation. No DNA integrated into the hESc genome, in contrast to HT1080 cells, where the HAC DNA frequently integrated into the host chromosomes.
25 More importantly, the HSV-1 HAC hESc retained their pluripotency and differentiation capabilities.

This is the first successful study to establish non-integrating, stable, gene expressing HACs using HSV-1 amplicons in human embryonic stem cells, and in
30 particular in cells of a human embryonic stem cell line, and is a significant step forward for the use of HACs in gene therapy.

The present invention provides an alternative efficient approach, based on direct HAC formation in two hESc lines utilizing HSV-1 amplicons for delivery. HACs can accommodate large introduced DNA sequences containing potentially therapeutic genes along with their regulatory sequences, and have successfully
5 been used as gene transfer vectors to complement genetic deficiencies in human cultured cells. HACs are also composed of heterochromatic and euchromatic regions in a similar pattern to endogenous chromosomes. While the presence of heterochromatin is necessary for the correct segregation of HACs, the euchromatin potentially ensures the prolonged expression of introduced DNA
10 sequences.

In the present invention input HAC DNA vectors were delivered to the HUES-2, HUES-10 and HT1080 cells and to induced pluripotent cells PF19.9, using HSV-1 amplicon mediated transduction. This technique is up to 10^4 times more efficient
15 at delivering large DNA than chemical transduction, in several different cell types.

FISH analysis revealed that HACs were formed in both HUES-2 and HUES-10 cells. This is the first report of *de novo* HAC formation in a karyotypically
20 normal primary human cell line, and is highly significant for developing HACs as gene expression vectors for gene therapy applications. Most importantly, the input HAC DNA vectors never integrated into the hESc genome, compared to HT1080, where integrations were found in most of the clones. It is possible that the presence in hESc of systems actively guarding genome integrity may prevent
25 or reduce the frequency of large vector integration events, thus giving an advantage to cells where the exogenous DNA forms a episomal HAC, in the presence of selective pressure. In this scenario, the smaller vector pHGNeo4 integrated in hES cells, because no other outcome would have allowed the cells to survive selection, as pHGNeo4 is incapable of forming a HAC. On the other hand,
30 HT1080 cells are tumour derived, and lack an efficient control of genome integrity, thus explaining why both HAC formation and integration events were observed.

The hESc HACs were fully stable, and formed an active centromere. This showed that although the HAC were not detected in 100% of the cells, it was not the result of instability. It is possible that the cloning procedure for hESc did not produce pure clones (hESc have a low survival rate as single cells, even when
5 ROCK inhibitor Y-27632 is used), and hence the HAC frequency in the cells was lower than 100%, as different sublines may have been present within the clones.

The analysis of HAC gene expression showed that the RNA levels of GFP reporter gene on or off selection were different in the clones characterized,
10 containing either HAC or integrated pHGNeo4 DNA. This confirmed the existence of clonal variability between different lines, possibly due to epigenetic effects. In one of the clones, MM40.2, GFP expression was highly stable, and did not change over prolonged time in culture, both on and off selection. Furthermore, the reporter gene expression was maintained following MM40.2
15 differentiation, although heterogeneity was observed in the GFP levels among the differentiated cells. The process of *de novo* HAC formation generally results in the multimerization of the input DNA, as shown by the alternate pattern of vector and alpha satellite signals observed in the FISH on chromatin fibres, and the possibility that more than one copy of the introduced DNA sequence is present on
20 the HAC. This ensures that copies of the introduced DNA sequence will be localized away from the centromeric heterochromatic area, and thus escape potential silencing. The spreading of heterochromatin, which is a stochastic event, may explain the variability observed in the GFP reporter gene expression following differentiation in clone MM40.2, and in prolonged culture in clone
25 SK40.19.

Importantly, neither the HSV-1 transduction nor the HAC formation led to a loss of pluripotency in the HUES-2 or HUES-10 cells, as suggested by the staining with hESc specific markers, expression of three germinal layer markers in EB-
30 derived cells, by the differentiation into neuronal types, and by MM40.2 teratoma formation. The HAC was present in the neuronal differentiated cells in both MM40.2 and SK40.19. The HAC frequency in MM40.2 neuronal cells was slightly lower than in the undifferentiated cells, yet the frequency in SK40.19 remained unchanged following differentiation.

The HSV-1 amplicon particles package DNAs up to 150kb. This is far larger than the size which can be accommodated in lentiviral, AAV or adenovirus based vectors. In this study, the p α 40 containing the shortest tract of 17 alpha satellite DNA (40 kb) was the most efficient vector at HAC formation in hESc, and since its total size is approximately 55 kb, it can accommodate a large introduced DNA sequence region containing gene loci of up to 100 kb. HSV-1 HAC replicate at the same rate as the endogenous chromosome, making the gene dosage level more easily controllable.

10

In summary, high capacity, gene-expressing HAC formation was observed in two hES cell lines, using a high efficiency delivery method based on the HSV-1 amplicon technology. The HACs obtained were stable and sustained long-term gene expression. The lack of integrated DNA in hESc following HSV-1 delivery is an important and exciting finding. These findings also suggest that HAC vectors may be used with induced pluripotent stem (iPS) cells.

15

RESULTS

Vector Construction

To generate HSV-1 based input HAC DNA vectors that were highly proficient at HAC formation, the BAC hBAC495J24 (containing 220 kb of chromosome 17 core α DNA) used in a previous study to construct an efficient input HAC DNA vector (pJM2256), was modified. As hBAC495J24 was approximately 70 kb larger than the 150 kb packaging limit of HSV-1. The aim was to reduce it while retaining its HAC-forming properties. Three derivatives of the BAC hBAC495J24 were generated, two of which arose spontaneously (containing 40 and 100 kb of 17 α DNA) during culture, and the third derivative (containing 60 kb of 17 α DNA) by utilizing the RED/ET recombination system. All three derivatives were modified by LoxP-Cre recombination with pHGNeo4 to include the essential HSV-1 elements and reporter genes (GFP), thereby generating p α 40 (55 kb, including 40 kb of 17 α DNA), p α 60 (75 kb, including 60 kb of 17 α DNA), and p α 100 (115 kb, including 100 kb of 17 α DNA) (Figure 4a and Figure 8a).

20
25
30

Introduction of Input HAC DNA Vectors into hES cells

Among the three HSV-1 HAC input DNA vectors that were generated, p α 40 (40 kb of 17 α DNA) was the most useful, as its size allowed up to 100 kb of introduced DNA sequence to be accommodated within the vector for packaging into HSV-1 amplicons (limit capacity of 150 kb). For this reason, the HAC forming efficiency was investigated by delivering p α 40 to HUES-2 (at passage 40) or HUES-10 (at passage 29) by HSV-1 amplicon transduction at multiplicity of infection (MOI) 2. In a parallel control experiment pHGNeo4 amplicons were also delivered to both cell lines. The other 17 α input HAC DNA vectors, p α 60 and p α 100, were delivered only to the HUES-2 line, in the same conditions described above.

The efficiency (%) of transduction, summarized in Table 2, was determined after 24 hours by FACS or counting GFP expressing cells. The results are also shown in Figure 4b and Figure 8b. In HUES-2, the average transduction efficiency was approximately 40% for both p α 40 and the control vector pHGNeo4. The two larger vectors, p α 100 and p α 60 were delivered to HUES-2 with an efficiency of 16% and 20% respectively (Table 2 and Figure 8b). In HUES-10, the delivery efficiency was 27% for both pHGNeo4 and p α 40 (Table 2, Figure 4b).

In parallel control experiments, the input HAC DNA amplicon vectors were delivered by HSV-1 mediated transduction to HT1080 cells, which efficiently form HAC. The delivery efficiency of the input HAC DNA vectors was similar to that observed in hESc (Table 2, Figure 4b and 8b).

Table 2. Average efficiency of HSV-1 amplicon transduction at MOI 2, and HAC formation in HUES-2, HUES-10 and HT1080 cells. NA, not applicable.

Cell Line	Input HAC DNA Vector	Efficiency of Transduction	No. of Clones Analyzed	HAC Positive Clones	% of HAC/Cells
HUES-2	p α 40	40%	10	5	10-70
	p α 60	20%	9	5	10-25
	p α 100	16%	7	3	10-20
	pHGNeo4	40%	5	NA	NA
	p α 40	27%	5	5	35-50

HUES-10	pHGNeo4	27%	1	NA	NA
HT1080	p α 40	19%	10	6	5-30
	p α 60	34%	3	1	20
	p α 100	25%	4	1	15
	pHGNeo4	28%	3	NA	NA

Viability of hESc Following Transduction with HSV-1 Amplicons

In addition to the experiments outlined above, to determine if the HSV-1 amplicon transduction affected hESc viability, 2.5×10^5 HUES-2 cells were transduced with pHGNeo4 amplicons at MOI 1, 2 and 5. The average efficiency of transduction was determined after 24 hours by FACS or counting GFP expressing cells, and found to be approximately 27% for MOI 1, 42% for MOI 2, and 48% for MOI 5. Since transduction efficiencies were similar at MOI 2 and 5, the HUES-2 cell lines probably reached transduction saturation in this range.

The cells were monitored for six days post transduction and the average growth rate was calculated by measuring the rate of population increase divided by the initial number of cells, and compared to that of an untreated control. The growth rate and morphology of HUES-2 were not affected post HSV-1 amplicon transduction, with a 5% reduction in viability, detected only for MOI 1.

Generation and Analysis of Stable Clones in hES Cells

Overall, 10 HUES-2 and 5 HUES-10 clones were isolated following G418 selection, derived from the p α 40 transduction (Table 2). In addition clones 9 and 7 were isolated from p α 60 and p α 100 respectively following transduction into HUES-2 cells. The stable clone formation efficiency of HSV-1 transduction was relatively high for both hESc lines, at 10^{-4} , as calculated by the ratio between the number of stable clones and GFP positive cells 24 hours post transduction. Chromosome metaphase spreads were prepared from stable clones, and analyzed by two colour FISH with vector and 17 α DNA probes (Table 2 and Figure 5a and Figure 8b). Following HSV-1 transduction with p α 40, HACs were detected in 5 of the 10 stable clones obtained in HUES-2 cells, and in all 5 clones isolated in HUES-10. The HACs were present in up to 70% of the cells from each clone. In

the HUES-2 cells transduced with either p α 60 or p α 100, HACs were detected in approximately half of the clones, with a frequency of up to 25% of the cells of each clone (Figure 8b). The lower HAC frequency per cell observed with p α 60 or p α 100 indicated that the p α 40 vector was the most efficient at HAC formation in HUES-2 following HSV-1 transduction. Most importantly, in none of the 31 clones (HUES-2 or HUES-10) analysed, the HAC DNA had integrated into the host genome. However, in all of the analyzed hESc clones derived from the HSV-1 delivery of the control vector pHGNeo4, the exogenous DNA had integrated into the host genome.

10

In HT1080, the HSV-1 amplicon transduction successfully generated several hundred clones from each of the 17 α HSV-1 HAC input DNA vectors, and several clones were selected for analysis from p α 40, p α 60 and p α 100 (Table 2). The stable clone formation efficiency of HSV-1 transduction was 5×10^{-3} . Positive clones were analyzed by two colour FISH with vector and 17 α DNA probes (Table 2). The HSV-17 α input DNA vectors generated HACs in most of the clones following transduction, but were present at a lower frequency in cells (up to 30%), and concomitant integrations in the HT1080 genome were found in all of the clones.

20

Karyotypic analysis of hESc HAC clones

Two hESc HAC clones were chosen for further studies. Clone MM40.2 derived from the HSV-1 mediated delivery of p α 40 into HUES-2, and contained a HAC in 70% of the cells. (Figure 5a). Clone SK40.19 was generated by the transduction of p α 40 in HUES-10 cells and contained a HAC in 50% of the analyzed metaphase (Figure 5a). The structure of the HACs present in both clones was analyzed further by FISH on extended chromatin fibres. The HAC contained repeat units of input DNA, arranged in an alternate fashion as seen in previous studies, but the vector backbone DNA was less abundant in clone SK40.19. (Figure 5b).

30

The percentage of GFP positive cells in the two HAC clones was established by FACS analysis, and found to be 33% for SK40.19 and 83% for MM40.2. These values are in concordance with the estimated HAC frequency.

HAC Stability and CENP C Staining

The mitotic stability of the HAC in MM40.2 and SK40.19 was monitored by FISH for 90 days (corresponding to approximately 30 passages) in the absence of selection. In both clones, the HAC frequency did not change significantly over this period, with a daily loss rate of 0.03% for MM40.2 and 0.24% for SK40.19 (calculated by the formula $N_n = N_0 \times (1-R)^n$, where N_0 is the number of metaphase chromosome spreads showing HAC in the cells cultured under selection, N_n is the number of HAC-containing metaphase chromosome spreads after n days of culture in the absence of selection, and R is the daily rate of loss).

To confirm that an active centromere was present on the HAC, cells from the MM40.2 and SK40.19 clones were stained with anti-centromere protein C (CENP C) antibody, coupled to FISH with HAC specific probes. A positive CENP C signal was identified on the HACs in both clones at a similar intensity of those observed on endogenous chromosomes (Figure 5c), thus confirming that an active centromere was present, similar to the HSV-1 HAC in HT1080 cells.

Pluripotency Studies

To determine if the MM40.2 and SK40.19 cells expressed pluripotency epitopes compared to the parental lines, cells from each clone (at approximately passage 70 and 60 respectively) were stained with antibodies against the embryonic stem cell markers Oct4, Sox2, Nanog and TRA-1-60 (Figure 6b). The HAC clones and the parental lines were stained similarly by the antibodies while HT1080 and MEF (mouse embryonic fibroblast) control cells were completely negative. The result indicate that the HACs containing hESC clones expressed embryonic stem cell markers.

The hESC markers expression was sustained over time in both HAC clones, as confirmed by RT-PCR analysis on RNA extracted from the MM40.2 and SK40.19 cells, over a period of 90 days, cultivated in the presence and absence of selection (Figure 6a).

To confirm that the HAC containing clones were pluripotent, differentiation of the three embryonic germ layers was induced through embryoid body formation of MM40.2 and SK40.19. Total RNA was extracted from the pool of differentiated cells and analysed for the presence of endoderm (α Feto-protein, HNF3 α , α 1 anti-trypsin), mesoderm (GATA-2) and ectoderm (CK-5, CK-14, high sulphur keratin, Pax6) specific transcripts. RT-PCR experiments revealed that mRNAs for all the markers were present (Figure 6c), indicating that both HAC clones retained pluripotency.

10 HAC Clones Directed Differentiation

Furthermore, neuronal differentiation was induced in clone MM40.2 and SK40.19 by treatment with medium containing noggin and fibronectin. After 25 days of directed differentiation, the cells were fixed in formaldehyde and stained with anti- β III tubulin antibody, a neuronal cell marker. Neuronal cells were detected that were highly positive for β III tubulin staining, in both HAC clones, ranging approximately between the 18% and the 40% of the treated cells (Figure 6d). On average, the SK40.19 was up to 2 times more efficient at forming neuronal cells, compared to MM40.2. The control cells (untreated MM40.2 and SK40.19; HT1080; MEF) never displayed positive cells. To determine if the HAC was still present in the cells following neuronal differentiation, FISH experiments were conducted on the interphasic cells stained with the anti- β III tubulin antibody. In clone SK40.19 the HAC frequency remained the same (50%) while a positive HAC signal was detected in 44% of MM40.2 the cells (Figure 8).

25 Teratoma Formation Assay

Since the MM40.2 cells contained the highest HAC frequency, teratoma were generated using these cells in immunodeficient mice, as this constituted the most rigorous test of pluripotency for human ES cells. Sub-cutaneous injection into immunodeficient mice generated tumours between the 5th and 7th week post-treatment. The subsequent histological analysis of haematoxylin/eosin stained tumour sections revealed the presence of ectodermal (neural tube), mesodermal (muscle and blood vessels) and endodermal (gut epithelium, alveoli, and glandular epithelium) structures (Figure 7a). Moreover, immunostaining using

suitable antibodies confirmed the presence of all three germinal layers in these sections (Figure 7b) thereby confirming the tumour growth as a teratoma.

As the sections prepared for the immunological staining were not suitable for FISH experiments, the presence of HAC in the teratoma cells was confirmed by Real Time qPCR on total genomic DNA extracted from the teratoma mass. HAC-specific primers (Neo and GFP) were used, and the HAC DNA content in the teratoma cells was normalized against that of the undifferentiated MM40.2 cells (Figure 10). No difference in the abundance of HAC sequences was detected between the undifferentiated MM40.2 and teratoma cells, suggesting that the HAC had not been lost during the in vivo cell replications. A similar analysis was conducted on the neuronal differentiated MM40.2 cells, previously analysed by FISH. The qPCR results closely reflected the FISH quantification of HAC distribution (Figure 10).

15

HAC Gene Expression Analysis

The expression level of the reporter gene for GFP from the MM40.2 and SK40.19 HAC was investigated by Real Time qPCR experiments on cDNA, in cells grown either on or off selection. The HAC gene expression was compared to that of a ubiquitously expressed gene (GAPDH). In SK40.19 the GFP expression was found to be decreased by about 50% after approximately 40 days on or off selection (Table 3). By contrast, in MM40.2, the GFP gene expression level remained constant for a prolonged period of time (60 days) on or off selection (Table 3). To further analyse the reporter gene expression clonal variability, the GFP relative amount in three pHGNeo4 stable clones was measured by Real Time qPCR. Compared to the MM40.2, one clone had approximately 5 times lower levels of GFP, and two had 5 times and 10 times more GFP respectively (Table S1).

Table S1. Analysis by qPCR of the GFP reporter gene expression in pHGNeo4 derived HUES-2 clones. The qPCR -fold difference values are expressed in reference to the MM40.2 clone, using GAPDH as internal control.

Clone	Relative GFP Expression	Standard Deviation
MM40.2	1	0.03
pHG1	0.21	0.13

pHG4	4.8	0.73
pHG6	11.3	1.52

Taken together, these data suggest that stochastic events may have affected the level of expression of the reporter gene in different clones. To confirm that the MM40.2 clone had very stable levels of reporter gene expression, GFP presence was further monitored by FACS over the course of 90 days. The percentage of GFP expressing cells remained relatively constant and ranged from 81-89%. The fold increase in the geometric mean of fluorescence intensity compared to the untreated control ranged from ~8-12, indicating that the average GFP expression per cell also remained relatively constant.

