

METHOD FOR DETECTING OR MEASURING THE IMPACT OF A VIRAL  
VECTOR COMPOSITION ON EUKARYOTIC CELLS AND BIOMARKERS  
USED THEREOF

5           The invention relates generally to the field of viral vector mediated gene therapy and recombinant gene expression. More particularly, the present invention relates to methods and compositions that utilize a particular panel of gene products ("biomarkers") and their differential expression patterns ("expression signatures"), wherein the expression patterns correlate with the  
10   impact a viral vector composition may have on any given target cell, in particular on any given eukaryotic cell. The invention is based on the identification of a specific set of biomarkers that are differentially expressed in viral vector transduced cells and which are useful in predicting the quality, i.e., purity and/or concentration, of a viral vector composition. The gene panel is also useful in  
15   designing specific adjuvant modalities with improved therapeutic efficiency.

          Virus-based vectors are a tool commonly used to deliver genetic material into cells. Such vectors were originally developed as an alternative to transfection of naked DNA for molecular genetics experiments and for therapeutic uses such as gene therapy or vaccines. Viral vectors fall into two  
20   main categories: integrating vectors, which insert themselves into the recipient genome and non-integrating vectors, which usually form an extra chromosomal genetic element. Integrating vectors such as gamma-retroviral vectors (RV) and lentiviral vectors (LV) are stably inherited. Non-integrating vectors, such as adenoviral vectors (ADV) and adeno-associated virus (AAV) vectors are quickly  
25   lost from cells that divide rapidly.

          In particular, retroviral vectors are derived from viruses belonging to the Retroviridae family that comprises enveloped RNA viruses with a complex macromolecular structure having an hydrodynamic diameter of approximately 150nm (Salmeen et al. 1975). Due to the large size the viral particles have low  
30   diffusivity ( $10^{-8}$  cm<sup>2</sup>/s); their density is about 1.15-1.16 g/cm<sup>3</sup> as determined by sucrose gradient ultracentrifugation (Coffin et al. 1997). They are composed by 60-70% protein, 30-40% lipid (derived from the plasma membrane of the

producer cells), 2-4% carbohydrate and 1% RNA (Andreadis et al. 1999). Retroviral particles consist of two identical copies of single-stranded positive sense RNA, plus integrase and reverse transcriptase enzymes, contained within a protein capsid surrounded by a lipid bilayer membrane. The lipid bilayer is  
5 studded with glycoprotein projections. Retroviral vectors are negatively charged particles in a broad pH range since their isoelectric point occurs at very low pH values. The envelope proteins and the lipid bilayer are probably the main contributors to the negative charge at the virus surface (Rimai et al. 1975).

The use of virus-based vectors has become a crucial delivery  
10 method for in vitro applications in drug discovery, for in vivo and ex vivo clinical assays, for gene therapy and animal model development. The lentiviral vectors, such as HIV-derived vectors, are currently the preferred tools for gene transfer, both in vitro and in vivo, due to their capacity to transduce immortalized cells or primary cells, both quiescent or proliferating cells, and to the resulting stable  
15 integration of the transgene in the genome. These advantages make them not only a valuable tool in both the therapeutic context, as well as in functional genomics, but also for the production of molecules of interest for human use. The benefit of such vectors for their use in gene therapy has been confirmed by the recent success obtained in the treatment of Adrenoleukodystrophy (N.  
20 Cartier, S. Hacein-Bey-Abina, et al. 2009) or human  $\beta$ -thalassaemia (Marina Cavazzana-Calvo et al. 2010).

Current research in regenerative medicine also uses virus-based vectors, especially for the production and differentiation of induced pluripotent stem cells (iPS) (Sommer et al. 2009). Although it is possible to reprogram  
25 somatic cells into iPS using crude vector compositions (Takahashi K. and Yamanaka S. 2006), the use of purified and concentrated vector compositions for reprogramming leads to a greater increase in yields and in clonal survival and quality (Vallier et al. 2009). Moreover, virus-based vectors are commonly used to generate cellular models as a part of the validation of therapeutic  
30 targets for drug discovery and production of recombinant proteins for therapeutic purposes. In this context, it is essential to finely characterize the changes induced in the transduced cells by the viral-based vectors. As

demonstrated by Banito A et al. (2009), cell manipulation like cell reprogramming may be slow and stochastic, suggesting the existence of barriers limiting its efficiency or stability. Here they identify senescence as one such barrier and show that ablation of different senescence effectors improves the efficiency of reprogramming. To that end, purified vectors allow one to avoid any negative impact of the viral-based vector compositions on target cells.

The cellular responses generated following infection with HIV have been extensively studied, including the transcriptional changes caused by the virus in immune cells (Giri et al. 2006). In contrast, there are few studies on transcriptional changes caused by HIV-derived lentiviral vectors. The studies to date have focused on the impact of certain stages of transduction, such as integration into the genome of the transduced cell. Some studies, i.e Zhao et al. 2004 and Mitchell et al. 2003, have characterized the global changes in expression profile induced in response to a lentiviral vector. However, all of these studies focus on specific conditions (transgene specific cell type) that do not allow one to draw a general conclusion. Particularly, no study has been conducted on the global impact and toxicity of viral vectors such as lentiviral vectors and the consequences of production process parameters, such as concentration and purification of viral vector composition, on the transcription profile of the transduced cells.

Factors influencing the choice of a particular vector include its packaging capacity, its host range, its gene expression profile, its transduction efficiency and its capacity to elicit immune responses, which is particularly problematic if repeated administrations or transductions are needed. Some of these parameters can be adjusted or controlled. One parameter or particular importance is the use of highly concentrated and highly purified vectors that allow for efficient cell transduction and avoidance of specific cell responses due to impurities in the viral vector compositions. Development of gene therapy, in vivo and ex vivo clinical assays, and drug discovery applications, highlight the need to have tools to measure or detect the effect of a viral vector composition on eukaryotic cells and also to increase the safety associated with the production of proteins for human uses through the use of such viral vectors.

Until now, the question of the toxicity associated with the use of viral vector compositions has not been globally explored nor solved. The present invention is related to the characterization and selection of transcriptional signatures, referred as biomarkers, related to potential deleterious effects in  
5 cells upon transduction by viral vector compositions. The present invention discloses assays designed to characterize the safety and quality of viral vector compositions obtained by any process as well as of genetically modified cells transduced by such viral vector compositions. The identification and modulation of genes influencing or interfering with transduction efficiency, which could  
10 affect the level of transgene expression, may be targeted to optimize the transduction process.

The present invention provides methods and compositions for characterization of global cellular changes in response to introduction of viral vector compositions into target cells and in particular into eukaryotic cells. Such  
15 global changes may be induced by the viral vector itself or may result from the environment in which the viral vector is found such as the level of purity or concentration of the viral vector compositions. As described in detail below, the methods and compositions of the invention utilize a particular panel of gene products ("biomarkers") and their differential expression patterns ("expression  
20 signatures"), wherein the expression patterns correlate with the quality of the viral vector composition impact on any given transduced eukaryotic cells.

The present invention provides a method to measure or detect the effect of a viral vector composition on eukaryotic cells by detecting modified gene expression profiles and by screening for modifications in the expression  
25 level of at least one biomarker in said cultured cells.

The present invention relates to methods and compositions that utilize a particular panel of biomarkers and their expression signatures, wherein the expression signatures correlate with the effect that a viral vector composition may have on any given target cell, in particular eukaryotic cell. The  
30 invention is based on the identification of a specific set of biomarkers that are differentially expressed in viral vector transduced eukaryotic cells and which are useful in predicting the quality of the viral vector composition. Accordingly, the

present invention provides methods of predicting the cellular response in cells transduced with a viral vector composition. Specifically, a method is provided for detecting a cellular response in target cells, in particular eukaryotic cells contacted with a viral vector composition comprising (i) measuring the expression level of one or more biomarkers in a eukaryotic cell that has not been contacted with a viral vector composition; (ii) measuring the expression level of one or more biomarkers in a eukaryotic cell following contact with a viral vector composition; and (iii) comparing the biomarker(s) expression in (i) to the biomarker(s) expression in (ii) wherein a change in biomarker expression level between (i) and (ii) indicates a cellular response and wherein said cellular response is correlated to the quality of the viral vector composition. In step (iii) the change in expression may be either an increase or decrease in biomarker expression. In a specific embodiment of the invention, the change in biomarker expression level between (i) and (ii) is established at a minimal absolute value of 1.3 FC (fold change  $\geq 1.3$ ) in expression, either a 1.3 fold increase in expression or a 1.3 decrease in expression. The FC value is correlated to the quality (concentration/purification) of the viral vector composition and to the conditions of transduction, in particular to the MOI.

The identified correlation between biomarker expression and viral vector transduction provides a method for determining the quality of the viral vector composition. The methods of the invention rely on measurement of the expression level of one or more predictive RNA transcripts or their expression products in a transduced cell wherein the predictive RNA transcript or their product is the transcript or product of one or more genes selected from the group consisting of the genes of Tables 2, 3, 4, 5, 6, 7 or 8.

In an embodiment of the present invention, the impact of a viral vector composition on target cells, in particular eukaryotic cells transduced with a viral vector composition is evaluated. In a specific embodiment of the present invention, the cultured cells are transduced by lentiviral vectors. In another embodiment of the invention, the eukaryotic cells are immortalized, primary or stem cells.

In one embodiment of the present invention, the expression level of a biomarker involved in senescence biological processes is measured. In a preferred embodiment, the biomarker involved in senescence is a SASP family member. In a more preferred embodiment, the senescence gene is one or more  
5 genes selected from Table 7.

In another embodiment of the present invention, the expression level of a biomarker involved in cell cycle is measured. In a preferred embodiment, the cell cycle biomarker is one or more genes selected from Table 2, Table 3, and Table 4.

10 In another embodiment of the present invention, the expression level of a biomarker selected from Table 8 is measured.

In another embodiment of the present invention, the invention further provides a method for assessing the quality of a viral vector composition for a transgene transfer into target cells comprising measuring the expression level of  
15 at least one biomarker selected in the group consisting of CXCL2 and EREG and/or of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK.

In a particular embodiment, the present invention provides a method  
20 for assessing the quality of a viral vector composition for a transgene transfer into target cells comprising:

(a) measuring the expression level of at least one biomarker selected in the group consisting of CXCL2 and EREG in target cells that has not been contacted with the viral vector composition;

25 (b) measuring the expression level of said at least one biomarker in target cells following contact with said viral vector composition; and

(c) comparing said biomarker(s) expression in (b) to the biomarker(s) expression in (a) wherein a significant upregulation in biomarker expression level in (b) compared to (a) indicates that the quality of the viral vector  
30 composition is insufficient.

More particularly, a significant upregulation at high MOI is a two fold upregulation compared to (a).

The method according to the present invention can further comprise a step of:

(d) measuring the expression level of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK in target cells that has not been contacted with the viral vector composition;

(e) measuring the expression level of said at least one biomarker in target cells following contact with said viral vector composition; and

(f) comparing said biomarker(s) expression in (e) to the biomarker(s) expression in (d) wherein a significant downregulation in biomarker expression level in (e) compared to (d) indicates that the quality of the viral vector composition is insufficient.

More particularly, a significant downregulation at optimal MOI is a 1,5 fold downregulation compared to (d).

It should be noted that alternatively, steps (d), (e), (f) can be performed before or simultaneously with step (a), (b) and (c).

The method according to the present invention can further comprise a step of:

(g) measuring the expression level of at least one biomarker selected in the group consisting of CXCL2 and EREG in target cells following contact with a control viral vector composition; and

(h) comparing said biomarker(s) expression in (g) to the biomarker(s) expression in (a), wherein no significant differential expression in biomarker(s) has to be detected in (g) compared to (a).

The method according to the present invention can further comprise a step of:

(j) measuring the expression level of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK in target cells following contact with a control viral vector composition; and

(k) comparing said biomarker(s) expression in (j) to the biomarker(s) expression in (d) wherein a potential downregulation in biomarker expression level in (j) compared to (d) has to be detected.

More particularly, the potential downregulation in biomarker expression level in (j) compared to (d) is at least 1,5 times less than the downregulation in biomarker expression level in (e) compared to (d) at high MOI.

The method according to the present invention can further comprise beforehand a step of titration of the viral vector composition and the control viral vector composition.

Advantageously, in the method for assessing the quality of a viral vector composition for a transgene transfer into target cells, the measures of the biomarker expression level are performed before the cells reach confluency.

Preferably, the target cells are eukaryotic cells and are transduced by a lentiviral vector composition.

The invention also provides kits for measuring the level of biomarker expression in a sample of transduced cells. The kits may include one or more reagents corresponding to the biomarkers described herein, e.g., antibodies that specifically bind the biomarkers, recombinant proteins that bind biomarker specific antibodies, nucleic acid probes or primers that hybridize to the biomarkers, etc. In some embodiments, the kits may include a plurality of reagents, e.g., on an array, corresponding to the biomarkers described herein. The kits may include detection reagents, e.g., reagents that are detectably labeled. The kits may include written instructions for use of the kit in predicting the quality of a viral vector composition, and may include other reagents and information such as control or reference standards, wash solutions, analysis software, etc.

In a preferred embodiment, the kit comprises one or more reagents able to measure the expression level of at least one biomarker selected in the group consisting of CXCL2 and EREG.

Advantageously, said kit further comprises one or more reagents able to measure the expression level of at least one biomarker selected in the

group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK.

5 Preferably, said kit can also further comprise a control viral vector composition.

The present invention further provides a biochip consisting of at least a biomarker selected in the group consisting of CXCL2 and EREG and at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1,  
10 OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK with optionally a ubiquitous gene.

The present invention still further provides a RT-qPCR plate comprising primers of at least a biomarker selected in the group consisting of CXCL2 and EREG and primers of at least one biomarker selected in the group  
15 consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK.

The present invention further relates to a biomarker composition useful for the measurement or detection of the effect of a viral vector  
20 composition on eukaryotic cells in order to determine the quality of such compositions. In a preferred embodiment of the present invention, the biomarker composition useful for the measurement or detection of the effect of a viral vector composition on eukaryotic cells comprises at least one of the products selected among the genes or the polypeptides present in Tables 2, 3,  
25 4, 5, 6, 7 and 8. In a more preferred embodiment, the biomarker composition comprises at least one of the products selected among the genes or the polypeptides present in Tables 3, 4, 7 and 8.

#### DETAILED DESCRIPTION OF THE INVENTION

30 The present invention provides a specific set of biomarkers that are differentially expressed in viral vector transduced cells. Such biomarkers, as

described in detail below, may be used in methods designed to predict the quality of viral vector compositions.

