Abstract: The invention provides a method of expressing a transgene in a host cell. The method comprises exposing a host cell, which comprises a genetic construct comprising a transgene, to conditions which substantially block mitosis in the host cell, such that the construct forms a mitotically stable episome resulting in expression of the transgene. The invention extends to the use of such constructs for expressing transgenes in a range of different medical and heterologous protein production applications.
GENE EXPRESSION FROM MITOTICALLY STABLE EPISOMES

The present invention relates to gene expression, and in particular to stable transgene expression in host cells using episomal vectors. The invention extends to the use of such vectors for expressing transgenes in a range of different medical and heterologous protein production applications.

Gene addition therapy attempts to correct a disease phenotype by introducing a functional copy of the non-functional gene into the patient, thereby restoring the synthesis of the functional gene product. One of the main challenges faced by gene therapy, however, is establishing stable transgene expression without integration into the host genome, because integration can often result in insertional mutagenesis and the activation of oncogenes.

Large-scale heterologous protein production is carried out by pharmaceutical companies by growing a host cell harbouring a vector-transgene construct usually in a continuous culture for many generations. A significant problem with growing host cells over numerous generations is the risk that the transgene will become integrated in the host's genome, which, as discussed above, can result in the generation of insertional mutations, which can have a deleterious affect on the culture and the resultant heterologous protein production.

Therefore, there is a need to provide improved and safer systems for the stable expression of transgenes in host cells, for use in a wide range of gene therapy applications and cell line biopharmaceutical production technologies.

According to a first aspect of the invention, there is provided a method of expressing a transgene in a host cell, the method comprising exposing a host cell, which comprises a genetic construct comprising a transgene, to conditions which substantially block mitosis in the host cell, such that the construct forms a mitotically stable episome resulting in expression of the transgene.

In a second aspect, there is provided a host cell comprising a genetic construct comprising a transgene, wherein the host cell has been exposed to conditions which substantially block mitosis in the host cell such that the construct forms a mitotically stable episome upon, and wherein the host cell is capable of expressing the transgene.
The host cell of the second aspect maybe prepared by the method according to the first aspect.

Advantageously, the inventors have overcome the problems of genomic insertional mutagenesis by using a construct in the method of the first aspect in order to establish mitotically stable, non-integrating stable episomes (including autonomous nuclear DNA elements) containing the transgene. This is achieved by blocking mitosis to induce cell cycle arrest in the host cell, which results in expression of the transgene. Although the inventors do not wish to be bound by hypothesis, they believe that applying the mitotic block to the host cell may allow time for epigenetic chromatinisation events to occur which allows the delivered episomal DNA from the construct comprising the transgene to establish permanency (i.e. mitotic stability), as well as features of replication and segregation. Accordingly, the mitotic block allows the establishment of mitotically stable replicating and segregating episomes, and so is an extremely important and non-obvious observation. Thus, the eposome may be capable of replicating and segregating into daughter cells.

As described in the Examples, the inventors have surprisingly demonstrated that, using the method of the first aspect, stable transgene expression can be established and maintained over many cell divisions, i.e. months of continuous culture amounting to over 100 cell doublings. In addition, the inventors have shown that it is possible to isolate stable cell clones and sub-clones expressing very high homogeneous levels of transgenes. They have also generated results which support the hypothesis that the constructs do indeed stably exist in the host cells and are replicating and segregating as nuclear episomal elements, i.e. (i) deep-sequencing of vector-related DNA sequences; (ii) deep-sequencing searched for vector-chromosome integrations and established that none were found; (iii) Q-PCR quantification of vector-related DNA sequences; (iv) Southern blotting analyses probing for the presence and configuration of vector sequences; and (v) FISH analyses quantifying episome-like elements in chromosome spreads.

It will be appreciated that episomes are the eukaryotic equivalent of bacterial plasmids. Generally, in eukaryotes, episomes are closed circular autonomous DNA elements that are replicated in the nucleus. Viruses are the most common examples of this, such as herpesviruses, adenoviruses and polyomaviruses. Other examples include aberrant chromosomal fragments, such as double minute chromosomes, that can arise during artificial gene amplifications or in pathologic processes (e.g., cancer
cell transformation). Episomes in eukaryotes behave similarly to plasmids in prokaryotes in that the DNA is stably maintained and replicated within the host cell. Cytoplasmic viral episomes (as in poxvirus infections) can also occur. Some episomes, such as herpesviruses, replicate in a rolling circle mechanism, similar to bacterial phage viruses. Others replicate through a bi-directional replication mechanism (Theta type plasmids). In either case, episomes remain physically separate from host cell chromosomes.

The method of the invention may comprise blocking host cell mitosis in order to induce cell cycle arrest. The rationale for inducing cell cycle arrest on the host cell is based on the assumption that stable nuclear DNA elements are anchored onto the nuclear matrix or scaffolding. Therefore, for an episome (i.e. the construct comprising the transgene) to become mitotically stable, an interaction maybe necessary between the newly introduced construct gene sequence and the nuclear matrix of the host cell, and such an interaction should be allowed sufficient time to take place prior to the dissipation of the nuclear membrane during mitosis, followed by potential diffusion of the vector genome. It was reasoned by the inventors that an induced cell cycle arrest may provide the opportunity for the construct to attach to the nuclear matrix.

Preferably, in the method of the invention, mitosis may be blocked between the G1 phase and S phase of the host cell cycle. In order to establish the stable replicative nature of nuclear episomes derived from cell transduction with the construct, a period of cell cycle arrest or quiescence may be induced either during or after transduction of the host cell with the transgene-containing construct. Thus, the mitotic block may be induced within the first 48 hours, 24 hours, 12 hours, 6 hours, 3 hours, 1 hour or 30 minutes upon transduction of the host cell with the construct. Preferably, however, the mitotic block is induced immediately upon transduction.

Cell cycle arrest may be allowed to proceed for a determined period, after which the host cell may then be returned to standard culture conditions, whereupon the usual cell cycle and mitosis of the host cell resumes. Accordingly, the mitotic block may be reversible. The inventors are of the view that, to date, no studies have ever reported the use of reversible cell cycle arrest as a means for encouraging the establishment of mitotically stable episomes.
It will be appreciated that the determined period for inducing the mitotic block will be dependent on the type of host cell being cultured, but it should be sufficiently long for epigenetic chromatinisation events to occur which would allow the construct comprising the transgene to establish permanency (i.e. replicate and segregate) within the host cell, but not so long that the host cell is actually killed. Thus, the period required to induce the mitotic block (i.e. cell cycle arrest) in the host cell may be between 1 hour and 10 days, or between 6 hours and 9 days, or between 12 hours and 8 days, or between 18 hours and 7 days. The period may be between 1 day and 6 days, or between 2 days and 5 days, or between 2 days and 4 days. The period required to induce the mitotic block arrest may be at least 1, 2, 3, 4, or 5 days, or more.

The mitotic block may be achieved through the means of addition or restriction or depletion of substances usually required in the medium in which the host cell is disposed. The medium may be physiological bodily fluid, or a growth culture medium. As the vast majority of research into the mechanisms of the cell cycle is in the field of cancer, many additives have been tested for their potential to stop the uncontrolled division of tumour cells, which may be used in the method of the invention. For example, types of chemotherapeutic drugs to induce G1/G0 arrest may include ribonucleotide biosynthesis inhibitors, as well as ionising and UV irradiation, which may be use in the method of the invention. As well as the addition of various inhibitors, cell cycle arrest may also be achieved using a range of physical, metabolic, cell culture or pharmacological stimuli, which may be selected from a group of stimuli consisting of metabolic restriction (such as serum starvation); hormone or cytokine treatment or withdrawal; pharmacological drug exposure or withdrawal; inhibition or activation of cell signalling pathways; and culture temperature or atmosphere stimuli.

In contrast to the use of cell cycle arrest as a method for inducing apoptosis in cancer cells, in the method of the invention, which may be used in a range of gene therapy applications, it is important that any damage to the host cell must be kept to a minimal. For this reason, therefore, a preferred method for inducing the mitotic block in the host cell may be methionine depletion of the host cell's growth media. Methionine depletion maybe preferred over chemical methods, which may damage the DNA and therefore increase the rate of random integrations and cell death. Methionine restriction in combination with serum depletion is a preferred means for inducing cell cycle arrest.
Thus, the concentration of methionine in the growth medium may be less than about 0.5 µM, or less than 0.4 µM or less than 0.3 µM. Preferably, there is substantially no methionine in the growth medium. Bovine serum contains methionine at an estimated concentration of 15 µM, resulting in a final methionine concentration of about 0.3 µM, which is referred to herein as being "methionine restricted".

The concentration of serum in the growth medium may be less than about 5%, 4%, 3% or 2% by volume. The serum concentration used in the Examples was 2% (v/v) and referred to herein as being "serum depleted". In a preferred embodiment, the growth medium used for inducing the mitotic block may comprise Dulbecco's Modified Eagle's Medium (DMEM) with no methionine and 4.5 g/l glucose (Invitrogen), 2% Foetal Calf Serum (Paa), 100 units/ml penicillin and 100 µg/ml streptomycin (Invitrogen). For CHO cells, the medium may comprise 0.02g/l L-proline.

The genetic construct may be in the form of an expression cassette, which is suitable for expression of the transgene in the host cell. The method may comprise introducing the genetic construct into the host cell. This may be achieved without the construct being incorporated in a vector. For instance, the genetic construct, which may be a nucleic acid molecule, may be incorporated within a liposome or a virus particle. Alternatively, a purified nucleic acid molecule (e.g. histone-free DNA, or naked DNA) may be inserted directly into a host cell by suitable means, e.g. direct endocytotic uptake. The genetic construct may be introduced directly into cells of a host subject (e.g. a eukaryotic or prokaryotic cell) by transfection, infection, electroporation, microinjection, cell fusion, protoplast fusion or ballistic bombardment. Alternatively, genetic constructs of the invention may be introduced directly into a host cell using a particle gun.

Preferably, however, the genetic construct may be harboured within a vector, for expression in a suitable host cell. The vector, which may be recombinant, may be a plasmid, cosmid or phage. Such vectors are useful for transforming host cells with the genetic construct, and for replicating the expression cassette comprising the transgene therein. The skilled technician will appreciate that genetic constructs of the invention may be combined with many types of backbone vector for expression purposes. For example, the vector comprising the transgene may be a virus. The
vector may be a member of the Retroviridae family, or of the Orthoretrovirinae Sub-family. The vector may be a member of the Lentivirus genus.

As a gene delivery vector, lentiviral vectors have key several advantages over other systems. Firstly, they have a large packaging capacity of at least 8 Kb of DNA, which is an important feature when packaging sizeable expression cassettes of tissue-specific promoters and transgenes. Secondly, they differ from simpler retroviruses not only in the genome organisation, but also in that they are able to transduce non-dividing cells, which is a very useful quality when considering application as a gene therapy vector to non-proliferating tissues such as muscle, neurons and haematopoietic stem cells. In addition, lentivectors have reduced immunogenicity compared to adenoviral vectors, making it possible to consider systemic delivery routes.

Lentiviral vectors are commonly produced by transient transfection of 293T cells and and harvesting of the virus-containing supernatant. Useful titres, typically $10^5-10^6$ TU/ml, are routinely achieved. Substituting the HIV-i envelope protein with the G glycoprotein of vesicular stomatitis virus (VSV-G) confers the viral particles with a highly stable capsid structure, allowing the vector preparations to be further concentrated by ultracentrifugation. VSV-G pseudotyped vectors also exhibit a dramatically increased tissue tropism compared to wild-type lentivirus vectors, which is why it is commonly utilised for both in vitro and in vivo applications.

HrV-i-derived vectors display a propensity for integrating within transcriptionally active genes, possibly as a result of interactions between the pre-integration complex (PIC) and chromatin-binding factors. Unfortunately, this also significantly increases the potential for insertional mutagenesis and subsequent oncogenesis. This may result as the integrating genome is more likely to interfere with an active part of the genome, which in turn causes major problems when considering the safety aspects of potential gene therapy vectors.

Accordingly, it is preferred that the vector is a non-integrating or integration-deficient vector. Preferably, the vector is an integration-defective lentiviral vector (IDLV), which, as described in the examples, has proven to be especially effective.

The development of integration-deficient lentiviral vectors (IDLVs) provides all of the advantages of previous generations of lentivectors together with a significantly improved safety profile. IDLVs utilise the natural tendency of lentiviruses to form
episomal circles as an intermediate or by-product during infection. The occurrence of these circular episomes can be greatly increased by impairing the ability of the virus to integrate. The most efficient way of achieving this may be to mutate one of the amino acids in the catalytic site of Integrase (IN), the enzyme which catalyses the integration of the viral cassette into the host genome. These Class I IN mutants, of which the D64V is one embodiment, are only defective in integration and the integrase retains its other functions, resulting in normal levels of viral DNA (Engelman et al, 1995, J Virol 69(5): 2729-2736).

The occurrence of these circular episomes can be greatly increased by impairing the ability of the virus to integrate. The most efficient way of achieving this may be to mutate one of the amino acids in the catalytic site of Integrase (IN), the enzyme which catalyses the integration of the viral cassette into the host genome. These Class I IN mutants, of which the D64V is one embodiment, are only defective in integration and the integrase retains its other functions, resulting in normal levels of viral DNA (Engelman et al, 1995, J Virol 69(5): 2729-2736).

The nucleic acid sequence for HIV integrase as represented below as SEQ ID No.1.

```
ttttagatggaataagctcaagaagaacacgaaaagtgtacctcataaatctga
gaccatggcaactgtttgtaacctctagaagggatatctctgtagctgaag
ctgtgataaatgtcagctaatgtcagctaatgtcagctaatgtcagctaatgtc
agcattttttttttataataggcagcgagctgtcctgtcctgtcctgtcctgt
```

[SEQ ID No.1]

Preferably, the construct comprises a mutation in the catalytic site of Integrase (IN) enzyme, preferably a D64V mutation, which is substantially set out as SEQ ID No: 2.

```
ttttagatggaataagctcaagaagaacacgaaaagtgtacctcataaatctga
gaccatggcaactgtttgtaacctctagaagggatatctctgtagctgaag
ctgtgataaatgtcagctaatgtcagctaatgtcagctaatgtcagctaatgtc
agcattttttttttataataggcagcgagctgtcctgtcctgtcctgtcctgt
```

[SEQ ID No.2]
Thus, the construct may comprise a nucleic acid sequence substantially as set out in either SEQ ID No.1 or SEQ ID No.2, or a functional fragment or variant thereof.

Advantageously, non-integrating lentivectors are also protected from epigenetic silencing which may otherwise occur upon integration into an inactive part of the genome. Expression levels both in vitro and in vivo have been significantly augmented by including the self-inactivating (SIN) deletion, which removes the negative regulatory effect arising from the full-length Long Terminal Repeat (LTR).

Accordingly, the construct may comprise a SIN deletion (Zufferey, R et al., 1997, Nat Biotechnol 15(9): 871-875.). Preferably, the nucleic acid sequence of the SIN mutation is substantially set out as SEQ ID No: 3.

tgggaaggctaatcctcctccaagtaagagatacctgggtctctctgtgtactgggcctctctggttagccagatctgagcctgggagctctctggtgtgtgtgtgactctggtactagagatccctcagacccttttactcagca

[SEQ ID No.3]

The construct may comprise one or more long terminal repeat (LTR). Preferably, the nucleic acid sequence of the LTR is substantially set out as SEQ ID No: 4.

actggaaggctaatcctcctccaagtaagagatacctgggtctctctgtgtactgggcctctctggttagccagatctgagcctgggagctctctggtgtgtgtgtgactctggtactagagatccctcagacccttttactcagca

[SEQ ID No.4]

Thus, the construct may comprise a nucleic acid sequence substantially as set out in either SEQ ID No.3 or SEQ ID No.4, or a functional fragment or variant thereof.