10

Table 3. Analysis by qPCR of the GFP reporter gene expression in clones MM40.2 and SK40.19, in the presence or absence of selection. The qPCR -fold difference values are expressed in reference to the day zero time-point, using GAPDH as internal control. SD, standard deviation. NA, not analyzed

		Day 0 (SD)	Day 30 (SD)	Day 60 (SD)
MM40.2	On selection	1.00 (0.03)	1.23 (0.03)	1.34 (0.05)
	Off selection	1.00 (0.03)	1.43 (0.15)	1.49 (0.07)
SK40.19	On selection	1.00 (0.02)	0.40* (0.01)	NA
	Off selection	1.00 (0.02)	0.52* (0.02)	NA

15 *37 days

Furthermore, the GFP reporter gene expression levels of the MM40.2 clone were characterized by qPCR following differentiation, and compared to the levels present in the undifferentiated MM40.2 parental. The GFP gene was still expressed in both EB derived-, and neuronal differentiated cells, at approximately 60-70% of the level present in the undifferentiated parental (Table 4). The level of GFP expression was consistent with the HAC frequency observed in the neuronal differentiated MM40.2 cells. Based on direct microscopic observation of the EB derived and neuronal cells, the derivative differentiated cells exhibited a heterogeneous GFP expression: while some cells were still highly GFP positive, in others the GFP fluorescence appeared reduced or absent (Figure 11). This

suggests that stochastic events in the early stages of hESc differentiation had an effect in the level of expression of the reporter gene during the later stages.

Table 4. Analysis by qPCR of the GFP reporter gene expression, in differentiated MM40.2 cells. The values are expressed in reference to the undifferentiated parental, using GAPDH as internal control. SD, standard deviation

	Undifferentiated hESc (SD)	EB Derived Cells (SD)	Neuronal Cells (SD)
GFP expression	1.00 (0.03)	0.70 (0.04)	0.60 (0.02)

HSV-1 amplicon dual transduction in human embryonic stem cells and induced pluripotent stem cells

To demonstrate that the simultaneous transduction with two different amplicons is feasible in human stem cells, two vectors, containing either GFP or RFP, were delivered to the human embryonic stem cells (hESc) HUES2 (Cowan et al. 2004 Engl. J. Med., 350, 1353-1366) and the induced pluripotent stem cells (iPSc) DF19.9 (WiCell Research Institute), at MOI 2. In control experiments, HT1080 cells, and the highly transducible human glioma cell line G16-9 (Moralli et al. 2006 EMBO Rep., 7, 911-918) were transduced with the same two vectors. The results, summarized in Figure 16, show that in hESc and iPSc the frequency of dual transduction events is comparable to that observed in the highly transducible cells G16-9, and is approximately 2-3 times higher of that observed in HT1080 cells. Examples are shown in Figure 17.

MATERIALS AND METHODS

Input HAC DNA Vector Assembly – HSV-1 Amplicon Vector Assembly

The pHGNeo4 vector, whose sequence is shown in Figure 15, carries the HSV-1 amplicon origin of replication (Ori) and packaging signal (pac), the reporter GFP gene, under control of the I/E promoter from HSV-1, and the G418 resistance gene (Neo), controlled by the SV40 promoter.

Vectors BAC17 α 40 and BAC17 α 100, containing respectively 40kb and 100kb of a satellite DNA from human chromosome 17 (17 α), were derived by spontaneous deletions of hBAC495J24. PAC17 α 60 was obtained by RED/ET recombination based on a commercial kit: Red/ET BAC subcloning kit, Gene Bridges GmbH, which transferred 60kb of 17 α DNA from hBAC495J24 to pCYPAC2 vector (commercially available from the BACPAC Resource Center, CHORI <http://bacpac.chori.org/pcypac2.htm>). Briefly, the pCYPAC2 vector was used as a template to generate a 9.5kb PCR fragment, using primers containing 50bp homologous tails to the alpha 17 satellite DNA consensus (Table S2). The linear 9.5kb pCYPAC2 PCR product was then transformed into *E. coli* cells containing hBAC495J24, and expressing the RED/ET system proteins.

Table S2. Primers used for PCR and Real Time qPCR amplification.

Primer	Sequence
pCYPAC2 Long F	5'- ATTCGTTGGAAACGGGATAATTTTCAGCTGACTAAACAGAAGCAGT CTCAGGATCCATTGACCCGGAACCTTAATATAAC-3' (Seq ID No: 5)
pCYPAC2 Long R	5'- ATATGCAATTGCAGATTCTGCAGAAAGTGTGTTTCTAAACTGCTAC ATCGGATCCAAAATCATTTAATTGGTGGTGCTGC-3' (Seq ID No: 6)
Vect 1F	5'-TGGTTTGTCCAAACTCATCAA-3'(Seq ID No: 7)
Vect 1R	5'-AGCAAAAACAGGAAGGCAAA-3' (Seq ID No: 8)
Neo7 F	5'-GAGTACGTGCTCGCTCGATG-3' (Seq ID No: 9)
Neo7 R	5'-CTTGCTCCTGCCGAGAAAGT-3' (Seq ID No: 10)
EGFP F	5'-TGCTCAGGTAGTGGTTGTCG-3' (Seq ID No: 11)
EGFP R	5'-AGAACGGCATCAAGGTGAAC-3' (Seq ID No: 12)
GAPDH F	5'- GAGTCAACGGATTTGGTCGT -3' (Seq ID No: 13)
GAPDH R	5'- GACAAGCTTCCCGTTCTCAG -3' (Seq ID No: 14)

The BAC17 α 40, PAC17 α 60, and BAC17 α 100 constructs were then retrofitted with pHGNeo4 by LoxP-Cre recombination (Moralli, D. et al., (2006) Rep., 7, 911-918), to generate three new HSV-1 based HAC vectors: pa40, pa60, and

pa100 respectively. The backbone sequence of vectors pa40 and pa100 is shown in Figure 13. The backbone sequence is all of the sequence of the vector excluding the α -satellite DNA sequence. The backbone sequence is the same for vectors pa40 and pa100 but pa40 has 40 kb of α -satellite DNA while pa100 has 100 kb of α -satellite DNA. Figure 14 shows the backbone sequence of the vector pa60. The backbone sequence is all of the sequence of the vector excluding the α -satellite DNA sequence. pa60 has 60 kb of α -satellite DNA. An example of the core α -satellite DNA is shown in Figure 12. A 40, 60 or 80 kb α -satellite DNA sequence may, for example, be made up from repeats of the sequence shown in Figure 12. Vectors comprising α -satellite DNA may be obtained from the BACPAC Resources Center (BPRC) at the Chori Institute (<http://bacpac.chori.org/>). The α -satellite DNA sequences used in the vectors described in this example derive from vector RP11-495J24, which is available from (BPRC).

15 Cell Culture

The human embryonic stem cell lines HUES-2 and HUES-10 were obtained from Douglas Melton (Harvard University, Cambridge, MA, USA) and grown under license from the UK Stem Cell Steering Committee as described (Cowan, C.A., Klimanskaya, I., *et al.* (2004) *N. Engl. J. Med.*, **350**, 1353-1366.; Karlsson, K.R., *et al.* (2008) *Exp. Hematol.*, **36**, 1167-1175.) on mitomycin C inactivated mouse embryonic fibroblasts (MEF) or SNL76/7 cells. Feeder independent HUES-2 and HUES-10 cells were grown on Matrigel (BD Biosciences) coated wells using the mTeSR medium (STEMCELL Technologies). TrypLE Express (Invitrogen) was used to enzymatically passage the hESc. Cells were maintained and passaged at high densities on Matrigel. To increase single cell survival, ROCK (Rho-associated kinase) inhibitor Y-27632 (Merck Biosciences) was added during each passaging step, at a final concentration of 10 μ M. The HT1080 (ATCC-CCL-121) cells were grown using standard techniques in DMEM medium (Invitrogen), supplemented with 10% FBS and 1% Penicillin/Streptomycin.

30

HSV-1 Amplicon Preparation

HSV-1 amplicons were prepared as described (Moralli, D. *et al.*, (2006) *Rep.*, **7**, 911-918.; Wade-Martins, *et al.* (2001) *Nat. Biotechnol.*, **19**, 1067-1070.). Briefly, the packaging cell line Vero 2-2 was transduced with each input HAC vector

DNA, fHSV Δ pac Δ 270 0+ and pEBHICP27 by Lipofectamine (Invitrogen) and Plus Reagent (Invitrogen). The cells were harvested, sonicated and then amplicons concentrated as described. The pellet was resuspended in 500 μ L of PBS. The titre of the HSV-1 amplicon preparation was determined by transducing the glioma cell line G16-9.

HSV-1 Amplicon Transduction

Transduction of HUES-2 (passage 40), HUES-10 (passage 29) and HT1080 cells with HSV-1 HAC amplicons, was carried out as described (Moralli, D. *et al.*, (2006) *EMBO Rep.*, **7**, 911-918., Wade-Martins, R., *et al.* (2001), *Nat. Biotechnol.*, **19**, 1067-1070). The HUES-2 and HUES-10 on Matrigel coated plates, and the HT1080 cells were seeded at 2.5×10^5 cells per 24 well plate the day before transduction. On the day of transduction, HSV-1 HAC amplicons were inoculated at a multiplicity of infection (MOI) of 1, 2 or 5 in 250 μ L of media (mTeSR or DMEM).

To improve the transduction efficiency, upon addition of the HSV-1 amplicons, the cells were centrifuged under low gravitational forces. The plates were covered with sterile adhesive films, to avoid aerosol escape during centrifugation, and centrifuged at 750 g for 45 minutes. Transient expression was monitored at 24 hours post transduction either under the microscope or by flow cytometry. Three days after HSV-1 HAC transduction, the HUES-2 cells were transferred onto G418 resistant inactivated MEF or SNL-76/7 cells. Two days later, 50 μ g/mL of G418 (Invitrogen) was added as selection. After seven days, individual hESc clones were observed and cells were removed from selection. The clones were allowed to grow for an extra seven days, then each clone was isolated and expanded on inactivated MEF. Upon reaching confluency in a 24 well dish, the cells were transferred to feeder-free growing conditions on Matrigel and mTeSR. The HT1080 cells were selected with 350 μ g/mL of G418.

30

Embryoid Body (EB) Formation and Germ Layer Differentiation

For germ layer differentiation of each hESc line, approximately 3×10^7 cells were used to form 192 uniform sized embryoid bodies (EB) of size $\sim 1.5 \times 10^5$ cells/EB. The hESc were seeded in non-adherent 96 well V-bottom plates (Nunc) in mTeSR

and 10 μ M ROCK inhibitor Y-27632. After 3 days the EBs were released into suspension on non-adherent plates for 8-10 days in DMEM, 20% FBS. The EBs were then plated on Matrigel coated plates and left to adhere and expand for a further 15-20 days.

5

Neuronal Differentiation

Neuronal differentiation was carried out as described (Iacovitti, L. *et al.* (2007) *Brain Res.*, **1127**, 19-25.) with minor modifications. Uniform sized EBs composed of approximately 4000 cells were formed using AggrewellTM400 plates (STEMCELL Technologies) as described by the manufacturer. After 2 days, the EBs were released and left in suspension in non-adherent plates for 5 days in DMEM F/12 (Invitrogen), 1% N2 supplement (Invitrogen), containing human plasma fibronectin (5 μ g/mL) (Sigma) and recombinant human noggin (200 μ g/mL) (RDI/Fitzgerald Industries). The EBs were then transferred onto Matrigel coated plates under the same medium for a further 8 days. Recombinant human bFGF (20ng/mL) (BD Biosciences) was then added to the medium, and the cells were incubated for further 8-10 days for expansion of neuronal rosettes. Neuronal rosettes were lifted using TrypLE Express (Invitrogen) and plated on Matrigel coated slides with the addition of ROCK Y-27632 inhibitor.

20

Teratoma Formation Assay

For the teratoma formation assay, 1x10⁶ Matrigel-grown MM40.2 cells were injected subcutaneously into immunodeficient mice (common gamma-chain^{-/-}, RAG2^{-/-}, C5^{-/-}). The mice were sacrificed between 5-7 weeks after injection and the teratoma was dissected and processed for haematoxylin/eosin staining and immunostaining with the following primary antibodies: anti-tubulin beta III isoform (Tuj1) (ectodermal derivatives), anti-alpha-actinin (mesodermal derivatives), anti-alpha fetoprotein (endodermal derivatives) (all from Millipore). Total genomic DNA was prepared from the teratoma mass by phenol/chloroform extraction. The HAC abundance was estimated by Real Time qPCR analysis of the DNA, using the GFP and GAPDH primers listed in Table S2, with the kit SYBR Green Supermix IQ (Quanta Biosciences), on an ICycler (Bio-Rad) machine. The DNA relative amount was measured using the 2^{- $\Delta\Delta$ Ct} method.

30

Fluorescence-Activated Cell Sorting (FACS)

Cells were fixed in 4% formaldehyde in PBS. Samples were then run through a FACS Calibur and a minimum of 10,000 events was captured for each sample. Acquired data was analyzed using FlowJo 7.6 software. The percentage of GFP
5 expressing cells was calculated by setting the gate at 0.5% for the negative control.

Fluorescence in situ Hybridization (FISH) and Immuno-FISH

Chromosome preparation and FISH analyses were carried out as described
10 (Moralli, D., *et al.*, (2006) *EMBO Rep.*, 7, 911-918; Moralli, D., and Monaco, Z.L. (2009) *PLoS One.*, 4, e4483 10.1371/journal.pone.0004483.; Moralli D. *et al.*, (2010) *Stem Cell Rev.* First published on December 29, 2010, 10.1007/s12015-010-9224-4.). For each experiment, up to 40 metaphases were scored, and the number of HAC containing cells was recorded. For HUES-2
15 karyotype analysis, metaphase spreads were subjected to FISH with whole chromosomes paint probes for chromosomes 12 and 17 (Aquarius Whole Chromosome Paint Probes, Cytocell), according to the manufacturer instructions. The HUES-10 karyotype was analyzed on Affymetrix Cytogenetics Whole-Genome 2.7M Arrays, following the manufacturer instructions.

20

The binding of Centromere Protein C (CENP C) to chromosome metaphase spreads was carried out by immuno-FISH as previously described (Moralli, D. *et al.*, (2006) *Rep.*, 7, 911-918), using an anti-CENP C primary antibody.

25 Cytological preparations were analyzed with an Olympus BX-51 epifluorescence microscope coupled to a JAI CVM4+ CCD camera, with CytoVysion software system (Genetix).

Immunofluorescence Staining of Fixed Cells

30 Actively growing cells were fixed in 2% formaldehyde in PBS. After permeabilization in PBS, 0.1% Triton X-100 the following antibodies were used: mouse-anti-TRA-1-60 (Abcam); rabbit-anti-Oct4 (Abcam); rabbit-anti-Nanog (Abcam); rabbit-anti-Sox2 (Abcam); mouse-anti- β III tubulin (R&D Systems),

followed by TRITC conjugated anti-rabbit or anti-mouse antibodies (Molecular Probes, Invitrogen).

5 The cells were analyzed with a wide-field inverted Nikon TE2000U fluorescence microscope. Images were acquired using the IPLab software, and pseudo-coloured using Adobe Photoshop.

RNA Preparation and Analysis

10 Total RNA was extracted from approximately 6×10^6 cells using the RNeasy kit (Qiagen), following the manufacturer's instructions. The RNA was treated with DNase I (Qiagen) and reverse transcribed into cDNA, using the RETROScript system (Ambion), with random decamer primers. Gene expression levels were quantified by Real Time qPCR analysis of the cDNA, using the GFP and GAPDH primers listed in Table S2, as described for the introduced DNA sequence
15 analysis. The RNA relative amount was measured using the $2^{-\Delta\Delta C_t}$ method.

CLAIMS

1. A method for providing a eukaryotic cell comprising a human artificial chromosome (HAC), the method comprising the step of:
5 transducing the eukaryotic cell with at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two viral vectors, wherein the at least one α -satellite DNA and at least one introduced DNA sequence are able to form one or more HACs inside the cell.
- 10 2. The method according to claim 1 wherein the at least two viral vectors are both HSV-1 amplicon vectors.
3. A method according to claim 1 or claim 2, wherein the cells are transduced with one α -satellite DNA sequence and one HAC is formed inside the cell.
- 15 4. A method according to any one of claims 1 to 3, wherein the at least one α -satellite DNA sequence and the at least one introduced DNA sequence are comprised on three or more vectors.
- 20 5. A method according to any one of claims 1 to 4 wherein the α -satellite DNA sequence and the at least one introduced DNA sequence are on different vectors.
6. A method according to any one of claims 1 to 5, wherein the at least two vectors also comprise DNA sequences for inducing the cell to become an induced pluripotent
25 stem cell.
7. The method according to any one of claims 1 to 6, wherein the eukaryotic cell is a human cell, preferably a human stem cell or a human induced pluripotent stem cell.
- 30 8. A method according to any one of the preceding claims wherein the cell is a human stem cell.
9. A method for providing a human cell comprising a human artificial chromosome (HAC) comprising the steps of:
35 providing a human pluripotent stem cell or an induced human pluripotent stem cell;

transducing the human stem cell with two HSV-1 amplicon vectors wherein one HSV-1 amplicon vector comprises an α -satellite DNA sequence and the other HSV-1 amplicon vector comprises an introduced DNA sequence, and wherein the α -satellite DNA and at least one introduced DNA sequence are able to form a HAC inside the cell.

10. A method according to claim 9 wherein the human stem cell is transduced with a further one or more HSV-1 amplicon vectors which comprise introduced DNA sequence and form part of the HAC inside the cell.

10

11. A composition comprising two or more viral vectors, preferably HSV-1 amplicon vectors as defined in any one of the preceding claims for use in medicine.

12. A composition comprising at least two viral vectors, preferably at least two HSV-1 amplicon vectors as defined in any one of claims 1 to 10 for use in the treatment of diseases preventable or treatable by introducing introduced DNA sequences into cells.

15

13. A eukaryotic cell as defined in any one of claims 1 to 10 for use in medicine.

14. A eukaryotic cell provided by the method according to any one of claims 1 to 10 for use in medicine.

20

15. A eukaryotic cell provided by the method according to any one of claims 1 to 10 for use in the treatment of diseases preventable or treatable by introducing introduced DNA sequences into cells.

25

16. A method of treating diseases treatable or preventable by introducing introduced DNA sequences into cells, the method comprising:
transducing the eukaryotic cell with at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two viral vectors, preferably at least two HSV-1 amplicon vectors, wherein the α -satellite DNA and at least one introduced DNA sequence are able to form one or more HACs inside the cell.

30

17. A method according to claim 16 further comprising the step of administering the transduced eukaryotic cells to a person.

35

18. A method of treating diseases treatable or preventable by introducing introduced DNA sequences into cells, the method comprising administering at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two viral vectors, preferably at least two HSV-1 amplicon vectors to the person, wherein the at least one α -satellite DNA and at least one introduced DNA sequence are able to form one or more HACs inside the cells of the person.
19. A kit comprising at least one α -satellite DNA and at least one introduced DNA sequence comprised on at least two viral vectors, preferably at least two HSV-1 amplicon vectors, wherein the at least one α -satellite DNA and the at least one introduced DNA sequence are able to form one or more HACs inside eukaryotic cells.
20. A method as described herein with reference to the figures.
21. A composition as described herein with reference to the figures.
22. A eukaryotic cell as described herein with reference to the figures.
23. A kit as described herein with reference to the figures.