In the present application, terms are employed with their usual meaning, except when precised otherwise.

5

### **Definitions**

The term 'viral vector composition' refers to any viral derived composition obtained from, but not limited to, retrovirus, lentivirus, adenoviral, adeno-associated virus and all compositions containing viral vectors thereof. In a preferred embodiment of the invention, the viral vector compositions are based on viruses belonging to the Retroviridae family that comprises enveloped RNA viruses including, for example, lentiviral (LV) and gamma-retroviral (RV) vectors. Viral vector compositions can be produced using any of the methods known to those of skill in the art. In a specific embodiment of the invention, viral vector compositions to be tested using the methods of the present invention may be obtained by processes such as those described in Figures 1A, 1B and 1C.

The term 'measuring or detecting the impact of a viral vector composition' means evaluating the expression of a biomarker following contact of a target cell or a eukaryotic cell with a viral vector composition.

The terms 'biomarker' or 'biological marker' mean an indicator of a biological state. It is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. As used herein a "biomarker" is a molecular indicator of a specific biological property and as used herein is a nucleic acid molecule (e.g., a gene or gene fragment) or an expression product thereof (e.g., a RNA, microRNA, a polypeptide or peptide fragment or variant thereof) whose differential expression (presence, absence, over-expression or under-expression relative to a reference) within a cell predicts the quality of a viral vector composition. An "expression product" as used herein is a transcribed sense or antisense RNA molecule (e.g., an mRNA), or a translated polypeptide corresponding to or derived from a polynucleotide

sequence. A "panel" of biomarkers is a selection of two or more combinations of biomarkers.

Biomarkers for characterizing a viral vector composition, according to the invention, include those listed in Tables 2, 3, 4, 6, 7 and 8. Such markers  
5 include genes that are found to be regulated by transduction of viral vector compositions into a cell. One or more of these biomarkers, or up to all of the biomarkers, may be used together in any combination in the methods according to the invention.

The terms 'nucleic acid' or "polynucleotide" are intended to include  
10 DNA molecules such as cDNA or genomic DNA, and RNA molecules, such as mRNA or any fragment of DNA or RNA of interest. These terms refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form including, for example, genomic DNA, cDNA, and mRNA. This term encompasses nucleic acid molecules of both natural and  
15 synthetic origin as well as molecules of linear, circular, or branched configuration representing either the sense or antisense strand, or both, of a native nucleic acid molecule. It is understood that such nucleic acids can be unpurified, purified, or attached, for example, to a synthetic material such as a bead or column matrix. The term also encompasses nucleic acids containing  
20 known analogues of natural nucleotides that have similar binding properties as the reference nucleic acid and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (e.g., degenerate codon substitutions), polymorphisms, alleles, and  
25 complementary sequences as well as the sequence explicitly indicated. The term nucleic acid is used interchangeably with gene, cDNA, and mRNA encoded by a gene or a fragment thereof which is selected as functionally equivalent to the complete molecule.

Accordingly, the present invention provides compositions comprising  
30 biomarkers, e.g., nucleic acid molecules and expression products thereof, or means for detecting said biomarkers, wherein the biomarkers are found to be

differentially expressed in viral vector transduced cells as compared to non-transduced cells.

5 Nucleic acid sequences encoding the biomarkers of the invention, are publicly available (for example, accessible in Genbank), known to those of skill in the art, and incorporated herein in their entirety. As described in detail below, such nucleic acid sequences may be used to design probes or primers for use in assays for measuring the levels of biomarker expression in a transduced cell.

10 Biomarkers according to the invention include substantially identical homologues and variants of the nucleic acid molecules and expression products thereof described herein, for example, a molecule that includes nucleotide sequences encoding polypeptides functionally equivalent to the biomarkers of the invention, e.g., sequences having one or more nucleotide substitutions, additions, or deletions, such as allelic variants or splice variants or species  
15 variants or molecules differing from the nucleic acid molecules and polypeptides referred to in the Tables 2, 3, 4, 6, 7, or 8 herein due to the degeneracy of the genetic code.

Other nucleic acids for use in the practice of the invention include those that have sufficient homology to those described herein to detect  
20 expression by use of hybridization techniques. Such polynucleotides preferably have about or 95%, about or 96%, about or 97%, about or 98%, or about or 99% identity with the biomarker sequences as described herein. The other polynucleotides for use in the practice of the invention may also be described on the basis of the ability to hybridize to polynucleotides of the invention under  
25 stringent conditions of about 30% v/v to about 50% formamide and from about 0.01M to about 0.15M salt for hybridization and from about 0.01M to about 0.15M salt for wash conditions at about 55 to about 65°C, or higher, or conditions equivalent thereto.

30 The terms 'polypeptide' or 'protein' refers to a polymer of amino acids without regards to the length of the polymer. Thus, peptides, oligopeptides and proteins are included within the definition of polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino

acid, polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

5 The term 'cultured cells' means eukaryotic cells grown under controlled conditions, generally outside of their natural environment, including primary cells, cells lines and transgenic cells. The cultured cells may, or may not be, transduced by a viral vector composition. The term 'target cells' means cells that are tested for expression of biomarkers by the method of the invention, or the kit of the invention. Target cells are preferably eukaryotic cells, more preferably immortalized cells, primary cells or stem cells. Same type, same  
10 culture conditions. For example, target cells are foreskin cells.

In particular, target cells can be permissive or non-permissive cells. The term « permissive cells » are target cells which are transduced with a viral vector composition at optimal Multiplicity of Infection (MOI) value less than or equal 40. The term « non-permissive cells » are target cells which are  
15 transduced with a viral vector composition at optimal MOI value more than 40.

Multiplicity of infection (MOI) is a frequently used term in virology which refers to the number of virions that are added per cell during infection.

The term "optimal MOI" means the appropriate MOI to transduce target cells. The optimal MOI is determined by a range of MOI on target cells,  
20 using a reporter gene expressing viral vector such as reporter gene expressing lentiviral vector. Reporter gene is but not limited to fluorescent reporter gene such as GFP or luminescent reporter gene such as luciferase.

The optimal MOI can be determined by the skilled person, based on several criteria such as, but not limited to:

- 25
- Percentage of transduced cells with a reporter gene expression level,
  - Viability of transduced cells, and/or
  - Reporter gene expression level.

Percentage of transduced cells and reporter gene expression level  
30 are determined by methods that are known to one ordinary skill in the art.

In a preferred embodiment of the invention, the optimal MOI is determined by the MOI corresponding to the highest percentage of transduced

cells, a reporter gene expression level sufficient to be detectable and a non-alteration of the viability of transduced cells compared to the viability of non-transduced cells.

For example, optimal MOI for fibroblast cells, such as foreskin  
5 fibroblast cells (also called "foreskin cells"), is 40.

The term "transfection" refers to the process of deliberately  
introducing nucleic acids into cells. The term is used strictly for non-viral  
methods in eukaryotic cells. Transfection is used in the process of viral vector  
production when gag-pol and env expressing plasmids are transfected on  
10 producer cells to get viral vectors in the supernatant.

The term "transduction" is the process of deliberately introducing  
nucleic acids into cells. The term is used for viral based methods in eukaryotic  
cells. Viral vectors are harvested from the producer cells and are contacted with  
the eukaryotic cells to obtain the finally transduced cells.

15 The terms "fold change" or "FC" represent the ratio between the  
expression level of cells transduced with a viral vector composition versus  
expression level of non-transduced cells or transduced with different batches of  
viral vector compositions.

The term 'cellular senescence' refers to stable cell cycle arrest  
20 accompanied by a set of characteristic morphological and physiological features  
that distinguish senescent cells from proliferating cells, as well as arrested  
quiescent or terminally differentiated cells (Kosar et al. 2011).

The term 'SASP' means "Senescence-Associated Secretory  
Phenotype" and is defined as a set of proteins secreted by cells undergoing  
25 cellular senescence.

The term "cell cycle" means the sequence of events within the cell  
between mitotic (cell) divisions. The cell cycle is conventionally divided into five  
phases: G0 (the gap); G1, (the first gap); S (the synthesis phase, during which  
the DNA is synthesized and replicated); G2 (the second gap); and M (mitosis).

30 The present invention provides a novel method for evaluating or  
measuring the effect of viral vector compositions on target cells, preferably  
eukaryotic cells by screening for modifications in the expression level of at least

one biomarker in said target or eukaryotic cells. The method of the invention can be used to assess the quality of the viral vector compositions by establishing whether the viral vector composition has an impact on cultured cells. The target or eukaryotic cells can be cultured cells transfected to produce  
5 the viral vector compositions or cultured cells transduced with the viral vector compositions. The biomarkers can be cellular nucleic acids and/or cellular proteins or their selected fragments. The goal is to determine if viral vector composition itself, rather than the transgene carried by the viral vector, affects cultured cells wherein said affect can depend on the concentration and purity of  
10 the viral vector composition. Indeed, to date methods have been lacking for the skilled artisan to compare the effect of viral vector compositions on the viability and/or toxicity of the target or eukaryotic cells.

Moreover, viral vector compositions can have different concentrations and titres. The titre of the composition is a very important  
15 parameter of the composition as it determines the Multiplicity of Infection (MOI) applied to cultured cells. The titre determination depends on several factors, i.e the measurement techniques and data processing. Typically, investigators have focused on vector pseudotyping or transduction protocol optimizations to improve the transduction efficiency (Janssens et al., 2003) although the use of  
20 higher MOI is the clue to reaching high transduction levels. However, since such a batch B-S induces cell toxicity (Selvaggi et al., 1997; Reiser, 2000; Baekelandt et al., 2003), the results of transduction efficiency with this type of product B-S are always a balance between the transduction level and the resulting toxicity on target cells. Furthermore, another drawback of published  
25 retroviral or lentiviral vectors concentrated by classical techniques is the inability of transduced stem cells, particularly for hematopoietic stem cells, to progress down differentiation pathways after transduction. Thus, the present invention provides methods for determining the effect of viral vector compositions, at different MOIs, on transduced cells. The present invention allows one to  
30 determine the range of MOI corresponding to a given viral vector composition and also optimize the protocols used for cell transduction. Particularly, the use of biomarkers of the present invention permits one to define the appropriate

range of MOI that allows an efficient transduction without inducing deleterious or undesirable effects on transduced cells, as cell cycle arrest or cellular senescence.

To study the mechanisms underlying the cellular response of transduced cells to viral vector compositions, a model system was developed based on the discovery of differentially expressed biomarkers in transduced cells, wherein the pattern of biomarker expression correlates with the quality of the viral vector composition (viral vector without cDNA). Accordingly, the present invention is based on the identification and validation of a number of biomarkers whose expression is modified after contacting with a viral vector composition. The invention is based on data generated with different viral vector compositions (different levels of concentration/purity) and different values of MOI to identify reliable biomarkers to measure the impact of viral vector compositions on cultured eukaryotic cells. In particular, the differences on expression levels of transduced cells genes has been compared between expression levels of non-transduced cells genes versus transduced cells genes with a viral vector without cDNA. Some genes identified as biomarkers are retested via microarrays experiments and/or RT-qPCR quantification of expression level in cells transduced with viral vectors compositions compared to non-transduced cells. For these validation experiments, a particular attention is given to the titration of viral vectors compositions. For example, viral vector compositions are titrated three times at different moments.

The present invention concerns a profile of biomarkers corresponding to molecules which are up or down regulated in transduced target cells, preferably eukaryotic cells, by a viral vector composition according to the invention. In a preferred embodiment of the invention, the change in biomarker expression level between non-transduced cells and transduced cells is established at a minimal absolute value of 1.3 FC (fold change  $\geq 1.3$ ) in expression, either a 1.3 fold increase in expression or a 1.3 decrease in expression. The FC value is correlated to the quality of the viral vector composition and to the conditions of transduction, in particular to the MOI. The profile of biomarkers is useful to characterize the biological specific state of the

transduced cells according to the invention. The profile comprises the biomarkers of the Table 2, Table 3, Table 4, Table 6, Table 7 and Table 8.

Comparison of gene expression levels between pre-transduced and post-transduced cells identified a gene expression signature composed of over-expressed and under-expressed genes. Accordingly, the present invention provides methods for assessing the quality of a viral vector composition. The methods of the invention rely on measurement of the expression level of one or more predictive RNA transcripts or their expression products in a transduced cell, normalized against the expression level of the RNA transcripts or their expression products in an untransduced cell, or against a reference set of RNA transcripts or their expression products, wherein the predictive RNA transcript is the transcript of one or more genes selected from the group consisting of the genes listed in Table 2, Table 3, Table 4, Table 6, Table 7, Table 8. While individual biomarkers are useful in assessing the quality of a viral vector composition, the combination of biomarkers as proposed herein, enables a more accurate determination of the quality of a biomarker composition.

Expression levels of the biomarkers in a sample may be determined by comparison to a suitable "control" or "reference" sample. For example, the relative expression level of markers in viral vector transduced cell may be determined with reference to the expression level of the same markers in a non-viral vector transduced cell. If the expression level of markers is greater or less than that of the reference, markers expression may be said to be "increased" or "decreased", respectively. Additionally, it is possible that the expression levels may remain constant between the control or reference and the sample.

The term "significant" change (upregulation or downregulation) is a change in the expression level which is important enough to be interpreted by the skilled person as meaningful. In particular, the change in the expression level is significant when the expression level of the transduced cells doesn't correspond to the expression level in the non-transduced cells taking into account the margin of error. In a preferred embodiment of the invention, a significant change in biomarker expression level between non-transduced cells and transduced cells is established at a minimal absolute value of 1.3 FC (fold

change  $\geq 1.3$ ) in expression, either a 1.3 fold increase in expression or a 1.3 decrease in expression. This minimal absolute value of 1.3 FC was considered for the first analysis with biochips. For validation experiment, independent t-tests were performed with Benjamini-Hochberg multiple test correction and a  
5 corrected p-value  $< 0.05$ . Probes with absolute value of fold changes (FC)  $\geq 1.5$  were retained as differentially expressed for both up and down-regulated probes.

Samples for analysis in such methods can be any target cell, preferably eukaryotic cell or eukaryotic cell extract. Such eukaryotic cells  
10 include cell lines, cultured cells, primary cells, stem cells.

As described in detail below, expression of the biomarkers within a cell may be evaluated by any suitable means. For example, expression may be evaluated using DNA microarrays. Alternatively, RNA transcripts may be measured using real time PCR, or, when RNA corresponds to a coding gene,  
15 protein products may be detected using suitable antibodies. Methods of determining expression levels of genes by these and other methods are known in the art.