Another useful addition to the construct maybe the woodchuck hepatitis virus post-transcriptional element (WPRE), which, when placed downstream of the transgene,
has been shown to increase transcription by more than five-fold (Zufferey, Ret al., J Virol 73(4): 2886-2892).

Therefore, the construct may comprise woodchuck hepatitis virus post-transcriptional element (WPRE), preferably disposed downstream of the transgene. Preferably, the nucleic acid sequence of the WPRE is substantially set out as SEQ ID No:5.

```
aatcaacctctggattacaaaatttgtgaaagattgactggtattcttaactatgttg
cctccttttaagctatagtggtttgtgatctcgtcttttcttcttgag
gagtgtggcccgtttgcagggcctagtggtgctgcacttggtttgtcgactgcaaa
ccccactgtggtggcccacacttgctacgctcttttccggaacttctccctc
cccctccctattgacccaggggaactctagcgcctgccccgtgcttgccagaca
ggggtccgcttgttgctgaactcaatctccggtggttctgtcgggagaagctgacgctct
tccatagctgcctgctgtggctacccgagcttcctctctcctcgtgctgcctgctgc
ccgctctttccgcccccaatctccaggggacccctctcttcgccctacagacagtgcagttcctcccttgggcccgcctcctccgcccgtctg

[SEQ ID NO.5]
```

Thus, the construct may comprise a nucleic acid sequence substantially as set out in SEQ ID NO.5, or a functional fragment or variant thereof.

The vector may or may not comprise a so-called Scaffold/Matrix Attachment Regions (S/MAR) element. Preferably, the nucleic acid sequence of a S/MAR element may be substantially set out as SEQ ID No: 6.

```
gatctaaataaactttataaatgtgagaaatattgcatgtctaaattgcaag
aacccggagacactatcactatatctgtaacacttaaagacctaatatcttgactac
ctctctccacataaattttaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaatctgtaagagctgtaaad
```
Advantageously, in embodiments of the vector comprising a S/MAR element, replication and segregation properties maybe induced. S/MARs are architectural components of genomic DNA that are involved in the organisation of chromatic into domains and also in the regulation of several DNA functions, such as replication and transcription. The S/MAR may be a miniMAR, as described in the Examples.

Preferably, the nucleic acid sequence of a miniMAR may be substantially set out as SEQ ID No: 7.

Thus, the construct may comprise a nucleic acid sequence substantially as set out in either SEQ ID No.6 or SEQ ID No. 7, or a functional fragment or variant thereof.

In the course of these studies, the inventors noted that gene delivery by IDLVs (with or without S/MARs), and followed by a period (preferably, 2-4 days) of induced cell quiescence (i.e. withdrawal from the cell cycle) lead to an establishment of stable gene expression in at least 15-20% of the transduced cell population.

The vector may include a variety of other functional elements including a suitable promoter to initiate transgene expression. For instance, the vector is preferably...
capable of autonomously replicating in the nucleus of the host cell. In this case, elements which induce or regulate DNA replication may be required in the recombinant vector. Suitable promoters may include the SV40 promoter. A terminator may also be provided. The vector may also comprise DNA coding for a gene that may be used as a selectable marker in the cloning process, i.e. to enable selection of cells that have been transfected or transformed, and to enable the selection of cells harbouring vectors incorporating heterologous DNA. For example, ampicillin resistance or green fluorescent protein (GFP) are envisaged. Alternatively, the selectable marker gene may be in a different vector to be used simultaneously with the vector containing the transgene. The vector may also comprise DNA involved with regulating expression of the transgene, or for targeting the expressed polypeptide to a certain part of the host cell.

Challenges, such as the potential for integration, vector mobilisation, and generation of replication competent lentiviruses are not entirely overcome by the use of IDLVs, although the risk for all of them is significantly reduced. The inventors have found that the method of the invention results in a surprisingly low amount of integration of the transgene into the host's genome. As described in the deep sequencing data in the Examples, the inventors observed only 1 or 2 integration events in over 10000 chromosomes. Thus, the amount of integration may be less than 1%, or less than 0.1%, or less than 0.01%. However, the inventors were pleased to observe that, as described in the examples, deep sequencing analysis of IDLV integration sites has shown that most integration events occur in non-coding regions and most likely via cell-mediated repair mechanisms rather than residual Integrase activity.

In a first embodiment, the construct (referred to herein as pRRLsc SV40 eGFP) may comprise a lentivector backbone, the nucleic acid sequence of which is substantially set out as SEQ ID No: 8.

tcgagatgtgttcagttaggggtgtggaaagttcccccaggttcctccccagccaggcagagtagta

gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
gcagggatggcattctcaattgtcgagaaccaggttgtggaaagttcccccaggttcctcccagggcagagtagta
Thus, the construct may comprise a nucleic acid sequence substantially as set out in SEQ ID No.8, or a functional fragment or variant thereof.