1/24

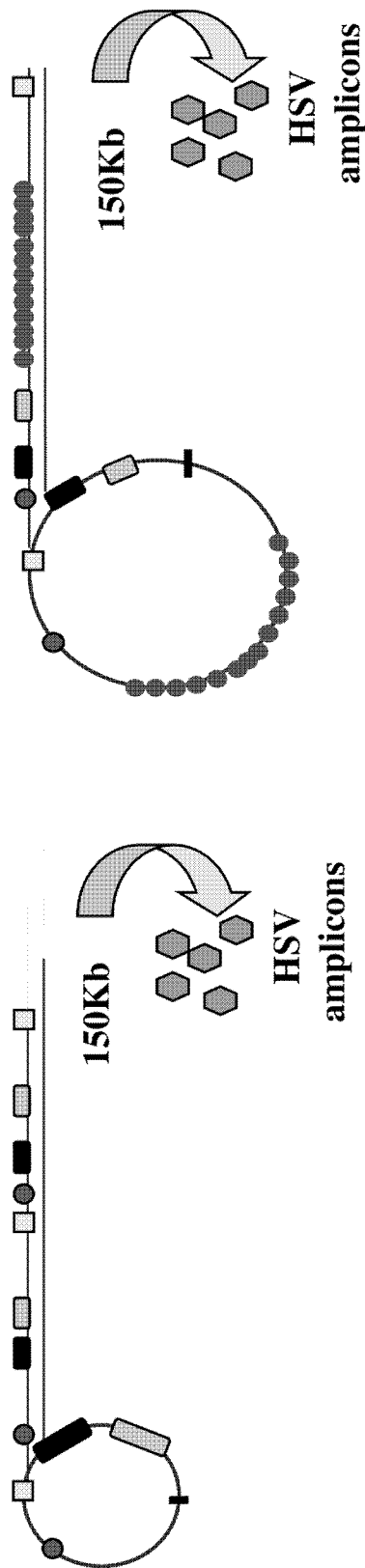


Fig. 1

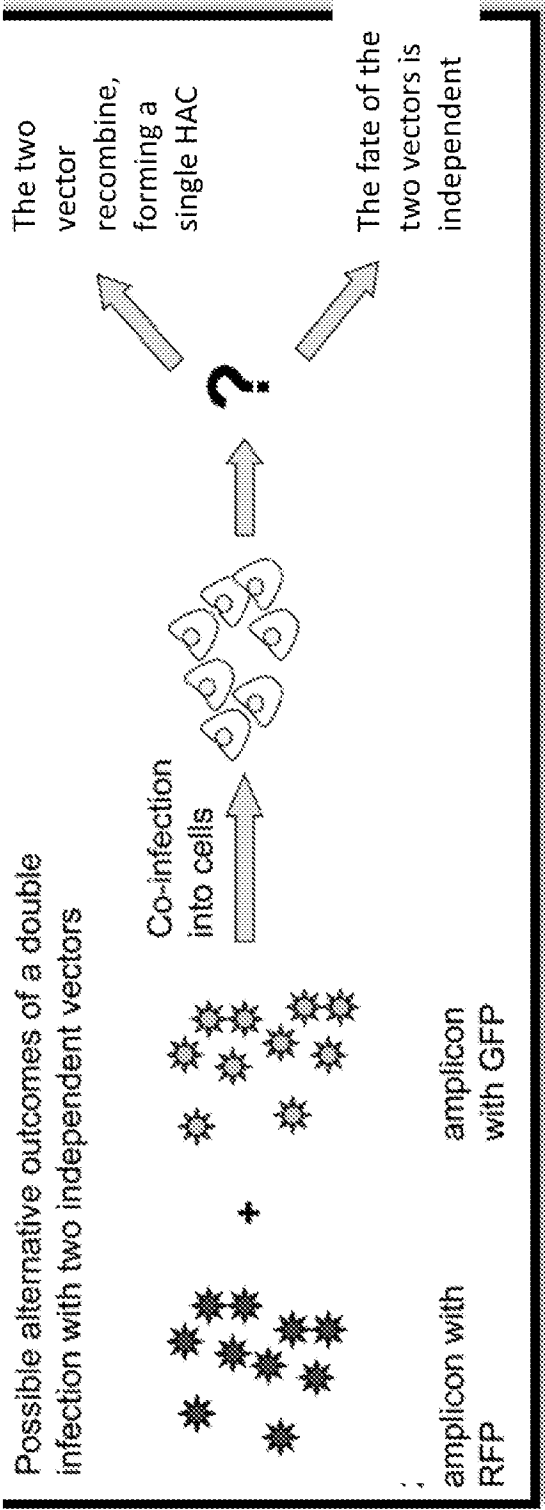


Fig. 2

3/24

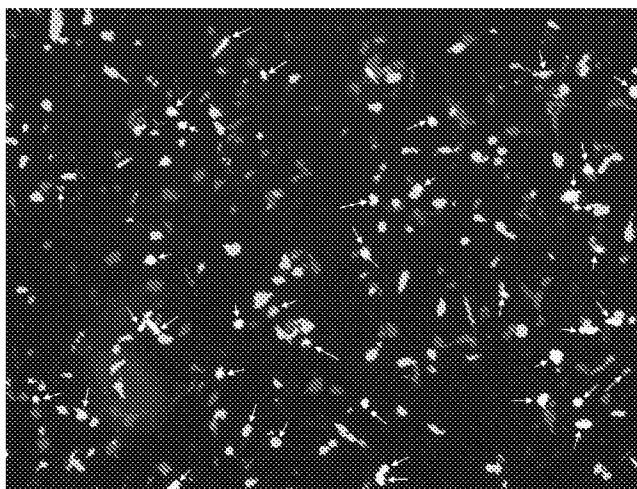
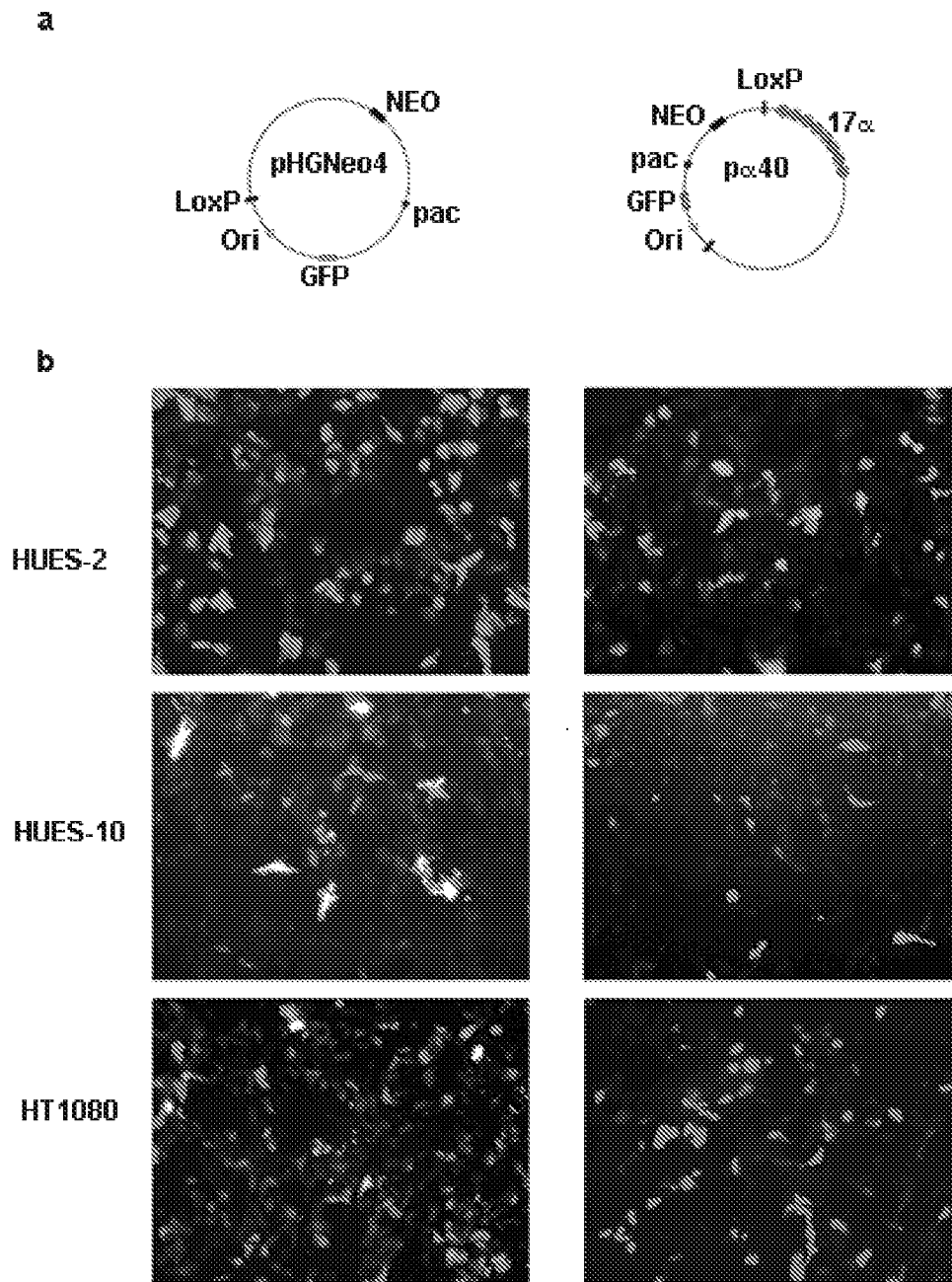
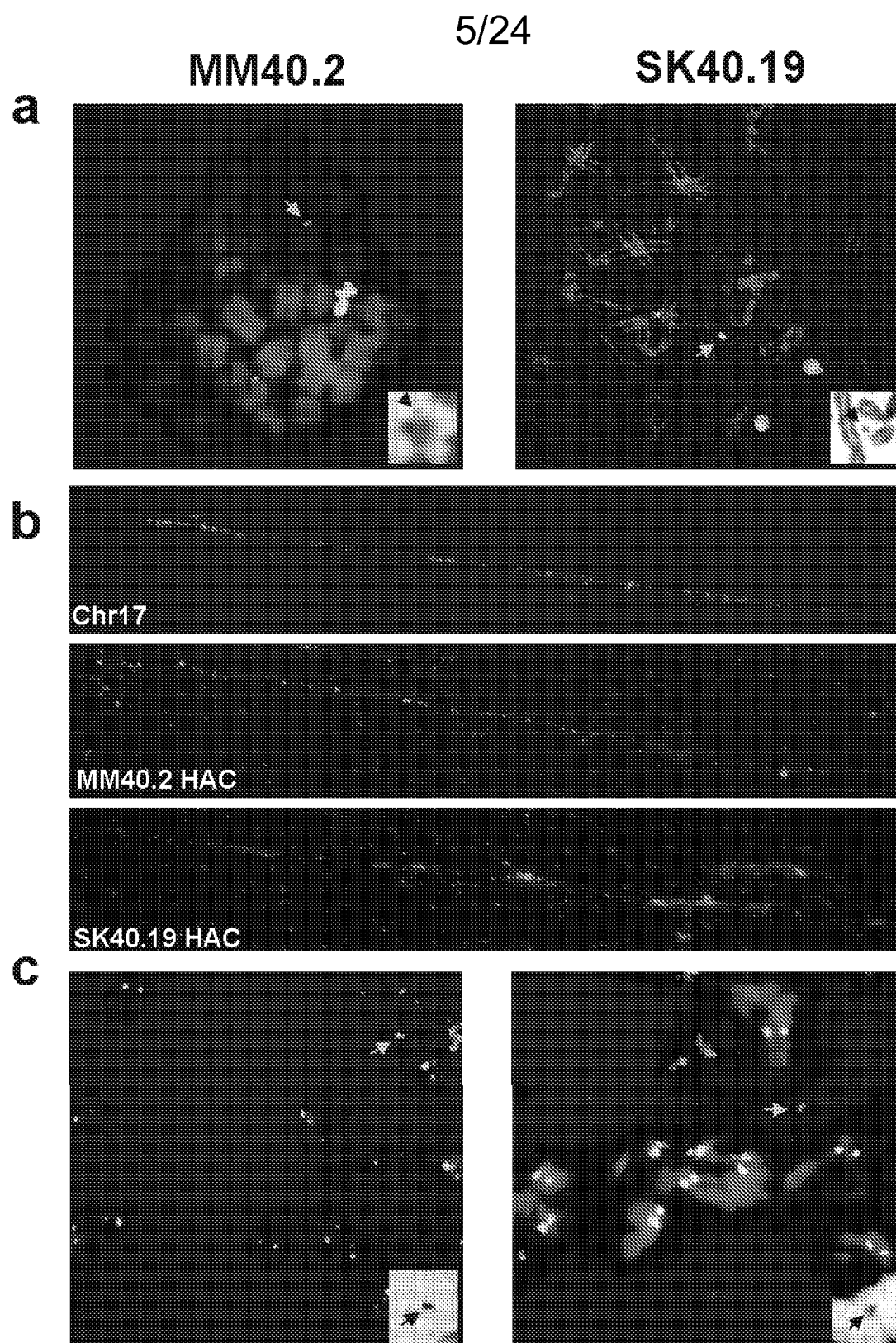