In the interest of brevity, Applicants are not expressly listing every possible combination of gene products suitable for use in the methods of the  
20 invention. Nevertheless, it should be understood that every such combination is contemplated and is within the scope of the invention. It is specifically envisioned that any combination of gene products listed in Tables 2, 3, 4, 6, 7 or 8 that were found to be differentially expressed between a control or reference, for example, untransduced cells, and the transduced cells may be particularly  
25 useful for analysis.

In a particular embodiment, the present invention further provides a method for assessing the quality of a viral vector composition for a transgene transfer into target cells comprising:

(a) measuring the expression level of at least one biomarker selected  
30 in the group consisting of CXCL2 and EREG in target cells that has not been contacted with the viral vector composition;

(b) measuring the expression level of said at least one biomarker in target cells following contact with said viral vector composition; and

(c) comparing said biomarker(s) expression in (b) to the biomarker(s) expression in (a) wherein an significant upregulation in biomarker expression level in (b) compared to (a) indicates that the quality of the viral vector composition is insufficient.

By "transgene", it is more specifically intended any nucleic acid of interest. A transgene is but not limited to reporter gene (GFP, luciferase...), any gene or combination of gene(s) portion(s) or sequence of interest such as shRNA (short hairpin RNA) or miRNA (micro RNA).

The term "insufficient" quality means that the viral vector composition to be tested modifies some characteristics of the transduced cells compared to non-transduced cells. Examples of modified characteristics are but not limited to the proliferation and/or the viability of the transduced cells compared to non-transduced cells. In particular, these characteristics are not modified when the viral vector composition is a concentrated and purified viral vector composition, such as a viral vector composition produced without serum, concentrated and purified by tangential ultrafiltration and diafiltration, such as obtained by the method described in patent application WO 2013/014537.

More specifically, said method for production of viral vector composition comprises the steps of :

- transfection of a producer cell, modified to complement deletions in the RNA viral genome upon which the viral vector is based, and culturing the producer cells under suitable conditions to permit the production of viral vector particles, wherein said culturing following transfection is conducted in serum free medium;
- collecting the supernatant containing said viral vector particles; and
- purifying the supernatant by tangential ultrafiltration and diafiltration, the ultrafiltration being preferably operated on polysulfone hollow-fiber cartridges.

In particular, this method of production of viral vector composition doesn't comprise any step of sodium butyrate induction.

Advantageously, the supernatant collection is performed by multiple steps comprised between 3 and 6, at specific time intervals. The supernatant  
5 collection is followed by clarification by centrifugation.

The method of production may further comprise a step of ion-exchange chromatography.

This method of production allows obtaining a purified viral vector composition comprising less than 2% of initial protein contaminants and 10 to  
10 30%, preferably less than 10%, of initial DNA contaminants, compare to the crude viral vector composition as present in the cell serum free medium.

In particular, said viral vector composition is capable of transducing target cells, in particular eukaryotic cells, without affecting cell viability and/or have little to no effect on cell proliferation, viability, and/or the ability of cells,  
15 such as stem cells to differentiate or such as primary cells to be reprogrammed into pluripotent cells.

More particularly, in the viral vector composition obtained by the above mentioned production method, the physical particles/transducing units (PP/TU) is usually comprised of between 100:1 up to 900:1, preferably between  
20 100:1 up to 600:1, more preferably between 100:1 up to 400:1.

In the present method for assessing the quality of a viral vector composition for a transgene transfer into target cells, a significant upregulation at high MOI is preferably a two fold upregulation compared to (a).

Advantageously, the expression level of the at least one biomarker  
25 can be measured by reverse transcription quantitative PCR (RT-qPCR)

The term "high MOI" means a very higher MOI than the optimal MOI. The high MOI is determined by a range of MOI using a reporter gene expressing viral vector such as reporter gene expressing lentiviral vector. Reporter gene is but not limited to fluorescent reporter gene such as GFP or  
30 luminescent reporter gene such as luciferase.

Preferably, the high MOI corresponds to:

- at least three times the optimal MOI for non-permissive cells,

- at least four times for permissive cells.

For example, high MOI for fibroblast cells, such as foreskin fibroblast cells, is at least 120, such as 150.

The method for assessing the quality of a viral vector composition for a transgene transfer into target cells can further comprise a step of:

(d) measuring the expression level of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK in target cells that has not been contacted with the viral vector composition;

(e) measuring the expression level of said at least one biomarker in target cells following contact with said viral vector composition; and

(f) comparing said biomarker(s) expression in (e) to the biomarker(s) expression in (d) wherein a significant downregulation in biomarker expression level in (e) compared to (d) indicates that the quality of the viral vector composition is insufficient.

More particularly, a significant downregulation at optimal MOI is a 1,5 fold downregulation compared to (d).

Advantageously, the expression level of the at least one biomarker can be measured by reverse transcription quantitative PCR (RT-qPCR).

The method according to the present invention can further comprise a step of:

(g) measuring the expression level of at least one biomarker selected in the group consisting of CXCL2 and EREG in target cells following contact with a control viral vector composition; and

(h) comparing said biomarker(s) expression in (g) to the biomarker(s) expression in (a), wherein no significant differential expression in biomarker(s) has to be detected in (g) compared to (a).

This significant differential expression in biomarker(s) has not to be detected even at high MOI.

The term "control viral vector composition" means a highly concentrated and purified viral vector composition. It can be obtained by, but

not limited to single or successive tangential ultrafiltration diafiltration, by ion-exchange chromatography, exclusion chromatography. In a preferred embodiment of the invention, the control viral vector composition is such as obtained by the process of production described in patent application WO  
5 2013/014537 (see for example "batch C" in this PCT application) and/or described above.

The method according to the present invention can further comprise a step of:

(j) measuring the expression level of at least one biomarker selected  
10 in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK in target cells following contact with a control viral vector composition; and

(k) comparing said biomarker(s) expression in (j) to the biomarker(s)  
15 expression in (d) wherein a potential downregulation in biomarker expression level in (j) compared to (d) has to be detected.

More particularly, the potential downregulation in biomarker expression level in (j) compared to (d) is at least 1,5 times less than the downregulation in biomarker expression level in (e) compared to (d) at high  
20 MOI.

The method according to the present invention can further comprise beforehand a step of titration of the viral vector composition and the control viral vector composition.

Titers of viruses in general, and lentiviral based vectors in particular,  
25 depend on the method and cells used for titration. The quantification of vector particles capable of achieving the steps of the transduction pathway from cell entry to gene integration and gene expression depends on the vector itself and cell characteristics.

Concerning the cells used for vector titration, it is important to ensure  
30 that the target cells are readily permissive, as it was demonstrated that the permissivity of all the cell types are not equivalent. Another point is that the transduction efficiency must be easily monitored for reliable quantification for

any transgenes and vectors over time. Here, in each titration experiment a standard GFP expressing lentiviral vector is quantified in terms of efficient units both by FACS (represented by the number of Transducing Units per ml TU/ml) and qPCR (represented by the number of Integrated Genome per ml IG/ml) after HCT116 transduction with serial dilutions of the vectors according to the material and methods as set forth above. Both results give a relative number of efficient particles for transduction but their respective absolute numbers do not give the same titer depending on the PCR itself and the target sequence used for amplification. These data show that it can be difficult to compare precisely these different approaches based on the functional titers in the absence of standardized methods. It can thus be of interest to include a control viral vector composition in the present method. Said control viral vector composition is thus a reference batch with preferably a known titer and a define target cell type.

In parallel, the determination of total particles is quantified with the P24 Elisa kit to estimate the total vector particles, even those that do not contain any genomic RNA and/or that are devoid of envelope proteins. Both titers are useful to determine the ratio between the physical particles PP that reflect the total particles and the biological titer that gives the real transduction ability. This ratio gives an estimation of the vector purity and integrity. Another ratio is used to reflect the vector integrity or infectivity and is expressed as the number of IG per ng P24 (1 ng of P24 corresponds to 107 PP).

Advantageously, in the method according to the invention, the measures of the biomarker expression level are performed before the cells reach confluency.

Preferably, the target cells are eukaryotic cells and are transduced by a lentiviral vector composition.

To determine the (increased, decreased) expression levels of the above described biomarkers in the practice of the present invention, any method known in the art may be utilized. In one preferred embodiment of the invention, expression based on detection of RNA which hybridizes to a "probe" or "primer" specific for the biomarkers described herein is used. A "probe" or "primer" is a single-stranded DNA or RNA molecule of defined sequence that

can base pair to a second DNA or RNA molecule that contains a complementary sequence (the target). The stability of the resulting hybrid molecule depends upon the extent of the base pairing that occurs, and is affected by parameters such as the degree of complementarity between the probe and target molecule, and the degree of stringency of the hybridization conditions. The degree of hybridization stringency is affected by parameters such as the temperature, salt concentration, and concentration of organic molecules, such as formamide, and is determined by methods that are known to those skilled in the art.

Probes or primers specific for the nucleic acid biomarkers described herein, or portions thereof, may vary in length by any integer from at least 8 nucleotides to over 500 nucleotides depending on the purpose for which, and conditions under which, the probe or primer is used. Probes or primers specific for the nucleic acid biomarkers described herein may have greater than 20-30% sequence identity, or at least 55-75% sequence identity, or at least 75-85% sequence identity, or at least 85-99% sequence identity, or 100% sequence identity to the nucleic acid biomarkers described herein. Probes or primers may be derived from genomic DNA or cDNA, for example, by amplification, or from cloned DNA segments, and may contain either genomic DNA or cDNA sequences representing all or a portion of a single gene from a single individual. Probes or primers may be chemically synthesized.

A probe or primer may hybridize to a nucleic acid biomarker under high stringency conditions as described herein. "Stringent hybridization conditions" as used herein mean conditions under which a first nucleic acid sequence (e.g., probe) will hybridize to a second nucleic acid sequence (e.g., target), such as in a complex mixture of nucleic acids. Stringent conditions are sequence-dependent and will be different in different circumstances. Stringent conditions may be selected to be about 5-10°C. Lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength pH. The  $T_m$  may be the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are

present in excess, at  $T_m$ , 50% of the probes are occupied at equilibrium). Stringent conditions may be those in which the salt concentration is less than about 1.0 M sodium ion, such as about 0.01-1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g., about 10-50 nucleotides) and at least about 60° C for long probes (e.g., greater than about 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal may be at least 2 to 10 times background hybridization. Exemplary stringent hybridization conditions include the following: 50% formamide, 5XSSC, and 1% SDS, incubating at 42°C, or, 5XSSC, 1% SDS, incubating at 65°C, with wash in 0.2XSSC, and 0.1% SDS at 65°C.

Probes or primers can be detectably-labeled, either radioactively or non-radioactively, by methods that are known to those skilled in the art. By "detectably labeled" is meant any means for marking and identifying the presence of a molecule, e.g., an oligonucleotide probe or primer, a gene or fragment thereof, or a cDNA molecule. Methods for detectably-labeling a molecule are well known in the art and include, without limitation, radioactive labeling (e.g., with an isotope such as  $^{32}\text{P}$  or  $^{35}\text{S}$ ) and nonradioactive labeling such as, enzymatic labeling (for example, using horseradish peroxidase or alkaline phosphatase), chemiluminescent labeling, fluorescent labeling (for example, using fluorescein), bioluminescent labeling, or antibody detection of a ligand attached to the probe. Also included in this definition is a molecule that is detectably labeled by an indirect means, for example, a molecule that is bound with a first moiety (such as biotin) that is, in turn, bound to a second moiety that may be observed or assayed (such as fluorescein-labeled streptavidin). Labels also include digoxigenin, luciferases, and aequorin.

Probes or primers can be used in biomarker detection methods involving nucleic acid hybridization, such as nucleic acid sequencing, nucleic acid amplification by the polymerase chain reaction (e.g., RT-PCR), single stranded conformational polymorphism (SSCP) analysis, restriction fragment polymorphism (RFLP) analysis, Southern hybridization, northern hybridization,

in situ hybridization, electrophoretic mobility shift assay (EMSA), fluorescent in situ hybridization (FISH), and other methods that are known to those skilled in the art.

5 A preferred biomarker detection method is reverse transcription quantitative PCR (RT-qPCR).

A preferred embodiment using a nucleic acid based assay to determine biomarker expression is by immobilization of one or more biomarker sequences identified herein on a solid support, including, but not limited to, a solid substrate as an array or to beads or bead based technology as known in  
10 the art. Alternatively, solution based expression assays known in the art may also be used. The immobilized sequence(s) may be in the form of polynucleotides as described herein such that the polynucleotide would be capable of hybridizing to a DNA or RNA corresponding to the biomarker sequence(s).

15 The immobilized polynucleotide(s) may be used to determine the biomarker expression signature in a sample of transduced and non-transduced cells. The immobilized polynucleotide(s) need only be sufficient to specifically hybridize to the corresponding nucleic acid molecules derived from the sample (and to the exclusion of detectable or significant hybridization to other nucleic  
20 acid molecules).

In embodiments where only one or a few biomarkers are to be analyzed, the nucleic acid derived from a sample isolated from cells may be preferentially amplified by use of appropriate primers such that only the genes to be analyzed are amplified to reduce contaminating background signals from  
25 other genes expressed in the cell. Alternatively, and where multiple genes are to be analyzed or where very few cells (or one cell) is used, the nucleic acid from the sample may be globally amplified before hybridization to the immobilized polynucleotides. Of course RNA, or the cDNA counterpart thereof may be directly labeled and used, without amplification, by methods known in  
30 the art.

A biochip may be used in the practice of the invention.

31

32

In particular, the present invention provides a biochip consisting of at least a biomarker selected in the group consisting of CXCL2 and EREG and at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK with optionally a ubiquitous gene.

The biochip may comprise a solid substrate comprising an attached probe or plurality of probes described herein. The probes may be capable of hybridizing to a target sequence under stringent hybridization conditions. The probes may be attached at spatially defined sites on the substrate. More than one probe per target sequence may be used, with either overlapping probes or probes to different sections of a particular target sequence. The probes may either be synthesized first, with subsequent attachment to the biochip, or may be directly synthesized on the biochip.

The solid substrate may be a material that may be modified to contain discrete individual sites appropriate for the attachment or association of the probes and is amenable to at least one detection method. Representative examples of substrates include glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon J, etc.), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses and plastics. The substrates may allow optical detection without appreciably fluorescing.

Biomarker expression may also be measured based on detection of a presence, increase, or decrease in protein levels or activity may also be used. Antibody based detection methods are well known in the art and include sandwich and ELISA assays as well as Western blot and flow cytometry based assays as non-limiting examples. Antibodies for use in such methods of detection include polyclonal antibodies and monoclonal antibodies that specifically bind to the biomarkers of Tables 2, 3, 4, 6, 7 and 8.