In a second embodiment, the construct (referred to herein as pRRLsc CMV eGFP) may comprise a lentivector backbone, the nucleic acid sequence of which is substantially set out as SEQ ID No: 9.

```plaintext
tggccttttgctgcccccttttgctgctccatgctttctctctccgggttatcccctgattctgtg
gataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccga
ggcgcagcgagtcagtgagcgaggaagcggaagagcgcccaatacgcaaaccgcctctcc
cgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactggaaagcg
ggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggcttt
cacctttatgcttccggctcgtatgttgtggaattgtgagcggatacaatttcaca
caggaaacacagtataacccagcagcagccattacaaacacctttttccggctcgtattg
gactgatgaatggtgagtaaagctgtagaagggcctgatgtatcacccttttagtctccggg
gcaacagcataatgactgcagcatggtaacgatgagttagcaacatgccttacaaggagagaaaaagcaccgtgcatgccgattggtggaagtaaggtggtgacgatcgtgccttattaggaaggcaacagacgggtctgacattgctcgtctgatctgattgtctgactgctagctgtttctgctgtgtgcccgtctgttgtgactctggtaactagagatccctcagacccttttagtccagtgtggaaaatctctagcagtggcgcccgaacagggacctgaaagcgaaagggaaaccagagctctctcgacgcaggactcggcttgctgaagcgcgcacggcaagaggcgaggggcggcgactggtgagtacgccaaaaattttgactagcggaggctagaaggagagagatgggtgcgagagcgtcagtattaagcgggggagaattagatcgcgatgggaaaaaattcggttaggccagggggaaagaaaaaatataaattaaaacatatagtatgggcaagcagggagctagaacgattcgcagttaatcctggcctgttagaaacatcagaaggctgtagacaaatactgggacagctacaaccatcccttcagacaggatcagaagaacttagatcattatataatcatagtagcaaccctctattgtgtgcatcaaaggatagagataaaagacaccaaggaagctttagacaagatagaggaagagcaaaacaaaagtaagaccaccgcacagcaagcggccgctgatcttcagacctggaggaggagatatgagggacaattggagaagtgaattatataaatataaagtagtaaaaattgaaccattaggagtagcacccaccaaggcaaagagaaggtgtgcacagacccaactgggtagataggtttctctgtttttggagcagcagaagcactattggggcagctcgaatgcattaaacgtatcagctcgtggggcaagggctatactgggtctctctggttagaccagatctgagcctgggagctctctggtctctctggttagaccagatctgagcctgggagctctctggtactgggtctctctggttagaccagatctgagcctgggagctctctggtactgggtctctctggttagaccagatctgagcctgggagctctctggtactgggtctctctggttagaccagatctgagcctgggagctctctggtactgggtctctctggttagaccagatctgagcctgggagctctctggtactgggtctctctggttagaccagatctgagcctgggagctctctggtactgggtctctctggttagaccagatctgagcctgggagctctctggtactgggtctctctggttagaccagatctgagcctgggagctctctttaaatggtggtggtttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
Thus, the construct may comprise a nucleic acid sequence substantially as set out in SEQ ID No.9, or a functional fragment or variant thereof.

The transgene may be any gene encoding a protein, which may have therapeutic or industrial utility. For example, the transgene may encode dystrophin, a blood coagulation factor, insulin or a cytokine receptor sub-unit.

The host cell may be a prokaryotic or eukaryotic cell. The host cell may be an animal cell. The host cell may be a cell from a vertebrate, mammal or domestic animal, and is preferably a human cell. Preferably, the host cell is a progenitor or stem cell, which may be either pluripotent or totipotent. The host cell may be a haematopoietic stem cell (HSC). The host cell may be either an embryonic, foetal or adult stem cell. The host cell may be in a cell culture, or may be in situ in a host organism, for example a subject to be treated.

As described in the Examples, CHO cells are preferred host cells. The Chinese hamster ovary cell line was originally derived from an in-bred female laboratory animal in 1957 (Tjio et al., 1958). The cells descend from a spontaneously immortalised population of cultured fibroblast cells. It has been hypothesised that the cell line is of clonal origin, as the first cells and all subsequently derived populations are deficient in proline synthesis. Cells require proline due to the absence of the gene for proline synthesis, the block in the biosynthetic chain lies in the step converting glutamic acid to glutamine gamma serialdehyde.

The chromosome number in wild type Chinese hamsters is very low at 22, which is why their cells were originally indicated as good candidates for cytogenetic research. However, in culture the cells were noted to have a labile karyotype right from the beginning, with variations in chromosome number as well as structural irregularities.

The cells were also found to proliferate rapidly and produce high amounts of protein, which is why they have since become popular for both in vitro research and
engineered protein production. Currently, CHO cells are the cell line of choice for the production of many therapeutic biopharmaceuticals, as they are capable of performing human-compatible post-translational modifications as well as yielding several grams of protein per litre. Another feature increasing the attractiveness of CHO cells for human protein production is their resistance to infection by many human viruses, including HSV and HIV-i, due to lack of viral entry receptors.

Regulation of cell cycle in CHO cells may be achieved by allowing them to reach a stationary growth phase, which can be used to induce an arrest in the G1 phase when prolonged for up to 80 hours (Tobey et al., 1970). The cells resume an exponential growth phase following dilution. The cells may be induced to re-enter cycling by the introduction of isoleucine and glutamine to the medium, which were then deemed as essential for growth of the cells.

The preferred host cell for the method of the invention may be CHO-Ki, which is a sub-clone of the parental CHO cell line created by single cell cloning in 1957 (Puck et al., 1958). The karyotype of this cell line is very different to the wild-type Chinese hamster, and only 8 of the 22 chromosomes in the cells are equivalent to wt chromosomes (Wurm et al., 2011). The modal chromosome number of this cell line is 20, although again this varies between sub-clones (Kao et al, 1970).

It will be appreciated that the method of the first aspect and the host cell of the second aspect maybe used as a means for expressing and producing proteins on a commercial scale.

For example, a culture of host cells may be grown in a reactor, which may be run in either batch or continuous culture to produce the protein. The protein may then be subsequently isolated. The specific culture conditions and growth medium will depend on the type of host cell being grown and on the type of transgene being expressed. Importantly, the period of time required to induce the mitotic block in the host cell and the means by which mitosis is blocked (e.g. methionine depletion) will depend on the type of host cell. For example, in the case of bone marrow cells, keratinocytes, muscle satellite cells or other stem and progenitor cell populations, the intervention to induce cell cycle arrest may vary in nature (e.g. cytokine withdrawal) and timing (e.g. cells may already quiescent).
The host cell may be used in therapy or diagnosis.

Thus, in a third aspect, there is provided the host cell according to the second aspect, for use in therapy or diagnosis.

Lentiviral vectors are currently the vector of choice for many primary immunodeficiency disorders, for example Wiskott-Aldrich syndrome, using autotransplantation of lentivector-transduced haematopoietic stem cells. Due to their ability to efficiently deliver genes to post-mitotic tissues, lentivectors are also the vector of choice for targeting the central nervous system, and for the treatment of diseases such as spinal muscular atrophy and Parkinson's disease.

Therefore, in a fourth aspect, there is provided use of the host cell according to the second aspect, for treating an immunodeficiency disorder, a blood clotting disorder, a central nervous system (CNS) disorder, spinal muscular atrophy, or Parkinson's disease.

In a fifth aspect, there is provided a method of treating, ameliorating or preventing an immunodeficiency disorder, a blood clotting disorder, a central nervous system (CNS) disorder, spinal muscular atrophy, or Parkinson's disease, in a subject, the method comprising administering, to a subject in need of such treatment, a therapeutically effective amount of the host cell according to the second aspect.

A clotting disorder may be treated by \textit{ex vivo} treatment of haematopoietic stem cells (HSCs). Previously, attempts have been made to treat the X-linked severe combined immunodeficiency (SCID-Xi) condition by HSC gene therapy using an integrating lentivirus vectors encoding the gamma-chain of the IL2 receptor. However, significant instances of haematological malignancies have arisen due to vector integration and insertional mutagenesis. Therefore, a therapeutic application using the methods of the invention involving the replicating episome (i.e. non-integrating vector) technology would be highly advantageous by using IDLVs expressing the gamma-chain of the IL2 receptor, coupled with cell cycle arrest to stably transduce the HSCs from SCID-Xi patients. The transduced stem cells may then be infused back into the patient to reconstitute the bone marrow and immune systems, but with a much reduced risk of integration and subsequent leukaemia induction.
Therefore, the host cell may be a stem cell, for example a haematopoietic stem cell. The construct may comprise a transgene, which encodes the gamma-chain of the IL2 receptor.

A "subject" may be a vertebrate, mammal, or domestic animal. Hence, the host cell according to the invention may be used to treat any mammal, for example livestock (e.g. a horse), pets, or may be used in other veterinary applications. Most preferably, however, the subject is a human being.

A "therapeutically effective amount" of the host cell is any amount which, when administered to a subject, is the amount of medicament or drug that is needed to treat the disease condition, or produce the desired effect.

It will be appreciated that the invention extends to any nucleic acid or peptide or variant, derivative or analogue thereof, which comprises substantially the amino acid or nucleic acid sequences of any of the sequences referred to herein, including functional variants or functional fragments thereof. The terms "substantially the amino acid/nucleotide/peptide sequence", "functional variant" and "functional fragment", can be a sequence that has at least 40% sequence identity with the amino acid/nucleotide/peptide sequences of any one of the sequences referred to herein, for example 40% identity with the nucleotide sequence identified as SEQ ID No:8 (i.e. pRRLsc SV40 eGFP DNA), or 40% identity with the nucleotide identified as SEQ ID No:9 (i.e. pRRLsc CMV eGFP DNA), and so on.

Amino acid/polynucleotide/polypeptide sequences with a sequence identity which is greater than 50%, more preferably greater than 65%, 70%, 75%, and still more preferably greater than 80% sequence identity to any of the sequences referred to are also envisaged. Preferably, the amino acid/polynucleotide/polypeptide sequence has at least 85% identity with any of the sequences referred to, more preferably at least 90%, 92%, 95%, 97%, 98%, and most preferably at least 99% identity with any of the sequences referred to herein.

The skilled technician will appreciate how to calculate the percentage identity between two amino acid/polynucleotide/polypeptide sequences. In order to calculate the percentage identity between two amino acid/polynucleotide/polypeptide sequences, an alignment of the two sequences must first be prepared, followed by calculation of the sequence identity value. The
percentage identity for two sequences may take different values depending on: (i) the method used to align the sequences, for example, ClustalW, BLAST, FASTA, Smith-Waterman (implemented in different programs), or structural alignment from 3D comparison; and (ii) the parameters used by the alignment method, for example, local vs global alignment, the pair-score matrix used (e.g. BLOSUM62, PAM250, Gonnet etc.), and gap-penalty, e.g. functional form and constants.

Having made the alignment, there are many different ways of calculating percentage identity between the two sequences. For example, one may divide the number of identities by: (i) the length of shortest sequence; (ii) the length of alignment; (iii) the mean length of sequence; (iv) the number of non-gap positions; or (iv) the number of equivalenced positions excluding overhangs. Furthermore, it will be appreciated that percentage identity is also strongly length dependent. Therefore, the shorter a pair of sequences is, the higher the sequence identity one may expect to occur by chance.

Hence, it will be appreciated that the accurate alignment of protein or DNA sequences is a complex process. The popular multiple alignment program ClustalW (Thompson et al., 1994, Nucleic Acids Research, 22, 4673-4680; Thompson et al., 1997, Nucleic Acids Research, 24, 4876-4882) is a preferred way for generating multiple alignments of proteins or DNA in accordance with the invention. Suitable parameters for ClustalW maybe as follows: For DNA alignments: Gap Open Penalty = 15.0, Gap Extension Penalty = 6.66, and Matrix = Identity. For protein alignments: Gap Open Penalty = 10.0, Gap Extension Penalty = 0.2, and Matrix = Gonnet. For DNA and Protein alignments: ENDGAP = -1, and GAPDIST = 4. Those skilled in the art will be aware that it may be necessary to vary these and other parameters for optimal sequence alignment.

Preferably, calculation of percentage identities between two amino acid/polynucleotide/polypeptide sequences may then be calculated from such an alignment as \((N/T)\times100\), where \(N\) is the number of positions at which the sequences share an identical residue, and \(T\) is the total number of positions compared including gaps but excluding overhangs. Hence, a most preferred method for calculating percentage identity between two sequences comprises (i) preparing a sequence alignment using the ClustalW program using a suitable set of parameters, for example, as set out above; and (ii) inserting the values of \(N\) and \(T\) into the following formula: Sequence Identity = \((N/T)\times100\).
Alternative methods for identifying similar sequences will be known to those skilled in the art. For example, a substantially similar nucleotide sequence will be encoded by a sequence which hybridizes to the sequences shown in SEQ ID No's: 1-9, or their complements under stringent conditions. By stringent conditions, we mean the nucleotide hybridises to filter-bound DNA or RNA in 3x sodium chloride/ sodium citrate (SSC) at approximately 45°C followed by at least one wash in 0.2x SSC/0.1% SDS at approximately 20-65°C. Alternatively, a substantially similar polypeptide may differ by at least 1, but less than 5, 10, 20, 50 or 100 amino acids from the sequences referred to herein.

Due to the degeneracy of the genetic code, it is clear that any nucleic acid sequence described herein could be varied or changed without substantially affecting the sequence of the protein encoded thereby, to provide a functional variant thereof. Suitable nucleotide variants are those having a sequence altered by the substitution of different codons that encode the same amino acid within the sequence, thus producing a silent change. Other suitable variants are those having homologous nucleotide sequences but comprising all, or portions of, sequence, which are altered by the substitution of different codons that encode an amino acid with a side chain of similar biophysical properties to the amino acid it substitutes, to produce a conservative change. For example small non-polar, hydrophobic amino acids include glycine, alanine, leucine, isoleucine, valine, proline, and methionine. Large non-polar, hydrophobic amino acids include phenylalanine, tryptophan and tyrosine. The polar neutral amino acids include serine, threonine, cysteine, asparagine and glutamine. The positively charged (basic) amino acids include lysine, arginine and histidine. The negatively charged (acidic) amino acids include aspartic acid and glutamic acid. It will therefore be appreciated which amino acids may be replaced with an amino acid having similar biophysical properties, and the skilled technician will known the nucleotide sequences encoding these amino acids.

All of the features described herein (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, maybe combined with any of the above aspects in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

For a better understanding of the invention, and to show how embodiments of the same may be carried into effect, reference will now be made, by way of example, to the accompanying diagrammatic drawings, in which:-
**Figure 1** shows embodiments of transfer plasmids used for making each of the three lentiviral vectors used in the method of the invention. The packaging plasmid during vector production contained the D64V mutation in the Integrase gene, resulting in integration-deficient lentiviral vectors (IDLVs). (A) was used to make IDLV SV40 eGFP, (B) was used for making IDLV SV40 eGFP mMAR, and (C) was used for IDLV CMV eGFP;

**Figure 2** shows stress-induced DNA duplex (SIDD) analysis of l-LTR and 2-LTR circular episomes derived from a standard IDLV vector does not reveal significant differences. The strand separation potential profiles are shown for l-LTR (A) and 2-LTR (B) circles for IDLV SV40 eGFP. Both forms can be expected to be formed in vitro after transduction. The length of the LTR region (shown in red) does not influence the strand separation potential for the remaining sections of the episome. G(x), average free energy of all states x in which the base pair separates against bp position for each episome. X axis, length of construct in bp starting at the beginning of the (first) LTR;

**Figure 3** shows SIDD analysis reveals the destabilising and condensing effects of S/MAR elements on l-LTR IDLV episomes. In (A), the promoters and Ori regions produce the highest destabilisation potential. After addition of a miniMAR element (B), the SIDD potential of these regions is reduced in favour of the miniMAR, particularly in the nearby SV40 promoter/ Ori (red arrow). Addition of the full-length S/MAR element (C) results in a typical periodically destabilised region (green line), whilst the SV40 promoter (red arrow) remains condensed to the same level as in construct B. G(x), average free energy of all states x in which the base pair separates against bp position for each episome. X axis, length of construct in bp;

**Figure 4** shows the results of plating CHO cells, transduced with IDLV vectors or mock control after 1 day, and then treated with methionine-free medium from Day 2 to Day 7, over a 5 day period. At D7, cultures were returned to normal growth medium and allowed to proliferate as normal with routine passaging as required. To confirm induction of transient cell cycle arrest, cell cycle phase analyses were performed by propidium iodide staining and flow cytometry at D7, at the end of the methionine depletion period. The coloured sections indicate the average percentages of cells in each cell cycle phase (i.e. G1, S and G2) as determined by PI staining followed by FlowJo analysis of the data. The cell cycle phases determined to differ significantly between treatments are denoted with *, as indicated by Student's t-test comparing each phase at D7. * = P < 0.05, ** = P < 0.01, *** = P < 0.001;

**Figure 5** shows the results of plating CHO cells into culture, and after 1 day transduced with IDLV-DG, IDLV-SGm and IDLV-CG vectors (MOI=i). Cells were
induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media at D2 (A and C) or allowed to proliferate freely (B and D). From D7, all cells were allowed to proliferate normally with routine passaging for a period up to D72. In (C) and (D) the stabilised transduction levels at the D72 time point are shown. Data are means ± s.e.mean, n=3. By D72 transduction levels in cultures subject to an initial cell cycle arrest were substantial, and significantly higher than for the cell cultures allowed to proliferate continuously (** P < 0.01 and *** P < 0.001; one-way ANOVA with Tukey’s post-hoc test). IDLV-SG = IDLV SV40 GFP; IDLV-SGm = IDLV SV40 GFP mMAR; IDLV-CG = IDLV CMV GFP;