Fig. 3

4/24

**Fig. 4**

**Fig. 5**

6/24

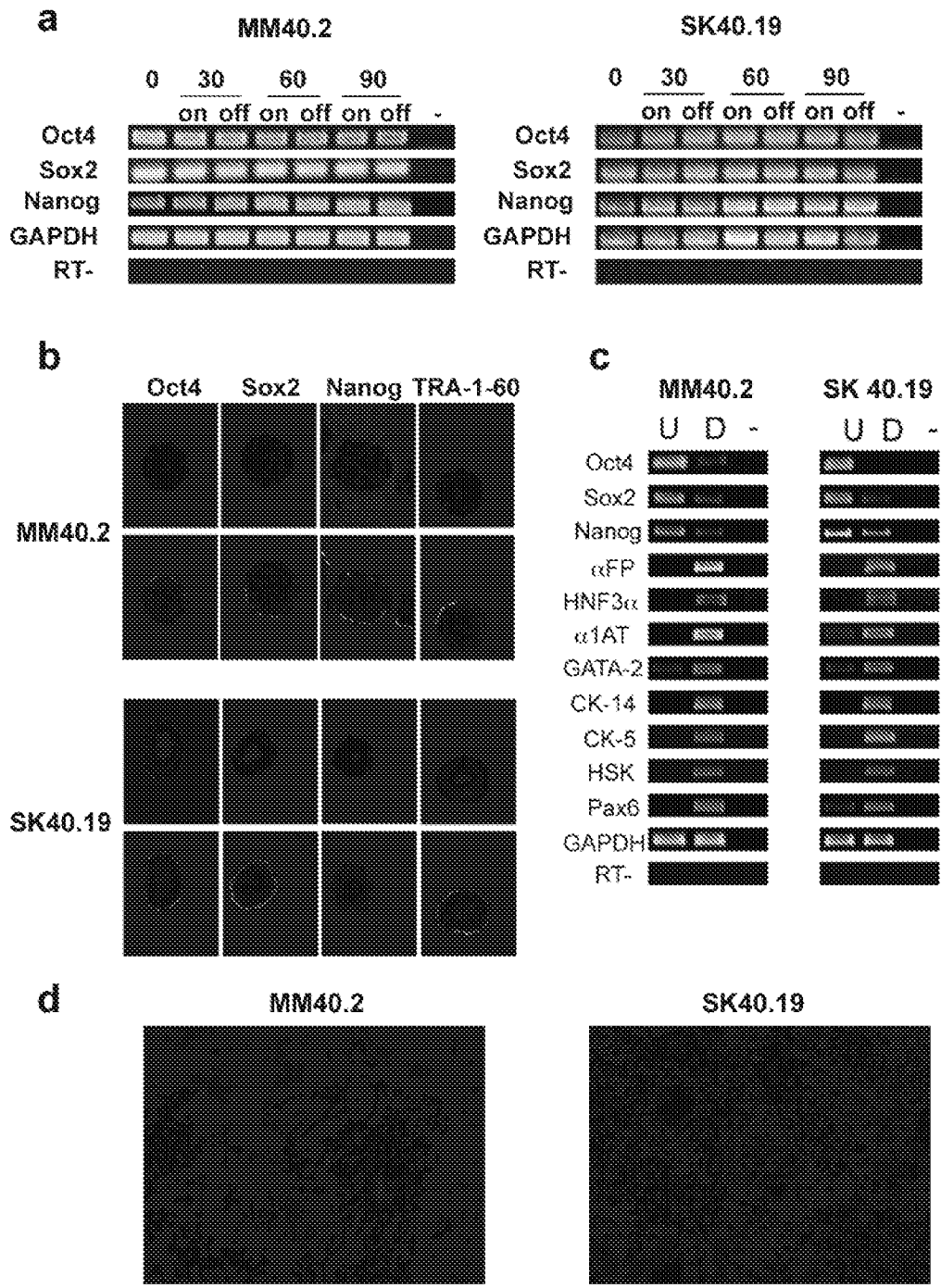


Fig. 6

7/24

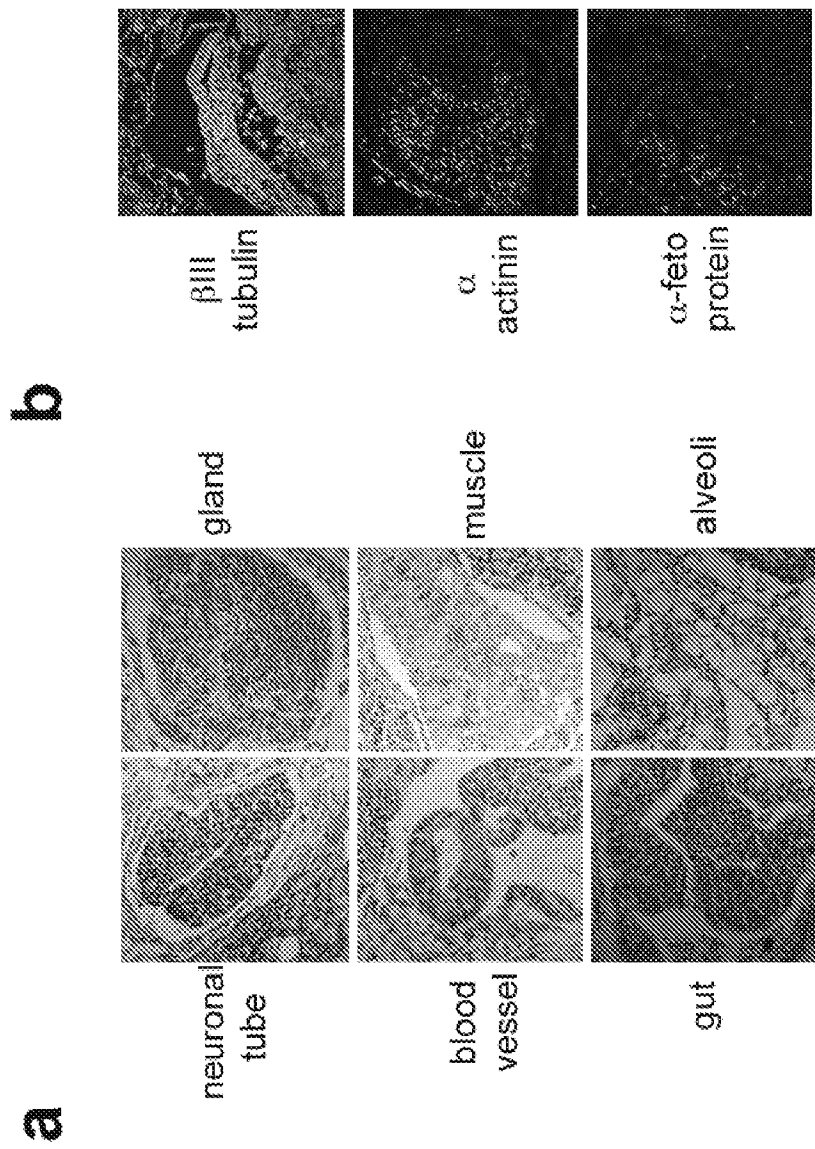
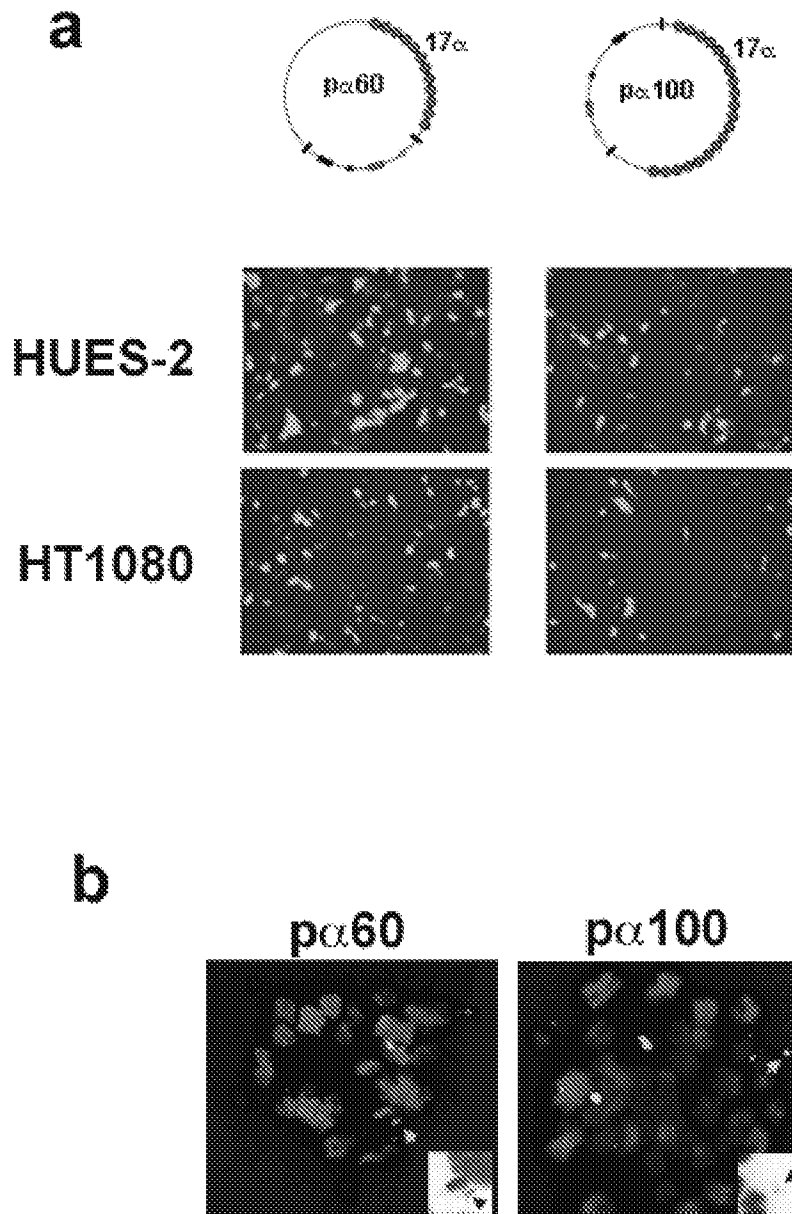


Fig. 7

8/24

**Fig. 8**

9/24

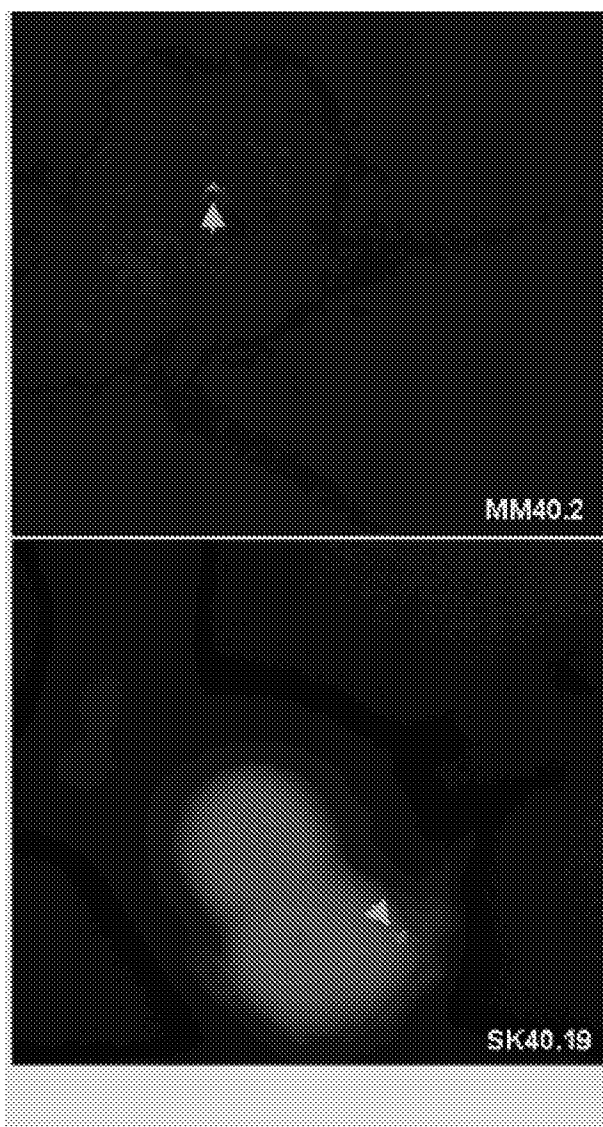


Fig. 9

10/24

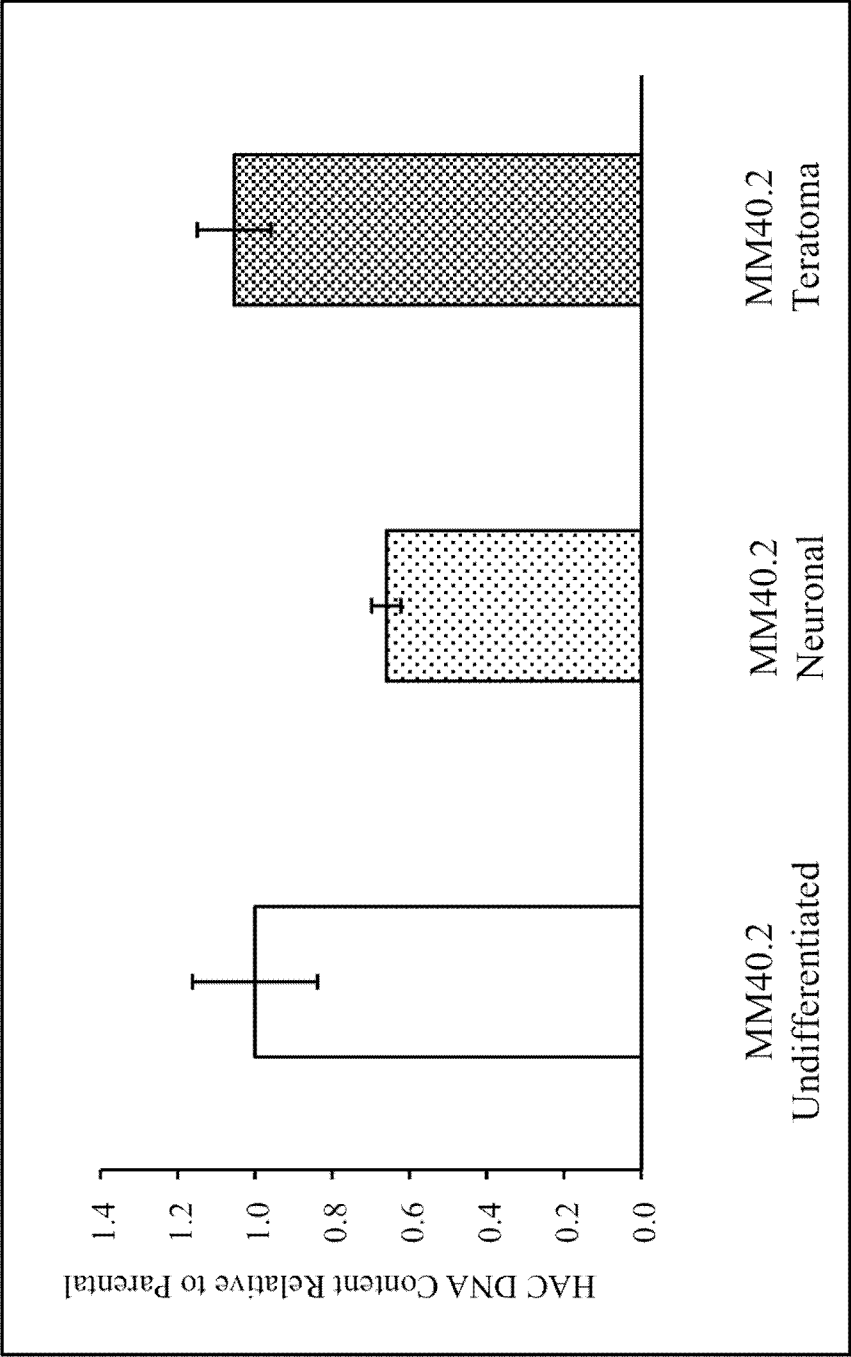


Fig. 10

11/24

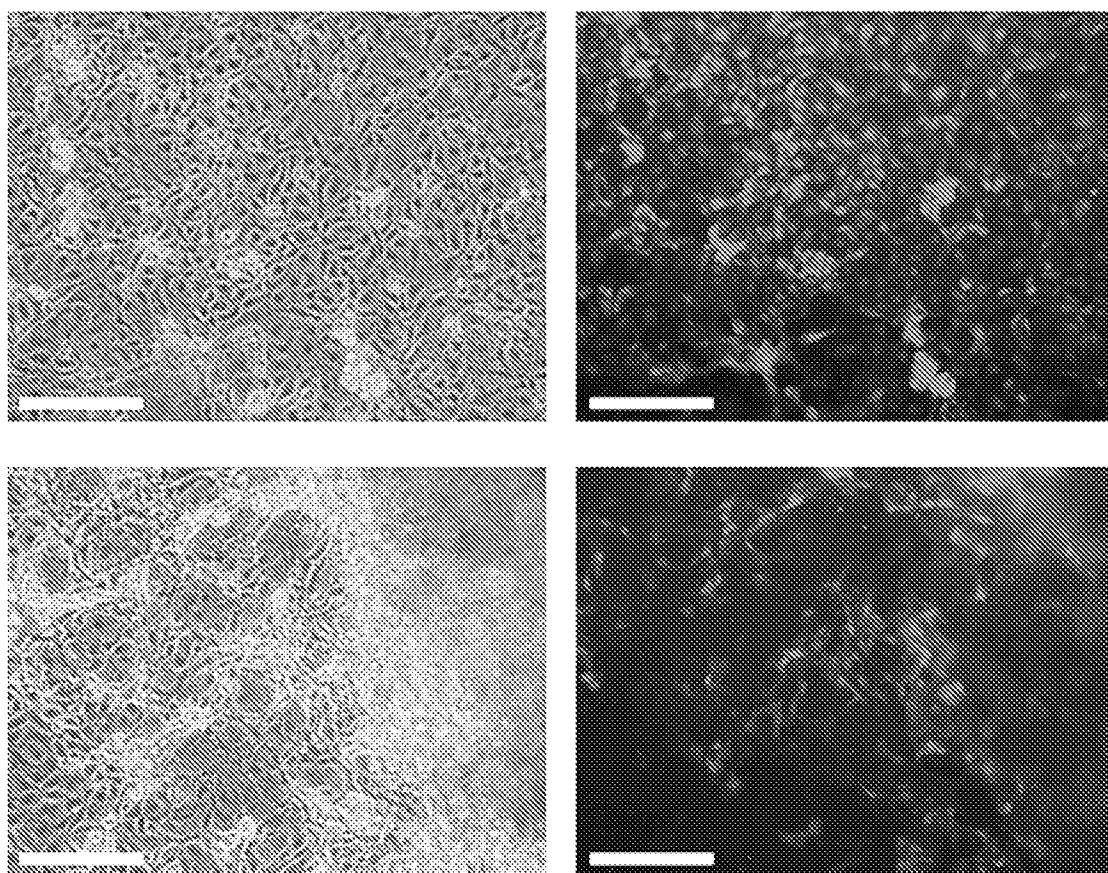


Fig. 11

12/23

AATTCGTTGGAACGGGATAATTTTCAGCTGACTAAACAGAAGCAGTCTCAGAATCTTCT
TTGTGATGTTTGCATTCAAATCCCCGAGTTGAACTTTCTTTCAAAGTTCACGTTTGAAA
CACTCTTTTTGCAGGATCTACAAGTGGATATTTGGACCACTCTGTGTCCTTCGTTTCGAA
ACGGGTATATCTTCACATGCCATCTAGACAGAAGCATCCTCAGAAGCTTCTCTGTGATG
ACTGCATTCAACTCACGGAGTTGAACTCTCCTTTTGAGAGCGCAGTTTTGAAACTCTCT
TTCTGTGGCATCTGCAAGGGGACATGTAGACCTCTTTGAAGATTTTCGTTGGAACGGAA
TCATCTTCACATAAAAACTACACAGATGCATTCTCAGGAACTTTTTGGTGTGTTTGTAT
TCAACTCCCAGAGTTGAACTTTCTTTTGAAAGAGCAGCTATGAAACACTCTTTTTCTA
GAATCTGCAAGTGGACGTTTGGAGGGCTTTGTGGTTTGTGGTGGAAAAGGAAATATCTT
CACCTAAATACTAGATAGAAGCATTCTCAGAAACTGCTTTGTGATGATTGCATTACCTC
ACAGAGTTGAACATTCTATTGATAGAGCAGTTTGAAACACTCTTGTTGTGGAATGTG
CAAGTGGAGATTTGGAGCGCTTTGAGGCCTATGGTAGTAAAGGGAATAGCTTCATAGA
AAAAGTAGACAGAAGCATTCTCAGAAAATACTTTGTGATGATTGAGTTTAACTCACAGA
GCTGAACATTCTTTGGATGGAGCAGGTTTGAGACACTCTTTTTGTACAATCTACAAGT
GGATATTTGGACCTCTCTGAGGATTTTCGTTGGAACGGGATAACTGCACCTAACTAAAC
GGAAGCATTCTCAGAAACTTCTTGGTGTGTTTGCATTCAAATCCCAGAGTTGAACCTT
CCTTTGATAGTTTCAGGTTTGAAACACTCTTTTTGTAGGATCTGCAAGTGGATATTTGGAC
CACTCTGTGGCCTTCGTTTCGAAACGGGTATATCTTCGCATAAAATCTAGACAGAAGCCT
TCTCAGAAACTTCTCTGTGATGATTGCATTCAACTCACAGAGTTGAACCTCCTATGGA
TAGAGCAGTGTTGAAACTCTCTTTTTGTGGAATCTGCAAGTGGATATGTGGACCTCTCC
GAAGATGTCTTTGGAACGGGAATATCTTCACATAAAAACTAAACAGAAGCATTCTCAG
AAACTTCTCTGTGATGTTTGTGTTCAACTCCCAGAGTTTCACATTGCTTTTCATAGAGTA
GTTCTGAAACATGCTTTTCGTAGTGTCTACAAGTGGACATTTGGAGCGCTTTCAGGCCT
GTGGTGGAAAACGAATTATGGTCACATAAAAACTGGAGAGAAGCATTGTCAGAACTTC
TTTGTGATGATTGCATTCAACTCACAGAGTTGAAGGTTCTTTTTCAAAGAGCAGTTTCCA
ATCACTCTTTGTGTGGAATCTGCAAGTGGATATTTGGACCTTTTTGAAGATTTTCGTTGGA
AACGGGAGAATCTTCACAGGAAAGCTAAACAGAAGCATTCTCAGAAACTTCTTTGTGAT
GCTTGCATTCAACTCACAGAGTTGAACTTTCTTTTCGAGAGAGAAGCTTTGAAACACTC
TTTTTCAGAAATCTGCAAGTGGACATTTGGAGGGCTTTGAGGCCTGTGGTGGAAAAGG
AATTATCTTCCCGTAAAAGCTAGATAGAAGCATTCTCAGAAACTACTTTGTGATGATTGC
ATTCAAGTCACAGAGTTGAACATTCCCTTTGACAGAGCAGTTTGGAACACTCTCTTTGTG
TAGAATCTGCAAGTGGAGATATGGACCGCTTTAGGCCTATGGTAGTAAAGGAAATAGCT
TCATATAAAAGCTAGACAGTAGCATTACAGAAAACCTTTGGTGACGACTGAGTTTAAC
TCACAGAGCTGAACATTCTTTGGATGGAGCAGTTTCGAAACACACTATTTGTAGAATG
TGCAAGTGGATATTTAGGCCTCTCTGAGGATTTTCGTTGGAACGGGATAAACCGCACA
GAACTAAACAGAAGCATTCTCAGAACCTTCTTCGTGATGTTTGCATTCAACTCACAGTG
TTGAACCTTTCTTTGATAGTTTCAGGTTTGAAACGGTCTTTCTGTAGAACTGCAAGTAGA
TATTTGGACCGCTCTGAGGATTTTCGTTGGAACGGGATAACCCGCACAGAAGCTAAACA
GAAGCATTCTCAGAACCTCTTTCGTGATGTTTGCATTCAACTCACAGTGCTGAACCTTT
CTTTGATAGTTTCAGCTTTGAAACACTCTTTTTGTAGAACTGCAAGTGGATATTTGGTCC
TCTCTGAGCATTTCGTTGGAACGGGATAAACTGCACAGAACTAAACAGAAGCACTCTC
AGAACCTTCTTCGTGATGTCTGCATTCAACTCACAGTGTTGGAACCTTTCTTTGATAGTTC
AGGTTTGAAACACTCTTTTTGTAGAACTGCAAGGGGATCATTGCACTCTTTGAGGAGT
ACCGTAGTAAAGGAAATAACTTCCTATAAAAAGAAGACAGAAGCTTTCTCAGAAAATTCT
TTGGGATGATTGAGTTGAACTCACAGAGCTGAGCATTCTTTCGATGTAGCAGTTTAGA
AACACACTTTCTGCAGAATCTGCAATTGCATATTTGGACCTCTTTGAGG