Preferably, the antibodies for use in such methods of detection include polyclonal antibodies and monoclonal antibodies that specifically bind to at least one biomarker selected in the group consisting of CXCL2 and EREG and/or at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK.

Such antibodies, as well as fragments thereof (including but not limited to Fab fragments) function to detect such biomarkers in cells by virtue of their ability to specifically bind to such polypeptides to the exclusion of other polypeptides to produce a detectable signal. Recombinant, synthetic, and hybrid antibodies with the same ability may also be used in the practice of the invention.

In a preferred embodiment, biomarker expression may also be measured by RT-qPCR. The present invention still further provides a RT-qPCR plate comprising primers of at least a biomarker selected in the group consisting of CXCL2 and EREG and primers of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK.

Example of primers that can be used in the RT-qPCR plate for EREG and CXCL2 are given in Table 10.

The present invention is based on the identification and the validation of target cell biomarkers that can be used to assess the quality of a viral vector composition.. In a preferred embodiment, the biomarkers are selected from those genes listed in Table 2, Table 3, Table 4, Table 6, Table 7, Table 8. As described in detail below, three families of biomarkers were identified: (i) biomarkers associated with senescence, (ii) biomarkers associated with the cell cycle, and (iii) others biomarkers found to be associated with the quality of the viral vector compositions.

To identify these biomarkers, gene expression was determined using different viral vector compositions obtained as follows. Producer cells were tri-transfected with plasmids constructs as described in Figure 1C, according to

standard techniques well known to those skilled in the art. Such techniques include, for example, the calcium phosphate technique, the DEAE-dextran technique, electroporation, methods based on osmotic shock, microinjection or methods based on the use of liposomes. In a specific embodiment of the invention, the cells may be transfected using a calcium precipitation method. Such a method is preferred when 293T cells are the producer cells of choice but equivalent cells may also be used.

Following transfection, the cells are incubated in serum free media for production of batch A, B C and D (Figure 1B). Batch B-S is produced with 10% serum (Figure 1A) Cells are incubated for a time sufficient to allow for the efficient production of viral particles. The incubation time following transfection, depends on a combination of factors including, for example, the type of viral vector used and the producer cell line of choice. In a specific embodiment of the invention, multiple harvests may take place following incubation. For example, four or more vector harvests may take place. To determine, the most productive incubation conditions, small batch experiments may be performed to determine optimized conditions for generating the highest titre and purest batch of viral particles.

The initial culture supernatant, containing viral vector particles, is referred to herein as, batch A. The batch B is obtained by commonly used prior art concentration methods (Figure 1A) such as a concentration step by ultrafiltration using centrifugation ready-to-use units on the post clarification harvest. The method of obtaining a viral vector composition may further comprise the step of tangential ultrafiltration diafiltration of the batch A product for further purification of viral vector particles. Such an ultrafiltration diafiltration step is a type of membrane filtration in which hydrostatic pressure forces a liquid against a semi-permeable membrane. Suspended solids and solutes of higher molecular weight than the membrane cut off are retained, while water and lower molecular weight than the membrane cut off solutes pass through the membrane. Ultrafiltration technique is carried out by tangential flow ultrafiltration using polysulfone hollow-fiber cartridges. Such a technique allows for monitoring and adapting the pressure to ensure the maintenance of vector

integrity and viability. Such a step provides for concentration of the vector particles, as well as acting as a purification step for removal of initial contaminants, such as host cell proteins and nucleic acids, from the collected batch. Such a batch is referred to herein as batch C.

5 In yet another embodiment of the invention, following the tangential ultrafiltration and/or diafiltration step, the method of obtaining a viral vector composition may further comprise the step of ion-exchange chromatography which may be performed to further concentrate and purify the viral vector particles. Such a batch is referred to herein as batch D.

10 Biomarkers have been identified by analyzing microarray experiments targeting the nucleic acids of cells, for example, human foreskin fibroblast cells transduced with batches B (obtained by process B), C (obtained by process C), B-S (obtained by process B with serum), UC or UC-S (obtained in presence of serum and concentrated by ultracentrifugation) or non-  
15 transduced (NT).

In a specific embodiment of the invention, genes involved in cellular senescence have been identified as useful biomarkers for determining the quality of a viral vector composition. Senescence is a permanent state of cell cycle arrest that, unlike quiescence, is unresponsive to growth factors (Young et  
20 al.). Originally described in terms of the replicative exhaustion of cultured fibroblasts, it has since been shown that senescence can occur prematurely upon a plethora of cellular stresses. Cellular senescence occurs in culture and in vivo as a response to excessive extracellular or intracellular stress. Induction of DNA damage response and chromatin remodeling of the INK4a/ARF locus  
25 are two of the mechanisms behind senescence induction. Li et al. (2009) demonstrated that cell culture conditions during reprogramming enhance the expression of the Ink4/Arf locus, further highlighting the importance of silencing this locus to allow proliferation and efficient reprogramming. Limiting senescence by using a highly purified vector suspension to overexpress  
30 combinations of factors such as Oct4, Sox2, Klf4, and c-Myc may have a profound positive effect on the efficiency of iPS cell generation, increasing both the kinetics of reprogramming and the number of emerging iPS cell colonies.

The present invention discloses the use of genes involved in the senescence phenotype to evaluate the impact of viral vector compositions on target cells. In a preferred embodiment of the present invention, the biomarker involved in cellular senescence phenotype listed in Table 5 or Table 6.

5           The senescent phenotype is not limited to an arrest of cell proliferation. A senescent cell is a potentially persisting cell that is metabolically active and has undergone widespread changes in protein expression and secretion, ultimately developing the senescence-associated secretory phenotype or SASP (Coppé et al. 2010). The SASP includes several families of  
10 soluble and insoluble factors. SASP factors can be soluble signaling factors (interleukins, chemokines, and growth factors), secreted proteases, and secreted insoluble proteins/extracellular matrix components (Coppé et al. 2010). Senescent cells develop altered secretory activities that may induce changes in the tissue microenvironment, relaxing its control over cell behavior and  
15 promoting tumorigenesis (Coppé et al. 2010). In cell culture, cell cycle arrest typically leads to senescence, because the cell is over-stimulated by serum, nutrients, oncogenes and so on (Blagosklonny, 2011). The present invention discloses the use of genes involved in SASP to evaluate the impact of viral vector compositions on target cells in contact with such compositions. In a  
20 preferred embodiment of the present invention, biomarkers associated with SASP, include those listed in Table 7.

In a specific embodiment of the invention, genes involved in cell cycle have been identified as useful biomarkers for determining the quality of a viral vector composition. Cell division consists of two consecutive processes,  
25 mainly characterized by DNA replication and segregation of replicated chromosomes into two separate cells. Originally, cell division was divided into two stages: mitosis (M), i.e. the process of nuclear division; and interphase, the interlude between two M phases. Stages of mitosis include prophase, metaphase, anaphase and telophase. Under the microscope, interphase cells  
30 simply grow in size, but different techniques revealed that the interphase includes G1, S and G2 phases. Replication of DNA occurs in a specific part of the interphase called S phase. S phase is preceded by a gap called G1 during

which the cell is preparing for DNA synthesis and is followed by a gap called G2 during which the cell prepares for mitosis. G1, S, G2 and M phases are the traditional subdivisions of the standard cell cycle. Cells in G1 can, before commitment to DNA replication, enter a resting state called G0. Cells in G0  
5 account for the major part of the non-growing, non-proliferating cells in the human body. The present invention discloses the use of genes involved in cell cycle to evaluate the impact of viral vector compositions on target cells in contact with such compositions. In a preferred embodiment of the present invention, biomarkers associated with cell cycle, include those listed in Table 2,  
10 Table 3, Table 4. In a more preferred embodiment of the present invention, the screening of the modification is realized on expression level of at least one biomarker associated to cell cycle, listed in Table 4.

The present invention further discloses a class of biomarkers whose expression level in the target cells is modulated depending on the quality, i.e.  
15 concentration and purification, of the viral vector compositions used. The impact of viral vector compositions having different qualities and obtained by different processes was investigated to select a novel class of biomarkers. The viral vector compositions were obtained by the processes described in Figure 1A and 1B. In a preferred embodiment of the present invention, the use of  
20 genes listed in Table 8, may be used to evaluate the impact of viral vector compositions on target cells in contact with such compositions.

The invention also provides kits for measuring the level of biomarker expression in a sample of transduced cells. The kits may include one or more reagents corresponding to the biomarkers described herein, e.g., antibodies  
25 that specifically bind the biomarkers, recombinant proteins that bind biomarker specific antibodies, nucleic acid probes or primers that hybridize to the biomarkers, etc. In some embodiments, the kits may include a plurality of reagents, e.g., on an array, corresponding to the biomarkers described herein. The kits may include detection reagents, e.g., reagents that are detectably  
30 labeled. The kits may include written instructions for use of the kit in predicting the quality of a viral vector composition, and may include other reagents and

information such as control or reference standards, wash solutions, analysis software, etc.

The present invention discloses a method for screening or detecting a nuclear or a cellular response in eukaryotic cells transduced with a viral vector composition comprising (i) measuring the expression level of one or more biomarkers in a eukaryotic cell that has not been contacted with a viral vector composition; (ii) measuring the expression level of one or more biomarkers in a eukaryotic cell following contact with a viral vector composition; and (iii) comparing the biomarker(s) expression in (i) to the biomarker(s) expression in (ii) wherein a change in biomarker expression level between (i) and (ii) indicates a cellular response and wherein said cellular response is correlated to the quality of the viral vector composition. In a preferred embodiment of the invention, the expressed one or more biomarker is a nucleic acid expressed within the eukaryotic cells in contact with a viral vectors composition.

In another embodiment of the invention, the expressed one or more biomarker is a polypeptide expressed within the eukaryotic in contact with viral vectors composition.

In a specific embodiment of the invention, the biomarker is a gene involved in the senescence biological process. Preferably, the biomarker involved in senescence is a gene of the SASP family. More preferably, the biomarker is selected from Table 7.

In a specific embodiment of the invention the biomarker is a gene of cell cycle family. Particularly, the biomarker is a gene selected from Table 2, Table 3, Table 4. More particularly, the biomarker is selected from the genes listed in Table 8.

In another embodiment of the invention, the biomarker is selected from Table 2, Table 3, Table 4, Table 7 and/or Table 8.

In one embodiment of the invention the viral vector composition is transduced into the eukaryotic cell. In another embodiment of the invention, eukaryotic cells are transduced by a lentiviral vector composition. In a specific embodiment of the invention, the eukaryotic cells are immortalized cells, primary cells or stem cells.

The present invention discloses a method for measuring or detecting the effects of a viral vector composition on eukaryotic cells according to claim 1 comprising:

- (i) contacting said cells with a viral composition of interest,
  - 5 (ii) measuring the level of biomarker expression in said cultured cells;
- and

(iii) optionally characterizing the biomarkers specific of the modifications of the nuclear or of the cytoplasmic expressions in said cells.

The present invention discloses a kit for measuring the expression of  
10 a biomarker associated with the transduction of a viral vector composition comprising one or more reagents corresponding to recognition of the biomarkers of Table 2, 3, 4, 5, 6, 7 or 8. Preferably, the reagent corresponds to the recognition of nucleic acids encoding the biomarkers of Table 2, 3, 4, 5, 6, 7 or 8. More preferably, the reagent corresponds to the recognition of  
15 polypeptides of Table 2, 3, 4, 5, 6, 7 or 8.

In a preferred embodiment, the kit comprises one or more reagents able to measure the expression level of at least one biomarker selected in the group consisting of CXCL2 and EREG.

Advantageously, said kit further comprises one or more reagents  
20 able to measure the expression level of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK.

Preferably, said kit can also further comprise a control viral vector  
25 composition.

The present invention discloses a biomarker composition useful for the measurement or detection of the effect of a viral vector composition on eukaryotic cells comprising at least one of the products selected among the genes or the polypeptides present in Table 2, Table 3, Table 4, Table 5, Table  
30 6, Table 7, Table 8. Preferably, the biomarker composition comprises at least one of the products selected among the genes or the polypeptides present in Table 3, Table 4, Table 7, Table 8.

Examples below are given with reference to the following figures and tables:

### FIGURES

**Figure 1A.** Common vector production method with serum by state of art processes. **Figure 1B.** Viral vector concentration and purification process used to obtain the viral vector compositions B and C. The different processes are sequential (from A to D corresponding to the obtaining of batches A to D) to meet the target cells concentration and purification requirements: immortalized cells (A), primary and stem cells (C) and *in vivo* injection (D). Process B (corresponding to obtained batch B) represents the state of the art process. **Figure 1C.** Plasmid harboring gag and pol genes, envelope expressing helper plasmid and transgene expression plasmid without cDNA.

**Figure 2A.** Summary of batch B and C titers used for the transcriptomics study. Characteristics of rLV-EF1 without cDNA batches B and C used for the transcriptomics study. (1) Transducing units (TU) were determined by qPCR. (2) Physical particles (PP) were quantified by HIV-p24 ELISA in order to determine PP/TU ratio. **Figure 2B.** Foreskin cells growth 48h post-transduction with batch B and C vectors. Foreskin cells 48 hours after transduction with an empty cassette carrying lentiviral vector (rLV-EF1 without cDNA) at MOI 40 or MOI 150. Batch B and C of rLV-EF1 were derived from the same crude harvest.

**Figure 3A.** Scatterplot of differential probes in cells transduced with batch B at MOI 150 vs non-transduced cells. Scatterplot representing differentially expressed probes, with an absolute value of  $FC \geq 1.5$ , in rLV-EF1 batch B transduced cells at MOI 150 versus non-transduced cells. X-axis represents normalized intensities for Non-Transduced (NT) cells, and Y-axis normalized intensities for Transduced (T) cells. Very light grey tone lines are fold change lines representing fold changes values of -2, 1 and 2. **Figure 3B.** Scatterplot of differential probes in cells transduced with batch C at MOI 150 vs non-transduced cells. Scatterplot representing differentially expressed probes, with an absolute value of  $FC \geq 1.5$ , in rLV-EF1 batch C transduced cells at MOI 150 versus non-transduced cells. X-axis represents normalized intensities for

Non-Transduced (NT) cells, and Y-axis normalized intensities for Transduced (T) cells. Very light grey tone lines are fold change lines representing fold changes values of -2, 1 and 2.

