

## **Modified sequence of adenovirus E2a gene for producing vectors based on adeno-associated viruses (AAV) (variants)**

### **Field of the invention**

The present invention relates to the field of genetic engineering, genetics and biotechnology. More specifically, the present invention relates to variants of a modified sequence of the adenovirus E2a gene, to a helper plasmid comprising the variants of the modified sequence of the adenovirus E2a gene for producing vectors based on adeno-associated virus, and to the use of the variants of the modified sequence of the adenovirus E2a gene and of the helper plasmids comprising these sequences for producing vectors based on adeno-associated viruses.

### **Background of the invention**

To date, various systems for drug delivery have been widely used, whereas, in the field of gene therapy, the delivery of gene-therapy constructs using adeno-associated viruses (AAV) is in most demand. AAV is a small non-enveloped virus from the Parvoviridae family characterized by a linear single-stranded DNA genome. The AAV genome includes sequences, flanked by inverted terminal repeats (ITRs), of the *rep* regulatory gene whose products are involved in the process of viral replication, and of the *cap* gene encoding the structural proteins VP1, VP2, VP3 forming the AAV capsid. While adeno-associated virus cannot replicate on its own, in order to effectively engage in its lytic life cycle, it requires simultaneous co-infection of the cell with the so-called helper virus, in particular with adenovirus (Ad) or herpes simplex virus (HSV) (Manuel AFV Gonçalves, Adeno-associated virus: from defective virus to effective vector, Virol J. 2005 May 6;2:43).

Adenoviruses are members of the Adenoviridae family and are characterized by the presence of an icosahedral capsid devoid of a lipoprotein envelope, as well as by the presence of a double-stranded DNA genome. Unlike AAV, adenoviruses have a significantly more complex genome, where the sequences of the virus gene and numerous regulatory elements are comprised in a size-limited nucleic acid due to sequence overlap. In general, the genome of members of the genus Mastadenovirus includes inverted terminal repeat (ITR)-flanked sequences of early genes, such as E1-E4, which encode proteins involved in replication and transcription processes, and sequences of late genes, such as L1-L5, which predominantly encode structural proteins of the virus. The use of helper adenovirus directly in the production of vectors based on adeno-associated viruses has not found wide application for the clinical development of gene-therapy drugs, since it led to the presence of all kinds of impurities, in particular in the form of adenoviral particles. Later, to overcome the above disadvantages of the method, developed was a three-plasmid, adenovirus particle-free, system for producing vectors based on adeno-associated viruses, since it was previously found that for the production of AAV it is necessary and sufficient to have functional sequences of adenovirus E1 genes, including E1a and E1b, E2a, E4 and VA (virus-associated) RNA (Juan Jose Aponte-Ubillus et al, Molecular design for recombinant adeno-associated virus (rAAV) vector production. Appl Microbiol Biotechnol, 2018; 102(3): 1045–1054). Accordingly, E1a is considered as a transactivator,

in particular facilitating the expression of AAV *rep* and *cap* genes; whereas E1b interacts with the adenovirus E4 gene facilitating viral mRNA intracellular transport in the appropriate moment. The latter is also involved in DNA replication. E2a and VA RNA genes provide for the stability of viral mRNAs during the process of efficient translation (Xiao Xiao et al, Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus. J Virol, 1998 Mar; 72(3): 2224–2232).

To date, the most common, adenoviral particle-free system for producing vectors based on adeno-associated viruses is a set of plasmids for transfecting animal cells comprising the exogenous sequence of the adenovirus E1 gene. Most typically, the set of transfection plasmids consists of a plasmid including a cassette with the gene of interest flanked by ITR sequences, a plasmid comprising sequences of AAV *rep* and *cap* genes, and a helper plasmid that includes regions of the adenovirus genome comprising sequences of the adenovirus E2, E4 and VA RNA genes and regulatory elements required for production of AAV. Many modern commercial companies possess such three-plasmid systems for producing AAV-based vectors.