Fig. 12

13/24

ttgacagctt	atcatcgaat	ttctgccatt	gcg	gccgccccggg	ccgtcgacca	attctcatgt
gcaaccaggc	gtttaagggc	accaataact	catccgctta	ttatcactta	ttcaggcgta	
ctcatcgcag	tactgttgta	attcattaag	gccttaaaaa	aattacgccc	cgccctgcca	
ggcatgatga	acctgaatcg	ccagcggcat	cattctgccg	acatggaagc	catcacaac	
gcccattggtg	aaaacggggg	cgaagaagtt	cagcaccttg	tcgccttgcg	tataatattt	
ggtgaaactc	acccagggat	tggctgagac	gtccatattg	gccacgttta	aatcaaaaact	
gaaataggcc	aggttttcac	cgtaacacgc	gaaaaacata	ttctcaataa	accctttagg	
ccggaaatcg	tcgtggtatt	cactccagag	cacatcttgc	gaatatatgt	gtagaaactg	
aacggtgtaa	caagggtgaa	cactatccca	cgatgaaaac	gtttcagttt	gctcatggaa	
acggaattcc	ggatgagcat	tcatcaggcg	taccaccage	tcaccgtctt	tcattgccat	
cttgtgctta	tttttcttta	cggctcttta	ggcaagaatg	tgaataaagg	ccggataaaa	
gttataggta	cattgagcaa	ctgactgaaa	aaaggccgta	atatccagct	gaacggtctg	
ggatatatca	acggtggtat	atccagtgat	tgccctcaaaa	tgttctttac	gatgccattg	
tgaaaatctc	gataactcaa	aaaatacgcc	ttttttctcc	attttagctt	ccttagctcc	
gttggaacct	cttacgtgcc	gatcaacgtc	cggtagtgat	cttatttcat	tatggtgaaa	
cccggtatca	acagggacac	caggatttat	tcattttcgc	caaaagttgg	cccagggtct	
tatttatctg	cgataagctc	atggagcggc	ttattctgcg	aagtgatctt	ccgtcacagg	
cgcggtatcg	ggaagtgacg	gacagaacgg	gtaaccgtcg	cacaggaagg	acagagaaag	
ccgctgctcg	tgacgggtg	acgttctctg	tcaggacctg	gattggggag	gcggttgccg	
cctatgcatg	gcacatgctg	tatgccggta	ttccggtcac	accacatacg	ttccgccatt	
gacataagtc	catcagttca	acggaagtct	taccgctgaa	agttctgcaa	agcctgatgg	
cccggcaccg	ggtgcagttt	gcatgcccgg	acacgaaggt	ttttgcgctg	gatgtggctg	
tatcctgaga	ataaatgcct	tggcctttat	agtctgatgc	ggttgcatg	ctgaaacaat	
tttttgtctg	ttaaacagag	aagctggctg	atggaaatgt	ggaactgagt	ggatatgctg	
ggaaaatctc	ccattatcgt	agagatccgc	ttatccactg	agaagcgaac	gaaacagtcg	
tataggaagt	agtgttctgt	catgatgcct	attattaatc	tcaggagcct	gtgtagcggt	
ccttcaggaa	caatagaaat	cttcgtgcgg	gcaagcggta	acgaaaacga	tttgaatatg	
gcaatggaca	gaacaacct	atgaacacag	tgttacgttg	aagtggagcg	gattatgtca	
agtagtgctc	gccgcagtcg	agcgacaggg	aaccatgatg	tggtctgtcc	ttttacagcc	
gggaacagca	cttatatatt	ctgcttacac	cgaagccctc	gagtgagega	ggaagcacca	
atccacttat	ccacggggat	atttttataa	acgatgcctg	aaaaaacttc	ccttgggggt	
tttttagagcg	ccttgtaggc	ctttatccat	ttattttttt	tatagttttt	agatcttctt	
tgccctttca	gtgtgacaaa	tcacctcaaa	gctggttcta	gagaagggtg	tgtgacaaat	
taacctgtg	acaaattgcc	ctcagaagaa	atgacagtc	tgtctgtgac	aaattgccct	
ttgactcttt	tttatttagt	gtgacaatct	gctgtttttt	cacaaagtta	tccttgctta	
catggcggaa	acagcgggta	tcaatcacaa	aaaaacttgt	cacacttcac	atggatctgt	
gtcaaacgac	ctcactgagg	cggcatatag	gaaacgtaaa	aatagcccgc	gaatcgtcca	
tctgttcggt	gaccagatca	gaaaatctga	tctctcccgg	gatcaaaaac	gtatgctgta	
cgagatccat	gttgctaaat	atgctgaaat	tgccacccta	caggaacatg	acggtatctg	
ggatatacgg	caggcattga	agagtttcgc	attcggattg	acctctgcgg	aagccagtaa	
agaggatgcc	ggcgatgaaa	aaggctatga	ggggaaggaa	gtgggttttt	atcgccctga	
cagtcacatc	agagggcttt	acagtgtaca	atcttttctt	tggtttatca	aacgtgcgca	
cgggttacag	aaccgggtta	cgcagtttcg	tatcaaccca	tatctcattc	ccttctttat	
gtatgccatg	cgttttatag	aatccctgtg	gcttagtgaa	acaaaagaaa	tcaccaatcc	
cgtctctctg	aaaatcgact	ggatcataga	tcagtatcgt	aagccggatg	gctcaggcat	
tatgcctgac	ttccgccgcc	gcttctctga	gcgttaccag	ctgcctcaaa	gttaccagcg	
tccaatgcgc	ctctcataca	ttgagaaaaa	ggtctgtgtt	aatgagatca	acagcagaac	
ttccttccgc	gatatacatt	ccatgacgac	gaaaggccgc	cagacgaetc	atatcgtatt	
			aggatagtct	gagggttatc	tgtcacagat	

Fig. 13a

14/24

```

ttgagggtgg ttctgcacat ttgttctgac ctactgaggg taatttgtca cagttttgct
gtttccttca gcttgcattg attttctcat actttttgaa ctgtaatttt taaggaagcc
aaatttgagg gcagtttgtc acagtttgatt tcccttctctc tcccttctgctc atgtgacctg
atatcggggg ttagttcgtc atcatttgatg aggggttgatt atcacagttt attactctga
attggctatc cgctgtgtga cctctacctg gagtttttcc caccggtggat atttcttctt
gcgctgagcg taagagctat ctgacagaac agttcttctt tgcttctctg ccagttcgtc
cgctatgctc ggttacacgg tctccttttg tagtggtgct cttattttta aataagtgc tgaggatgt
tgacttttgc attttgttgt tgctttgcag taaattgcaa gatttaataa aaaaacgcaa
agcaatgatt aaaggatgtt cagaatgaaa ctcatggaaa cacttaacca gtgcataaac
gctggtcatg aaatgacgaa ggctatcgcc attgcacagt ttaatgatga cagcccgga
cgagagaaaa taacccggcg ctggagaata ggtgaagcag cggatttagt tggggttct
tctcaggcta tcagagatgc cgagaaaagca gggcgactac cgcacccgga atctgaaatt
cgaggacggg ttgagcaacg tgttggttat acaattgaac aaattaatca tatgctgat
gtgtttggtg cgcgattgct acgtgctgaa gacgtatttc caccggtgat cggggttgct
gcccataaag gtggcgttta caaaacctca gtttctgttc atcttgcctc ggatctggct
tacgtgtttt tacgtgtttt gctcgtggaa ggtaacgacc cccagggaac agcttctat
tatcacggat gggtaaccaga tcttcatatt catgcagaag acactctctc gccttctat
cttggggaaa aggacgatgt cacttatgca ataaagccca ctgtctggcc gggcttgac
attattcctt cctgtctggc tctgcacctg attgaaactg agttaatggg caaatttgat
gaaggtacgg tgcccaccga tccacacctg atgctccgac tggccattga aactgttgct
catgactatg atgtcatagt tattgacagc gcgcctaacc tgggtatcgg cactgattaat
gtcgtatgtg ctgctgatgt gctgattggt gctgattggt ctgagttggt tgactacacc
tccgcactgc agtttttctga tatgtctctg tttgcttacc gatctgctca agaacttga tcttaagggg
ttcagacctg atgtacgtat tttgcttacc aaatacagca atagtaatgg ctctcagttc
ccgtggatgg aggagcaaat tcgggatgcc tggggaagca tgggtctaaa aaatgttgta
cgtgaaacgg atgaagtggg taaaggtcag atccggatga gaactgtttt tgaaacaggcc
attgatcaac gctcttcaac tgggtgcctgg agaaaatgctc tttctatttg ggaacctgtc
tgcaatgaaa ttttcgatcg tctgattaaa ccacgctggg agattagata atgaagcgtg
cgctgttat tccaaaacat acgctcaata ctcaaccggg tgaagatact tcgttatcga
caccagctgc cccgatgggt gattcggttaa ttgctgcgct aggagtaatg gctcgcggta
atgccattac tttgcctgta tgtggtcggg atgtgaagtt tactcttgaa gtgctccggg
ttactaggtg tgagaagacc tctcgggtat ggtcaggtaa tgaacgtgac caggagctgc
ttactgaggc cgcactggat gatctcatcc cttcttttct actgactggt caacagacac
cgcggttcgg tcgaagagta tctggtgtca tagaaaattgc cgtatgggagt cgcctgcgta
aagctgctgc acttaccgaa agtgattatc gtgttctggt tggcgagctg gatgatgagc
agatggctta ttgggtaacg ttgcagaatg attategccc aacaagtgtg tatgaagcgtg
gtcagcgtta tgcaagccga ttgcagaatg aatttgctgg aaatatctct gcgctggctg
atgcggaaaa tatttcacgt aagattatta cccgctgtat caacaccgcc aaattgccta
aatcagttgt tgcctttttt tctcaccctg gtgaactatc tgcccgggtc ggtgatgcac
ttcaaaaaag ctttacagat aaagaggaat tacttaagca gcaggcatct aaccttcatg
agcagaaaaa agctgggggt atatttgaag ctgaagaagt tatcactctt taacttctg
tgcttaaaac gtcactgtca tcaagaacta gtttaagctc acgacatcag tttgctctg
gagcgacagt attgtataag ggcgataaaa ttcttaagga acttgaaaag tctcgtgttc
caactgagtg tatagagaaa attgaggcca tgtctttact cctggacagg tttgctctg
gatcgagcca cgttttatg cactgtttat ctgggcccac cctggacagg tttgctctg
cagaaagcat aactggcctg aatattctct ctcgatcgtg tgttccactt gtatcgctg
tctgataatc agactgggac caggttccca ctcgtatcgt cgggtctgatt gttccactct
gaccacggtc ccactcgtat cgtcggctctg attattagtc tgggaccacg gttccactct
tatcgtcggg ctgataatca gactgggacc acgggtcccac tctgtatcgt ggtctgatta
ttagtctggg accatgggtc cactcgtatc gtcgggtctga ttattagttc gggaccacgg
tcccactcgt atcgtcggtc tgattattag tctggaacca cgggtcccact cgtatcgtcg
gtcctgattat tagtctggga ccacgggtccc actcgtatcg tgggtctgat tattagctg
ggacacgat cccactcgtg ttgtcgggtc gattatcggg ctggggaccac gttcccactt
gtattgtcga tcagactatc agcgtgagac tacgattcca tcaatgcctg tcaagggcaa
gtattgacat gtcgtcgtaa cctgtagaac ggagtaacct cgggtgtcgg ttgtatgcct
gctgtggatt gctgctgtgt cctgcttacc cacaacattt cgggtgtcgg tatgtggaca
aaataacatg ttaccaggc cgtgccggca cggttaaccg gctgcatccg atgcaagtgt
gtcgtgtcgc acgagctcgc gagctcggac atgaggttgc gtgtcgtga gtgtcgtga
tttgtattgt ctgaagtgtg ttttacgtta agttgatgca gatcaattaa taccgacgt
gcgtcataat tgattatttg acgtggtttg atggcctcca cgcacgttgt gatattgaga
tgataatcat tatcacttta cgggtccttt cgggtgatcc gggcgggcga gttcttcttc
cctcgcgggt tttcgtctat tatgaaaatt ttccggttta aggcgtttcc gttcttcttc
gtcataactt aatgttttta tttaaaatac cctctgaaaa gaaaggaaac gacaggtgct
gaaagcgagc tttttggcct ctgtcgtttc ctttctctgt ttttgcctg ggaatgaaca
atggaagtcc gagctcatcg ctaataactt aatttatnnn aagttatnnn gtcattggctg
nattaagggt tccggatccc ccatacactc cccgtgacg cgccctgacg ggttgtctg
cgccccgaca cccgccaaca cccgtgacg cgccctgacg ggttgtctg

```

Fig. 13b

15/24

```

ccgcttacag acaagctgtg accgtctccg ggagctgcat gtgtcagagg ttttcaccgt
catcaccgaa acgcgcgagg cagccggcca ttatgcacga ccccgccccg acgcccggc
gcccggggcc cgtggccgcg gcccggttgg ttagtcacga gcccgcgcca tccgcgccat
ctgccatggg cggggcgcgga gggcggttgg gtccgcgccc cgccccgcac ggcattctcat
taccgcccga tccggcggtt tccgcttcg tccgcgatgc taacgaggaa cgggcagggg
gcggggcccc ggccccgact tcccggttcg gcggtaatga gatacagacc ccgcgcgccc
gttgcccgct cccgggcccc ccggtcccg ccccggaagc cggggacca cgggacggcg
ggcggcccaa gggcgccccg ccttgccgc ccccgattgg cgggcggggc ggaccgcccc
aagggggcgg ggccgcgggg taaaagaagt gagaacgcga agcgttcgca cttcgtccca
atatatatat attattaggg cgaagtgcga gcactggcgc cgtgccccgac tccgcgcggg
ccccgggggg gggcccgggg ggccgggggg ggggtctctc ggcgcacata aaggccggcg
gcgaccgagc cccgcagacg cgcgcggcca cgaacgacgg gagcggtgc ggagcacgg
gaccgggagc gggagtcgca gaggccgctc ggagcggagc gcgtcggcat cgcgacgccc
cggctcggga tcgggatcgc atcggaagg gacacgcgga cgcggggggg aaagacccgc
ccacccccac cacgaaacac aggggacgca ccccgggggc ctccgacgac agaaaccac
cgttcgcct ttttgcaag ggttggtggc ttgggtggg ggaggaggg ggagcgggg
ggtggaggag gggggacgcg gggcgagg gggggggacg cggggggcga ggagggggga
cgcgggggcg gaggaggggg gacgcgggg gggggggcga ggctcaccg cgttcgtgcc
ttccgcagg aggaacgtcc tcgtcgataa gctgatccat cgccaccatg gtgagcaagg
cgcggggggc gttcaccggg gtgtgtccca tectgttcga gctggacgg gactaaacg
gccacaagtt cagcgtgtcc ggcgagggcg agggcgatgc caccacggc aagctgaccc
tgaagtccat ctgcaccacc ggcgaagctg ccgtgcccgt gccaccctc gtgaccaccc
tgacctcgg catgcgcgaa ttacagcgtc agggagcgac catgaaagc cagacttct
gcaactacaa gacccgcgcc gagggtgaag gaggagcgac catcttctc aaggacgag
actgaaggg catcgacttc aaggaggagc tccgagggcg caccctggtg aaccgcatcg
atcacacag ccacaacgtc tatatcatgg gcaacatcct ggggcacaag ctggagtaca
acttcaagat ccgccacaac atcgaggacg cgcagcgtgc gctcggcgg cactaccagc
agaacacccc catcggcgac ggccccgtgc tgctgcccga caaccactac ctgagacccc
agtccgcctc gagcaagac cccaacgaga agcgcgatca catggtcctg ctggagtctg
tgaccgcgc cgggatcact ctcggcatgg acgagctgta caagtaaac ggccaacttg
ttattgcat cttataatgg ttacaaataa agcaatagca tcacaaatit cacaataaaa
gcatttttt cactgcattc tagttgtgtg agcaatagca tcacaaatit atcttatcat
gtctggatcg gtttaaacg atccactagt aacggccgcc agtaatagca tcttcaatgt
aattccagc cgagcgcggt cgtaccatt accagttggt agtgtgtctg aattaattaa
aaccgggcag gccatgtctg cccgtatttc cggtaggaa ctggtgtcaa aaataataat
aaaacacaaa cttttggtg ttcggtttat gctgtttctt atccattatg tactatttaa
cctacttccc gtttttccc atttggttac tcttcttctt ttacttttt atcatgggag
cgggtattat ttttgccgt caaactcgga acttgtttat tttttcactg cttatgggag
tgctttctga aatttcacaa ataaagcatt tttttcactg cttatgggag aaaaagtata
tagcatcaca atgtatctta tcatgtctgg cgagggagcg attccaaccg gttttggtg
aaactcatca cgaaccccc gccggggagg gctggggcg gggaggcctg gggcactgag
tagagatgcc gcccggggagg gggggcggg gctggggagg gggggcggg tgggcccggg
gagggctggg tggggcggg gggggcggg gctggggagg gggggcggg gggctggggc
cggggagggg gggggcggg gggggcggg gctggggagg gggggcggg gggaggggg
ggtggtgtgt agcaggagcg gtgtgttgc cgggggtacg tctggaggag cgggaggtgc
gcgtgacgt gtggatagg aacaggagtt gttgcgcggt gagttgtcgc tgtgattgt
gtgttgggg aggtgtggt gatgacgtga cgtgtgacgt gctgtgtcgc tgtgattgt
ttggtttcac ctgtggcagc ccgggcccc cgcggcgctg gggggcggg gggggcggg
cggcggtcc gggggaggag agcgggggg gggggcggg gggggcggg gggggcggg
ggggaggagg cggggggagg agcgggggg gggggcggg gggggcggg gggggcggg
ggggaggagg cggggggagg agcgggggg gggggcggg gggggcggg gggggcggg
cggggaggag cggggggagg agcgggggg gggggcggg gggggcggg gggggcggg
cccaaaacac accccccggg ggtcgcgcgc ggccctttaa agcgggtggc gggggcaggc
cgggcccccc gcggccgaga ctacgagtt agacaggcaa gcaactactg cctctgcacg
cacatgcttg cctgtcaaac tctaccacc cggcacgctc tctgtctcca tggcccggcg
ccgcccgcac gcggcccccc gccgccccg gcccggggg attcttgaag cggggcggg
cgcacagtc caggtaacga gctcgaatta taatggttc ttagggtcga cgttccaaac
gcctatttt ataggttaat gtcataataa taatggttc ttagggtcga cgttccaaac
aagcttgata aaaaacctc ctagcgttaa cctgaacctg aaacataaaa tgaatgcaat
tgttgtgtt aactgttta ttgcagctta taatggttac aataaaagca atagcatcac
aaatttcaca aataaagcat ttttttact gcattctagt tgtggtttgt ccaaactcat
caatgtatct taacgcgtaa attgtaagcg ttaatatatt gttaaaatt gcgttaaat
tttgttaaat ttttaacca tggccgaaat aggccgaaat cttataaat ccttataat
caaaagaata gaccgagata cggccgaaat ttggaaacag atgtccactat
taaagaacgt ggactccaac ctatcagggc gatggccac gatggccac
tacgtgaacc atcacccata tcaagtttt tggggtcgag gtgcccgtaaa gcactaaac