**Figure 6** shows the results of transducing CHO cells with IDLV-SG (Figure 6.1), IDLV-SGm (Figure 6.2) and IDLV-CG vectors (Figure 6.3) (MOI=i). The cells were subjected to 5 days of methionine and serum depletion, allowed to proliferate for 100 days, and then subjected to dilution cloning in 96-well plates. After 14 days, 12 GFP+ clones were expanded to establish cell lines, and evaluated by flow cytometry. Data is shown for the initial parent populations (A) and for the populations derived by dilution cloning (B). The clones which exhibited a true clonal distribution with close to 100% GFP+ cells and a Gaussian MFI distribution were selected for further detailed evaluations, which were 4, 7 and 10 for IDLV-SG, clones 5, 8 and 11 for IDLV-SGm, and clones 2, 10 and 12 for IDLV-CG;

**Figure 7** shows the results of CHO cell lines derived by dilution cloning of mixed parent populations transduced with IDLV-SG vector and subjected to transient cell cycle arrest (clones 4, 7 and 10). The cells were maintained in continuous proliferating culture for up to 60 days with routine passaging approximately 3 times weekly at a 1:5 split. The clonal populations were examined every 10 days by flow cytometry and fluorescence microscopy. (A) GFP fluorescence micrograph for the parental population of transduced CHO cells prior to dilution cloning. (B) Flow cytometry plots at D16 and D50 and GFP fluorescence micrograph (D40) for clones 4, 7 and 10;

**Figure 8** - Genomic DNA was prepared from the stably GFP+ CHO clones derived from transduction and transient cell cycle arrest with IDLV-SG (tracks 1, 2 & 3), IDLV-SGm (tracks 4, 5 & 6) and IDLV-CG (tracks 7, 8 & 9), and from a polyclonal population transduced with integrating LV-SG (track 10). The DNA was subject to LAM-PCR as described in Fig 5.11, and the final products were analysed by agarose-EtBr gel electrophoresis. Track 1 shows a 100 bp marker ladder. Dominant bands at 280 bp (green boxes) correspond to LTR to vector backbone junctions present in all vector forms, 180 bp bands (blue boxes) correspond to LTR-LTR junctions only
present in 2-LTR circles. The integrating clonal control (track 10) shows a smear of bands of varying sizes, as expected;

**Figure 9** shows the majority of reads generated by high-throughput sequencing contain only vector sequences. Sequences from samples resulting in > 400 independent sequence reads were analysed and all SV40 promoter-containing clones were found to contain mostly vector DNA. IDLV3 clones 2 & 12 and the positive ICLV control also contained non-vector DNA. Vi: IDLV SV40 GFP, V2: IDLV SV40 GFP mMAR, V3: IDLV CMV GFP;

**Figure 10** shows Quantitative Real-Time PCR results indicate vector genomes persist in copy numbers of 1-15 per cell. (A) 4 independent qRT-PCR reactions were run to quantify the amount of episomes present in transduced CHO cells. Primer binding sites are shown with black arrows. From the absolute quantities of these signals, the comparable amounts of i-LTR, 2-LTR and backbone signals per cell were calculated and are shown in (B). The numbers of backbone signals (grey bars) can be used as an estimate for the number of vector genomes per cell. The 3 controls (IDLV and ICLV harvested 24 post-transduction and ICLV clonal population in 3 lanes on right) resulted in very high backbone (BB) and i-LTR signal. The 3 clones transduced with IDLV CMV eGFP (Vector 3) produced high levels of 2-LTR signal. Vector 1: IDLV SV40 eGFP, vector 2: IDLV SV40 eGFP mMAR, vector 3: IDLV CMV GFP. PC, polyclonal. Data shown are the mean of 3 reactions performed in triplicate (n = 9);

**Figure 11** shows Southern blotting of DNA from IDLV-transduced CHO cells indicates several clones containing episomal vectors. High molecular weight DNA was harvested from cells transduced with vectors, and digested with either EcoRI or XhoI (mMAR-containing vector only). 10 µg digested DNA was loaded per lane, and probed with \(^{32}\)P-labeled GFP probe. The expected band sizes for each of the 2 episomal conformations are shown in the table (B). Grey squares indicate the presence of bands corresponding to the expected size(s); and

**Figure 12** shows fluorescent in situ hybridisation analysis of transduced cells supports the hypothesis that a majority of vector genomes remain episomal. A statistical analysis comparing the ratios of signals occurring inside nuclei per total nuclei in each sample shows a significant difference between each sample and negative control (t-test, \(P < 0.0001\), n=20 fields/sample). Average diameter of a metaphase, 30 \(\mu\)m (partial metaphases shown above). (A) Clonal CHO cell line transduced with ICLV SV40 eGFP (integrating control), (B) CHO cell line transduced with IDLV SV40 eGFP, Clone 4, and (C) CHO cell line transduced with IDLV SV40 eGFP mMAR, Clone 11.
Examples

Materials & Methods

Lentivector manufacture, concentration and titration

5 Manufacture of lentivectors by transient transfection

293T cells were seeded at 3 x 10^6 cells per plate on 15 cm² plates, and were transfected 2 days later having reached ~ 90% confluency. The media was replaced with 20 ml fresh media (DMEM + glucose + 10% FBS + 2% pen/ strep) 2 h before transfection. The DNA mix was prepared in a molar ratio of 1:1:1:2 using 12.5 µg pMDLg/pRRE or pMDLg/pRREintD64V, 6.25 µg pRSV-REV, 7 µg pMD2.VSV-G and 32 µg of either pRRLsc SV40 eGFP, pRRLsc SV40 eGFP mMAR or pRRLsin PPT CMV eGPF WPRE transfer plasmid (all plasmids from the Yanez group).

Diagrammatic representation of the transfer plasmids is shown in Figure 1. This was made up to a total volume of 112.5 µl with TE buffer and then 1012.5 µl water was added to make 0.1 x TE. 125 µl 2.5 M CaCl₂ was added, the solution vortexed and incubated for 5 min at room temperature. Then, 1250 µl 2 x HBS was added dropwise while vortexing the DNA/CaCl₂ mix at full speed. This mixture was then added to cells and the cells returned to incubators set at 37°C, with 5% CO₂. 16 h after transfection the media was removed and replaced with 18 ml fresh media per plate.

The supernatant containing lentivector particles was harvested on days 2 and 3 post-transfection.

Harvested medium was centrifuged at 2500 rpm for 10 min at RT and the supernatants filtered through a 0.22 µm Nalgene filter. Filtered medium was transferred to high speed polyallomer centrifuge tubes and spun at 50 000 x g for 2 h at 4°C, using SW32.1T1 rotor (Beckman). Supernatant was discarded and tubes kept upside down on tissue paper for a few minutes to drain the remaining supernatant. 50 µl of DMEM (without supplements) was added per tube, pipetted up and down for 15 times to resuspend the concentrated viral pellet and transferred to an eppendorf tube. This was centrifuged for 10 min at 4000 rpm, RT to remove debris. The supernatant was then transferred to a new eppendorf tube and adjusted to 10 mM MgCl₂, 5 U/ml DNase I was added to the preps to remove plasmid contamination and incubated at 37°C for 30 min. The vector stocks were then aliquoted and frozen at -80°C until use.
**Lentiprep titration by eGFP flow cytometry**

HeLa cells were chosen for the IDLV titration by flow cytometry for GFP expression, as the expression kinetics from IDLV episomes is slower than that from ICLVs and a fast-growing cell line may result in excessive dilution of the gene product (Wanisch, K, Yanez-Munoz, RJ (2009), *Mol Ther* 17(8): 1316-1332). HeLa cells are further slowed down by the polybrene used to aid transduction, therefore they were used for all GFP titrations of IDLVs in this work. The cells were seeded at $10^5$ cells per well in 2 ml medium on 6-well plates. The next day, 10-fold vector dilutions were made in full DMEM, ranging from $10^{-3}$ to $10^{-6}$.

Medium was removed from wells and 1 ml of DMEM with 16 µg/ml polybrene (Sigma) was added per well, as it has been shown to increase transduction efficiency by neutralising the charge repulsion between virions and the cell surface [Davis *et al.*, 2004].

The cells were then transduced with 1 ml of each virus dilution/well, and 1 ml DMEM for mock. 72 h after transduction, cells were harvested, resuspended in 300 µl PBS 4% paraformaldehyde and run through a flow cytometer, using the mock to set a negative population. Titre was calculated using dilution with 1-10% green cells, using the formula [eGFP transducing units (TU)/ml = % green cells x $10^5$ (cells/well on day 0) x l/vector dilution]. Only vectors with a titre higher than $10^7$vp/ml were used in the *in vitro* transduction experiments.

**Lentivector titration by qrt-PCR**

6-well plates were seeded with $10^5$ HeLa cells per well in 2 ml medium. The cells were transduced the next day with 1 ml of $5x10^{-4}$ and $5x10^{-5}$ dilutions of vector preps (Multiplicity of infection (MOIs) 0.5 and 0.05 for a $10^9$/ml vector) and 1 ml of DMEM with 16 µg/ml polybrene (Sigma) and a mock control. 24 h after transduction, cells were harvested and DNA was prepared with DNeasy tissue kit (Qiagen). Quantitative Real-time PCR reactions were set up using ABI universal master mix (2X) and corresponding primer/probe concentrations. Standards (in triplicate) were made in mock cell extract. Samples were measured in duplicate. In the negative control (duplicate) Qiagen DNA elution buffer AE was used instead of cell extract.

Human β-actin qPCR:
Forward primer: 5'-TCACCCACACTGTGCCCATCTACGA-3' (SEQ ID No: 10);
Reverse primer 5'-CAGCGGAACCGCTCATTGCCAATGG-3' (SEQ ID No: 11),
Probe 5'-(VIC)-ATGCCCTCCCCATGCCATCCTGCGT-(TAMRA)-3' (SEQ ID No: 12).
Reaction volume was 25 µl, sample volume 5 µl, primer concentration 900 nM,
probe concentration 200 nM. Standards: 10-10^5 copies of pRTBACTSTD or "HT1080
cell equivalents"/reaction. Standard cycling: 50°C 2 min, 95°C 10 min, 50X (95°C 15
sec, 60°C 1 min). Lentiviral backbone qPCR (Bushman’s late reverse transcript
reaction (Miller et al, 1997)):

Forward primer 5'-TGTGTGGCCCTGTGGTTGT-3' (SEQ ID No: 13), reverse primer
5'GAGTCCCTGTCGAGAGAGC-3' (SEQ ID No: 14).

Reaction volume 25 µl, sample volume 6.25 µl, primer concentration 300 nM,
Standards: 10-10^9 copies of pHRSIN-cPPT-SEW/reaction.
Standard cycling: 50°C 2 min, 95°C 10 min, 50X (95°C 15 sec, 60°C 1 min).

In vitro evaluation of IDLV stability

IDLV transduction Protocol for CHO cells
2 x 10^5 CHO cells per well were seeded on 6-well plates, as this was found to be the
optimal cell density for successful induction of quiescence. After 24 h, the media was
removed and replaced with 1 ml media containing 16 µg polybrene/well. The cells
were then transduced by adding 1 ml media containing vector dilution to MOI 50 as
calculated from the qPCR titres, corresponding approximately to MOI 0.5 by flow
cytometry. The low MOI was chosen in order to avoid non-specific integration
events, which may increase as higher MOIs of vector particles are used. 24 hours
post-transduction, one well transduced with each virus and a control were analysed
for GFP expression by flow cytometry, and a PI stain was done on the same cells for
cell cycle analysis. For the cells designated to undergo quiescence, the media in the
remaining wells was changed for DMEM without methionine, with proline and 2%
FBS. The media was replaced with fresh each day for the next 4 days. At the end of
the quiescent period, one well for each virus and control was used for GFP
expression analysis and cell cycle analysis. The remaining cells were split 1:5 and
seeded in DMEM with proline and 10% FBS. Cell cycle analysis was repeated 3 days
later. After this, cells were split 3 times a week and GFP expression was measured by
flow cytometry at least once a week. The cells were stored in liquid N_2 after ~ 70
days.
For the control cells not undergoing quiescence, transduction was done as above. 24 h post-transduction, one well per vector was analysed for GFP expression and cell cycle phase by flow cytometry. Cell cycle and GFP expression analyses were repeated 4 days later, and following this the cells were cultured normally by splitting 3 times a week and measuring GFP expression by flow cytometry at least once a week. Cultures were discontinued after ~70 days and cells stored in liquid N₂.

Dilution cloning of IDLV-transduced CHO cells

To obtain clonal populations of CHO cells transduced with IDLV vectors and stably expressing GFP, 1 cell population for IDLV SV40 eGFP, IDLV SV40 eGFP mMAR and IDLV CMV eGFP each was retrieved from liquid N₂ and cultured for ~14 days prior to cloning. Cells were seeded on 96-well plates at a density of 0.7 cells/well in 100 µl media containing 50 µl fresh DMEM + Pro + PenStrep +10% FCS, and 50 µl CHO cell conditioned media. Conditioned media was taken from 24-hour old untransduced CHO cell cultures and filtered through 0.2µm syringe filters. After 7 days, 100 µl of the same 50-50 media was added to wells. After a further 7 days in culture, wells containing GFP positive clonal populations were identified by FACS, and 12 clones for each vector were trypsinised and moved onto 12-well plates. For IDLV CMV eGFP, only 3 eGFP positive wells were identified, and at least 2 of these appeared to contain more than 1 clone. Cells from all 3 wells were trypsinised, mixed together and re-seeded at a density of 0.7 cells/well. This time, 12 positive clones were obtained. When a sufficient amount of cells was available for each clone, clonal eGFP expression was confirmed by flow cytometry.

Propidium Iodide stain for cell cycle analysis

CHO-Ki cells were prepared for cell cycle analysis by staining with PI, as follows: approximately 1x10⁶ cells were trypsinised and spun down at 800 rpm for 4 min. The supernatant was discarded and the cells resuspended in the remaining drops of medium by gently flicking the tube. 1 ml of Solution I was added into each tube while shaking and cells incubated at room temperature in the dark for 30 min. 1 ml of Solution II was then added into each tube whilst shaking and incubated at 4 °C for 1-2 h. The stained nuclei were analysed with FACSCantoII flow cytometer to determine the amount of DNA in each nucleus. The FACS data were analysed using FlowJo software (Treestar Inc.)
qrt-PCR of Lentiviral episomes in CHO cells

The quantification of episomes in transduced CHO cells was carried out by Real-Time PCR using the Roche LightCycler 480 system. 100 ng of high molecular weight genomic DNA in 2 µl was used as a template in each reaction. 10 µl of reaction mix contained 300 nM of each primer and 5 µl Absolute Blue qPCR SYBR Green 2X Buffer (Thermo Scientific). Primers were as described in Table 1 below.

Table 1 - Primer sequences used in qRT-PCR and LAM-PCR for lentivectors

<table>
<thead>
<tr>
<th>Primer name</th>
<th>Experiment(s)</th>
<th>Reaction(s)</th>
<th>Primer sequence</th>
<th>SEQ ID No:</th>
</tr>
</thead>
<tbody>
<tr>
<td>4nr</td>
<td>qrt-PCR</td>
<td>1-LTR</td>
<td>5’ GTCTGAGGGATCTCTAGTTAC</td>
<td>15</td>
</tr>
<tr>
<td>2nr</td>
<td>qrt-PCR</td>
<td>1-LTR</td>
<td>5’ TCTCTGGTTAGACGATCTCG</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Inverse PCR</td>
<td>1st PCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>U3 lenti 8</td>
<td>qrt-PCR</td>
<td>2-LTR</td>
<td>5’ GTTGGGAGTGAATTAGCCCT</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Inverse PCR</td>
<td>2nd PCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH535</td>
<td>qrt-PCR</td>
<td>2-LTR</td>
<td>5’ AACTAGGGAACCACCTGCTTAAG</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>LRT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MH532</td>
<td>qrt-PCR</td>
<td>LRT</td>
<td>5’ GAGTCCTGCGTGAGAGAGC</td>
<td>19</td>
</tr>
<tr>
<td>Actin F</td>
<td>qrt-PCR</td>
<td>B-actin</td>
<td>5’ TGGCATCCACGAACTACAT</td>
<td>20</td>
</tr>
<tr>
<td>Actin R</td>
<td>qrt-PCR</td>
<td>B-actin</td>
<td>5’ TGGTACCAACCAGACAGACT</td>
<td>21</td>
</tr>
<tr>
<td>HIV 543 bio</td>
<td>LAM-PCR</td>
<td>Linear PCR</td>
<td>5’ AGTGCTTTCAAGTAGTGCTG</td>
<td>22</td>
</tr>
<tr>
<td>SK LTR 1</td>
<td>Inverse PCR</td>
<td>1st PCR</td>
<td>5’ GAGCTCTTCTGGCTAACTAG</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd PCR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SK4 LTR bio</td>
<td>LAM-PCR</td>
<td>1st exp. PCR</td>
<td>5’ AGTAGGTGTGCCCCTCCTG</td>
<td>24</td>
</tr>
<tr>
<td>LC I</td>
<td>LAM-PCR</td>
<td>1st exp. PCR</td>
<td>5’ GATCTGAATTCAGTGCCACAG</td>
<td>25</td>
</tr>
<tr>
<td>HIV LTR 566</td>
<td>LAM-PCR</td>
<td>2nd exp. PCR</td>
<td>5’ GTCTGGTGCTGACTCTGGTAAC</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC II</td>
<td>LAM-PCR</td>
<td>2nd exp. PCR</td>
<td>5’ AGTGCCACAGCAGTTAGG</td>
<td>27</td>
</tr>
</tbody>
</table>

Standards for the i-LTR, 2-LTR and LRT reactions were made by diluting a plasmid containing each target sequence so the copy number ranged from 10^2 to 10^9 per reaction. The standard for the β-actin reaction was based on DNA concentration, and was calculated as follows: lxio? cells were counted and the DNA was extracted
using the high molecular weight protocol. On average, this resulted in 300 ng of DNA in a volume of 200 µl. Hence, the amount of DNA expected to be obtained was 3 pg/cell.

Each sample was analysed twice in triplicate, and any outliers or replicates producing abnormal melting curve were excluded. Averages from all experiments and duplicates were used in further analyses. The reaction conditions were standard qRT-PCR cycling: 50°C 2 min, 95°C 10 min, 50X (95°C 15 sec, 60°C 1 min).

Linear amplification-mediated PCR (LAM-PCR)