**Figure 4.** Venn diagram: downregulated probes in cells transduced with batch B vs non-transduced cells and downregulated probes in cells transduced with batch C vs non-transduced cells MOI 150. Venn diagram showing the intersection between the list of downregulated probes ( $FC \leq -1.5$ ) in cells transduced with batch B at MOI 150 compared to non-transduced cells, and the list of downregulated probes ( $FC \leq -1.5$ ) in cells transduced with batch C at MOI 150 compared to non-transduced cells.

**Figure 5.** Profile plot of cell cycle probes in cells transduced with batch B versus non-transduced cells at MOI 150 with  $FC \leq -3$ . Profile plot representing differential probes having  $FC \leq -3$  when comparing batch B transduced cells at MOI 150 with non-transduced cells. A baseline transformation was applied on intensity values before representing data. Normalized intensity values are plotted in the following conditions: batch B transduced at MOI 150 cells, batch C transduced cells at MOI 150, non-transduced cell.

**Figure 6A.** RT-qPCR validation of the downregulation of 5 cell cycle genes in cells transduced with rLV-EF1 batch B at MOI 150 and cells transduced with rLV-EF1 batch C at MOI 150 compared to non-transduced cells. RT-qPCR was performed on RNA from rLV-EF1 batch B and rLV-EF1 batch C transduced cells at MOI 150 and RNA from non-transduced cells. RT-qPCR fold changes, calculated from threshold cycle ( $C_t$ ) values in each condition versus values in non-transduced cells, using the  $2^{-\Delta\Delta C_t}$  method, are plotted in the 3 conditions (NT: Non-Transduced cells, Batch C: rLV-EF1 batch C transduced cells at MOI 150, Batch B: rLV-EF1 batch B transduced cells at MOI 150). **Figure 6B.** Comparison of Fold Changes values obtained with RT-qPCR and from microarray experiments for the 5 previously cited validated genes. All fold changes are calculated comparing expression values in cells transduced with rLV-EF1 batch B or batch C at MOI 150 versus expression values in the non-transduced cells. (1) HUGO Gene symbol, (2) Batch used for

transduction at MOI 150, (3) Fold Change value calculated from Ct values using the  $2^{-\Delta\Delta CT}$  method and transformed into equivalent negative value ( $-1/FC$ ), (4) Fold change value from microarray experiment.

**Figure 7A.** Cells transduced with batch B obtained with serum (batch B-S) compared to non-transduced cells. Foreskin cells 48 hours after transduction with an empty cassette carrying lentiviral vector (rLV-EF1 without cDNA) produced in the presence of serum at MOI 40 and MOI 150. **Figure 7B.** Characteristics of batch B-S of rLV-EF1 without cDNA used for the transcriptomics study. (1) Transducing units (TU) were determined by qPCR. (2) Physical particles (PP) were quantified by HIV-p24 ELISA in order to determine PP/TU ratio.

**Figure 8.** Scatterplot of differential probes in cells transduced with batch B obtained with serum (B-S) versus non-transduced cells at MOI 150. Scatterplot representing differentially expressed probes, with an absolute value of  $FC \geq 1.5$ , in rLV-EF1 without cDNA batch B-S transduced cells at MOI 150 versus non-transduced cells. X-axis represents normalized intensities for Non-Transduced (NT) cells, and Y-axis normalized intensities for Transduced (T) cells. Very light grey tone lines are fold change lines representing fold changes values of -2, 1 and 2.

**Figure 9.** Scatterplot of differential probes in cells transduced with batch B obtained with serum (B-S) versus non-transduced cells at MOI 40. Scatterplot representing differentially expressed probes, with an absolute value of  $FC \geq 1.5$ , in rLV-EF1 without cDNA batch B-S transduced cells at MOI 40 versus non-transduced cells. X-axis represents normalized intensities for Non-Transduced (NT) cells, and Y-axis normalized intensities for Transduced (T) cells. Very light grey tone lines are fold change lines representing fold changes values of -2, 1 and 2.

**Figure 10.** Venn diagram: Modulated probes in cells transduced with batch B obtained with serum (B-S) at MOI 150 vs non-transduced cells and modulated probes in cells transduced with batch B obtained with serum (B-S) at MOI 40 vs non-transduced cells. Venn diagram showing the intersection between the list of differentially expressed probes (absolute value of  $FC \geq 1.5$ )

in cells transduced with batch B-S at MOI 150 compared to non-transduced cells, and the list of differentially expressed probes (absolute value of  $FC \geq 1.5$ ) in cells transduced with batch B-S at MOI 40 compared to non-transduced cells.

**Figure 11.** Profile plot representing probes impacted in cells transduced with batch B obtained with serum (B-S) at MOI40 and 150 and not differential in cells transduced with batches B and C, at MOI40 and 150. Profile plot representing probes that were differentially expressed in cells transduced with rLV-EF1 without cDNA batch B-S at MOI 40 and MOI 150 compared to non-transduced (NT) cells, and that were not differential in cells transduced with rLV-EF1 without cDNA batch B and C transduced cells at MOI 40 and 150 compared to non-transduced cells. By “not differential”, is meant that the FC absolute value is  $< 1.3$ . A baseline transformation was applied on intensity values before representing data. B-40: cells transduced with batch B at MOI 40. B-150: cells transduced with batch B at MOI 150. C-40: cells transduced with batch C at MOI 40. C-150: cells transduced with batch C at MOI 150. NT: non transduced cells. B-S-40: cells transduced with batch B obtained with serum at MOI 40. B-S-150: cells transduced with batch B obtained with serum at MOI 40.

**Figure 12.** Characterization of rLV-EF1-GFP vectors batch used (B, C, B-S and UC).

**Figure 13.** Photos of Foreskin fibroblast cells transduced with rLV-EF1-GFP vectors. Cells were observed 48 hours after transduction.

**Figure 14.** Probe selected in figure 5 (downregulated cell cycle genes with  $FC \leq -3$  with B-MCS-MOI150 vs NT) represented in the following conditions: NT, B-GFP-MOI40, C-GFP-MOI40.

**Figure 15.** Profile plot showing behavior of probes corresponding to table 3 after transduction with rLV-EF1-GFP batch B and C at MOI 40 versus Non Transduced (NT) cells.

**Figure 16.** Profile plot representing the probes corresponding to the 18 genes from table 4 and MKI67 after transduction with rLV-EF1-GFP vector at MOI 40.

**Figure 17.** Profile plot representing the 18 cell cycle genes from table 4 and MKI67 after transduction with rLV-EF1-GFP vector (B, B-S and C) at MOI 40.

**Figure 18. Validation of some cell cycle genes as biomarkers by RT-qPCR.** mRNA relative quantification values are represented as mean  $\pm$  SD calculated from 3 samples per condition. Fold change values for each condition compared to NT are indicated above each diagram bar when differences are significant (t-test unpaired, p-value < 0.05).

**Figure 19. Validation of CXCL2 and EREG genes as biomarkers by RT-qPCR.** mRNA relative quantification values are represented as mean  $\pm$  SD calculated from 3 samples per condition. Fold change values for each condition compared to NT are indicated above each diagram bar when differences are significant (t-test unpaired, p-value < 0.05).

## 15 TABLES

**Table 1.** Primers used for RT-qPCR validations. For each tested gene and the GAPDH reference gene (represented by their respective gene symbol), the sequences of the forward and reverse primers are shown.

**Table 2.** Cell cycle genes downregulated in cells transduced with rLV-EF1 without cDNA batch B and batch C at MOI 150 compared to non-transduced cells. (1) HUGO gene symbol (2) Gene description from NCBI (National Center for Biotechnology Information) (3) Agilent Probe Identifier, (4) Nucleic sequence Accession Number from the NCBI database RefSeq RNA, (5) Gene Ontology (GO) category "M phase of mitotic cell cycle", (6) GO "G1 phase" category, (7) GO "G2 phase" category, (8) GO "S phase" category, (9) GO "G1/S transition of mitotic cell cycle" category, (10) GO "G2/M transition of mitotic cell cycle" category, (11) GO "M/G1 transition of mitotic cell cycle" category. Genes are sorted by increasing values of Fold Change in batch B transduced cells at MOI 150 versus non-transduced cells.

**Table 3.** Genes annotated in the "cell cycle" Gene Ontology category and which are upregulated in cells transduced with rLV-EF1 batch B and batch C at MOI150 compared to non-transduced cells. (1) HUGO gene symbol (2)

Gene description from NCBI (3) Agilent Probe Identifier, (4) Nucleic sequence Accession Number from the NCBI database RefSeq RNA, (5) Proteic sequence Accession Number from the NCBI database RefSeq Protein. Genes are sorted by decreasing values of Fold Change.

5                   **Table 4.** Genes annotated in the “cell cycle” Gene Ontology category, which are downregulated in cells transduced with rLV-EF1 batch B at MOI 40 and 150 and which are downregulated in cells transduced with rLV-EF1 batch C compared to non-transduced cells at MOI150 but not at MOI 40. (1) HUGO gene symbol (2) Gene description from NCBI (3) Agilent Probe Identifier,  
10 (4) Nucleic sequence Accession Number from the NCBI database RefSeq RNA, (5) Proteic sequence Accession Number from the NCBI database RefSeq Protein.

**Table 5.** List of genes related to cellular senescence, either being part of the “Senescence and autophagy” pathway referenced as WP615 in the  
15 publically accessible Wikipathway database or as extracted from the literature as being characteristic as the Senescence-Associated Secretory Phenotype (SASP). (1) HUGO gene symbol (2) Gene part of the SASP (3) Gene included into the “Senescence and autophagy” Wikipathway (4) Gene extracted as related to the SASP from Coppé *et al.*, 2010 (5) Gene extracted as associated  
20 to the SASP from Young *et al.*, 2009. [ If we are merely citing to Wikipathway as the source of our current list of used genes that’s okay---our genes are listed in the table. We just can’t rely on Wikipathway for future searching. Whatever you cite to in the specification has to be permanent as of the filing date]

**Table 6.** Cellular senescence-associated genes (either being part of  
25 the human “senescence and autophagy pathway » referenced WP615 in the publically accessible Wikipathway database or part of the list of SASP genes extracted from the literature (Table 5)) which are differentially expressed genes in cells transduced with rLV-EF1 batch B-S at MOI 150 compared to non-transduced cells, which are not differential in cells transduced with rLV-EF1  
30 batch B and batch C transduced cells at MOI 150 compared to non-transduced cells. By “not differential”, is meant that the FC absolute value is < 1.3. (1) HUGO gene symbol (2) Gene description from NCBI (3) Agilent Probe Identifier,

- (4) Nucleic sequence Accession Number from the NCBI database RefSeq RNA,  
 (5) Proteic sequence Accession Number from the NCBI database RefSeq Protein.

**Table 7.** Ten cellular senescence-associated biomarkers. These  
 5 genes were selected from Table 6 as being differentially expressed in cells  
 transduced with rLV-EF1 batch B-S at MOI 40 compared to non-transduced  
 cells and not differential in rLV-EF1 batch B or C transduced cells at MOI 40. By  
 "not differential", is meant that the FC absolute value is  $< 1.3$ . (1) HUGO gene  
 symbol (2) Gene description from NCBI (3) Agilent Probe Identifier, (4) Nucleic  
 10 sequence Accession Number from the NCBI database RefSeq RNA, (5) Proteic  
 sequence Accession Number from the NCBI database RefSeq Protein. (6) to  
 (11) fold change values in each transduced condition compared to the non-  
 transduced control condition. ND: not statistically differential. FC B-S vs NT MOI  
 150: Fold Change obtained between cells transduced with batch B-S at MOI  
 15 150 versus non-transduced cells. FC B vs NT MOI 150: Fold Change obtained  
 between cells transduced with batch B at MOI 150 versus non-transduced cells.  
 FC C vs NT MOI 150: Fold Change obtained between cells transduced with  
 batch C at MOI 150 versus non-transduced cells. FC B vs NT MOI 40: Fold  
 Change obtained between cells transduced with batch B at MOI 40 versus non-  
 20 transduced cells. FC C vs NT MOI 40: Fold Change obtained between cells  
 transduced with batch C at MOI 40 versus non-transduced cells. FC B-S vs NT  
 MOI 40: Fold Change obtained between cells transduced with batch B-S at MOI  
 40 versus non-transduced cells.

**Table 8.** Selection of genes not impacted with a high-quality vector.  
 25 (1) HUGO gene symbol (2) Gene description from NCBI (3) Agilent Probe  
 Identifier, (4) Nucleic sequence Accession Number from the NCBI database  
 RefSeq RNA, (5) Proteic sequence Accession Number from the NCBI database  
 RefSeq Protein. (6) to (11) fold change values in each transduced condition  
 compared to the non-transduced control condition. ND means not statistically  
 30 differential. FC B-S vs NT MOI 150: Fold Change obtained between cells  
 transduced with batch B-S at MOI 150 versus non-transduced cells. FC B vs NT  
 MOI 150: Fold Change obtained between cells transduced with batch B at MOI

150 versus non-transduced cells. FC C vs NT MOI 150: Fold Change obtained between cells transduced with batch C at MOI 150 versus non-transduced cells. FC B vs NT MOI 40: Fold Change obtained between cells transduced with batch B at MOI 40 versus non-transduced cells. FC C vs NT MOI 40: Fold Change obtained between cells transduced with batch C at MOI 40 versus non-transduced cells. FC B-S vs NT MOI 40: Fold Change obtained between cells transduced with batch B-S at MOI 40 versus non-transduced cells.

**Table 9.** Validation of cell cycle genes of table 4 as biomarkers with rLV-EF1-GFP MOI 40.

**Table 10.** Sequence of primers used in the validation experiment.

### EXAMPLES

The examples below are provided to help better understand the invention although the invention is not to be limited to these examples.

#### **Material and Methods**

**Plasmid construction.** Three plasmids were used to produce a recombinant virion or recombinant retrovirus. A first plasmid provides a nucleic acid encoding a viral gag and pol gene (Figure 2A). These sequences encode a group specific antigen and reverse transcriptase, (and integrase and protease-enzymes necessary for maturation and reverse transcription), respectively, as discussed above. A second plasmid provides a nucleic acid encoding a viral envelope (env) (Figure 2B), such as VSV-G (Vesicular Stomatitis Virus G). A third plasmid provides the *cis*-acting viral sequences necessary for the viral life cycle (Figure 2C). This third plasmid also contains a cloning site for a heterologous nucleic acid sequence to be transferred to a target cell. A schematic illustration of a suitable vector is shown in Figure 2C with the GFP as a transgene but which can be replaced by any gene or sequence of interest such as cDNA, shRNA or miRNA.