```

Fig. 13c

16/24

```

ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagccggcg aacgtggcga
gaaaggaagg gaagaaagcg aaaggagcgg ggcctagggc gctggcaagt gtagcgggtca
cgctgcgcgt aaccaccaca cccgcgcgcg ttaatgcgcc gctacagggc gcgtcagggtg
gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa atacattcaa
atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat tgaaaaagga
agagtccctga ggcgaaagaa accagctgtg gaatgtgtgt cagttagggt gtggaaagtc
cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt cagcaaccag
gtgtggaag gtccaggtcc cccagcagg cagaagtatg caaagcatgc atctcaatta
gtcagcaacc atagtccgc ccctaactcc gccatcccg cccctaactc cgcccagttc
cgccattct ccgccccatg gctgactaat tttttttatt tatgcagagg ccgaggccgc
ctcggcctct gagctattcc agaagtagtg aggaggcttt tttggaggcc taggcttttg
caaagatcga tcaagagaca ggatgaggat cgtttcgc atgtgaacaa gatggattgc
acgcaggttc tccggcgcgt tgggtggaga ggctattcgg ctatgactgg gcacaacaga
caatcggtc ctctgatgcc gccgtgttcc ggctgtcagc gcagggcgcg ccggttcttt
ttgtcaagac cgacctgtcc ggtgccctga atgaactgca agacgaggca gcgcggctat
cgtggctggc cagcaggggc gttccttgcg cagctgtgct cgacgtgtgc actgaagcgg
gaagggact gctgctattg ggcgaaagtgc cggggcagga tctcctgtca tctcaccttg
ctcctgccga gaaagtatcc atcatggctg atgcaatgcg gcggctgcat acgcttgatc
cggtacctg cccattcgac caccaagcga aacatcgcat cgagcgagca cgtactcgga
tggaagccgg tcttgctgat caggatgatc tggacgaaga gcatcagggg ctgcgccag
ccgaactgtt cgccaggctc aaggcgagca tgcccgacgg gctcagggc gtcgtgaccc
atggcgatgc ctgcttgccg aatatcatgg tggaaaatgg ccgcttttct ggattcatcg
actgtggccg gctgggtgtg gcggaccgt atcaggacat agcgttggct acccggtata
ttgctgaaga gcttggcggc gaatgggctg accgcttctt cgtgctttac acccggtata
ctcccgatc gcagcgcac gccttctatc gccttcttga cgtgctttac gggtatcgccg
tctggggttc gaagtcccta tactttctag gcttcttga tgagcgggac tgagcgggac
ngcaagtggc acttttcggg tactttctag agaataggaa cttcgatca attcatcgat
acattcaaat atgtatccgc gaaatgtgcg cggaaaccct ttttctaaat ttttctaaat
aaaaaggaag agtatgagta ttcaacattt taataatgg taaatgcttc aataatgg
attttgctt cctgtttttg ctacccaga aacgctgggt cttattccct tttttgcggc
tcagttgggt gacagagtgg gttacatcga actggatctc aaagtaaaag atgctgaaga
gagttttcgc cccgaagaag gatgagcaet actggatctc aacagcggta agatccttga
cgcggtatta tcccgtagt agagcaactc ttaaaagtte tgctatgtgg
tcagaatgac ttggttgagt cacagaaaag ggtcgccgca tactactatc
agtaagagaa ttatgcagt ctgccataac catcttacgg atggcatgac
tctgacaacg atcggaggac gatgagcaet ccaacttact ccaacttact
tgtaactcgc cttgatcgtt ggaaccgga aacactgcgg tgggggatca
tgacaccag atgcctgtag ggcggggca aactatttaa acgagcggc acgacgagc
acttactcta gcttcccggc caatggcaac gagcgggata aggttgacgg
accacttctg cgctcgcccc ttccggctgg agactggatg ctggagccgg
tgagcgtggg tctcgcggtg agctgggtat ctggagccgg cctcccgat
cgtagtattc tacacgacgg ggagtcaggc aactatggat gatggtaaac cctcccgat
tgagatagg gcttccactga ttaagcattg gtaactgtca gaacgaaata gacagatcgc
acttttagatt gatttaaaac ttcatTTTTT gtaactgtca gaccaagttt actcatatat
tgataatctc atgacaaaaa ttcatTTTTT atttaaaagg atagcctttt agatcctttt
cgttgatgat accgctgcct tactgggtgc ttagcttttc ttccactgag cgtcagaccc
taatccgaag tggtcagact ggaaaatcag atttagccagt ctgaatgacc tgtcacggga
agatagcacc acatagcaga cccgccataa agggcaggaa ctgctgaaca gcaaaaagtc
cttgatttat tttgagtttc cccgccataa aacgccctga acgggctttt cgggctttt
ccccattcac tgccagagcc cttgcatgaa tccataaaag gcgctgtag tgccatttac
tcaggtgcct gatggtcggg gtagcgcgag cgaactgaat agacagcgac agacagcgac
aagatccgca gttcaacctg gacaaaagga atattcagcg ttgcccggc ttgcccggc
ctctcatgtt taacgtacta ttgatagtac gtactaagct acgtactaag acgtactaag
actaagctct catggctaac gtactaagct cteatgtttc tctaaagttt tctaaagttt
tgaacaataa aattaatata aatcagcaac ttaaatagcc agctttttaag gtttaacggg
aagaaaaaaa agaataatata aggtttttaag ttagagcctc tcaaagcaat gtttaacggg
acaggaacac ttaacggctg tgagaagccc ccatgggcta tcaaagcaat gtttaacggg
ttcgctgacc atttccgggg tgcggaacgg cgttaccaga aactcagaag ggtctgcttc
ccaaaccgac tctgacggca gtttacgaga gagatgatag ggtctgcttc aggtctacaat
atgctacaca attaggcttg tacatattgt cgttagaagc aatacaagct aatacaagct
ccttatgtat catacacata cgatttaggt cgttagaagc aatacaagct aatacaagct
caggtcgagg gacctataa cttcgtatag cgttagaagc aatacaagct aatacaagct
gatgcggccc cnnnnn

```

Fig. 13d

17/24

	attc	gttggaaacg	ggataatttc	agctgactaa	acagaagcag
tetcaggatc	cattgacccg	gaacccttaa	thnnnntgca	tgaatccata	aaaggcgcct
gtagtgccat	ttacccccat	tcactgccag	agccgtgagc	gcagcgaact	gaatgtcacg
aaaaagacag	cgactcaggt	gcctgatggt	cggagacaaa	aggaatattc	agcgatttgc
ccgattgcgg	ccgcaagatc	cgcagttcaa	cctgttgata	gtacgtacta	agctctcatg
tttcacgtac	taagctctca	tgtttaacgt	actaagctct	catgtttaac	gaactaaacc
ctcatggcta	acgtactaag	ctctcatggc	taacgtacta	agctctcatg	tttcacgtac
taagctctca	tgtttgaaca	ataaaattaa	tataaatcag	caacttaaatt	agcctctaag
gttttaagtt	ttataagaaa	aaaaagaata	tataaggctt	ttaaagcttt	taaggtttaa
cggttgtgga	caacaagcca	gggatgtaac	gcactgagaa	gcccttagag	cctctcaaa
caattttcag	tgacacagga	acacttaacg	gctgacagaa	ttagccatgg	gctagcgatg
accctgctga	ttggttcgct	gaccatttcc	gggggtgcgga	acggcggttac	cagaaactca
gaaggttcgt	ccaaccaaac	cgactctgac	ggcagtttac	gagagagatg	atagggtctg
cttcagtaag	ccagatgcta	cacaattagg	cttgtaacata	ttgtcgttag	aacgcggcta
caattaatac	ataaccttat	gtatcataca	catacgattt	aggtgacact	atagaataca
agcttgcatg	cctgcaggtc	gagggacccta	ataacttcgt	atagcataca	ttatacgaag
ttatnnnnnn	nnnnattaag	ggttccggat	ccccataca	ctccgctatc	gctacgtgac
tgggtcatgg	ctgcgccccg	acacccgcca	acacccgctg	acgcgcctgt	acgggcttgt
ctgctccccg	catccgctta	cagacaagct	gtgaccgtct	ccgggagctg	catgtgtcag
aggttttcac	cgtcatacc	gaaacgcgcg	aggcagccgg	ccattatgca	cgacccccgc
ccgacgcggg	cacgcggggg	gcccgtggcc	gcggcccggt	ggtcgaaccc	ccggccccgc
ccatccgcgc	catctgccat	gggcggggcg	cgagggcggg	tgggtccgcg	ccccgccccg
catggcatct	cattaccgcc	cgatccggcg	gtttccgctt	ccgttccgca	tgctaacgag
gaacgggacg	ggggcggggc	ccgggccccg	acttccccgt	tcggcggtaa	tgagatacga
gccccgcgcg	cccgttgccc	gtccccgggc	cccccggtcc	cgcccccgcg	acgcccggac
caacgggacg	gcgggcggcc	caaggcgccg	cgcccttgcc	gcccccccat	tgcccggcg
gcgggaccgc	cccaaggggg	cggggccgcc	gggtaaaaga	agtgagaacg	cgaagcggtc
gcacttcgtc	ccaatatata	tatattatta	gggcgaagtg	cgagcactgg	cgccgtgccc
gaetccgcgc	cggccccggg	ggcgggcccc	ggcggcgggg	ggcgggtctc	tccggcgcac
ataaaggccc	ggcgcgaccg	acgcccgcag	acggcgccgg	ccacgaacga	cgggagcggc
tgccgagcac	gcccagccgg	agcgggagtc	gcagagggcc	gtcggagcgg	acggcgctcg
catcgcgacg	ccccggctcg	ggatcgggat	cgcatcggaa	agggacacgc	ggagcggggg
gggaaagacc	cgcccccccc	acccacgaaa	cacaggggac	gcacccccgg	ggcctccgac
gacagaaacc	caccgggtccg	ccttttttgc	acgggttaagc	accttggttg	ggcggaggag
ggggggacgc	gggggttgag	gaggggggac	gcggggggcg	aggagggggg	acgcgggggc
ggaggagggg	ggacgcgggg	gcggaggagg	ggggacgcgg	gggcggagga	gggggctcac
ccgcgttcgt	gccttcccg	aggaggaacg	tcctcgtcga	taagctgac	catcgccacc
atggtagaca	agggcgagga	gctgttcacc	gggggtgtgc	ccatcctggt	cgagctggac
ggcgacgtaa	agccacacaa	gttcagcgtg	tcgggcgagg	gcgagggcga	tgccacctac
ggcaagctga	ccctgaagtt	catctgcacc	accggcaagc	tgcccggtgc	ctggccccacc
ctcgtgacca	ccctgacct	cggcgtgcag	tgcttcagcc	gctacccccg	ccacatgaag
cagcacgact	tcttcaagtc	cgccatgccc	gaaggetacg	tccaggagcg	caccatcttc
ttcaaggacg	acggcaacta	caagacccgc	gccgaggtga	agttcgaggg	cgacaccctg
gtgaaccgca	tcgagctgaa	gggcatecgac	ttcaaggagg	acggcaacat	cctggggcac
aagctggagt	acaactacaa	cagccacac	gtctatatca	tgcccgacaa	gcagaagaac
ggcatcaagg	tgaacttcaa	gatccgccac	aacatcgagg	acggcagcgt	gcagctcgcc
gaccactacc	agcagaacac	ccccatcggc	gacggccccg	tgctgctgcc	cgacaaccac
tacctgagca	cccagtcgcg	cctgagcaaa	gaccccaacg	agaagcgcg	tcacatggtc
ctgctggagt	tcgtgaccgc	cgccgggatc	actctcgcca	tggacgagct	gtacaagtaa
agcggccaac	ttgtttattg	cagcttataa	tggttacaaa	taaagcaata	gcatacaaaa
tttcacaaat	aaagcatttt	tttactgca	ttctagtgtg	ggtttgtcca	aactcatcaa
tgtatcttat	catgtctgga	tcggttttaa	ccgatccact	agtaacggcc	gacagtgtgc
tggaaattaat	taaaattcca	gctgagcgcc	ggtcgctacc	attaccagtt	ggtctggtgt
caaaaataat	aataaccggg	caggccatgt	ctgcccgtat	ttcgcgtaag	gaaatccatt
atgtactatt	taaaaaacac	aaacttttgg	atgttcgggt	tattcttttt	cttttacttt
tttatcatgg	gagcctactt	cccgttttcc	ccgatttggc	tacatgacat	caaccatata
agcaaaagtg	atacgggtat	tattttttgcc	gctattttctc	tgttctcgct	attattccaa

Fig. 14a

18/24

```

ccgctggttg gtctgctttc tgacaaactc ggaacttggt tattgcagct tataatgggt
acaaataaag caatagcatc acaaatattca caaataaagc atttttttca ctgcattcta
gttggtggtg tccaaactca tcaatgtatc ttatcatgtc tggatctctg acctgagatt
ggcggcactg aggtagagat gcccgaaacc ccccgaggga gcgcgggacg cgccggggag
ggctggggcc ggggagggct ggggcccggg agggctgggg ccggggaggg ctggggccgg
ggagggctgg ggccggggag ggctggggcc ggggagggct ggggctgggg agggctgggg
ctggggaggg ggcggtgggt tgtagcagga gcggtgtgtt gcgcccgggt acgtctggag
gagcgggagg tgctgtggat cgtgtggatg aggaacagga gttgttgccg ggtgagttgt
cgctgtgagt tgtgttgggt ggcagggtgt gtggatgacg tgacgtgtga cgtgcggagt
gcgcccgtgt ctgttggttt cactgtgtgc agcccgggce ccccgcgggc gcgcgcgccg
gcaaaaaagg cgggcccggg tccgggcccg gtgcgcgcgc gcggcgggcg tggggggcgg
ggcgcgcggg gcggggggag gagcgggggg aggagcgggg ggaggagcgg aagagcggag
ggggggagga gcggggggag gagcgggggg aggagcgggg ggaggagcgg ggggaggagc
ggggggagga gcggggggag gagcgggggg aggagcgggg ggaggagcgg ggggaggagc
aaaacggggc ccccccacaa ccccgggccc agactagcga gttagacagg caagcactac
gcggcgggca cgcacatgc acgcgcgcgc ccccgcccgc ccccgcccgc gggcccacgg
tcgctctgtc ccatggccc cgcgcgtccc aaccgcacag tccacggtaa ttagcgaagg
ggcctcgtga tacgcctatt tttatagggt aatgtcatga taataatggt ttcttagggg
cgacggatgc gataagcttg ataaaaaacc tcccacacct cccctgaac ctgaaacata
aaatgaatgc aattgttgtt gtttaactgt ttattgcagc ttataatggt tacaataaaa
gcaaatgtgc cacaatttcc catcaatgta tcttaacgcg ttttttaacc actgcatctc
tgtccaaact atttttggtt aatcagctca ttttttaacc aataggccga agtggttcc
atcccttata aatcaaaaga atagaccgag atagggttga gtgtgttcc aggcgaacac
aagatgcac cactacgtga accatcaccc taatcaagtt ttttggggtc gagcttgacg
ggcgaatggc atcggaaccc taaagggagc ccccgattta gcggttgacg cgggcgctag
aaagcactaa cgagaaagga agggaaagaa acaccgcgcg cgcttaatgc ggcctatttg
gcgcgctcag gtggcacttt tcggggaaat gtgcgcggaa cctgataaat gtggaatgtg
taaatacatt caaatatgta tccgctcatg tgaggcgga agaaccagct gcagaagtat
tattgaaaaa ggaagagtcc tgcccagcag tccccaagc gctcccagc aggcagaagt
gggtgtgaaa gtcccagga cagggtgtga accatagtc cgcacctaac tccgcccac
tgcatctcaa ttccgcccag tctccgccc tctgagctat tccagaagta gtgaggaggc
ctccgcccag cgcctcggcc ttgcaaaagt cgatcaagag gatcggttcg tttttggag
agcctaggct ttgcaagat tgcaagcagg ttctcgggcc gcttgggttg catgattgaa
caagatggat agacaatcgg ctgctctgat gccgcccgtg tccggctgtc cggctatgac
tgggcacaa cgcgccgttc tttttgtcaa gaccgacctg tccggtgcc tgaatgaact agcgcagggg
ggcgcggcgc tategtggct ggccacgacg ggcgttcctt tgccggggca gcaagacgag
gtcactgaag cgggaaggga ctggctgcta tccatcatgg gtcagcagtg gctcagctgt
tcatctcacc ttgtctctgc ctgcccattc gaccaccaag atctggacga gcatgcccg
catacgtttg ggatggaagc ggtcctgttc ctcaaggcga gcatgcccg
gggctcgcgc cccatggcga tgcctgcttg cccaatatca tgggtgaaaa
ctcgtcgtga tctggattca cgggtcgggt gtggcgggac gctatcagga
gctaccgctg atattgctga agagcttggg ggcgaatggg gctgaccgct
tacggtatcg ccgctcccga ttccgagcgc atcgcttctc atcgcttctc
ttctgagcgg gactctgggg ttccaagtgc ctatacttcc tagagaatag
tcaattcacc gatngcaagt ggcacttttc ggggaaatgt gcgcggaacc
tatttttcta aatacattca aatatgtatc cgctcatgag acaataacc
ttcaataata ttgaaaaagg aagagtatga ttccgtgtc tttccgtgtc
ccttttttgc ggcattttgc ctccctgttt ttgctcacc agaaacgctg
aagatgctga agatcagttg ggtgcacgag tgggttacat cgaactggat
gtaagatcct tgagagtttt cgccccgaag aacgttttcc aatgatgag
ttctgctatg tggcgcggtg ttatcccgtg ttgacgccc gcaagagcaa
gcatacacta ttctcagaat gacttgggtg agtactcacc agtcacagaa
cggatggcat gacagtaaga gaattatgca aacctagagt aacctagagt
cgcccaactt acttctgaca acgatcggag gactaaccgt gctaaccgt
acatggggga tcatgttaac cgccttgatc gttgggaacc ggagctgaat
caaacgacga gcgtgacacc tagcaatggc aacaacgttg aatagactgg
taactggcga actacttact ctgacttccc ggcaacaatt