Linear amplification-mediated PCR (LAM-PCR) is a highly sensitive method that can be used to identify unknown sequences flanking known sequences (Schmidt et al., 2007). The protocol consists of a linear PCR with a biotinylated primer, and the resulting product is purified using streptavidin-coated magnetic beads. A second strand is synthesised using random hexanucleotide primers and Klenow fragment of DNA polymerase, and the double-stranded product is digested with a suitable restriction enzyme that cuts within ~200 bp from the 5' primer into the vector sequence and at unknown sites into the genomic DNA. Next, a unidirectional double-stranded linker is ligated to the sticky end created by the restriction enzyme. The DNA is denatured using NaOH and subsequently subjected to two exponential nested PCR reactions to achieve maximum amounts of specific product. The first set of primers bind to the linker cassette and 5'biotinylated vector, the second set bind to linker cassette and vector sequence (LTR). The products can be visualised on an agarose gel and compared to expected fragment lengths, and sequenced either by extracting the bands from the gel or by conventional shotgun cloning and sequencing (Gabriel et al., 2009; Schmidt et al., 2007).

Here, LAM-PCR was performed on all clonal cell line hmw DNA, with integrating clone 6 as a positive control and DNA from CHO cells transduced with IDLV SV40 eGFP and harvested 24 h post-transduction as a negative control. The protocol was performed as previously published (Schmidt et al., 2007). The LAM-PCR reactions were carried out by the author in Christof Von Kalle's laboratory at the National Centre for Tumour Diseases at the German Cancer Research Institute in Heidelberg, Germany. The subsequent Deep sequencing was carried out in the same laboratory using the Roche 454 sequencing system. The bioinformatic analysis of the results was performed by Dr Uwe Appelt.
Linear PCR
Each reaction contained 2 µl of hmw DNA ([c] 250 ng/µl), 2.5 µl 10X Taq Buffer (Qiagen), 1 µl dNTPs (10 µM each, Fermentas), 0.25 µl Taq (Qiagen), 0.5 µl Primer LRT 543 bio (0.5 pmol/µl), and water to a total volume of 25 µl. The PCR reaction was performed as follows: 95°C for 2 min, 95°C for 45 s, 58°C for 45 s, 72°C for 1 min, steps 2-4 repeated 50 times, 0.25 µl Taq added, Steps 2-4 repeated 50 times, 72°C for 5 min.

Restriction enzyme digest
The product was digested with TSP509I, which cuts ~210 bp into the vector sequence from the 1st primer in a LTR-backbone or LTR-genome junction, and -105 bp into the second LTR sequence in a LTR-LTR junction.

1st & 2nd Exponential PCR
Each reaction contained 2 µl of product from the linear PCR reaction, together with 2.5 µl 10X Taq Buffer (Qiagen), 0.5 µl dNTPs (10 µM each), 0.25 µl each primer (SK4 LTR bio and LC-i), 0.25 µl Taq (Qiagen) and water to 25 µl. The reaction conditions were the same as above, but repeated only for 35 cycles. For the 2nd nested PCR, 1 µl product from previous reaction was used as a template for primers HIV LTR566 and LC II.

The products of the reaction were visualised on a Spreadex gel (Elchrom Scientific). They were also subjected to sequencing using the Roche 454 system, creating an average of ~1200 independent sequence reads per sample.

Southern Blotting
Southern blotting is a method for detecting specific DNA sequences present within a DNA sample. The DNA samples are usually digested using restriction enzymes, and the fragments are separated by size by agarose gel electrophoresis. The DNA fragments are transferred onto a membrane using capillary transfer, and the membrane is then exposed to a hybridisation probe which is labelled to enable detection. The probe is designed to be complimentary to the DNA sequence of interest (Southern, 1975).

Digestion of mammalian genomic DNA
5 µg of high molecular weight DNA was diluted with water and 10X enzyme buffer for a final digestion volume of 60 µl (ie. 6 µl ofiox buffer + water + DNA = 56 µl) and
allowed to stand at RT for 30 min. Spermidine was added to a final concentration of 2 mM (1.2 µl of 100 mM) to increase enzyme activity and the mixture incubated at RT for 1 h. 3 µl of an appropriate restriction enzyme was added and the digest incubated overnight at the optimum temperature for the enzyme. After 12 h, 1 µl of enzyme was added, and the incubation continued for further 12 h or until viscosity was gone. A 0.7 % agarose gel was run with 1 µl of digests to estimate amounts to be loaded in Southern Blot gel.

Preparation of hot ladder
A mixture was made up of 1 µg DNA ladder, 2 µl 10x NTB, 4 µl CTG, 1 µl (10 µCi [α-32P]dATP, 1 µl Klenow, water was added to 20 µl, and the mix incubated at RT for 90 min. 40 µl of TE was then added to stop reaction. A Microspin S-300HR column was packed by spinning at 3003g for 1 min. Sample was loaded on column and centrifuged at 3000 g for 2 min. Approximately 15 Geiger counts of ladder were loaded per lane. Remaining hot ladder was stored at -20°C.

Agarose gel for Southern Blot
A 0.7% agarose gel was cast in 1x TAE with ethidium bromide. The gel was loaded with 10 µg of digested genomic DNA per lane and hot DNA ladder. The gel was run in TAE with ethidium bromide overnight in at 35V, constant voltage.

Salt transfer and UV fixation protocol
Gel was agitated in 0.25 M HCl for 10 min and washed twice with water. It was then agitated in denaturing solution for 30 min and washed twice with water. Lastly it was agitated in neutralising solution for at least 30 min. A GeneScreen Plus membrane was cut to the size of the gel and pre-wet in water, followed by equilibrating in IOXSSC for 15 min. A capillary blot was then set up using IOxSSC. Order from bottom: Inverted gel, GeneScreen Plus membrane, two IOxSSC - soaked Whatman 3MM paper pieces, stack of blotting paper about 10 cm high, glass plate, book weighing about 500 g. The blot was left to transfer overnight. The following day, the membrane was retrieved and agitated in 0.4 M NaOH for 1 min followed by 0.2 M Tris-CIH pH 7.5, 1x SSC for 1 min. It was washed with 2xSSC and DNA cross-linked to membrane using Stratagene UV crosslinker.

PCR for eGFP Probe
PCR was performed to create a 717 bp fragment of the eGFP gene. Plasmid pRRLsc SV40 eGFP was used as template. Forward primer: 3’ATG GTG AGC AAG GGC GAG
(SEQ ID No: 28), reverse primer: 3’CTT GTA CAG CTC GTC CAT (SEQ ID No: 29).
The reaction was set up with an annealing temperature of 56 °C and a total of 35 cycles.

5  **Probe labelling by random priming**
100 ng of DNA fragment to be labelled was made up to 4 µl with TE, 1 µl of 60 ng/µl dN6 was added, the mixture boiled for 5 min, and cooled on ice for 2 min.
12 µl of the following mixture was then added to the denatured DNA: 12 µl dH2O
3 µl 1x random priming buffer, 3 µl CTG. Finally 2 µl (20 µCi [α-32P] dATP and 1
µl Klenow were added and the reaction incubated at RT for 1 h. After incubation, 40
µl TE buffer was added, the sample was loaded onto a pre-packed Microspin column
at 3000 rpm for 2 min.

**Hybridisation**
Membrane was pre-hybridised with 15 ml Church mix at 68 °C for at least 1 h.
Labelled probe was boiled for 5 min, transferred to ice for 2 min, and added to hybridisation tube. The reaction was left to hybridise overnight. The membrane was washed with 2xSSC, 0.5% SDS at 65°C 3 times quickly and 3 x 10 min washes. Finally it was washed with 2xSSC and exposed to Phosphoimager screen.

**Fluorescent in Situ Hybridisation**
**Cell Fixing Protocol**
Approximately 1.5 x 10⁶ CHO cells were detached using trypsin/EDTA. Cells were transferred into a 15ml Falcon tube and centrifuged at 1000 rpm for 5 minutes to pellet. The supernatant was decanted and the pellet resuspended in remaining drops of media. 7 ml of prewarmed, hypotonic KC1 solution was added, tubes inverted to mix and incubated at RT for 5 min. This was done in order to enlarge the cells by osmosis, so that the nuclei would lie further apart on the resulting slide. Tubes were centrifuged at 1200 rpm for 6 min, supernatant decanted, and the pellet again resuspended in the remaining drops by flicking. To fix the cells whilst preventing shrinkage, 5 ml of 5% acetic acid was added, the tubes were inverted to mix, and centrifuged at 1200 rpm for 6 min. Supernatant was removed and pellets resuspended in 5ml of freshly made fixative. Any clumps were removed with a
Pasteur pipette. Cells were washed with fixative twice more to remove any traces of media. Finally, the fixed preparation was centrifuged at 1200 rpm for 6 min and stored at -20 °C.

5 FISH Probe labelling
FISH probes were prepared for each vector by labelling the corresponding transfer plasmid, each containing around at least 3 kb complementary DNA, with SpectrumGreen and SpectrumOrange fluorochromes (Abbott Molecular). The plasmid DNA was labelled using Nick Translation System according to the manufacturer's instructions, using 1 μg plasmid and labelling with both SpectrumGreen and SpectrumOrange dUTPs. The reaction was incubated at 15°C for 1.5 h and the product stored in the dark at -20 °C.

Probes mixes were prepared for FISH by adding Cot-i DNA to each labelled probe (50X the weight of probe) and then 2x the total volume 100% ethanol. The probes were pelleted and dried using Speedy Vac apparatus and Savant Centrifuge/concentrator. The dry pellets were resuspended in Abbott LSI/WCP Hybridisation buffer at a concentration of 6.7 ng/μl and stored in the dark at -20°C.

20 FISH Run Procedure
Fixed cells were dropped onto slides pre-soaked in fixative and the area containing cells was marked using a diamond pen. Slides were soaked in PBS pH 7.4 for 5 minutes and then rinsed twice with sterile water. They were then dehydrated by incubating for 2 min each in 70%, 90% and 100% ethanol. The slides were air-dried, and 15 μl (song) probe was applied onto a cover slip which was then placed onto the slide and sealed with rubber solution (Weldtite Ltd). The slides were incubated at 72°C for 6 min and then at 37°C overnight. The next day, the coverslips were removed by washing in 4x SSC and 0.05% Tween pH 7.0. A stringent wash was performed for 5 min at 72°C in 0.4X SSC pH 7.0. After this, the slides were incubated for 2 min at RT in 4x SSC/0.05% Tween pH 7.0, and then rinsed 3 times with PBS for 2 min each time. After the final PBS wash, the slides were air dried and 15 μl of the nuclear stain DAPI (Cytocell,i60 ng/ml) was applied together with a new cover slip, which was then sealed with nail varnish.

35 Imaging and analysis of FISH signals
Although the cell cultures from which the FISH slides were derived were not blocked in metaphase, at least 50 spontaneous metaphases were present on all the slides.
Images were taken for ~20 metaphases per slide using 100χ magnification and the Zeiss Axioplan 2 microscope and Metasystems Isis 2 software. Three slides were designated for detailed analysis: mock transduced control, and clonal cell lines transduced with either integrating or non-integrating SV40 GFP vector. For each slide, 80 fields using 40χ magnification lens were processed by counting the number of nuclei in the field, the number of signals outside nuclei, number of signals inside nuclei, and number of nuclei without signals. The ratio of signals inside nuclei versus outside nuclei was calculated for each field, to account for variation within different regions of the slide. The number of signals per nucleus was calculated as the number of "in" signals per field minus the number of nuclei without signals. Statistical analyses were performed on A) the ratio of signals inside vs outside nuclei, using one-way ANOVA and comparing "true" samples to the negative control, and B) the number of signals per nucleus on each slide using one-way ANOVA.

Results
Example 1 - Computational duplex destabilisation analysis of the episomal IDLV structures
Both integrating and non-integrating lentiviral genomes give rise to circular products during cellular infection. Circularisation can occur by homologous recombination between the two LTRs, or the linear DNA can be circularised by non-homologous end-joining. In this way, circles containing either one or two copies of the LTR are formed. 1-LTR circles have been found to be several-fold more common than 2-LTR circles in vivo, although non-integrating vectors may produce comparatively more 2-LTR forms (Bayer et al., 2008). After comparison of the SIDD profiles for all S/MAR-containing and non-S/MAR-containing episomal forms of both viruses, no significant difference was found between the alternative 1-LTR and 2-LTR forms of the same cassette (Figure 2). Both vectors have identical destabilisation profiles excepting the slightly longer condensed LTR region in the 2-LTR form, which suggests that neither form should be favoured for episome establishment based on the LTR conformation as other parts of the vector are not affected. In Figure 3, only 1-LTR forms of the episomes are shown.

Since S/MAR elements have been shown to need read-through transcription to function effectively, there were two theoretically possible locations within the vector cassette; immediately after the eGFP gene, or after the Gag gene but relying on residual promoter activity from the LTR. Two different S/MAR elements were
analysed for strand separation potential; the full-length β-interferon S/MAR and its shortened mutant daughter element, miniMAR.

The inclusion of both S/MAR elements was found to condense the SV40 promoter adjacent to the transgene, with energy required for strand separation increasing from < 6 kcal/mol to 7 kcal/mol (Figure 3). However, there was no significant difference in the promoter destabilisation between the early and late locations of the S/MARs. This together with the fact that transcription is expected to be strongest from the SV40 promoter lead the inventors to design the constructs using the late location for the S/MAR elements.

Example 2 - Optimisation of induced cell cycle arrest for lentiviral transduction
Following the IDLV transduction, half of the cells were induced by methionine and serum depletion to enter a reversible period of cell cycle arrest. The cell cycle phase analysis was performed using propidium iodide staining of the nuclei followed by flow cytometric analysis and computer analysis of the resulting data.

CHO cell cultures transduced with various IDLVs were induced to undergo transient cell cycle arrest by methionine and serum depletion of the culture media. Cells were plated, transduced after 1 day, and then treated with methionine-free medium from D2 to D7 over a 5 day period. At D7, cultures were returned to normal growth medium and allowed to proliferate as normal with routine passaging as required. To confirm induction of cell cycle arrest, cell cycle phase analyses were performed by propidium iodide staining and flow cytometry at the D2 and D7 time points, effectively before and after the methionine depletion period.

As shown in Figure 4 both for cells transduced with IDLV vectors and mock control, methionine depletion of culture media effectively increased the proportion of cells in G1 phase (P < 0.05), and decreased the proportion of cells in S-phase (P < 0.01) and in G2 phase (P = 0.0001). This confirms the action of methionine depletion to induce a transient cell cycle arrest, and the independence of the cell cycle arrest of IDLV transduction.

Example 3 - Optimisation of in vitro IDLV transduction of CHO cells
To obtain information on the minimum multiplicity of infection (MOI) of IDLV vector needed to transduce ~100% of CHO cells, a range of low MOIs were tested.
The purpose of the experiment is to avoid both suboptimal transduction levels and high incidence of integration resulting from excessive MOIs used.

1 x 10⁵ CHO cells were seeded on 6-well plates in 2 ml media. 24 hours after seeding, the media was removed and replaced with 1 ml fresh media containing virus particles corresponding to 1, 2, and 5 virus particles per cell (MOIs 1, 2 and 5), as calculated from the flow cytometry titres. The virus/media mixture included polybrene at a final concentration of 8 µg/ml, to aid transduction by neutralising the charge repulsion between the virions and the cell surface (Davis et al., 2004). The cells were incubated at 37°C, and GFP expression in the transduced cells was measured 24 h and 72 h post-transduction by flow cytometry. MOI of 1 was found to be sufficient to give > 90% eGFP positive CHO cells.

Example 4 - A period of induced cell cycle arrest results in high stable transduction of CHO cells bVIDLVs

To explore the impact of including the miniMAR element on episome retention, an in vitro transduction experiment was conducted using 3 different VSVg-pseudotyped integrase-deficient lentivectors. All constructs contained the eGFP gene, either driven by the CMV promoter or SV40 promoter with or without the additional miniMAR element.

All lentiviral infections resulted in 92-100% transduction efficiency. Two days after transduction, cells transduced with IDLV SV40 eGFP and IDLV SV40 eGFP mMAR were 98% eGFP positive and those transduced with IDLV CMV eGFP were 91% eGFP positive (Figure 5). In the cells placed in induced cell cycle arrest, the eGFP expression increased to 100% during 5 days of non-division.

In the cells in continuous culture, eGFP expression declined steadily, reaching levels of under 3% after 14 days (Figure 5, B and D). The average percentage of transduced cells still expressing eGFP above background after 72 days in culture for IDLV SV40 eGFP was 0.8%, for IDLV SV40 eGFP mMAR 0.5% and for IDLV CMV eGFP was slightly higher at 1.4%. (N = 3). The difference between vectors is not statistically significant (One-way ANOVA).

The GFP expression pattern in the cells subjected to an initial period of cell cycle arrest was distinctly different from those in continuous culture. Although the proportion of eGFP expressing cells did decline somewhat when cells were released
from the mitotic block into continuous culture, the difference in the stable expression levels was approximately 10-fold compared with the cells not held in cell cycle arrest. The population reached stable expression levels after approximately 14 days in culture in accordance with the cells in the control group, but holding the cells in cell cycle arrest post-transduction increased the stable levels of eGFP expression from less than 0.5% above background to 6% - 45% (Figure 5 A and C).