**Viral vectors manufacturing processes. Cell lines and culture conditions.** Viral vectors were produced using a Human Embryonic Kidney (HEK293T) cell line. A Human colon carcinoma (HCT116; ATCC N° CCL-247) adherent cell-line is used for quantification of infectious particles. All cells were

provided by the American Type Culture Collection (ATCC) and cultured in Dulbecco's Modified Eagle's Medium (DMEM, Gibco, Paisley, UK) supplemented with 10% FCS; 1% penicillin/streptomycin and 1% ultraglutamine (PAA) at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> in air. For the production of viral vector supernatants, DMEM was only supplemented with 1% penicillin/streptomycin and 1% ultraglutamine (PAA).

**Viral vectors production.** Viral vector production was performed in a 10-layer CellSTACK (6320cm<sup>2</sup>, Corning). HEK293T cells were seeded at 9.5x10<sup>3</sup> viable cells/cm<sup>2</sup> in DMEM supplemented with 10% FCS; 1% penicillin/streptomycin and 1% ultraglutamine (PAA) and placed at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> in air. Four days after seeding, the supernatant was discarded and replaced by fresh DMEM without FCS supplemented with 1% penicillin/streptomycin and 1% ultraglutamine (PAA) before transfecting the cells.

The tri-transfection mix was composed by the following three plasmids: pENV, pGagPol (viral DNA construct contained in the bacterial host deposited at CNCM Collection respectively under the accession number CNCM I-4487 and CNCM I-4488), and pLV-EF1 (viral DNA construct derived from that contained in the bacterial host deposited at CNCM Collection under the accession number CNCM I-4489). The final concentration was adjusted to 40mg/ml-1 using sterile water. CaCl<sub>2</sub> (2.5M) was then dripped to the plasmid-water mixture under soft checking to reach a final concentration of 500mM. The obtained mixture was then dripped to an equivalent volume of Hepes Buffered Saline (HBS 2X) and incubated at room temperature for 20 minutes. After incubation, the transfection mixture was added to the cell culture media and incubated for 24 hours at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> in air.

After 24 hours post-transfection, the supernatant was discarded and replaced with fresh non-supplemented DMEM and the cells were incubated at 37°C in a humidified atmosphere of 5% CO<sub>2</sub> in air. After medium exchange, the supernatant was collected several times (32h, 48h, 56h and 72h post transfection). Some fresh and no supplemented media were added and the

cells were incubated prior to further harvests at 37°C in a humidified atmosphere of 5% CO<sub>2</sub>.

Each harvest was clarified by centrifugation for 5min. at 3000g before being microfiltered through 0.45µm pore size sterile filter unit (Stericup, Millipore). The whole set of harvest were then pooled to supply the crude harvest to obtain the Batch A (crude viral vectors composition).

**Viral vectors concentration and purification.** The viral vector compositions used to identify biomarkers of the present invention are obtained by a standard and commonly used concentration process based on either ultracentrifugation or centrifugation on central units (corresponding to the obtaining of batch B), or by concentration and/or purification processes (C and D) associated with a serum free production process (corresponding to the obtaining of batch A), as described in the PCT application WO 2013/014537 incorporated by reference in its entirety herein. Another batch UC (or UC-S) is produced for validation experiments obtained in presence of serum and concentrated by ultracentrifugation. The different batches correspond to different purification strategies going from no purification to several purification steps based on ultrafiltration and chromatography. The concentration and purification of the crude harvest was first performed by tangential flow ultrafiltration using polysulfone hollow-fiber cartridges. The supernatant was then diafiltered for 20 diavolumes in a continuous mode diafiltration against DMEM or TSSM buffer. Once the diafiltration performed, the retentate was recovered and further concentrated on ultrafiltration disposable units. The hollow fiber filtration (HFF) retentate was then benzonase treated by addition of Benzonase (250U/µl) for a final concentration of (72U/ml), and MgCl<sub>2</sub> (1.0mM) for a final concentration of 1µM, before being incubating at 37°C for 20 minutes.

The post HFF material can be further purified by ion exchange chromatography (IEX) on Sartobind Q75 (Sartorius) disposable membrane using an AKTA purifier system (GE Healthcare). The ion exchange membrane would be equilibrated with 5 column volumes of non-supplemented DMEM (or TSSM) at 2ml/min. The viral supernatant would be then loaded on the membrane at 2ml/min using a sampling loop. The flow through would be

collected. The following step gradient can be applied to the AKTA system: 0M, 0.5M, 1.2M and 2M NaCl. The elution pic (collected with the 1.2M NaCl step gradient) would be immediately 10x diluted in the following buffer: 20mMTris + 1.0% w/v Sucrose + 1.0% w/v Mannitol, pH7.3 and further concentrated on ultrafiltration disposable units.

**Viral vectors compositions.** Viral vectors compositions are obtained by the method described in the PCT application WO 2013/014537 incorporated by reference in its entirety herein. The resulting compositions are

- Batch B obtained after centrifugation on central units of batch A;
- Batch C obtained after tangential ultrafiltration diafiltration of batch A;
- Batch C-S obtained after tangential ultrafiltration diafiltration of batch A produced in the presence of 10% Fetal Bovine Serum (BIOWEST). This batch is only used for figure 13;
- Batch B-S obtained in the presence of 10% Fetal Bovine Serum (BIOWEST) by the same process as the batch B without serum;
- Batch UC or UC-S obtained in presence of 10% Fetal Bovine Serum (BIOWEST) and concentrated by ultracentrifugation.

The processes used to obtain these batches are described in Figures 1A and 1B.

**Functional particle quantification using qPCR.** Transduction unit titration assays were performed as follows. HCT116 cells are seeded in 96-wells plate at 12500 cells per well and 250µL of DMEM supplemented with 10% FCS; 1% penicillin/streptomycin and 1% ultraglutamine (complete medium). 24h later, five serial dilutions are performed with complete medium for each vector sample and for a known a rLV-EF1-GFP as an internal standard (viral DNA construct contained in the bacterial host deposited at CNCM Collection under the accession number CNCM I-4489). The cells are transduced by these serial dilutions in the presence of 8µg/mL Polybrene® (Sigma). For each sample series, one well of non-transduced cells is added for control. Three days post-transduction, cells are trypsinized and each cell pellet is taken up with 250µL of PBS, genomic DNA are extracted and submitted to qPCR. Results are normalized with the known rLV-EF1-GFP internal standard that was previously

titrated by FACS, using 100µL of the cell suspension. The titre is expressed by transducing units/ml (TU / mL) using the internal standard whose titre was previously determined by FACS (Canto II) using standard conditions by considering the percentage of positive cells.

5                   **Physical particle quantitation by p24 ELISA assays.** The p24 core antigen is detected directly on the viral supernatant with a HIV-1 p24 ELISA kit provided by Perkin Elmer. The kit is used as specified by the supplier. The captured antigen is complexed with biotinylated polyclonal antibody to HIV-1 p24, followed by a streptavidin-HRP (horseradish peroxidase) conjugate. The  
10                   resulting complex is detected by incubation with ortho-phenylenediamine-HCl (OPD) which produces a yellow color that is directly proportional to the amount of p24 captured. The absorbance of each microplate well is determined using microplate reader and calibrated against absorbance of an HIV-1 p24 antigen standard curve. The viral titer expressed in physical particles per ml is  
15                   calculated from the amount of p24 knowing that 1pg of p24 corresponds to  $10^4$  physical particles.

**Empty cassette vector production for microarray analyses.** Lentiviral vector without cDNA (rLV-EF1) was produced at different purities for microarray studies. Batches B and C of rLV-EF1 vectors were purified from the  
20                   same crude harvest. An additional production was achieved in the presence of 10% Fetal Bovine Serum (BIOWEST) in order to generate a B batch containing serum, hereinafter mentioned as B-S batch.

**GFP expressing lentiviral vector production.** Independent batches B, C, B-S of GFP expressing lentiviral vectors were produced.

25                   In order to provide another type of low quality concentrated vector, ultracentrifugation method was used to concentrate vectors produced in the presence of 10% serum (Batch UC or UC-S).

**Culture of foreskin cells.** Human foreskin fibroblast cells were obtained from the American Type Culture Collection (N° CRL-2097) and  
30                   cultured in EMEM (Earl's Minimum Essential Medium, GIBCO) supplemented with 10% Fetal Bovine Serum (BIOWEST), 1% penicillin/streptomycin (PAA)

and 2mM glutamine (PAA). Cells were maintained at 37°C in the presence of 5% CO<sub>2</sub> and passaged twice a week at 5 000 cells/cm<sup>2</sup>.

**Transduction of foreskin cells for transcriptomics analysis.**

Human foreskin fibroblasts were seeded at 5000 cells/cm<sup>2</sup> in T25-flasks 24 hours before transduction. Cells were transduced in quadruplicate at MOI 40 and 150 using the batches B, C and B-S of rLV-EF1 vector in a final volume of 5 mL and in the presence of 4 µg/mL of Polybrene® (Sigma). A non-transduced control only received 4 µg/mL of Polybrene®. The transduction supernatant is removed after approximately 16 h. Cells were trypsinized 54 hours post-transduction, washed with 1x PBS, centrifuged and the pellets were kept at -80°C. Pictures were taken 48 hours post-transduction.

Human foreskin fibroblasts were seeded at 5000 cells/cm<sup>2</sup> in 6 well-multiplate 24 hours before transduction. Cells were transduced in quadruplicate at MOI 40 and 150 using the batches B, C, B-S and UC (or UC-S) of rLV-EF1-GFP vector in a final volume of 5 mL and in the presence of 4 µg/mL of Polybrene® (Sigma). A non-transduced control only received 4 µg/mL of Polybrene®. The transduction supernatant is removed after approximately 16 h. Cells were trypsinized 54 hours post-transduction, washed with 1x PBS, centrifuged and the pellets were kept at -80°C. Pictures were taken 48 hours post-transduction.

**RNA extractions.** Total RNA samples were extracted from cell pellets using the TRIzol® Plus RNA Purification System (Life Technologies) according to manufacturer's instructions. Total RNA concentration and purity were determined using a Nanodrop 1000 spectrophotometer (Nanodrop Technologies). RNA quality and integrity were checked with the Agilent 2100 Bioanalyzer (Agilent Technologies, USA) and were conform to Agilent microarrays' requirements.

**DNA microarray experiments.** Microarray experiments were performed at the Biochips Platform of Genopole, University of Toulouse, INSA, UPS, INP, CNRS & INRA (Toulouse, France) according to manufacturer protocols. Briefly, after addition of a dilution of exogenous RNA from the one color RNA Spike-In Kit (Agilent Technologies) for quality control check, 100 ng

of total RNA were converted to cRNA, amplified and cyanine 3-labeled using the Agilent Low Input Quick Amp kit. 1650 ng of cyanine 3-labeled cRNA were hybridized at 65°C for 17 hours at 10 rpm to Agilent Whole Human Genome Oligo Microarrays 4x44K version 2, containing 44 000 probes (consisting of 60-mer length oligonucleotides) targeting 27 958 genes. Hybridized arrays were washed and scanned on the Agilent high-resolution scanner G2505C and the images were analyzed using Feature Extraction 10.10 (Agilent Technologies). After quality control based on Feature Extraction QC reports, 3 or 4 replicates were retained per condition. Concerning the validation experiment with rLV-EF1-GFP, the Feature Extraction 11.5 version was used.

**Microarray data statistical analyses.** Raw datasets from Feature Extraction were imported into GeneSpring® GX 12 Software (Agilent Technologies) and normalized using the 75th percentile methods. Probes were then filtered by flag values attributed by GeneSpring® when importing Feature Extraction data (for each probe, one of the following flag is affected: “detected”, “not detected” or “compromised”, using GeneSpring® default parameters). Probes detected and not compromised in more than 60% of replicates in at least one condition were retained (eliminating undetected or compromised spots). Baseline transformation of intensity values to median of all samples was applied for profile plot representations. It means that, for each probe, the median of the log summarized values from all the samples is calculated and subtracted from each of the samples. In order to identify differentially expressed probes between each condition and the control condition, independent *t*-tests were performed with Benjamini-Hochberg multiple test correction and a corrected *p*-value < 0.05. Probes with absolute value of fold changes (FC) ≥ 1.5 were retained as differentially expressed for both up and down-regulated probes.

**Microarray data functional analyses.** Annotations provided by Agilent and included in GeneSpring® are based on a dataset called ‘technology’ in GeneSpring® and named 26652 version 2012.1.10. For each probe, different type of annotations are provided including Gene symbol, Description, Gene Ontology terms, RefSeq RNA accession number among other data. It should be

noted that only one RefSeq transcript is associated with each probe, although a probe can target several alternative transcripts of the same gene. The gene ontology (GO) option on GeneSpring® GX 12 was used to determine the most significant biological processes (corrected  $p$ -value  $< 0.1$ ) represented in differentially expressed probe lists, compared with the human whole genome. Pathway analysis was used to find direct relationships between entities of interest. This was performed in GeneSpring® with the "Single Experiment Analysis" algorithm. The selected human pathway sources were curated pathways referenced in WikiPathways included by default in GeneSpring® GX 12. Pathways with a  $p$ -value  $< 0.05$  and a minimal number of 5 genes were retained.

**Relative Quantitative RT-PCR (RT-qPCR).** A total of 1  $\mu$ g of total RNA from each sample was reverse transcribed using the Superscript III RT cDNA synthesis kit (Life Technologies) and oligo(dT)<sub>12-18</sub> according to manufacturer's instructions. cDNA products were then mixed with SYBR® GreenER™ qPCR SuperMixes for ABI PRISM (Life Technologies) and specific primers synthesized by Eurogentec (Belgium). GAPDH was used as an internal control to normalize transcript levels. All primers were designed using Primer 3 software version V.0.4.0 and their characteristics are summarized in table 1. Real time PCR was performed in duplicate, from at least two independent samples, using a StepOne instrument (Applied Biosystems) and relative quantification was calculated by the  $2^{-\Delta\Delta CT}$  method (Livak *et al.* 2001). For the validation experiment, additional primers listed in table 10 were used and RT-qPCR was performed in duplicate from three independent samples. To assess significance of gene differential expressions, unpaired Student  $t$ -tests were performed to compare  $\Delta\Delta Ct$  values between conditions, with a  $p$ -value cut-off of 0.05.

## Results

### Candidate biomarkers identification

**Impact on proliferation of cultured cells transduced with a highly purified viral vector composition versus an ordinary concentrated viral vector composition.** In order to evaluate viral vector transduction effects

according to the purity level and independently from any transgene, foreskin fibroblast cells were transduced at MOI 40 and 150 with two rLV-EF1 (without cDNA) compositions, batch B and batch C (described above) derived from the same crude harvest (batch A) and whose characteristics are summarized in Figure 1A. Cells were observed 48 hours after transduction as presented in Figure 2B.