For the treatment of genetic diseases *in vivo*, the obvious advantage will be the use of drugs with a high concentration of delivery particles, which, in turn, will allow achieving a therapeutic effect. To produce highly concentrated formulations of gene-therapy drugs, it is necessary to possess production methods that provide for a high yield of the drug and/or its components, including delivery particles. The use of large plasmid vectors in experimental work may be a factor complicating the work of the experimenter, and also may directly affect the productivity of the process of producing vectors based on adeno-associated viral particles.

Thus, currently, there is a constant need to develop new systems for producing vectors based on adeno-associated viruses, for example, which comprise a set of plasmids that differ from the existing components of the system by the fact that they comprise smaller nucleic acid sequences encoding adenovirus proteins, which would allow, while maintaining complete functionality of the system components, to increase productivity in the production of AAV as compared to the use of conventional larger plasmids from commercially available systems.

#### **Disclosure of the invention**

One of the urgent goals of research in the field of development of effective gene therapy is to create novel methods and to improve the existing methods for producing gene-therapy drugs in order to increase product yield.

The authors of the invention have developed variants of a modified sequence of the adenovirus E2a gene and helper plasmids for producing vectors based on adeno-associated viruses, comprising modified sequences of the adenovirus E2a gene. The authors of the invention have found that the use of variants of the modified sequence of the adenovirus E2a gene surprisingly leads to increased level of productivity during generation of AAV.

#### *Definitions and general methods*

Unless defined otherwise herein, all technical and scientific terms used in connection with the present invention will have the same meaning as is commonly understood by those skilled in the art.

Furthermore, unless otherwise required by context, singular terms shall include plural terms, and the plural terms shall include the singular terms. Typically, the classification used and methods of cell culture, molecular biology, microbiology, genetics described herein are well known and widely used by those skilled in the art. Enzyme reactions and purification methods are performed according to the manufacturer's guidelines, as is common in the art, or as described herein.

The terms "naturally occurring", "native", or "wild-type" are used to describe an object that can be found in nature as distinct from being artificially produced. For example, a protein or nucleotide sequence present in an organism, including in a virus, which can be isolated from a source in nature and that has not been intentionally modified by a person in the laboratory, is naturally occurring.

As used in the present description and claims that follow, unless otherwise dictated by the context, the words "include" and "comprise", or variations thereof such as "includes", "including", "comprises", or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

### *Detailed description of the invention*

#### *Nucleic acid*

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in the sequence of SEQ ID NO:1 or in a sequence homologous to the sequence of SEQ ID NO:1 with a degree of identity of at least 63%.

The terms "nucleic acid", "polynucleotide", "oligonucleotide", "polynucleotide sequence" used interchangeably in the present description, mean a precise sequence of nucleotides, modified or not, determining a fragment or a region of a nucleic acid, containing unnatural nucleotides or not, and being either a double-strand DNA or RNA, a single-strand DNA or RNA, or transcription products of said DNAs.

As used in the present description, polynucleotides include, by way of non-limiting examples, all nucleic acid sequences which are obtained by any means available in the art, including, as non-limiting examples, recombinant means, i.e. the cloning of nucleic acid sequences from a recombinant library or a cell genome, using ordinary cloning technology and PCR and the like, and by synthetic means.

It should be mentioned that the present invention does not relate to nucleotide sequences in natural chromosomal environment thereof, i.e. in a natural state. Accordingly, isolated nucleic acids produced by

way of recombinant genetics, for example, using host cells, or produced by way of chemical synthesis should be included in the present invention.

Unless otherwise indicated, the term "nucleotide sequence" encompasses its complement. Thus, a nucleic acid having a particular sequence should be understood as one which encompasses the complementary strand thereof with the complementary sequence thereof.

The term "homologous sequence" refers to two or more nucleotide sequences having a degree of identity of at least 61 percent, in particular 61 percent, 62 percent, 63 percent, 64 percent, 65 percent, 66 percent, 67 percent, 68 percent, 69 percent, 70 percent, 71 percent, 72 percent, 73 percent, 74 percent, 75 percent, 76 percent, 77 percent, 78 percent, 79 percent, 80 percent, 81 percent, 82 percent, 83 percent, 84 percent, 85 percent, 86 percent, 87 percent, 88 percent, 89 percent, 90 percent, 91 percent, 92 percent, 93 percent, 94 percent, 95 percent, 96 percent, 97 percent, 98 percent, 99 percent, 100 percent. Thus, according to the present invention, homologous nucleotide sequences may have a common evolutionary origin or be produced by way of codon optimization in the initial sequence.