The highest retention levels were achieved by the IDLV SV40 eGFP vector, with an average of 25% of cells remaining GFP-positive after 72 days in culture. The same vector but including the mMAR element resulted in slightly lower levels of eGFP expression, at an average of 15% positive cells after 72 days. The differences in transgene expression levels for the IDLV SV40 eGFP and IDLV SV40 eGFP mMAR vectors, between the cells held quiescent and in continuous culture, were highly statistically significant (P < 0.001, One-way ANOVA and Tukey's Test for pairwise comparison, including the last 3 data points, n = 3.) For IDLV CMV eGFP the difference was also significant, P < 0.01.

Comparing cell populations which were held in cell cycle arrest but transduced with different vectors, there are also significant differences. The greatest difference is between IDLV SV40 eGFP and IDLV CMV eGFP (P < 0.001). The difference between the S/MAR-containing and non-S/MAR-containing SV40 vector is also significant (P < 0.01), using One-way ANOVA followed by Tukey's post-test. The proportion of cells expressing eGFP in all of the transduced cell lines fluctuated slightly during ~70 days in culture. In one cell population not held in cell cycle arrest and transduced with IDLV CMV eGFP, the transgene-expressing percentage began to rise after ~35 days in culture, increasing from an average of 1% to 4% over the course of a month. Similarly, after extended culture the three cell populations simultaneously transduced with IDLV SV40 eGFP and initially held quiescent had diverged in the percentage of GFP-expressing cells by more than 3-fold and contained 45%, 20% and 12% GFP positive respectively.

**Example 5 - High transduction of CHO cells by IDLVs following a period of cell cycle arrest is a clonally stable phenomenon.**

To obtain more information about the stably expressing cell populations, dilution cloning was performed for cells transduced with each of the 3 vectors. 14 days after seeding, 35 out of 82 colonies (43%) for IDLV SV40 eGFP and 17 out of 124 colonies (14%) for IDLV SV40 eGFP mMAR were expressing eGFP. Since the total cell
population transduced with IDLV CMV eGFP was initially only 11% positive, two rounds of dilution cloning was required to obtain a sufficient number of positive clones. Twelve eGFP-positive colonies for each vector were expanded to obtain clonal populations, and analysed for eGFP expression and Mean Fluorescence Intensity (MFI) by flow cytometry.

The FACScan analysis of the percentage of eGFP expressing cells and the MFI was used to conclude whether a given population had likely risen from a single cell (Figure 6). Some eGFP-positive populations appeared to be of mixed origin, as both eGFP-positive and negative cells were found in many of the populations. For further analysis, 3 clones with close to 100% eGFP positive cells, 14 days post-cloning and with a Gaussian MFI distribution with different mean MFIs were chosen for each of the 3 vectors. The 9 clonal populations were kept in continuous culture and screened every 10 days by flow cytometry to ascertain whether the MFI and percentage of cells expressing eGFP remained stable (Figure 7). All of the clones retained the majority of eGFP expressing cells, as well as exhibiting consistent MFI levels over 50 days in culture. The level of eGFP expression remained most stable in cells transduced with SV40 promoter driven vectors; in all 3 clonal cell lines transduced with IDLV SV40 eGFP, the eGFP expression varied less than 1% over 35 days and remained above 95% for all clones. In the IDLV SV40 eGFP mMAR-transduced cells, the variation between each measurement was less than 3.5% over the same time period, and whilst clones 8 and 11 were over 97% eGFP positive, clone 5 remained consistently 82-85% positive.

Clonal cell lines derived from the population transduced with IDLV CMV eGFP showed a tendency towards a decline in eGFP expression. Initially, the clonal populations were 63%, 93% and 99% GFP positive, and by day 50 post-cloning the expression had increased up to 76% in the first clone, and declined to 83% in both other clones. This is in line with the observations from the polyclonal population, where one population also showed an increase and the other two showed a decrease in eGFP expression.

Example 6 - Evaluation of episomal status of IDLV genomes in CHO cell clones stably transduced following a period of induced cell cycle arrest

All of the clonal, stably eGFP-expressing populations were derived from cells transduced with integrase-deficient lentivectors, and hence the transgene cassette is unlikely to have integrated into the host genome via Integrase-mediated pathways.
We decided to employ a battery of techniques to ascertain whether the transgene in these stably expressing cells remains episomal or has been integrated into the host genome.

There are several methods which can be employed to assess the episomal status of the vector, each of them with their own caveats. To overcome such limitations associated with each technique, we decided to combine several of them and use the combined results to draw final conclusions. To this end, vectors in the transduced nuclei were visualised using Fluorescent In Situ hybridisation (FISH), however the small size of the vector cassette meant operating on the outer limit of the sensitivity of the technique. Southern blotting was used to assess vector digest patterns in the highly expressing clonal populations, but it is not sensitive enough to be used for the polyclonal populations. On molecular level, linear amplification-mediated PCR was used to amplify regions adjacent to the LTRs, which were then subjected to sequencing to distinguish the vector's genome from the host's genome. Although each technique has their own limitations, used together they can provide evidence regarding the integration status of the vectors.

Example 7 - Agarose gel analysis and high throughput sequencing of LAM-PCR products amplified from DNA from IDLV-transduced CHO cell clones

Linear amplification-mediated PCR (LAM-PCR) is a method whereby unknown DNA sequences adjacent to a known sequence can be amplified for sequencing. It is generally used for analysing lentivector integration sites, and in the case of IDLVs it can be used to detect the proportional frequency of integrants and episomal forms. This method was used to analyse the 3 distinct clonal populations for each of the 3 vectors shown in Figure 7. Each sample was subjected to an initial linear PCR reaction starting from the LTR, followed by second strand synthesis. The strands are then digested with a restriction enzyme that cuts approximately at 210 bp if the strand was synthesised from an LTR towards the vector genome, or -105 bp if it was synthesised from an LTR joined to another LTR, and at a variety of lengths when the LTR is joined to the host genome.

These fragments were ligated to a linker at the overhang created by the restriction enzyme, and the resulting double-stranded DNA molecules with known sequences at either end were subjected to 2 cycles of nested PCR with primers binding to LTR and linker sequences. The products of the second exponential PCR reactions were visualised on a gel to obtain information on the fragment sizes prior to sequencing.
(Figure 8). All clones of IDLV SV40 eGFP produced a band consistent in size (280 bp) with a fragment produced as the primer bound to an LTR is extended towards the vector genome, and all clones of IDLV CMV eGFP produced a band consistent with the primer binding an LTR adjoining another LTR (180 bp). Two of the IDLV SV40 eGFP mMAR clones (8 and 11) also gave bands corresponding to the LTR-vector genome junction, although clone 11 also has a second band indicating a possible integrant. The single band in the lane for IDLV SV40 eGFP mMAR clone 5 does not correspond to either the 1-LTR or the 2-LTR band sizes.

The PCR products from LAM-PCR were subjected to high throughput sequencing, resulting in an average of 1200 reads per sample. Three samples resulted in less than 400 sequence reads (IDLV SV40 eGFP clone 4, IDLV SV40 eGFP mMAR clone 5, and IDLV CMV eGFP clone 10), and were discounted from further analysis. The threshold for a required number of sequence reads in order to be included in the analysis was set at 400 reads per sample, defined as ≥1/3 of the average. The sequencing results were divided into 2 main categories, Internal Bands (IB) containing only vector DNA, and non-IBs containing vector DNA and other sequences (Figure 9).

The majority of reads for all SV40-containing vectors contained only vector sequences. To investigate the sequence reads classed as non-vector (red sections of bars in Figure 9), they were blasted against the mouse and rat genomic sequences in the absence of an annotated sequence for the Chinese hamster genome. The non-vector sequences in IDLV SV40 eGFP clone 7 and IDLV SV40 eGFP mMAR clone 11 did not contain a single BLAST result. This suggests that the non-vector sequences in these samples were the result of contamination from surrounding lanes containing HrV-transduced human cells as the samples were sequenced in a facility usually processing such samples. The vector-genome sequences in IDLV SV40 eGFP clone 10 and IDLV SV40 eGFP mMAR clone 8 did result in some BLAST hits, however these were excluded after further analysis established them as too partial and fragmented to be likely true matches. Of all the IDLV samples, only IDLV CMV eGFP clone 2 contains a likely integrant, at a site corresponding to the murine chromosome 4. The integrating control sample resulted in high read counts for 2 integration sites, one corresponding to the murine chromosome 1 and the other to chromosome 2. The 1 read for the same site discovered in IDLV CMV eGFP clone 12 is likely to be contamination from the adjoining lane.
Following the publication of a draft genomic sequence for the Chinese hamster, the sequences corresponding to loci in the murine chromosomes 1, 2 and 4 were verified by BLAST analysis to be present also in the CHO cells. As the genome is not annotated, no information could be retrieved regarding the location of these sites (Xu et al., 2011).

BLAST analysis of sequences generated by high-throughput sequencing of CHO clones transduced with IDLVs verifies very few integration events.

Table 2 - High throughput sequencing results

<table>
<thead>
<tr>
<th>Clone</th>
<th>Total Reads</th>
<th>Vector Reads</th>
<th>Verified Host</th>
<th>Host location (mouse)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDLV1 clone 7</td>
<td>1191</td>
<td>1122</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>IDLV1 clone 10</td>
<td>906</td>
<td>852</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>IDLV2 clone 8</td>
<td>1269</td>
<td>1196</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>IDLV2 clone 11</td>
<td>1010</td>
<td>651</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>IDLV3 clone 2</td>
<td>1500</td>
<td>1275</td>
<td>1</td>
<td>Chr 4</td>
</tr>
<tr>
<td>IDLV3 clone 12</td>
<td>1299</td>
<td>946</td>
<td>1</td>
<td>Chr 1</td>
</tr>
<tr>
<td>ICLV control</td>
<td>1279</td>
<td>403</td>
<td>248</td>
<td>Chr 1, 239 reads; Chr 2, 9 reads</td>
</tr>
</tbody>
</table>

Genomic DNA from the stably GFP+ CHO clones and cells transduced with an integrating LV was subject to LAM-PCR. The products were analysed by high-throughput deep sequencing followed by BLAST analysis of all generated sequences against the vector genome and the mouse genomic sequence as the closest available sequenced relative of the Chinese hamster. The table shows numbers of successful sequencing reads per sample (Total Reads), and numbers of reads containing only vector sequences (Blue) and other sequences verified by BLAST analysis against the murine genome (Red). The green column indicates the corresponding location of the BLAST result in the mouse genome.

Example 8 - Copy number and conformation analysis of vector genomes IDLV-transduced CHO cells by Quantitative Real-Time PCR

To gather additional data on the episome numbers and integration status, the polyclonal samples and 3 clonal samples per vector were also analysed using quantitative real-time PCR (qRT-PCR). qRT-PCR allows for a simultaneous
amplification and quantification of target DNA using a fluorescently labelled reporter. Here, the method was used to quantify 3 different vector sequences: i) vector backbone adjacent to the LTR, which should be present once in all vector-containing samples; ii) LTR, which should be present once or twice in every sample; and iii) LTR to LTR junction, which should be present in samples containing 2-LTR circles (Figure 10 A). The quantities of these 3 sequences were normalised to the amount of cellular DNA present in each sample by quantifying the amount of CHO β-actin.

High molecular weight DNA from both polyclonal and clonal cell populations was subjected to all 4 qRT-PCR reactions, and the copy number of each of the vector sequences was then calculated per cell using β-actin data to estimate number of cell equivalents per sample (Figure 10 B). Each vector genome is expected to produce one backbone (BB) signal, one or two LTR signals, and one or none LTR-to-LTR signals, depending on the conformation and integration status. The amount of BB signal can be used to estimate vector copy number and the ratios of the 3 independent vector signals can be used to infer conformation and integration status.

Figure 10 B shows the amounts of the 3 independent vector signals as calculated per cell equivalent. Most samples resulted in low amounts of vector signal, corresponding to 1-15 vector genomes per cell. The 3 polyclonal samples indicated average vector copy numbers of 0.6/cell for IDLV CMV eGFP, 1.1/cell for IDLV SV40 eGFP mMAR and 1.6/cell for IDLV SV40 eGFP, which reflects the differing percentages of eGFP-expressing cells in these populations at the time of investigation (Figure 4 C). The clonal populations transduced with IDLV SV40 eGFP averaged 2.8 (Clone 4), 0.6 (Clone 7) and 1.1 (Clone 10) backbone copies per cell. For cells transduced with IDLV SV40 eGFP mMAR, the average backbone copy numbers were 0.1 (Clone 5), 4.8 (Clone 8) and 6.0 (Clone 11) per cell. The average BB copy numbers for clonal populations derived from IDLV CMV eGFP -transduced cells were 4.4 (Clone 2), 1.8 (Clone 10) and 1.4 (Clone 12) signals per cell.

In addition to data on vector copy numbers, the qPCR data was used to provide information on vector genome conformation. This was based on the following principles:
If the vector has integrated via the integrase-mediated pathway, two copies of the LTR signal should be present for each of the backbone signals, with no LTR-to-LTR signal present (first row of Table 3).

If the vector is present as episomal 1-LTR circles, the ratio of LTR to backbone signals should be 1:1 with no LTR-to-LTR signal present (second row of Table 3).

If the vector is present mostly as 2-LTR episomes, each vector genome should produce one LTR-LTR signal, two LTR signals, and one backbone signal (third row of Table 3).

The ratios of 1-LTR and 2-LTR signals per copy of backbone signal were calculated for each vector (Table 3). Only one sample produced a ratio suggesting a possible Integrase-mediated integration event; IDLV SV40 eGFP clone 4 gave a ratio of 2:1 of LTR:BB (2.07:1 averaged over all data points). The remaining clones produced ratios indicative of either 1-LTR or 2-LTR episomes.

Table 3 - Ratios of vector signals suggest the majority of vectors persist in episomal forms.

<table>
<thead>
<tr>
<th>Clone</th>
<th>BB</th>
<th>1-LTR</th>
<th>2-LTR</th>
<th>Inferred</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTEGRATED</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPISOMAL 1-LTR CIRCLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPISOMAL 2-LTR CIRCLE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IDLV SV40 eGFP Clone 4</td>
<td>1</td>
<td>2.07</td>
<td>0.01</td>
<td>Integrated</td>
</tr>
<tr>
<td>IDLV SV40 eGFP Clone 7</td>
<td>1</td>
<td>0.65</td>
<td>0.07</td>
<td>1-LTR Episomal</td>
</tr>
<tr>
<td>IDLV SV40 eGFP Clone 10</td>
<td>1</td>
<td>1.12</td>
<td>0.42</td>
<td>1-LTR &amp; 2-LTR Episomal</td>
</tr>
<tr>
<td>IDLV SV40 eGFP mMAR Clone 5</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>IDLV SV40 eGFP mMAR Clone 8</td>
<td>1</td>
<td>1.45</td>
<td>0.01</td>
<td>1-LTR Episomal</td>
</tr>
<tr>
<td>IDLV SV40 eGFP mMAR Clone 11</td>
<td>1</td>
<td>0.96</td>
<td>0.01</td>
<td>1-LTR Episomal</td>
</tr>
<tr>
<td>IDLV CMV eGFP Clone 2</td>
<td>1</td>
<td>2.35</td>
<td>0.23</td>
<td>2-LTR Episomal</td>
</tr>
<tr>
<td>IDLV CMV eGFP Clone 10</td>
<td>1</td>
<td>5.02</td>
<td>1.07</td>
<td>2-LTR Episomal</td>
</tr>
<tr>
<td>IDLV CMV eGFP Clone 12</td>
<td>1</td>
<td>5.14</td>
<td>1.80</td>
<td>2-LTR Episomal</td>
</tr>
</tbody>
</table>
The expected ratios of signals for each of the 3 possible vector conformations are shown in the first 3 rows (blue). Absolute amounts of qPCR signals were normalised for cell number and used to calculate ratios of i-LTR and 2-LTR signals per copy of backbone signal in each sample. The predominant conformations for each sample as inferred from the signal ratios are shown in the right hand column. N=9. BB: Vector backbone.

Example 9 - Integration status analysis of vector genomes in stably transduced CHO cell clones by Southern blotting

Southern blotting is a method that can be used to detect the presence and abundance of a known DNA sequence within a sample, as well as inspecting the integration status of the sequence by probing DNA digested with a restriction enzyme. Here, high molecular weight DNA extract from transduced cells was subjected to a restriction enzyme digest that cuts once within the vector genome and separated on an agarose gel. The DNA was then transferred onto a membrane and probed with a radioactive probe. If episomal, the resulting bands should correspond to the expected size of the episome; if integrated, each integration event should produce a band not likely to correspond to the expected episome size. The results show multiple integration sites in the clonal control transduced with integration-competent SV40 eGFP control vector (Figure 11, lane 1). The controls in lanes 2 and 3 contain DNA was harvested 24 hours after transduction with IDLV SV40 eGFP and ICLV SV40 eGFP, and both show bands corresponding to i-LTR and 2-LTR circles.

Lanes 4-6 contain DNA from clonal populations transduced with IDLV SV40 eGFP; clone 4 shows bands corresponding to both episomal forms and clone 7 shows a faint 2-LTR sized band. Of the 3 clonal populations transduced with IDLV SV40 eGFP mMAR in lanes 7-9, clone 11 shows a band corresponding to the i-LTR episome (lane 9). The DNA in the other two clones has not digested sufficiently to enter the gel.