A slight growth retardation was visible at MOI 40 with batch B transduced cells compared to non-transduced cells, although no growth difference was noticeable after batch C transduction at the same MOI.

Thus, at usual MOI (MOI 40 is commonly used for foreskin transduction), highly purified viral vector composition (batch C) does not induce a visible effect on transduced cells growth, whereas ordinary concentrated viral vector composition (batch B) seems to give a negative impact on transduced cells growth.

At MOI 150, a strong proliferation arrest could be seen with batch B transduced cells compared to non-transduced cells, whereas we only observed a moderate growth retardation with batch C.

**Impact on cell the transcriptome of cells transduced with a highly purified viral vector composition versus an ordinary concentrated viral vector composition.** As an example, the ordinary concentrated viral vector composition means the batch B obtained with serum (B-S). To explore underlying changes at the transcriptional level, these cells were collected 54 hours post-transduction. This post-transduction delay of 54 hours was determined as appropriate from a preliminary study as it was between the time when a growth delay appeared in transduced cells versus non-transduced cells, and the time when non-transduced cells reached confluence (data not shown). RNA were extracted and used to perform Agilent whole human genome microarrays allowing the quantification of nearly all human transcripts.

**Transcriptional changes observed at MOI 150.** First, RNA levels from cells transduced with rLV-EF1 batch B and rLV-EF1 batch C at MOI 150 were compared to RNA from non-transduced cells. After statistical analyses,

probes upregulated or downregulated 1.5-fold or more were retained for each comparison.

As shown in Figures 3A and 3B, a number of 1027 probes were differentially expressed in transduced cells after batch B transduction and a number of 906 probes were differentially expressed in transduced cells after batch C transduction, compared to non-transduced cells. Downregulated probes were almost twice as numerous as upregulated probes (703 and 650 downregulated probes, respectively for transduced cells with batch B and batch C).

Comparison of the downregulated genes at MOI 150 shows that the majority of downregulated genes were common to the analysis of the transcriptome of cells transduced with batch B versus the transcriptome of non-transduced cells transcriptome and analysis of the transcriptome of cell transduced with batch C versus the transcriptome of non-transduced cells tr, except for a set of batch-specific genes which are specifically impacted in the transcriptome of cells transduced with batch B or the transcriptome of cells transduced with batch C. As shown on the Venn diagram in Figure 4, 560 downregulated probes represent the intersection of the two lists of downregulated probes. Thus, there is a pool of 560 probes commonly impacted in transcriptome of cells transduced with batch B or C versus non-transduced cells transcriptome.

A Gene Ontology (GO) analysis with GeneSpring® on these 560 probes revealed that cell cycle genes were significantly overrepresented with 239 probes (representing 204 distinct genes among the 1004 human genes comprised in the "cell cycle" GO category). These genes are presented in Table 2, sorted by increasing FC values. Numerous other GO terms were significantly over-represented, the majority being linked to cell cycle. In particular, all GO categories corresponding to each cell cycle phases and transitions were significantly impacted.

In order to go further with the analysis of the 560 common downregulated probes, a pathway analysis was performed on this list using GeneSpring®. The first resulting human pathway was the human "Cell cycle"

pathway referenced WP179 in the publically accessible Wikipathway database on July 2012 with 37 genes whose expression level is impacted-. "G1 to S cell cycle control" (WP45 reference in the publically accessible Wikipathway database on July 2012 database), "Mitotic G2-G2/M phases" (WP1859 reference in the publically accessible Wikipathway database on July 2012) and "Mitotic M-M/G1 phases" (WP1860 reference in the publically accessible Wikipathway database on July 2012) human pathways were also significantly impacted. Hence, all cell cycle phases seem to be impacted with major downregulations. Cell cycle arrest at the G2-M checkpoint was confirmed by downregulation of CDC25C, Cyclin B1 and CDC2 (HUGO gene nomenclature) associated with an upregulation of p21 (Chiu *et al* 2011). Other blockages need to be confirmed.

Among genes annotated in the GO category "cell cycle", the proliferation marker MKI67 was highly downregulated, with an average FC value of -8,6 and -5,6 (average FC value obtained with the values of the 3 probes representing this gene) on transcriptome of cells transduced respectively with batch B and batch C, versus non-transduced cells transcriptome . These values correlate with the observed proliferation retardation which was more pronounced after batch B transduction compared to batch C transduction.

A similar FC difference was almost systematically observed between the two conditions (i.e batch B and batch C), with an average 30% lower FC for these genes on transcription level in cells transduced with batch B versus transcription level in non-transduced cells compared to transcription level in cells transduced with batch C versus transcription level in non-transduced cells, as illustrated in Figure 5. This observation is based on selection of probes impacted in cells transduced with batch B at MOI 150, having FC values  $\leq -3$ . The FC difference was at least of 10%, except for 5 genes of this selection Downregulations of cell cycle genes are more pronounced in cells transduced with batch B versus non-transduced cells compared to cells transduced with batch C versus non-transduced cells.

Remarkably, 6 genes annotated in GO as “cell cycle” were upregulated with the two batches (i.e batch B and batch C) compared to non-transduced cells: CDKN1A, MDM2, TP53INP1, TGFB2, CDH13, RASSF2 according to HUGO gene nomenclature (as presented in Table 3). It could be noted that CDKN1A, encoding p21 protein, corresponds to a cell cycle inhibitor, and that its overexpression is stronger in cells transduced with batch B versus non-transduced cells than cells transduced with batch C versus non-transduced cells.

**Transcriptional changes observed at MOI 40.** Then, RNA levels from cells transduced with rLV-EF1 batch B and C at MOI 40 were compared to RNA of non-transduced cells. After statistical analyses, probes upregulated or downregulated 1.5-fold or more were retained for each comparison.

Among the 239 cell cycle downregulated probes at MOI 150 with rLV-EF1 batch B or C versus non-transduced cells, only 31 probes (representing 28 genes) were also under-expressed at MOI 40 with batch B versus non-transduced cells, and among them, 10 probes were downregulated with batch C versus non-transduced cells at the same MOI. FC were comprised between -1.5 and -2 at MOI 40, although they were comprised between -1.5 and -19 at MOI 150, showing that the impact on cell cycle was deeply stronger at MOI 150 compared to MOI 40.

Finally, there were 18 cell cycle genes impacted with FC values  $\leq -1.5$  only with batch B, FC being above the -1.5 cut-off for batch C at MOI 40. These genes are: ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A, TTK according to HUGO gene nomenclature (Table 4), correspond to the first cell cycle genes impacted in response to cell contact with a viral vector, and their deregulation happens earlier with a low quality vector batch compared to a highly purified vector batch. Thus, such genes could be used as early markers of an impact of a viral vector composition on the cell cycle of target cells.

**Quantitative PCR validations.** This Example describes subsequent technical validation of cell cycle genes underexpression by RT-qPCR. In order

to confirm differential expression values obtained from microarrays experiments, a set of 5 cell cycle genes (E2F8, MKI67, NEK2, AURKB, CENPA according to HUGO gene nomenclature) was chosen among the 10 more under-expressed probes in cells transduced with batch B compared to non-transduced cells at MOI 150, RT-qPCR was performed on RNA from batch B and batch C transduced cells at MOI 150 versus non-transduced cells. The results, presented in Figure 6A and 6B, confirm the under-expression of these genes, and the stronger downregulation resulting from batch B transduction compared to batch C transduction.

**Impact of a viral vector batch produced with serum on cell transcriptome.** In order to assess the effects of vector medium composition after production of the viral vector composition with serum, rLV-EF1 vector (without cDNA) was produced in the presence of 10% serum and concentrated using process B, giving a batch B-S, whose characteristics are summarized in Figure 7B. This batch was used to transduce foreskin cells at MOI 40 and MOI 150.

Cells were observed 48 hours after transduction, as shown in Figure 7A. A growth arrest of cells transduced with batch B-S compared to non-transduced cells. This growth arrest is stronger at higher MOI (MOI 150) compared to MOI 40. Remarkably, Aggregates could be observed in cells transduced with batch B-S, and their volume increases with MOI. These cells were collected 54 hours after transduction for RNA extractions and microarray hybridizations. Surprisingly, during trypsinization, the cells transduced with batch B-S were more difficult to detach than cells transduced with batch B or C or non-transduced cells. RNA levels of cells transduced with batch B-S at moderate or higher MOI were compared to RNA of non-transduced cells using Agilent whole human genome microarrays. After statistical analyses, probes upregulated or downregulated 1.5-fold or more were retained for each comparison.

**Transcriptional changes observed in cells transduced with batch B-S at MOI 150 versus non-transduced cells.** 1019 probes were significantly differential in cells transduced with batch B-S at MOI 150 compared

to non-transduced cells as shown in Figure 8. GeneSpring® Pathway analysis on this list revealed that the human “senescence and autophagy” Wikipathway (WP1267 reference in the Wikipathway database was significantly impacted with 10 differential genes : BMP2, COL1A1, CXCL1, GABARAPL1, HMGA1, IGF1, IL1B, MMP14, PLAT, SERPINB2 according to HUGO gene nomenclature (other impacted pathways: cell cycle, MAP kinase, focal adhesion ...).

A number of supplementary genes associated with the Senescence-Associated Secretory Phenotype (SASP) but not included in the senescence and autophagy pathway of Wikipathway were selected for testing. The list was extracted from the data of the literature (Coppé *et al.* 2010 and Young *et al.* 2009). A defined relevant list of cellular senescence associated genes from “senescence and autophagy” pathway and literature, is represented in Table 5.

Finally, 20 genes belonging to the pathway “senescence and autophagy” or associated with the SASP (AREG, BMP2, COL1A1, CXCL1, CXCL2, EREG, GABARAPL1, HMGA1, ICAM1, IGF1, IL1B, MMP1, MMP14, MMP3, NRG1, PLAT, PLAUR, SERPINB2 and TNFRSF10C according to HUGO gene nomenclature, not represented in Tables) were selected because they appeared to be differentially expressed in cells transduced with batch B-S compared to non-transduced cells. Moreover, several collagen genes were downregulated, and PTGS2 gene was upregulated: these genes also participate into the senescence biological process (Coppé *et al.* 2010).

Among the 20 genes belonging to the pathway “senescence and autophagy” or associated with the SASP identified above, 10 genes are not impacted in cells transduced with batch B or C versus non-transduced cells, at MOI 150 (AREG, BMP2, EREG, HMGA1, ICAM1, MMP1, MMP14, NRG1, PLAT and PLAUR according to HUGO gene nomenclature). By “not impacted”, it is meant that the FC absolute value is < 1.3.

6 other genes were also impacted in cells transduced with batch B or C versus non-transduced cells but in an opposite way as the change happening during cellular senescence (COL1A1, CXCL1, CXCL2, GABARAPL1, IGF1 and PLAUR according to HUGO gene nomenclature). Finally, 16 genes present an expression profile characteristic of the apparition of cellular senescence only in

response to batch B-S, compared to other batches at MOI 150. These 16 genes are presented in table 5.

**Transcriptional changes observed in cells transduced with batch B-S at MOI 40 versus non-transduced cells.** Differentially expressed probes with the B-S batch at MOI 40 were examined. 2841 probes were significantly differentially expressed compared to the non-transduced control as shown in Figure 9. As shown in Figure 10, 631 probes were still differential among the 1019 differential probes at higher MOI.

In order to identify probes associated with vectors produced with serum, probes were selected that were differentially expressed in cells transduced with batch B-S (at both MOI) versus the non-transduced condition, and that were not differential in cells transduced with batches B and C (at both MOI). The corresponding set of 235 genes is represented in the profile plot shown in Figure 11. By "not differential", it is meant that the FC absolute value is < 1.3.

Within this list of 235 genes, cellular senescence-associated genes were selected in order to identify early biomarkers of the apparition of a senescent phenotype in cells in contact with a viral vector. A list of 10 cellular senescence-associated genes was obtained, which could be candidate biomarkers revealing a negative impact of a viral vector composition on cells to be transduced. These genes are GABARAPL1, IGF1, PLAUI, BMP2, EREG, MMP1, MMP14, NRG1, PLAT and PLAU according to HUGO gene nomenclature (more details in table 7).

**Selection of restricted list of genes not impacted with a high-quality vector.** This example discloses a selection of genes not impacted by batch C (both MOI) and impacted with other batches. Whereas most of differentially expressed genes are commonly expressed in cells transduced with batch B versus non-transduced cells and cells transduced with batch C versus non-transduced cells at MOI 150, a few genes show different expression patterns.

In order to select genes specifically impacted in cells in contact with a low quality vector composition, probes that were not differentially expressed

with batch C (FC chosen between -1.2 and 1.2) and differentially expressed in cells transduced with batch B versus non-transduced cells at MOI 40 and 150 were selected. The corresponding selection results in one gene: CXCL2. Two other genes share the same profile, except for reduced FC values for B versus NT at MOI 40 (absolute values comprised between 1.3 and 1.5): FOXQ1 and ZNF547 (Agilent Probe ID: A\_33\_P3352822, RefSeq mRNA accession number: NM\_173631, RefSeq protein accession number: NP\_775902).

Interestingly, CXCL2 and FOXQ1 are also differential when comparing cells transduced with batch B-S versus non-transduced cells at MOI 150 but with an opposite evolution, as they are over-expressed cells transduced with batch B-S versus non-transduced cells, although they are under-expressed in cells transduced with batch B versus non-transduced cells.

Another gene presents an interesting expression profile: MAP3K8. The corresponding probe was not initially selected as it is not statistically significantly differential at MOI 40 when comparing cells transduced with batch B versus non-transduced cells, but a slight difference could be seen when examining intensity values, corresponding to a FC value of -1.2, thus confirming the downregulation tendency observed at higher MOI. This gene is downregulated in cells transduced with batches B and B-S versus non-transduced cells and not differential in cells transduced with batch C at both MOI.

These three genes (CXCL2, FOXQ1 and MAP3K8 according to HUGO gene nomenclature), shown in Table 8, are hence biomarkers for use in the practice of the invention as their expression is specifically impacted when transducing cells with a low quality vector batch, and not affected by high quality vector transduction.