The term "identity" refers to the comparison of a nucleotide sequence of two nucleic acids. Identity is determined based on a reference sequence. The sequence of the adenovirus E2a gene is used as a reference sequence according to the present invention. Algorithms for sequence analysis are widely known in the art.

One of the properties of the genetic code is degeneracy, i.e. the ability of different codons (trinucleotides) to encode the same amino acid. Such codons that are translated to the same amino acid are called synonymous codons. In natural sequences, one of the synonymous codons is selected randomly in the course of evolution, but the frequencies of usage of synonymous codons are different: each amino acid has more and less preferred ones. Codon optimization is a widely used technique to amplify the production of protein molecules, which provides a rational mapping of one of suitable synonymous codons to each amino acid in a protein sequence. One of the common principles of codon optimization involves the usage of the most frequent codons, whereas other approaches were introduced later, such as harmonization (reproduction of distribution of codon usage frequencies), but they do not always increase productivity. In addition to codon frequencies, the sequence GC content (ratio of guanine and cytosine to the total length of the sequence) may affect the production efficiency, in particular, it was shown that high GC content is associated with increased mRNA levels in mammalian cells (Grzegorz Kudla ET AL., High Guanine and Cytosine Content Increases mRNA Levels in Mammalian Cells, June 2006, Volume 4, Issue 6, e180, pp. 933-942). It is further worth noting that stable secondary structure elements of mRNA, i.e. those having a low free folding energy, may reduce the efficiency.

The term "adenovirus" refers to a virus of the Adenoviridae family. Viruses of this family are viral particles that include a double-stranded DNA genome and an icosahedral capsid lacking a lipoprotein envelope. The Adenoviridae family can be divided into 6 genera and, in particular, comprises the genus of Mastadenoviruses infecting mammals. The criteria for demarcation in the family are the results of phylogenetic analysis, features of organization of the genome, as well as infection of certain animals by the virus (Mária Benkő et al, ICTV Virus Taxonomy Profile: Adenoviridae 2022, J Gen Virol. 2022 Mar;103(3):001721). In the genus Mastadenovirus, distinguished is a group of primate adenoviruses, which

combines seven subgroups or species of adenoviruses (HAdV-A-G). In view of the presence of different biological characteristics, serotypes of viruses have also been characterized in these subgroups of adenoviruses, such as serotypes 2, 5 in the species Mastadenovirus C, serotype 54 in the species Mastadenovirus D and others. Adenoviruses infect the respiratory tract, gastrointestinal tract, kidneys, and mainly use the CAR receptor (coxsackievirus and AdV receptor) to penetrate the cell (Urs F. Greber, Adenoviruses - Infection, pathogenesis and therapy, FEBS Lett. 2020 Jun;594(12):1818-1827). More detailed information about various members of the family Adenoviridae has been provided in the literature (Balázs Harrach et al, Adenoviruses across the animal kingdom: a walk in the zoo, FEBS Lett. 2019 Dec;593(24):3660-3673).

The organization of the adenovirus genome is characterized by significant variability between genera both in terms of the content of functional and regulatory elements and in terms of genome size. In general, the adenovirus genome includes sequences of early (E1-E4), intermediate (transcripts IX and IVa2) and late (L1-L5) genes flanked by inverted terminal repeats (ITRs), which comprise a conservative sequence and serve as the origin of replication of the virus. Depending on the type of adenovirus, virus-associated RNA (VA RNA) genes may also be represented in the genome (Thomas Lion, Adenovirus Infections in Immunocompetent and Immunocompromised Patients, Clin Microbiol Rev. 2014 Jul; 27(3): 441–462).