Clone 2 of IDLV CMV eGFP (lane 10) shows a band corresponding to i-LTR and 2-LTR episomes in addition to several other bands. All IDLV CMV eGFP clones have the same band at ~2200 bp, which is too small to contain the entire episome and may therefore indicate either rearrangement or integration.
Example 10 - Conformation and copy number analysis of vector genomes using Fluorescent In Situ Hybridisation (FISH) has been applied commonly to scan for large chromosomal abnormalities, but some successes have also been reported with small single-copy targets (Rupprecht et al., 2009). In our case, it was successful in providing images for vectors containing the SV40 promoter. The images were analysed in 2 different ways: firstly, a qualitative analysis of metaphase images was used to obtain auxiliary information on episomal status. Secondly, statistical analysis was applied on numbers of signals counted in large numbers of fields to account for variation within slides and minimise observer bias.

The metaphase images provide supportive evidence for the presence of episomes in the cells transduced with non-integrating vectors, as only 1 doublet signal was found on all metaphase images on the IDLV slides. In contrast, 8 clear doublets were observed in metaphases of the 3 different ICLV clones investigated (Figure 12 A). The samples included in the statistical analysis were clonal IDLV SV40 eGFP, IDLV SV40 eGFP mMAR and ICLV SV40 eGFP cell lines, as well as a negative control. Signals and nuclei were counted in 20 fields throughout the slide, and the ratio of signals inside nuclei versus outside nuclei was calculated for each field. A Student's t-test shows that each sample slide is significantly different to the untreated slide (P < 0.0001), which supports the conclusion that the signals arise from the target sequence rather than background. The average numbers of signals are similar for all the sample slides, varying from ~1.6 per nucleus for both IDLV and ICLV SV40 eGFP, to ~1.9 per nucleus for IDLV SV40 eGFP mMAR.

Discussion
High levels of retention of IDLV genomes achieved in CHO cells following a period of induced cell cycle arrest.

In the polyclonal populations which underwent a period of cell cycle arrest following transduction, the expression levels for each of the 3 vectors fluctuated somewhat in the course of ~100 days in culture. The proportion of eGFP-expressing cells remained most stable in SV40 promoter-driven vectors incorporating the mMAR element, where the GFP positive population did not change by more than 5 percentage points throughout 100 days under observation. In the cells transduced with IDLV SV40 eGFP without the mMAR element, expression declined in one population and increased in another, but the average remained the highest throughout the experiment (54% at ~100 days post-transduction). This maybe a
function of the promoter activity influenced by the intracellular location of the transgene, or simply genetic drift.

Cells transduced with IDLVCMV eGFP, on the other hand, exhibited a gradual decline in the proportion of GFP positive cells. This may result from the gradual shutting down of the promoter, as has been demonstrated in some tissue types including liver. In the polyclonal transduced cells populations that did not undergo cell cycle arrest, eGFP expression declined to background levels of ≤ 3 % in all populations by 14 days post-transduction. Nevertheless, over the course of 72 days post-transduction, the proportion of eGFP-expressing cells in one population spontaneously increased. This population was transduced with IDLVCMV eGFP, and the percentage of GFP positive cells slowly increased from 1% to 4% between 40 and 70 days post-transduction. This can be hypothesised to have arisen from a spontaneous integration event followed by genetic drift favouring the cells containing the integrated eGFP gene, as the spontaneous integration frequency of IDLV vectors has been estimated at ~1 %. Although further investigations into this cell population were outside the remit of this project, the integration status of the vector genome in this cell population could be investigated by obtaining a clonal population by dilution cloning and performing integration analyses by LAM-PCR and Southern blotting. The expression levels of the other two replicates which had undergone the same treatment remained low, o - 0.3% above background. No change was observed in the populations transduced with the other 2 IDLV vectors.

The effect of the S/MAR element on episome retention

The presence of the miniMAR-element in the vector cassette did not influence the level of retention of the vectors in any of the cell types, either with cell cycle arrest or in continuous culture. This suggests that even if association of vector DNA with the nuclear scaffolding took place, it happened independently of the S/MAR element and was not assisted by its presence.

The effect of a period of induced cell cycle arrest on episome retention

The transduction of CHO cells followed by a 5-day period of cell cycle arrest induced by methionine depletion resulted in a very high percentage (15-50%) of cells expressing the transgene in a mitotically stable fashion. One of the most consistent characteristics of CHO cells appears to be genome instability, as demonstrated by several studies into the genome structure of cell lines in laboratory use. Karyotypic rearrangements have been found both on the basis of banding (Deaven et al., 1973)
and, more recently, upon constructing a chromosomal map by using CHO genomic BAC library clones as FISH hybridisation probes (Omasa et al., 2009). However, confirming rearrangements at a sequence level, whether in CHO-Ki cells in general or our clonal cell lines in particular is not currently possible since the Chinese hamster genome remains to be sequenced in its entirety. This propensity for genomic rearrangements may also indicate a tendency towards random integration of any episomal material present. Interestingly, gene amplification in another CHO-derived cell line has been found to result in the amplified gene nesting between two inverted genomic regions (Omasa et al., 2009).

One method by which large genomic rearrangements could be examined is FISH. The pre-requisite for constructing probes to look for such events is knowledge of the genomic sequence and chromosome structure of the cells under scrutiny. With the recent publication of the Chinese Hamster draft genomic sequence and the annotating and arrangement into chromosomes underway, such an approach may soon become possible.

Analysis of the integration status of vector genomes in clonal IDLV-transduced CHO cell populations

There are two possible explanations for the unprecedentedly high retention of the transgenes in the transduced cells initially held in cell cycle arrest: either the viral episomes became associated with the nuclear matrix during cell cycle arrest and have become mitotically stable entities, or the quiescent period induced a higher than usual rate of integration. Since all the IDLVs used carried a specific incapacitating D64V mutation in the IN enzyme that usually facilitates lentiviral integration, it is unlikely that such integration events were iV-mediated (Gaur et al., 1998). The mitotic stability of gene expression must therefore either stem from stable episomes or integrants facilitated by cellular mechanisms. Although approximately 1% of cells transduced with IDLVs can be estimated to contain integrants (Yanez-Munoz et al., 2006), the rate of spontaneous integration events observed in the cells in induced cell cycle arrest would have had to increase 20 to 30-fold to account for the percentage of stably expressing cells. It was hypothesised that the induced cell cycle arrest either helped the episomes to become associated with the nuclear matrix and therefore mitotically stable, or otherwise significantly increased the activity of the cellular DNA repair mechanisms which promoted integration.
The previously reported levels of residual integration from IDLVs vary from 0.3% - 2% depending on the cell type, and in HeLa cells remained between 1.1% and 1.3% even when a much higher vector dose of MOI 25 was used (Wanisch et al., 2009). Estimates of integration frequency in IDLVs with the D64V mutation such as the ones used in this work set it approximately 3 logs lower than wild-type vector. In a comprehensive study using LAM-PCR-based methods to detect IDLV D64V integration frequencies in the rat eye cup 2.5 months post-injection, only one integration event was detected despite approximately 30-50% of vector genome having been scanned for vector-genome junctions (Yanez-Munoz et al., 2006). In this work, only vector-genome junctions containing LTR sequences will have been detected, and it is therefore possible that vector genome breakage followed by integration has occurred elsewhere and would not have been detected by the LAM-PCR-based methods. Integration-defective mutants may exhibit a tendency towards integration through non-viral pathways, and the genomes may integrate in oligomeric tandem repeats similar to those formed after plasmid transformation (Hagino-Yamagishi et al., 1987). However, based on the data from the previous more comprehensive study, this is unlikely. The same study also noted a high frequency of 2-LTR junctions, which is in correlation with the results observed here for IDLV CMV GFP.

LAM-PCR and deep sequencing suggest the majority of IDLV vector genomes are not integrated.

For deep sequencing, an average of 1200 amplicons was generated by the LAM-PCR protocol for each clone. A subsequent BLAST analysis of the amplicons confirmed that the vast majority of the sequences amplified corresponded to the vector. Of the amplified sequences not corresponding to the vector, only one was confirmed by BLAST analysis to be a likely integrant, in IDLV CMV eGFP clone 2.

The non-vector amplicons obtained for IDLV SV40 eGFP clone 10 and IDLV SV40 eGFP mMAR clone 8 did not retrieve any BLAST results from the mouse or the rat genome, making it likely they were the result of contamination from neighbouring sequencing lanes containing HIV-i samples in human cells. The possibility that at least some of them resulted from integrations in the CHO genome in a rare section that does not correlate well with the mouse and rat genomes cannot be completely excluded, however, it is highly unlikely considering the almost perfect alignment of the other integration sites with one or both the rat and mouse genomes.
In a similar way, the one BLAST-positive read for sample IDLV CMV eGFP clone 12 is likely contamination from the neighbouring lane, which contained the positive control, as it is unlikely for integration to have occurred at exactly the same site in two independent samples. However, the possibility of this cannot be entirely discounted, as the site in question could be a hotspot for rearrangement and therefore exhibit a tendency towards non-integrase-mediated integration.

One factor that may prevent integration sites from being detected by LAM-PCR is amplicon length bias. If the restriction site is further than ~400 bp away from the LTR in the host genome, the resulting fragment will be too long for downstream processing and deep sequencing. It is therefore possible that some integration sites were missed because of this. Another factor introducing inherent bias in the LAM-PCR experiment is the positioning of the primers in the LTR. Although this is the best method to detect LTR-mediated integrations, in our case it is possible that some of the integration events were not LIV-mediated and therefore the vector-genome junction is present elsewhere in the genome. These would not have been picked up by the protocol used. However, in a previous study several sets of primer were used to scan a large part of the vector genome for vector-host junctions following an IDLV transduction, and only one such event was discovered (Yanez-Munoz et al., 2006). It would be interesting to perform a more thorough scan of the clonal populations here; however this was not possible within the scope of the current project.

**Quantitative RT-PCR provides data on intracellular vector conformation**

To obtain information on the integration status of the vector genomes, quantitative real-time PCR results were analysed by investigating the LTR-to-backbone (BB) signal ratios in each sample. Ideally, in 1-LTR circles, the ratio should be 1:1, and in 2-LTR circles it should be 2:1:1 accompanied by a 2-LTR signal. A ratio of 2:1 without 2-LTR signal is indicative of integration (See Table 2). In IDLV SV40 GFP clone 4, the LTR-to-BB signal ratio is close to 2:1 and in IDLV SV40 GFP mMAR clone 8 the ratio is 1.5:1, both of which are potentially indicative of integration. In all IDLV CMV GFP clones the LTR-to-BB ratio is high, ranging from 2:1 to 5:1, but as the 2-LTR signal is also high this can be expected to arise from 2-LTR circles. The 1-LTR to BB ratio may be affected if the reverse transcription during transduction has been incomplete, as the BB signal is measured from the part which is the last one to be transcribed. It is also possible that the LTR to BB ratios are affected by the method of DNA purification, which may damage the circular episomes and the breakage may result in the BB signal being reduced.
Also, as with any very sensitive PCR-based methods, different primers may have slightly differing binding efficiencies, so obtained ratios may deviate somewhat from expected. Ratios may also be affected if the method of DNA purification causes the DNA to become fragmented, although it should not be the case here since the protocol used was specifically designed to keep the DNA molecules as intact as possible.

Southern Blotting produces results consistent with several clones containing episomal vectors.

In the Southern blotting experiment, 2 out of 3 clones containing IDLV SV40 eGFP and at least 1 clone containing IDLV SV40 eGFP mMAR produced bands consistent with the expected size for an episomal vector genome. The remaining 2 clones transduced with IDLV SV40 eGFP mMAR did not produce unambiguous results; DNA from cells transduced with IDLV SV40 GFP mMAR was digested with a different enzyme than for the other 2 vectors (Xhol instead of EcoRI) due to the presence of several EcoRI sites within the miniMAR-element. The resulting longer DNA fragments did not enter the gel at the correct level for IDLV SV40 eGFP mMAR clones 5 and 8. It is therefore possible that these 2 samples, digested with a different enzyme, could produce bands corresponding to the expected episome size.

All IDLV CMV eGFP -transduced samples show a band which is shorter than the expected L-LTR episome. It is possible that some recombination events have resulted in a shorter vector, and the presence of the same band for all 3 clones suggests the existence of a recombination hotspot within the vector genome. This may also explain the failure for the FISH probe to work for this vector, since if the CMV sequence is sufficiently rearranged the plasmid-based probe may no longer hybridise to the target sequence. A possible explanation is episomal recombination and rearrangement, especially since all 3 independent clones exhibit the same form. In the case of an integrant, it is unlikely for the integration to have occurred in the same place in all 3 clones. However, two rounds of dilution cloning were done for the IDLV CMV eGFP -transduced cells, after only 3 possibly polyclonal and partially positive populations were obtained after the first round. A mixture of these 5 populations was utilised in the second round of dilution cloning, this time obtaining 12 positive clones. Therefore the possibility that clones 10 and 12 originated from the same clone cannot be discounted.
Bands corresponding to the expected sizes of the 1-LTR and 2-LTR episomes can be taken as strong evidence for the presence of the vector DNA in episomal forms. To date, no evidence has been found of a tendency of lentiviral episomes to form concatemeric forms which would give the same blotting pattern once integrated. As the copy number of genomes is estimated to be well below 10 per cell, integrated head-to-tail concatenated genomes should produce 1-2 additional restriction bands flanking the integration site, and these should be visible on the Southern blot. On the other hand, episomes present in the nuclear environment for several hundred generations may well be assumed to recombine, forming circles not corresponding to the original 1-LTR and 2-LTR sizes. Again, no data is available on the frequency of such recombination events in long-term cultures, as usually IDLV episomes are transient in dividing cells.

Fluorescent In Situ hybridisation provides supporting data on the copy number and intracellular location of the episomes

FISH was in this case limited to supportive evidence due to the small size of the target sequence. Statistical analysis of the signals supports the theory that the signals observed arise from real vectors present in the cells, and may be used to infer approximate copy numbers per cell. Here, copy numbers were estimated for IDLV SV40 GFP clone 4 and IDLV SV40 GFP mMAR clone 11, and both had the same average of 1.6 transgenes per cell.

FISH signals were only observed in slides containing SV40 promoter-driven vectors, which were all analysed using the same probe. The slides made with cell lines transduced with IDLV CMV eGFP did not produce quantifiable signals. The probe used for these cells differed from the probe used for the SV40-containing cell lines, and hence it may be assumed that the reason for the weak signal was a suboptimal probe. Putative reasons for the FISH probe for the IDLV CMV GFP vector not working include problems with target binding; the CMV sequence may hybridise very weakly and is removed during stringent washes. Another possibility is that the probe may bind to genomic sequences resulting in high weak-level background, drowning out the true signals. While the technique could be optimised and the quality of signals and images potentially improved, due to time constraints this was not possible within the scope of this project.

Studies of other episomal systems using FISH have indicated that episomes result in single sharp signals of low intensity, whereas integration events give doublet signals
of higher fluorescence intensity (Rangasamy, 2010). This is in keeping with our observations, as although doublet signals cannot be observed on all metaphases transduced with integrating vectors, there is a tendency towards higher signal intensity in the integrating control (seen as larger dots for the vectors in Figure 12) than in slides transduced with non-integrating vectors.