#### **Candidate biomarkers validation.**

In order to validate candidate biomarkers response depending on lentiviral vector batch quality, an independent experiment was performed. Foreskin fibroblast cells were transduced at MOI 40 and 150 with five rLV-EF1-GFP compositions: batch B, batch C, batch C-S, batch B-S and batch UC (or UC-S) (described above) and whose characteristics are summarized in Figure

12. These batches were deeply characterized in order to ensure the most precise comparisons between conditions. Cells were observed 48 hours after transduction as presented in Figure 13. At MOI 40, a growth retardation with batch B transduced cells compared to non-transduced cells was confirmed, as well as the absence of visible retardation with batch C. All the other batches produced with serum, induced cell proliferation delays, similarly to batch B. Thus, the proliferation rate with batch B is better than with batch B-S, and the one with batch C is better than with batch C-S. At MOI 150, the only one batch which does not impact the cell proliferation is batch C. A strong proliferation arrest is seen with batch B transduced cells compared to non-transduced cells, confirming previous results. Batches B-S and UC also induced strong growth retardations. The impact on cell growth observed with batch C-S was lower than with batches B, B-S and UC, being just slightly more intense than with batch C.

**Impact on the transcriptome of transduced cells.** To explore underlying changes at the transcriptional level, these cells were collected 54 hours post-transduction. RNA were extracted and used for subsequent validation experiments.

**Transcriptional changes observed at MOI 40.** RNA from MOI 40 transduced cells with batch B, C and B-S vectors and from NT cells were used to perform Agilent whole human genome microarrays. RNA levels from cells transduced with rLV-EF1-GFP batch B, C and B-S at MOI 40 were compared to RNA from non-transduced cells. After statistical analyses, probes upregulated or downregulated 1.5-fold or more were retained for each comparison.

#### **Cell cycle genes category.**

Within downregulated probes for B vs NT and C vs NT, cell cycle probes are still predominant, with 209 probes on 496 downregulated probes for C vs NT, and 236 on 591 downregulated probes for B vs NT.

Probes corresponding to cell cycle genes previously selected on figure 5 as being downregulated with  $FC \leq -3$  with rLV-EF1 batch B at MOI150 vs NT, were still downregulated with rLV-EF1-GFP batch B and batch C at MOI40 (fig. 14 and table 9). Except for 4 probes, the downregulation was

stronger with batch B than with batch C, confirming the stronger impact on cell cycle with a non-highly purified vector batch.

Similarly, cell cycle upregulated genes corresponding to table 3 were still upregulated with the rLV-EF1-GFP vector batch B or batch C, as shown on figure 15 and table 9. This upregulation is stronger with batch B than batch C compared to NT, except for CDKN1A which exhibit a 1.6-fold upregulation with the 2 batches.

The 18 cell cycle genes selected in table 4 as candidate biomarkers exhibit a downregulation after transduction with rLV-EF1-GFP batch B as well as batch C vectors at MOI 40 compared to NT (NB: RRM2 is only downregulated with a -1.3 FC with batch C vs NT). The profile plot presented in figure 16 confirms a stronger downregulation of these genes with batch B compared to batch C. The proliferation marker MKI67 is also represented in figure 16 and exhibit a more pronounced downregulation with batch B than with batch C compared to NT, confirming previous results.

**Cell cycle genes behavior with B-S batch.** As shown on the profile plot presented in figure 17, downregulation of the 18 cell cycle genes from table 4 and MKI67 are similar with batch B-S compared to batch C (except for one gene: RRM2, which is downregulated in the same proportion with batches B and B-S, and nearly not impacted with batch C).

We can hypothesize that this last observation is linked to the presence of growth factors from the serum, which could be concentrated conjointly with the vector, and thwart the cell cycle genes downregulation, even if the cells are not growing as well as batch C transduced cells.

## 25 RT-qPCR validations

RT-qPCR were performed on MKI67 and E2F8 cell cycle genes with three objectives:

1/ validation of differential gene expressions obtained at MOI 40 for batches B, C and B-S on microarrays with an independent mRNA quantification technique,

2/ validation of the specific behavior of these genes after transduction with another low quality batch: the UC batch,

3/ confirmation of the downregulation of these 2 genes at MOI 150.

As presented in figure 18, for these 2 genes at MOI 40, a stronger downregulation with batch B compared to batch C was confirmed with respective FC compared to NT of -4.6 and -2.4 for MKI67, and -3.1 and -1.7 for E2F8. In agreement with microarrays' results, the downregulation was not higher with batch B-S than with batch C (FC -2.9 and -1.7 for MKI67 and E2F8 respectively). The UC batch produces the same downregulation level as B batch at this MOI (FC -4.7 for MKI67 and -3.0 for E2F8).

At MOI 150, downregulation for these 2 genes are stronger in each condition, but the ratio between batch B and batch C downregulations remains the same (approximately 2 fold). B-S batch generates an equivalent or lower downregulation than batch C, which is in agreement with FC obtained with microarrays. With UC batch, the downregulation reaches an intermediate level between those of batches B and C.

To conclude, we identified and validated a set of cell cycle genes that are early response genes in response to transduction and whose downregulation is stronger at the same MOI with a low quality vector produced without serum than with a highly purified vector. However, these genes cannot be used as unique quality biomarkers as they behave similarly with low quality batches produced with serum than with highly purified batches, certainly due to compensation mechanisms, through cell cycle genes activation by serum growth factors.

#### **Genes presenting a batch specific behaviour**

Among the 16 genes from table 6, only 2 were validated in this independant experiment as presenting a different behavior according to batch quality: CXCL2 and EREG.

Sequences of the primers used in the RT-qPCR validation are given in Table 10.

CXCL2 was significantly upregulated after transduction with B-S batch at MOI 40 (FC B-S vs NT: 1,9) although it was downregulated with batch C (FC C vs NT: -1.75). RT-qPCR validation experiments on the same samples (fig. 19) confirm the downregulation with batch C, and show a slight

upregulation with B-S compared to NT but it is not statistically significant. However, the UC batch exhibits a 1,6-fold upregulation compared to NT.

At MOI 150, RT-qPCR results shown in figure 19B, confirm a specific behavior of this gene depending on vector batch quality. Indeed, only low quality batches induce an upregulation of CXCL2, although the highly purified batch C provokes a slight downregulation of this gene (FC vs NT : 1,4). Noticeably, the B batch which induces no differential expression at MOI 40 (with even a slight downregulation), generates a strong upregulation at MOI 150. This could explain the different results observed with rLV-EF1 batch B which induced CXCL2 downregulation at MOI 40 and 150, maybe due to lower MOI, compared to rLV-EF1-GFP vector. Finally, at high MOI, CXCL2 upregulation appears to be specific of a low quality vector batch.

EREG exhibits a slight upregulation on microarrays with batches B (FC vs NT : 1,5) and B-S (FC vs NT : 1,4), and is not impacted with batch C. RT-qPCR validations were performed with RNA obtained from MOI 150 experiment (fig. 19C). A strong and significant overexpression was confirmed with batches B and B-S (respective FC vs NT: 3.0 and 2.9), while the absence of impact after batch C transduction was confirmed. It can be noticed that UC batch does not induce a significant overexpression.

## REFERENCES

Andreadis ST, Roth CM, Le Doux JM, Morgan JR, Yarmush ML.  
1999. Large-scale processing of recombinant retroviruses for gene therapy.  
5 Biotechnol Prog 15(1):1-11.

Banito A, Rashid ST, Acosta JC, Li S, Pereira CF, Geti I, Pinho S,  
Silva JC, Azuara V, Walsh M, Vallier L, Gil J. Senescence impairs successful  
reprogramming to pluripotent stem cells. Genes Dev. 2009 Sep 15;23(18):2134-  
10 9.

Baekelandt V, Eggermont K, Michiels M, Nuttin B, Debyser Z.  
Optimized lentiviral vector production and purification procedure prevents  
immune response after transduction of mouse brain. Gene Ther. 2003  
15 Nov;10(23):1933-40.

Blagosklonny MV. Cell cycle arrest is not senescence. Aging (Albany  
NY). 2011;3(2):94.

Coffin JM, Hughes SH, Varmus HE (eds). Retroviruses. Cold Spring  
20 Harbor Laboratory Press, 1997.

Cartier, N., Hacein-Bey-Abina, S., et al., 2009. Hematopoietic Stem  
Cell Gene Therapy with a Lentiviral Vector in X-Linked Adrenoleukodystrophy.  
Science, 326(5954), p.818-823.

25 Cavazzana-Calvo, Marina et al., 2010. Transfusion independence  
and HMGA2 activation after gene therapy of human  $\beta$ -thalassaemia. Nature,  
467(7313), p.318-322.

Coffin JM, Hughes SH, Varmus HE (eds). Retroviruses. Cold Spring  
30 Harbor Laboratory Press, 1997.

Chiu SC, Wang MJ, Yang HH, Chen SP, Huang SY, Chen YL, Lin SZ, Harn HJ, Pang CY. Activation of NAG-1 via JNK signaling revealed an isochailulactone-triggered cell death in human LNCaP prostate cancer cells. BMC Cancer. 2011;11:146

5

Coppé JP, Desprez PY, Krtolica A, Campisi J. The senescence-associated secretory phenotype: the dark side of tumor suppression. Annual Review of Pathological Mechanical Disease. 2010;5:99–118.

10

Giri, M.S. et al., 2006. Microarray data on gene modulation by HIV-1 in immune cells: 2000-2006. Journal of Leukocyte Biology, 80(5), p.1031-1043.

15

Kosar M, Bartkova J, Hubackova S, et al. Senescence-associated heterochromatin foci are dispensable for cellular senescence, occur in a cell type- and insult-dependent manner and follow expression of p16 ink4a. Cell Cycle. 2011; 10(3):457–468.

20

Mitchell, R., Chiang, C.Y., et al., 2003. Global analysis of cellular transcription following infection with an HIV-based vector. Molecular Therapy, 8(4), p.674–687.

25

Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>(-Delta Delta C(T))</sup> Method. Methods. 2001;25(4):402–408.

30

Rimai L, Salmee I, Hart D, Liebes L, Rich MA, McCormick JJ. 1975. Electrophoretic mobilities of RNA tumor viruses. Studies by Doppler-shifted light scattering spectroscopy. Biochemistry 14(21):4621-7.

Salmeen I, Rimai L, Liebes L, Rich MA, McCormick JJ. 1975. Hydrodynamic diameters of RNA tumor viruses. Studies by laser beat frequency light scattering spectroscopy of avian myeloblastosis and Rauscher murine leukemia viruses. *Biochemistry* 14(1):134-41.

Selvaggi TA, Walker RE, Fleisher TA. Development of antibodies to fetal calf serum with arthus-like reactions in human immunodeficiency virus-infected patients given syngeneic lymphocyte infusions. *Blood*. 1997 Feb 1;89(3):776-9.

Sommer, C.A. et al., 2009. Induced Pluripotent Stem Cell Generation Using a Single Lentiviral Stem Cell Cassette. *Stem Cells*, 27(3), p.543-549.

Takahashi K and Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663-676, August 25, 2006.

Vallier L, Touboul T, Brown S, Cho C, Bilican B, Alexander M, Cedervall J, Chandran S, Ahrlund-Richter L, Weber A, Pedersen RA. Signaling pathways controlling pluripotency and early cell fate decisions of human induced pluripotent stem cells. *Stem cells*, 2009, 27 (11) : 2655-2666.

Young ARJ, Narita M. SASP reflects senescence. *EMBO Rep*. 2009;10(3):228-230.

Zhao, Y., Azam, S. & Thorpe, R., 2005. Comparative studies on cellular gene regulation by HIV-1 based vectors: implications for quality control of vector production. *Gene therapy*, 12(4), p.311-319.

## CLAIMS

1. A method for assessing the quality of a viral vector composition for a transgene transfer into target cells comprising (a) measuring the expression  
5 level of at least one biomarker selected in the group consisting of CXCL2 and EREG in target cells that has not been contacted with the viral vector composition; (b) measuring the expression level of said at least one biomarker in target cells following contact with said viral vector composition; and (c) comparing said biomarker(s) expression in (b) to the biomarker(s) expression in  
10 (a) wherein an significant upregulation in biomarker expression level in (b) compared to (a) indicates that the quality of the viral vector composition is insufficient.

2. The method of claim 2, wherein a significant upregulation at high MOI is a two fold upregulation compared to (a).

15 3. The method according to any of claim 1 to 2, further comprising a step of (d) measuring the expression level of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK in target cells that has not been contacted  
20 with the viral vector composition; (e) measuring the expression level of said at least one biomarker in target cells following contact with said viral vector composition; and (f) comparing said biomarker(s) expression in (e) to the biomarker(s) expression in (d) wherein an significant downregulation in biomarker expression level in (e) compared to (d) indicates that the quality of  
25 the viral vector composition is insufficient.

4. The method of claim 3, wherein a significant downregulation at optimal MOI is a 1,5 fold downregulation compared to (d).

5. The method according to any of claims 1 to 4, further comprising a step of (g) measuring the expression level of at least one biomarker selected in  
30 the group consisting of CXCL2 and EREG in target cells following contact with a control viral vector composition; and (h) comparing said biomarker(s)

expression in (g) to the biomarker(s) expression in (a), wherein no significant differential expression in biomarker(s) has to be detected in (g) compared to (a).

6. The method according to any of claims 1 to 5, further comprising a step of (j) measuring the expression level of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK in target cells following contact with a control viral vector composition; and (k) comparing said biomarker(s) expression in (j) to the biomarker(s) expression in (d) wherein a potential downregulation in biomarker expression level in (j) compared to (d) has to be detected.

7. The method according to claim 6, wherein the potential downregulation in biomarker expression level in (j) compared to (d) is at least 1,5 times less than the downregulation in biomarker expression level in (e) compared to (d) at high MOI.

8. The method according to any of claims 5 to 7, comprising beforehand a step of titration of the viral vector composition and the control viral vector composition.

9. The method according to any of claims 1 to 8, wherein the measures of the biomarker expression level are performed before the cells reach confluency.

10. The method according to any of claims 1 to 9, wherein target cells are eukaryotic cells and are transduced by a lentiviral vector composition.

11. A kit comprising one or more reagents able to measure the expression level of at least one biomarker selected in the group consisting of CXCL2 and EREG.

12. The kit of claim 11, further comprising one or more reagents able to measure the expression level of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK.

13. The kit of claim 11 or 12, further comprising control viral vector composition.

14. A biochip consisting of at least a biomarker selected in the group consisting of CXCL2 and EREG and at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK with optionally a ubiquitous gene.

15. A RT-qPCR plate comprising primers of at least a biomarker selected in the group consisting of CXCL2 and EREG and primers of at least one biomarker selected in the group consisting of ASPM, AURKB, CENPA, CENPF, CKS1B, E2F8, ERCC6L, FAM83D, KIFC1, MKI67, NEK2, NUSAP1, OIP5, PRC1, RRM2, SGOL1, SPC25, TOP2A and TTK.