```

Fig. 14b

19/24

```

ataaagtgtgc aggcaccactt ctgcgctcgg cccttcgggc tggctggttt attgctgata
aatctggagc cgggtgagcgt ggggtctcgc gtatcattgc agcactgggg ccagatgggt
agccctcccc tategtagtt atctacacga cggggagtcg ggcaactatg gatgaacgaa
atagacagat cgctgagata ggtgcctcac tgattaagca ttggtaactg tcagaccaag
tttactcata tatactttag attgatttaa aacttcattt ttaatttaa aggatctagg
tgaagatcct ttttgataat ctcattgacca aaatccctta acgtgagttt tctgtccact
gagcgtcaga ccccggtgat gataccgctg ccttactggg tgcattagcc agtctgaatg
acctgtcacg ggataatccg aagtggtcag actggaaaat cagagggcag gaactgctga
acagcaaaaa gtcagatagc accacatagc agaccgccca taaaacgccc tgagaagccc
gtgacgggct tttcttgat tattttgagt ttcctataac ttcgtataat gtatgctata
cgaagtattt aggtccctcg actacgtcgt taaggccgtt tctgacagag taaaattctt
gagggaactt tcaccattat gggaaatggt tcaagaaggt attgacttaa atccactcaa
atggtcaggt cattgagtgt ttttaatttg ttgtatttt tttttttaga gaaaaatctc
caatatataa attaggaatc atagtttcat gattttctgt tacacctaac tttttgtgtg
gtgccctcct cctgtcaat attaatgtta aagtgcattt ctttttctt atcactgtga
gccattagta tcaatttgct tacctgtatt cctttacatc ctccttttct tcttcttga
taaatgtatg tagattgcgt atatagtttc gtctacccta tgaacatatt ccattttgta
attcgtgtgc gtttctatta tgaatttcac ttataaagtt tatgtacaaa tatcataaaa
aaagagaate ttttaagca aggattttct taacttcttc ggcgacagca tcaccgactt
cggtgtgact gttggaacca cctaaatcac cagttctgat acctgcaccc aaaactttt
taactgcac tccaatggcc ttaccttctt caggcaagtt caatgacaat ttcaacatca
ttgcagcaga caagatagtg gcgatagggt tgacctatt ctttggcaaa tctggagcag
aacgtggca tggttcgtac aaaccaaatg cgggtgttct gtctggcaaa gaggccaagg
acgcagatgg caacaaaccc aaggaaactg ggataacgga ggcttcacgc gagatgat
caccacacat gttgctgggt attataatac catttaggtg ggttgggttc ttaactagga
tcattggcgc agaatcaatc aattgatgtt gaaccttcaa tgtagggaa tctgttctga
tggtttccct cacagtttt ctcataatc ttgaagaggc caaaacatta gcttttcca
aggacaaat aggcaatggt ggctcatgtt gtaggggccat gaaagcgccc atcttctga
ttcttgcac ttctggaacg gtgtattgtt cactatccca agcgacacca tcaccatcgt
cttctttct cttaccaaag taaatacctc ccactaatc tctgacaaca acgaagtcag
tacctttagc aaattgtggc ttgattggag ataagtctaa aagagagtcg gatgcaagt
tacctgtct taagtggcg tacaattgaa gttctttacg gatttttagt aaacctgtt
caggtctaac actaccggt aacgcgttgc ggaagcatca gcaaataagg ccagcacagc
cagcgcagtt gccgctttgg ttcttgattc tggttctgat gaattaaaca aagcggcaca
gtaacaaagg acttcattga taattttct tcaaggaggaa gacatgtcat tctttctac
gttaaaaaa gctttgtctt tgaaggagaa acttgcgtc actggtgttc ttgttctgat
ttgcgcactt gttgtgctg ggtttgcatg ggaacgtcat cagctaaagc aagccataga
gaaaattggc agtcttgatc aggtgttaa ggaacgtgat aagtcaataa tggatcttaa
ccagaccatt gagacgatga acaaaagcaga gcaacatttt cacagccagg aagtgaaaaa
tgaatcagaa caagccaagt atgctgacag gcaaatggaa cgaaaagctg aagttcagaa
acaactgggt gcggcgggta atgttcgcca gcgtattcct gctgacactc agcggttgct
ccgggagtcg atcagcgaat ttaacgccga cgccgacaaa ggttaaccac cctgccccca
aaagtgcatt tatgtgtagg atgcccagag tttagcagtg atattttgat gatctgcag
cgtatatcct cgatacagaa acgatcgtga tggggattaa caggaagaat gcgaagttta
atgattacaa ccgagctatt agcggtaact aaaagggatt tttatgtctg ataaagtac
agtaaaagcaa actatcaaca aagcgacttc aatctacaaa attgagcaaa tcaactgttg
caagccagga tctgaacaat accgtcgtgc tttcagctt gccgatcagc ttggtttaaa
acaccggat tgcattgagc atgtatttcc gacctatgct gatgagcaat gtactcatgt
tcttaccgaa gaggattttt tcagcactga agaacgagaa ggcgttgatc gctgcattgg
tgtgatttgt tcttcggtaa gtgatgagtt attcccta atgtccctgaat atgggtggat
tggataccaa ttctgtacg agggcgatga gcttaaatgc tatgaacatg gtcttctcat
cgaaagcgta gaataatac actcccttc aaccggctac gttggccggt ttttcaacta
tccacattat ccaactggata gatccaataa tcaggtccat acagatccca attagatcca
tatagatccc tgatcgttgc aggccgcgcc acgtctggct tagaagtga tgcgcatgtg
tgctggagg gggatatagg atgtgtgttg gataaaaaatg tgtgctgacg ggttgcta
gtgtgtggc gggatatagg atgtgtgttg acgggaaagc ttgggtagtt atcaccactt
ataaaaacta tccacacaa tccgaaaaag taatatgaat caatcatita tctccgat
tctttacgca gacattgaaa gtaaggcaaa agaactaaca gtttaattcaa acaactagt
gcagcctgta gcgttgatgc ccttgggggt attcgtgccg aagccatcaa agagcaaagg
agaaagtaaa gagattgatg ccaccaaaagc gttttccag ctggagatag ctaaagccga
gggttacgat gatattaaaa tcaccggtcc tcgactcgat atggatactg atttcaaac
gtggatcggt gtcattctac cgttcagcaa atacggcttg tectcaaaaca ccatccagtt
atcgtttcag gaattcgcta aagcctgttg tttccctca aaacgtctgg atgcgaaact
gcgtttaacc attcatgaat cacttgagc cttgcgtaac aagggtatcg cttttaagcg
cgaaaagat gctaaaggcg gctatcagac tggctgtctg aaggtcgggc gttttgatgc
tgacctgat ctgatagagc tggagggctg attcgaagtt gtgggagctg ttccagcttg
attatcgct tetgtgcaa caccagcct tgcgtgccct tccgaagaaa gaagctgcac
aagccattta cactttcatc gaaagccttc cgcagaaccc gttgccccta tcttccgcgc

```

Fig. 14c

20/24

```

gaatccgtga ggcgcctggct ttgcagtcag ctgttggcga gcaaaaccgt atcattaaga
aagcgataga acagcttaaa acaatcggct atctcgcactg ttctattgag aagaaagggcc
gggaaagttt tgtaatcgtc cattctcgca atccaaagct gaaactcccc gaataaggtg
gtgctggagg gaaaccgcat taaaaagatg tgtgctgccg ggaaggcttg tccaatttcc
tggtttttgaa tgtgcgctgg agggggacgc cctcagttt gccagactt tcctccagc
acacatctgt ccatccgctt ttccctccag tgcacatgta attctctgce ttccctcca
gcacacatat ttgataccag cgatccctcc acagcacata attcaatgcg acttccctct
atcgcacatc tttagactttt attctccctc cagcacacat cgaagctgcc gggcaagccg
ttctcaccag ttgatagaga gtgaagcttg gctgcccatt gaagcaggaa atcaccaaaa
tgattcaggc tacaacctga acgtagaaga aatccgcgtc ctttatgcgt ggagatgcc
aaagcatggt gtgacacact tggcaaagga gtaagcatgc agagaatgct atgtacaagc
atctacgcat acattattat tttatgcagc atttttaatt aaattcaaaa atacagcata
aaggatgact ttcgatgagt gattccagcc agcttcacaa ggttgctcaa agagcaaaca
gaatgctcaa tggtctgact gaacaagtac agttgcaaaa ggatgagcta caccggaacg
agttttacca ggtctatgcy aaagcggcac tggcaaaatt gccctactg actcgagcga
acgttgacta tgccgtaagt gaaatggaag aaaaggggta tgttttcgat aaacgcctg
ctggctcttc aatgaaatat gcgatgtcaa ttcagaacat cattgacata tatgaacatc
cgggagtgcc aaaataccgg gatcgctaca gcgaacgtat gtgattttca tctccaatct
taaaggcgtt gtgtcaaaaa ctgtatcgac ggtttctctg gcgcatgcaa tgcgtgtca
ccctcatctt cttatggagg atttaaggat tctggttatt gacctgac cgcaatcttc
agcaacgatg tttttaagcc ataacactc tattggtatc gtaaacgcaa catctgcaca
ggctatggtg cagaatgtaa gccgtgaaga gctgttagag gattttattg ttcttctgt
tgtaccctgg gttgacgtta tgccctgcgtc gattgacgat gcccttattg catccgattg
gagagagctg tgcaatgagc atctaccggg tcagaacatc catgctgtcc tgaagaaaa
tgtgattgat aagctgaaga gcgattatga ctttatectc gttgatagtg gtccctacct
tgacgccttc ctgaaaaatg ctttggcctc ggccaatata ctgtttacac ctctgcgc
agcaactgtc gatttccact catcgcttaa atacgttgcc cgccttctg agttggtgaa
ctcatttctg atgaagctg cgagtgccag cttgcgacta acattggtt tatgtccaag
ttgagtaaca aggcagacca taagtattgc catagcctgg ctaaagaagt gttcgggtgg
gatattgctt atgtcttctt cctcgcctt gacgggtttg aaacgctgcg ggcgattt
tgacactggt atttcagcta acccggaac gtatgttggg agtgctgatg cattgaagaa
cgcgcgaaat gccgcggaag attttgctaa agcagttttt gaccgtattg aatttatcag
atctaaactg ggagtaagaa acccccatgt caaagaaaaa cagaccaaca attggcgaa
cccttaatcc ttcaatatta agcggatttg atagttcttc agcctctggc gatcgaagtcg
agcaggtatt caagttatca actggtcgcc aggccacatt tattgaagag gtaatacctc
cgaaccaggt agaaagcgat acctttgttg atcagcataa gaccaggcat gaccaggcat
ctcttacgcc aaaatcatta aaaagtatcc gaagcactat taagcatcag caattttacc
ctgcaatagg tggttagacgg gctacaggga aaattgaaat tttggatggt tcccggcgtc
gagcttctgc catcttagag aacgtagggg tgccgggtttt agtcacggac caggagatca
gcgttcagga agcgcaaaat ttagcgaaag acgttcagac agcattgcag cacagcatc
gagaaatagg tctgcgtttg atgcgaatga aaaatgatgg gatgagtcag aaggatattg
cagccaaaga agggctgtct cagycgaagg tcacgcgtgc tctccaggca gctcttatc
cggaagaatt agtcgcctt ttccctgtgc agtggaatt aaccttttcg gactacaaaa
cgctttgtgc tgttggcgac gaaaatgggt acaagaattt agagtttgat cagcttatc
aaaacatac cccggaaata aacgacatct tatccattga agaaatggcc gaagatgaag
ttaaaaaata aatcctgcgc ttgataacaa actgaattat gataaagggt gataaagggt
ctaaagataa gtccgtagtt acattttctt atgagtttaa tgcactctca aaagagttac
ggaagcgctt gaaaggccgt cgacaggatg attgggcata gagcctcgat aaaaagccga
aggaagaact tttcgccatt caaggtgaaa ctaatttcac agtaaatcag cctccctcc
agccttaaac ttcgacacgt ttgcttgcac egagaaagtc aacaagtgc tttcaataaa
tctcttccg aataaaaaaa ttgcttgcac ccttgtgtt aaggaagagt aaattgagta
atctctccg aaaagggatt cacacaagtg ccttgtgtt aaggaagagt aattgagta
acttacgca ataccgaat cgtattgcag ataccgaaa acgctctaaa gctgtgctt
gctgggcaa cactgcgag ttcggtactg ataaccaatt cattaaagat gatgcccgcg
gtgccgcac tatccttgaa gctgcacgta aagacccggt ttttgcgggt atctctgata
atgccaccgc tcaaatcgct acagcggtgg caagtcgact ggctgactac gccgcagcac
ataaatctat gccgcgtccg gaaattcttg cctctgcc cagacgctg gaaaactgcc
tgatagagtc caccgcgaat agcatggatg ccactaataa agcgatgctg gaactctgtc
cagcagagat gatgagcgtt tctgacggtg ttatgcgtct gcctttattc ctgcgatga
tcttgctgt tcagttgggg gcagctaccg ctgatgcgtg taccttcatt ccggttacgc
tgaccagtc gaagtcttta acgtggcagg ttcattcttt cggttcttatg
ctgctgggtg tggtctggac atgcaatccg cgggtgtgta ttcatctttt cgtcgcgct
atgtgctggt ggcaagctcc gatggcacca gcaaaaccgc cagccagtta atggaagact
tcgaaggcca gaatgtacca atccgaaaag gtcgcactaa aaccttcaag aaccgtatta
agtctgttgg tgataacggg tccggcagcc gtttactaat gtttactaat gctgctgtg
agcaaatcac tggtacctgc tctctgaact caattgagac atgtgagcc
ccaagcgcc ggataaaggc actgagatcg caattgagac ggaaatcaat attgaagccg

```

Fig. 14d

21/24

```

ctcctgagct gatccccgtg atcaaccacg aaatgaagaa atacacctg ttcccaagtc
agttcggttat cgcggctgag cacacggtag aggcggcgta tgaagcacag cgtgaatttg
gtctggacct gggttcccta cagttccgca ccctgaagga atacctgtct catgaacagg
atatgctgcg tcttcgcatc atgatctggc gcactcttgc gaccgacacc tttgacatcg
ctctgccggg taaccagttc tttgatgtat gggcaaccat cattcgtggc aaattccaga
ctgtatatcg cgacattatt gagcgcgtta aatcttctgg tgcgatgggg atgtttgctg
gtgctgatgc agcatctttc ttcaaacagt tgccgaagga tttcttccag ccagccgaag
actatatcca gactccgtat gttcactaca tcggtagccc atttaggacc acccacagca
cctaacaaaa cggcatcagc cttcttggag gcttccagcg cctcatctgg aagtggaaaca
cctgtagcat cgacctgcag gggggggggg gcgctgaggt ctgcctcgtg aagaagggtg
tgctgactca taccaggcct gaatcgcccc atcatccagc cagaaagtga gggagccacg
gttgatgaga gctttgttgt aggtggacca gttggtgatt ttgaactttt gctttgccac
ggaacggtct gcgttgctcg gaagatgcgt gatctgatcc ttcaactcag caaaagttcg
atztatcca caaagccgcc gtcccgtaaa gtcagcgtaa tgctctgcca gtgttacaac
caattaacca attctgatta gaaaaactca tcgagcatca aatgaaactg caatttatc
atatcaggat tatcaatacc atatttttga aaaagccggt tctgtaatga aggagaaaac
tcaccgaggc agttccatag gatggcaaga tcctggtatc ggtctgcat tccgactcgt
ccaacatcaa tacaacctat taatttcccc tcgtcaaaaa taagggttat aagtgagaaa
tcaccatgag tgacgactga atccggtgag aatggcaaaa gcttatgcat ttctttccag
acttgttcaa caggccagcc attacgctcg tcatcaaaat cactcgcac aaccaaacgg
ttattcattc gtgattgcgc ctgagcgaga cgaaatacgc gatcgctgtt aaaaggacaa
ttacaaacag gaatcgaatg caaccggcgc aggaacactg ccagcgcac aacaatattt
tcacctgaat caggatattc ttctaatacc tggaaatgct ttttcccggg gatcgcatg
gtgagtaacc atgcacatc aggagtacgg ataaaaatgct tgatggctcg aagaggcata
aattccgtca gccagtttag tctgaccatc tcactctgtaa catcattggc aacgctacct
ttgccatgtt tcagaaacaa ctctggcgca tcgggcttcc catataatcg atagattgtc
gcacctgatt gcccgacatt atcgcgagcc catttatacc catataaatc agcatccatg
ttggaattta atcgcggcct cgagcaagac gtttcccggt gaatatggct cataacaccc
cttgatttac tgtttatgta agcagacagt tttattgttc atgatgatat atttttatct
tgtgcaatgt aacatcagag attttgagac acaacgtggc tttccccccc cccctgcag
gtcgatagca gcaccaccaa ttaaattgatt ttggatccga tgtagcagtt tagaaacaca
ctttctgcag aatctgcaat tgcatat