Summary of episomal status analysis

The episomal status of the IDLV-introduced transgenes in CHO cells was analysed using several different techniques, each with their own limitations. Here, a summary of all results from the different experiments is used to analyse the integration status of each clone. The conclusions are also tabulated in Table 4.

Table 4 - Summary of the experimental evidence to evaluate the episomal status of IDLV genomes in CHO cell lines derived by dilution cloning from transduced cell populations subjected to a transient cell cycle arrest

<table>
<thead>
<tr>
<th>Vector transduction</th>
<th>IDLV-SG</th>
<th>IDLV-SGm</th>
<th>IDLV-CG</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHO Cell Clone Identifier #</td>
<td>4</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>LAM-PCR amplicon size (agarose gel)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>LAM-PCR amplicon Sequence</td>
<td>?</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>qRT-PCR for 1-LTR, 2-LTR &amp; backbone sequence</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Southern blotting analyses</td>
<td>✓</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>FISH analyses</td>
<td>✓</td>
<td>?</td>
<td>✓</td>
</tr>
<tr>
<td>Overall majority consensus</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

The table shows the results from 5 independent experimental techniques to evaluate the episomal status of IDLV genomes in clonal CHO cell lines derived by dilution cloning from transduced cell populations subjected to early transient cell cycle arrest. Green tick indicates that the experiment provides evidence to support the existence of replicating episome structures. The red cross indicates that the experiment does not provide evidence to support the existence of replicating episome structures. The "?" indicates unavailability of information where the experiment was either not performed or judged not to provide reliable data.
The linear amplification-mediated PCR reaction was designed to reveal any existing vector-genome junctions. The products of the reaction were first visualised on an agarose gel, as products arising from exclusively episomal vector are expected to correspond to defined sizes. Here, all clones except IDLV-SGm clones 5 and 11 exhibited only bands corresponding to the expected lengths for i-LTR and 2-LTR episomes (Figure 11). High throughput sequencing of the LAM-PCR products and subsequent BLAST analysis of the sequence reads revealed one likely integrant in IDLV-CGclone 2, corresponding to a locus on the murine chromosome 4 (Table 4). Although some of the other samples initially produced reads not corresponding to the vector genome, these were excluded after further BLAST analysis revealed no homology to either the rat or the mouse genome. Two samples, IDLV-SGclone 10 and IDLV-SGm clone 8 did not produce successful sequence reads and could not be analysed.

The quantitative real-time PCR experiment produced data on the amount of LTR signals, LTR to LTR junctions, and vector backbone present in the cells. The ratio of the amount of signal is expected to be different in cells with episomes and integrants, if the integration has been LTR-mediated. The signals were calculated as the amount of each type of product present per cell equivalent, and expressed as ratios of i-LTR and 2-LTR (LTR to LTR junction) signals for each backbone (BB) signal. IDLV-SG clones 7 and 10, and IDLV-SGm clone 11 produced LTR to BB signal ratios close to 1:1 (0.7 to 1, 1.1 to 1 and 0.96 to 1 respectively), which is expected to arise from i-LTR circles. All IDLV-CGclones produced high amounts of LTR signal as well as high amounts of 2-LTR signal, indicating the presence of 2-LTR circles. The ratios of i-LTR: 2-LTR: BB in the clones were 2.4 to 0.2 to 1 for clone 2, 5.0 to 1.1 to 1 for clone 10, and 5.1 to 1.8 to 1 for clone 12 (see Table 114 and Figure 11). The Southern blotting experiment indicated bands corresponding to the expected i-LTR and 2-LTR episomes for cells transduced with IDLV-SG, clones 4 and 7, and cells transduced with IDLV-SGm, clone 11. Of these, clones 7 and 11 contain a single band, indicating that the cell line is indeed clonal and contains the transgene in a single conformation.

In conclusion, at least two of the clonal transduced CHO cell lines appear to retain the transgene in an entirely episomal configuration, IDLV-SGclone 7 and IDLV-SGm clone 11. These two clones showed no evidence for integration in any of the experiments. Furthermore, several other clones exhibit mostly episomal evidence, including IDLV-SGclone 4, in which the only evidence against episomal vector
genome is the LTR to backbone ratio in the qRT-PCR. For IDLV-CG clones 10 and 12, the Southern blotting pattern does not correspond to the expected episome size, but no other experiments found any evidence for integration. This points to either a single integration event resulting from a breakage in the vector genome followed by a non-LTR mediated integration, or rearrangement of the episome.

Conclusions
The potential for generating stable episomes using IDLVs has been investigated by transducing IDLV vectors in vitro into CHO-Ki cells and inducing a period of reversible cell cycle arrest for 4-5 days following transduction, to assist in the establishment of the episomes in the nuclear environment. The cell cycle arrest was achieved by methionine restriction and serum depletion, and accompanied by a control group of transduced cells which were not induced to undergo cell cycle arrest. The integration status of the vector genomes in the stably transduced cells has also been investigated. Cells containing episomes were analysed using four independent methods: LAM-PCR, quantitative Real-Time PCR, Southern blotting and Fluorescent In Situ Hybridisation, to ascertain the intranuclear location of the vectors and to identify potential integration events.

One of the main questions that remains unanswered is how the IDLV-generated episomes are able to replicate; the minimal requirement for mitotic retention is a replication signal, and as the IDLV episomes without S/MAR elements performed as well as the ones containing such elements, the S/MAR cannot be the replication signal responsible for their retention. The long terminal repeats (LTRs) are the main feature of the IDLVs that distinguish them from other types of episomes investigated in this research, and it is possible they could be acting to initiate replication.

CHO cells were originally derived from a tissue sample from an inbred laboratory animal. They are noted for their chromosome instability. It is possible that the stress caused by methionine depletion induces further DNA instability, chromosome breaks and integration or any exogenous DNA present in the nucleus. However, very little evidence for integration events was found during extensive investigations, favouring instead the hypothesis that the quiescent period encouraged the episomes to become permanently associated either with chromosomes or the nuclear matrix.
References


Claims

1. A method of expressing a transgene in a host cell, the method comprising exposing a host cell, which comprises a genetic construct comprising a transgene, to conditions which substantially block mitosis in the host cell, such that the construct forms a mitotically stable episome resulting in expression of the transgene.

2. A method according to claim 1, wherein a period of cell cycle arrest is induced either during or after transduction of the host cell with the transgene-containing construct.

3. A method according to either claim 1 or claim 2, wherein the mitotic block is induced within the first 48 hours, 24 hours, 12 hours, 6 hours, 3 hours, 1 hour or 30 minutes upon transduction of the host cell with the construct.

4. A method according to any preceding claim, wherein the mitotic block is reversible.

5. A method according to any preceding claim, wherein the period required to induce the mitotic block in the host cell is between 1 hour and 10 days, or between 6 hours and 9 days, or between 12 hours and 8 days, or between 18 hours and 7 days.

6. A method according to any preceding claim, wherein the period required to induce the mitotic block in the host cell is between 1 day and 6 days, or between 2 days and 4 days.

7. A method according to any preceding claim, wherein the mitotic block is achieved by means of addition of a ribonucleotide biosynthesis inhibitor, ionising or UV irradiation, or by using a physical, metabolic, cell culture or pharmacological stimulus, which is selected from a group of stimuli consisting of metabolic restriction (such as serum starvation); hormone or cytokine treatment or withdrawal; pharmacological drug exposure or withdrawal; inhibition or activation of cell signalling pathways; and culture temperature or atmosphere stimuli.

8. A method according to any preceding claim, wherein the mitotic block is induced in the host cell by methionine depletion of the host cell's growth media.
9. A method according to claim 8, wherein the concentration of methionine in the growth medium is less than about 0.5µM, or less than 0.4µM or less than 0.3µM.

10. A method according to any preceding claim, wherein the mitotic block is induced in the host cell by serum depletion.

11. A method according to claim 10, wherein the concentration of serum in the growth medium is less than about 5%, 4%, 3% or 2% by volume.

12. A method according to any preceding claim, wherein the genetic construct is in the form of an expression cassette, which is suitable for expression of the transgene in the host cell.

13. A method according to any preceding claim, wherein the genetic construct is harboured within a vector, for expression in the host cell.

14. A method according to claim 13, wherein the vector comprising the transgene is a virus.

15. A method according to claim 14, wherein the vector is a member of the Retroviridae family, or of the Orthoretrovirinae Sub-family.

16. A method according to claim 15, wherein the vector is a member of the Lentivirus genus.

17. A method according to any one of claims 13-16, wherein the vector is a non-integrating or integration-deficient vector.

18. A method according to claim 17, wherein the vector is an integration-defective lentiviral vector (IDLV).

19. A method according to any preceding claim, wherein the construct comprises a SIN deletion.

20. A method according to any preceding claim, wherein the construct comprises one or more long terminal repeat (LTR).
21. A method according to any preceding claim, wherein the construct comprises woodchuck hepatitis virus post-transcriptional element (WPRE).

22. A method according to any preceding claim, wherein the vector comprises a so-called Scaffold/Matrix Attachment Regions (S/MAR) element, optionally wherein the S/MAR is a miniMAR.

23. A method according to any preceding claim, wherein the construct comprises a nucleic acid sequence substantially as set out in any one of SEQ ID No.1-9, or a functional fragment or variant thereof.

24. A method according to any preceding claim, wherein the transgene encodes dystrophin, a blood coagulation factor, insulin or a cytokine receptor sub-unit.

25. A method according to any preceding claim, wherein the host cell is a eukaryotic cell.

26. A method according to any preceding claim, wherein the host cell is an animal cell.

27. A method according to any preceding claim, wherein the host cell is from a vertebrate, mammal or domestic animal, and is preferably a human cell.

28. A method according to any preceding claim, wherein the host cell is a progenitor or stem cell, which may be either pluripotent or totipotent.

29. A method according to any preceding claim, wherein the host cell is a haematopoietic stem cell (HSC).

30. A method according to any preceding claim, wherein the host cell is an embryonic, foetal or adult stem cell.

31. A method according to any preceding claim, wherein the method comprises expressing and producing proteins on a commercial scale.

32. A host cell comprising a genetic construct comprising a transgene, wherein the host cell has been exposed to conditions which substantially block mitosis in the
host cell such that the construct forms a mitotically stable episome upon, and wherein the host cell is capable of expressing the transgene.

33. A host cell according to claim 32, wherein the host cell is prepared by the method according to any one of claims 1-31.

34. The host cell according to either claim 32 or claim 33, for use in therapy or diagnosis.

35. Use of the host cell according to either claim 32 or claim 33, for treating an immunodeficiency disorder, a blood clotting disorder, a central nervous system (CNS) disorder, spinal muscular atrophy, or Parkinson’s disease.

36. Use according to claim 35, wherein the immunodeficiency disorder is X-linked severe combined immunodeficiency (SCID-Xi).

37. Use according to claim 36, wherein the host cell is a haematopoietic stem cell.

38. Use according to claim 36, wherein the construct comprises a transgene, which encodes the gamma-chain of the IL2 receptor.
Figure 2C – D

Diagram of genetic elements in three segments labeled c) and d).

In c):
- dLTR PBS
- GAG (glycosylated alpha 1,3-galactosyltransferase)
- RRE (Rev response element)
- cPPT
- SV40 Promoter
- eGFP (enhanced green fluorescent protein)
- PPT

Diagram in d):
- dLTR dLTR PBS
- GAG
- RRE
- cPPT
- SV40 Promoter
- eGFP
- PPT

Graph G(x) with bp (base pairs) on the x-axis.

Substitute Sheet (Rule 26)
Figure 3A - B

A) No S/MAR

B) MiniMAR
Figure 3C

C) Beta-IFN S/MAR

---

SUBSTITUTE SHEET (RULE 26)
Figure 4

Cell cycle arrested vs. Controls
Figure 5A - B

A) GFP expression in transduced CHO cells with cell arrest

B) GFP expression in transduced CHO cells in continuously proliferative culture
Figure 5C - D

C) GFP expression in transduced CHO cells with cell arrest

D) GFP expression in transduced CHO cells in continuously proliferative culture
Figure 6.1 (A)

A) CHO IDLV-SG parent population

![Graph showing FITC-A distribution with a peak at 54%]

Count

FITC-A

Log scale
Figure 6.1 (B) - Clones
Figure 6.1 (B) - Clones

4

5

6
Figure 6.1 (B) - Clones

[Diagram depicting data with axes and labeled P4]

SUBSTITUTE SHEET (RULE 26)
**Figure 6.2(A)** – CHO IDLV -SGm parent population
Figure 6.2 (B) – Clones

1

2

3

SUBSTITUTE SHEET (RULE 26)
Figure 6.2 (B) – Clones

### Figure 4

![Graph 4](Image)

### Figure 5

![Graph 5](Image)

### Figure 6

![Graph 6](Image)
Figure 6.2 (B) – Clones

Graph 7

Graph 8

Graph 9

SUBSTITUTE SHEET (RULE 26)
Figure 6.2 (B) – Clones

Graph 10

Graph 11

Graph 12
Figure 6.3(A) – CHO IDLV – CG parent population
Figure 6.3(B) – Clones

1

2

3

SUBSTITUTE SHEET (RULE 26)
Figure 6.3(B) – Clones

**Figure 4**

**Figure 5**

**Figure 6**
Figure 6.3(B) – Clones

Count

7

P4

8

P4

9

P4

SUBSTITUTE SHEET (RULE 26)
Figure 6.3(B) – Clones

Graph 10

Graph 11

Graph 12
**Figure 11A - B**

A) 

5000 bp  
4000 bp  
3000 bp  
2000 bp  

B) 

<table>
<thead>
<tr>
<th>Lane</th>
<th>Sample</th>
<th>Expected 1-LTR band</th>
<th>Expected 2-LTR band</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ICLV clonal control</td>
<td>3049 bp</td>
<td>3283 bp</td>
</tr>
<tr>
<td>2</td>
<td>IDLV SV40 GFP 24h control</td>
<td>3049 bp</td>
<td>3283 bp</td>
</tr>
<tr>
<td>3</td>
<td>ICLV 24h control</td>
<td>3049 bp</td>
<td>3283 bp</td>
</tr>
<tr>
<td>4</td>
<td>IDLV SV40 GFP clone 4</td>
<td>3049 bp</td>
<td>3283 bp</td>
</tr>
<tr>
<td>5</td>
<td>IDLV SV40 GFP clone 7</td>
<td>3049 bp</td>
<td>3283 bp</td>
</tr>
<tr>
<td>6</td>
<td>IDLV SV40 GFP clone 10</td>
<td>3049 bp</td>
<td>3283 bp</td>
</tr>
<tr>
<td>7</td>
<td>IDLV SV40 GFP mMAR clone 5</td>
<td>3782 bp</td>
<td>4016 bp</td>
</tr>
<tr>
<td>8</td>
<td>IDLV SV40 GFP mMAR clone 8</td>
<td>3782 bp</td>
<td>4016 bp</td>
</tr>
<tr>
<td>9</td>
<td>IDLV SV40 GFP mMAR clone 11</td>
<td>3782 bp</td>
<td>4016 bp</td>
</tr>
<tr>
<td>10</td>
<td>IDLV CMV GFP clone 2</td>
<td>3771</td>
<td>4005</td>
</tr>
<tr>
<td>11</td>
<td>IDLV CMV GFP clone 10</td>
<td>3771</td>
<td>4005</td>
</tr>
<tr>
<td>12</td>
<td>IDLV CMV GFP clone 12</td>
<td>3771</td>
<td>4005</td>
</tr>
</tbody>
</table>
INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C12N

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

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

EPO-Internal, WPI Data, BIOSIS, Sequence Search, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>KLAUS WANISCH ET AL: &quot;Integrating on-deficiency Lentiviral Vectors: A Slow Coming of Age&quot;, MOLECULAR THERAPY, vol. 17, no. 8, 1 August 2009 (2009-08-01), pages 1316-1332, XP055021370, ISSN: 1525-0016, DOI: 10.1038/mt.2009.122</td>
<td>32</td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td>34-38</td>
</tr>
<tr>
<td>A</td>
<td></td>
<td>1-31,33</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Further documents are listed in the continuation of Box C.

See patent family annex.

Date of the actual completion of the international search: 6 June 2013

Date of mailing of the international search report: 18/06/2013

Name and mailing address of the ISA/Authorized officer:
European Patent Office, P.B. 5818 Patentlaan 2
NL-2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-2016
Rutz, Berthold

Form PCT/ISA/210 (second sheet) (April 2009)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>

Form PCT/IB/210 (continuation of second sheet) (April 2009)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>W0 2010/055413 AI (SAN RAFFAELE CENTRO FOND [IT]; FOND TELETHON [IT]; ANN0NI ANDREA [IT];) 20 May 2010 (2010-05-20), page 44, line 5 - line 26; claim 4</td>
<td>1-38</td>
</tr>
</tbody>
</table>
### INTERNATIONAL SEARCH REPORT

**Box No. I  Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   **a. (means)**
   - [X] on paper
   - [X] in electronic form

   **b. (time)**
   - [X] in the international application as filed
   - [X] together with the international application in electronic form
   - [ ] subsequently to this Authority for the purpose of search

2. [ ] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. **Additional comments:**
<table>
<thead>
<tr>
<th>Patent document cited in search report</th>
<th>Publication date</th>
<th>Patent family member(s)</th>
<th>Publication date</th>
</tr>
</thead>
<tbody>
<tr>
<td>US 2011218234 A1</td>
<td>08-09-2011</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WO 2010055413 A1</td>
<td>20-05-2010</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>