```

Fig. 14e

22/24

gaacccagagtcctcgctcagaagaactcgtcaagaaggcgatagaaggcgatgcgctgcgaatcgggag
 cggcgataccgtaaaagcagaggaagcggctcagccattcgcgcgaagctcttcagcaatatacgggt
 agccaacgctatgtcctgatagcggctccgccacaccagccggccacagtcgatgaatccagaaaagcgg
 ccattttccaccatgatattcggcaagcaggcatcgccatgggtcacgacgagatcctcgcgcgtcgggca
 tgcctgccttgagcctggcgaaacagttcggctggcgagcccttgatgctcttcgtccagatcatcctg
 atcgacaagacggcttccatccgagtacgtgctcgcctgatgcgatgtttcgccttggtgggtcgaatggg
 caggtagccggatcaagcgtatgcagccgcgcattgcatcagccatgatggatactttctcggcaggag
 caaggtgagatgacaggagatcctgccccggcacttcgccccaatagcagccagtcctctcccgcttcagt
 gacaacgtcgagcacagctgcgcaaggaaacgcgcgtggtggccagccacgatagccgcgtgccttcgtct
 tgcagttcattcagggcaccggacagggtcggtcttgacaaaaaagacggggcgccctgcgctgacagcc
 ggaacacggcgccatcagagcagccgattgtctgtttgtgcccagtcatagcgaatagcctctccacca
 agcggcgaggaaacctgcgtgcaatccatcttgttcaatcatgcgaacgatcctcatcctgtctcttgga
 tcgacttttgaaaaagcctaggcctccaaaaaagcctcctcactactcttggaatagtcagaggccag
 gcggcctcggcctcgcataaaaaaattagtcagccatggggcgaggagaatgggaggaacgggagc
 gagttagggggcggaatggggcgagttagggcgggactatgggtgctgactaattgagatgcatgctttg
 catacttctgctgctggggagcctggggactttccacacccgtgtgctgactaattgagatgcatgctt
 tgcatacttctgctgctggggagcctggggactttccacacccgtgtgctgactaattgagatgcatgctt
 tcttccgcctcaggactcttcttttcaatattattgaagcatttatcagggttattgtctcatgagc
 ggatacatatttgaaatgtatttagaaaaataaacaataaggggttcgcgcacatttcccgaaaaagtc
 cacttgacgcgccttgtagcggcgcaatgaagcgcggcggtgtggtgggttacgcgcagcgtgacgcctac
 acttgcagcgccttagcgcgccttcttctgcttcttcccttcttctcgcgcagcgttcgcgcgctt
 ccccgctcaagctctaaatcggggcctccctttagggttccgatttagtgctttacgcttgcagcctcga
 aaaaacttgattaggggtgatgggttcacgtagtgggccatcgccctgatagacgggttttctgccttttgac
 gtggaggtccacgttcttaataagtggtactcttgttccaaactggaacaacactcaaccctatctcggtc
 tattcttttgattataagggattttgcgatttcggcctattgggttaaaaaatgagctgatttaacaaa
 aatttaacgcgaatttttaacaaaatatttaacgcttgaacatttaacgcgttaagatacattgagtggtg
 acaaacacaaactagaatgcagtgaaaaaaatgctttattttgtgaaatttgtgatgctattgctttattt
 gtaaccattataagctgcaataaaacagtttaacaacaacaattgcattcattttatgttttcagggttcagg
 gggaggtgtgggaggttttttatcaagcttatcgataccgctcgacccctaagaaaccattattatcatgac
 attagctataaaaaataggcgctatcacgagggccttctgcttcaagaatttaattcgagctcgtttacgt
 ggactgtgctggtttgggacgggctcccgctggggcccgggcgccggggcgggggcgggcggtggcggg
 cggcgggccatggagacagagagcgtgcccgggtggtagagtttgacaggcaagcatgtgctgacaggg
 cgagtagtgcttgcctgtctaaactcgttagtctcggcgcggggggcgccgggctgtccgcgcgcacccgt
 ttaaggggcgcgcgcgacccccgggggtgtgttttggggggggcccggttttcggcgcttcggcgctc
 tccccgcgtctcccccgctctctcccccgctctctcccccgctctctcccccgctctctcccccgct
 ctctcccccgctctctcccccgctctctcccccgctctctcccccgctctctcccccgctctctcccc
 cgctccccgggccccggccccccagcccgccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc
 tgc
 acgtcacacgtcacgtcatccaccacacctgcccacaacacaaactcacagcgacaactcacccgcgcaac
 aactcctgttctcatccacacgtcacccgcgcacctccccgcctctccagacgtaccccggcgaacacac
 cgctcctgtctacacacacccgcgcctccccagcccccagccctccccagcccccagccctccccggcccc
 cccctcccgccccagccctccccggcgctccccgcgcctccccggccccagccctccccagccctcccc
 cccggccccagccctccccggcgctccccgcgcctccccgggggggttcgggcatctctacccagtgcc
 gccaatctcaggtcagagatccagacatgataagatacattgatgagtttggaacaacacaactagaatg
 cagtgaaaaaaatgctttattttgtgaaatttgtgatgctattgctttattttgtaaccattataagctgca
 ataaacaagtctcaggtttgtcagaagcagacacagcggttggaataatagcagagaacagagaaat
 agcggcaaaaaataataccgtatcacttttgcctgatattggttgatgctatgtagcgaatcgggaaaaac
 gggaggtaggctccatgataaaaaagtaaaagaaaaagaataaacggaacatccaaaagtttgtgtttt
 ttaaatagtagataatggatttcttaccgcgaataacgggcagacatggcctgcccgggttattattttt
 ttgacaccagaccaactggtaattggtagcagccggcgctcagctggaatttttaatttaattccagcacact
 ggccggcgttactagtggtatgggttttaaccgatccagacatgataagatacattgatgagtttggaaca
 accacaactagaatgcagtgaaaaaaatgctttattttgtgaaatttgtgatgctattgctttattttgtaa
 ccattataagctgcaataaaacagttggcgcgtttactttgtacagctcgtccatgcccagagatgattccg
 gggcggtcaggaactccagcaggaccatgtgatcgcgtctctcgttgggtcttttgctcagggcggaact
 ggggtgctcaggttagtggttgcggcgagcagcagggcgctcgcgcgatgggggtgttctgctggttagtg
 gtcggcgagctgcacgctgcgcgtctcgtatgttggcggtatcttgaagttcaccttgatgcccgttcttc
 tgccttgcggccatgatataagacgttgcgtggtgtagttgtactccagcttgcctccaggtatgttgc
 cgtcctccttgaaatcgatgccccctcagctcgtatgcggttaccaggggtgctgccttcgaacttcacctc
 ggcgcgggtcttgcgttgcgtctccttggaagaagatgggtgcgcctcctggagctagcctcggcagtg
 gcggacttgaaagagtcgtgctgcttcatgtggtcggggtagcggctgaagcactgcacgcgtaggtca
 ggggtggtcacgaggggtggggcagggcacgggcagcttgccggtggtgagatgaacttcaggggtcagctt
 gccgtaggtggcatcgccctcgcctcgcggacacgctgaacttgggcccgtttacgtcgcgcgtccagc
 tccagaggtgggacaccccggtgaacagctcctcgccttgcctaccatgggtggcagatggatcagc
 ttatcgacagaggttctctctcgtgggaaggcacgaacggggtgagccccctctccgcccccgcgtc
 cccccctctcgc
 ctctccacccccgcgtccccccctctctcgcgccacccaaggtgcttaccggtgcaaaaaaggcggaaccg

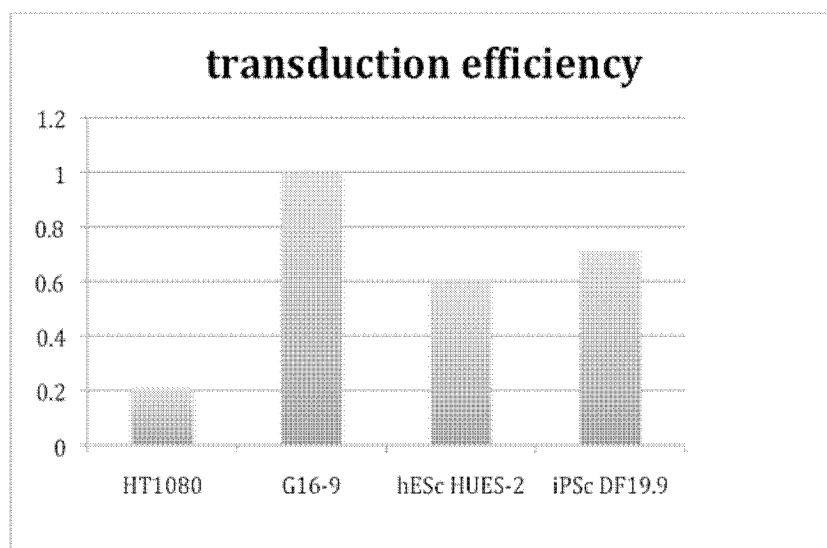
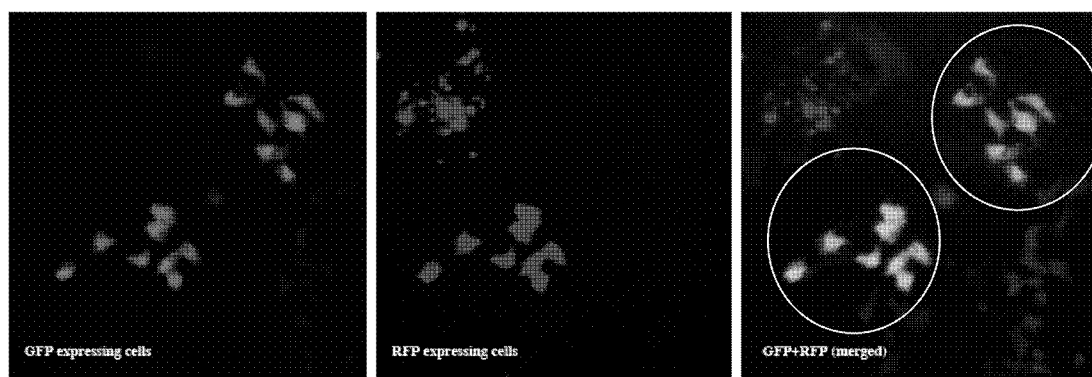
Fig. 15a

23/24

gtgggtttctgtcgtcggaggccccgggggtgcgtccctgtgtttcgtgggtgggggtgggcgggtcttt
 cccccccgctccgctgtccctttccgatgcgategccgatcccgagccggggcgctcgcatgccgacgc
 cgtccgctccgacggccctctgagactcccgctcccggtccgctgtctccgcagccgctcccgctgttcg
 tggccggcgccgtctgctggggcgctcggtcgcgcggggcctttatgtgctgctggagagacccgcccccgcc
 gccccggccccgccccggggcgggcgggagtcgggcacggcgccagtgctcgcacttcgcccataataat
 atatatatatattgggacgaagtgcgaacgcttcgcgttctcacttcttttaccggcgcccgccccctt
 gggcggttcccgcccgccgccaatggggggggcggaaggcgggcgcccttggggcgcccgccgctcccg
 ttggtcccggttccggcgggcgggacccggggggcccggggacggccaacgggcgcgcggggctcgtatc
 tcatcaccgcccgaacgggaagtccggggcccgggcccccgccccctgcccgttccctcgttagcatgcggaa
 cgggaagcggaaaccgcccggatccggcggttaattgagatgccatgcggggcgggggcgccgacccacccgccc
 tcgcgccccgcccattggcagatggcgcggtatggcgggggcggggggttcgaccaacgggcgcgcggccacg
 ggcccccgccgtgcggcgctcgggcggggtcgtgcataaatggccggctgcctcgcgcgtttccgggtgatg
 acggtgaaaacctctgacacatgcagctcccgagacgggtcacagcttctgtgtaagcggatgcggggag
 cagacaagcccgctcagggcgctcagcgggtgtttggcggggtgtcggggcgccagccatgacccagtcacgt
 agcgatagcggagtgatgggggatccggaaccccttaataataacttcgtataatgtatgctatacgaagt
 tattaggtccctcgacctgcaggcatgcaagcttcttatagtgctacctaataatcgtatgtgtatga
 tacataaggttatgtatttaattgtagccgcttctaacgacaataatgtacaagccataattgtgtagcatc
 tggcttactgaagcagaccctatcactctctcgttaactgcgctcagagtcgggttgggttggaacgaacc
 ttctgagtttctggttaacgcccgttcgcaccccggaatgggtcagcgaaccaatcagcagggctcatcgt
 agcccatggctaattctgtcagccgttaagtgttccctgtgtcactgaaaattgctttgagaggtctctaag
 ggcttctcagtgcttaccatccctggcttgttgtccacaaccggttaaaccttaaaagctttaaaagcctt
 atatatctctttttttcttataaaacttaaaaccttagaggctatttaagttgctgatttatattaattt
 tattgttcaacatgagagcttagtacgtgaaacatgagagcttagtacgttagccatgagagcttagta
 cgttagccatgaggggttagttcgtttaaactgagagcttagtacgtttaaactgagagcttagtacgtg
 aaacatgagagcttagtacgtactatcaacaggttgaactgcggatcttgcggccgcaatcgggcaaatc
 gctgaatattcttttctcgcaccatcaggcacctgagtcgctgtcttttctgtgacattcagttcgc
 tgcgtcacggctctggcagtgaaatgggggtaaatggcactacaggcgccctttatggattcatgcaagg
 aaactcaaaataatacaagaaaagcccgctcagggccttcacgggcgttttatggcgggctctgtatgtg
 gtgctatctgactttttgtgttcagcagttccctgcctctgattttccagtcctgaccacttcggattat
 cccgtgacaggtcattcagactggctaattgcacccagtaaggcagcgggtatcatcaacgggggtctgacgc
 tcagtggaaacgaaaactcacgttaagggttttgggtcatgagattatcaaaaaggatcttcacctagatc
 ctittaaattaaaaatgaagttttaaatcaatctaaagtataatagtaaaccttggctcagagttacc
 aatgcttaatcagtgaggcacctatctcagcgatctgtctattttcgttcacccatagttgcctgactccc
 cgtcgtgtagataaactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagac
 ccacgctcacggctccagatttatcagcaataaaccagccagccggaaggccgagcgcagaagtggct
 ctgcaactttatccgctccatccagtcctattaattgttgcgggaagctagagtaagtagttcggcag
 taatagtttgcgcaacgttgttgcattgtctacaggcatcgtgggtgtcacgctcgtcgtttggatggct
 tcatcagctccgggttcccaacgatcaaggcgagttacatgatccccatgttgtgcaaaaaagcgggta
 gctccttcgggtcctccgatcgttgtcagaagtaagttggccgcagtggttatcactcatgggttatggcagc
 actgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactgggtgagtaactcaaccaag
 tcattctgagaatagtgtafcggcgaccgagttgctcttggccggcgtaatacgggataataaccgcgc
 cacatagcagaactttaaaagtgtctatcattggaaaacgttcttcggggcgaaaactctcaaggatctt
 accgctgtttgagatccagttcgtatgaacccactcgtgcacccaactgatcttcagcatcttttactttc
 accagcgtttctgggtgagcaaaaacaggaaggcaaaatgcccgaaaaaagggaataaggcgacacgga
 aatgttgaatactcatactcttcttttcaatattattgaagcatttatcaggggtattgtctcatgag
 cggatacatatttgaaatgtatttagaaaaataaacaataagggttccgcgcacatttccccgaaaagtgc
 ccacttgcnatcgtatgaattgatccgaagttccctattctctagaagtataggaacttc

Fig. 15b

24/24

**Fig. 16****Fig. 17**

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2012/051075

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/85 C12N5/0735 C12N5/074
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, BIOSIS, MEDLINE, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LARIN Z ET AL: "Advances in human artificial chromosome technology", TRENDS IN GENETICS, ELSEVIER SCIENCE PUBLISHERS B.V. AMSTERDAM, NL, vol. 18, no. 6, 1 June 2002 (2002-06-01), pages 313-319, XP027303358, ISSN: 0168-9525 [retrieved on 2002-06-01]	1-5,7-19
Y	the whole document	1-19
X	EP 1 536 007 A1 (JAPAN SCIENCE & TECH AGENCY [JP]) 1 June 2005 (2005-06-01)	1,3-5,7,8,11-19
Y	the whole document	1-19
	----- -/--	



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

14 September 2012

Date of mailing of the international search report

27/09/2012

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Wimmer, Georg

INTERNATIONAL SEARCH REPORT

International application No

PCT/GB2012/051075

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	OSHIMURA M ET AL: "Transfer of human artificial chromosome vectors into stem cells", REPRODUCTIVE BIOMEDICINE ONLINE, REPRODUCTIVE HEALTHCARE LTD, GB, vol. 16, no. 1, 1 January 2008 (2008-01-01), pages 57-69, XP027052086, ISSN: 1472-6483 [retrieved on 2008-01-01]	1-5,7,8
Y	the whole document	1-19
Y	----- DANIELA MORALLI ET AL: "A novel human artificial chromosome gene expression system using herpes simplex virus type 1 vectors", EMBO REPORTS, vol. 7, no. 9, 1 September 2006 (2006-09-01), pages 911-918, XP55037829, ISSN: 1469-221X, DOI: 10.1038/sj.embor.7400768 the whole document	9-19
Y	----- EPSTEIN A L: "Progress and prospects: Biological properties and technological advances of herpes simplex virus type 1-based amplicon vectors", GENE THERAPY, vol. 16, no. 6, June 2009 (2009-06), pages 709-715, XP002683402, ISSN: 0969-7128 the whole document	7-19
Y	----- WO 2009/033653 A1 (MINITUEB ABFUELL UND LABORTECH [DE]; SCHINDELHAUER DIRK [DE]; WOLF ECK) 19 March 2009 (2009-03-19) the whole document	7-19
X,P	----- M. A. MANDEGAR ET AL: "Functional human artificial chromosomes are generated and stably maintained in human embryonic stem cells", HUMAN MOLECULAR GENETICS, vol. 20, no. 15, 18 May 2012 (2012-05-18), pages 2905-2913, XP55037807, ISSN: 0964-6906, DOI: 10.1093/hmg/ddr144 the whole document	1-19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2012/051075

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 20-23
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
see FURTHER INFORMATION sheet PCT/ISA/210
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box II.2

Claims Nos.: 20-23

As laid out in Rule 6.2 PCT, claims shall not refer to the figures unless in exceptional cases where this is absolutely necessary. The references in claims 20-23 do not represent such an exceptional case, and it is moreover open to widest interpretation what the scope of the corresponding methods or entities "with reference to the figures" might be. This lack of clarity (Art. 6 PCT) is such that a meaningful search of claimed matter is impossible.

The applicant's attention is drawn to the fact that claims relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure. If the application proceeds into the regional phase before the EPO, the applicant is reminded that a search may be carried out during examination before the EPO (see EPO Guideline C-VI, 8.2), should the problems which led to the Article 17(2) declaration be overcome.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/GB2012/051075

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1536007	A1	01-06-2005	AU 2003261863 A1 29-03-2004
			CA 2501708 A1 18-03-2004
			EP 1536007 A1 01-06-2005
			JP 4293990 B2 08-07-2009
			US 2007004002 A1 04-01-2007
			US 2008235817 A1 25-09-2008
			WO 2004022741 A1 18-03-2004

WO 2009033653	A1	19-03-2009	DE 102007043131 A1 12-03-2009
			WO 2009033653 A1 19-03-2009