Adenovirus replication involves the pre-terminal protein (pTP), adenovirus polymerase (AdV Pol), and DNA-binding protein, encoded by the E2 gene, as well as some cellular transcription factors. At the initial stages of replication, these proteins bind to various ITRs of the virus, which determines the formation of the preinitiation complex and subsequently determines the formation of a new duplex genome (Rob C. Hoeben et al., Adenovirus DNA Replication, Cold Spring Harb Perspect Biol. 2013 Mar; 5(3): a013003). Adenovirus proteins encoded by the E1 gene are necessary for the activation of early gene promoters during the lytic cycle of the virus; furthermore, transcription is activated also by means of the physical interaction of E1 and E4 gene products. In turn, E4 encodes a complex of protein products facilitating transcription and involved in RNA splicing and transport (M Bondesson, Adenovirus E4 open reading frame 4 protein autoregulates E4 transcription by inhibiting E1A transactivation of the E4 promoter, J Virol. 1996 Jun;70(6):3844-51).

The terms "adenovirus" or "Ad" are used interchangeably herein.

Production of rAAV vectors by using the transfection of mammalian cells such as HEK293 with a three-plasmid system is the most efficient and widely used method. The first one in the system of such plasmids comprises an expression cassette comprising a gene of interest (GOI) operably linked to expression control sequences, wherein these elements are flanked by inverted terminal repeats (ITRs) of the virus. ITRs are essential components for viral replication and packaging of said expression cassette of interest into the virus capsid. The second plasmid is one comprising the nucleotide sequences of the *rep* and *cap* genes of the adeno-associated virus, which encode the replication enzymes of the virus and structural proteins of its capsid of a desired serotype. The third plasmid, also called a helper plasmid, comprises adenovirus genes facilitating replication and packaging of the recombinant AAV genome

(Parminder Singh Chahal et al., Production of adeno-associated virus (AAV) serotypes by transient transfection of HEK293 cell suspension cultures for gene delivery, J Virol Methods. 2014 Feb; 196: 163–173).

The resulting variants of the modified sequence of the adenovirus E2a gene with regulatory elements which comprise a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in the sequence SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63% led to an increased level of productivity during the generation of vectors based on AAV as compared to the sequence of SEQ ID NO: 1 or to a sequence homologous to SEQ ID NO: 1 with a degree of identity of at least 63%. Deletions of said regions do not negatively affect the expression of the E2a gene and the function of the respective protein.

SEQ ID NO 1 is the sequence of the wild-type gene with regulatory elements, which encodes the adenovirus E2a protein and corresponds to the genome sequence of adenovirus serotype 2 (NCBI Ref. No. AC\_000007.1) from 22393 bp to 27214 bp.

In some embodiments of the invention, the sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 1, or a codon-optimized variant of the sequence of SEQ ID NO: 1.

Those skilled in the art will readily appreciate that the genome sequences of adenoviruses of different serotypes can have significant similarity and be characterized by a high percentage of identity, especially, for example, the genome sequences of different adenovirus serotypes of the same genus. Accordingly, adenovirus serotypes 1, 5, 6, 7, 11, 54 and 89 relate to the genus Mastadenovirus, whereas the sequence of SEQ ID NO: 1 is homologous to the sequence of the E2a gene with regulatory elements of adenovirus serotype 1, which is SEQ ID NO: 32, with 98% identity; to the sequence of the E2a gene with regulatory elements of adenovirus serotype 5, which is SEQ ID NO: 33, with 97% identity; to the sequence of the E2a gene with regulatory elements of adenovirus serotype 6, which is SEQ ID NO: 34, with 99% identity; to the sequence of the E2a gene with regulatory elements of adenovirus serotype 7, which is SEQ ID NO: 35, with 64% identity; to the sequence of the E2a gene with regulatory elements of adenovirus serotype 11, which is SEQ ID NO: 36, with 64% identity; to the sequence of the E2a gene with regulatory elements of adenovirus serotype 54, which is SEQ ID NO: 19, with 63% identity; to the sequence of the E2a gene with regulatory elements of adenovirus serotype 89, which is SEQ ID NO: 37, with 98% identity. The sequence of human adenovirus serotype 2 in the case at hand serves as a reference sequence with

respect to nucleic acid positions for all other adenovirus genome sequences mentioned herein. Respective adenovirus nucleic acid positions can be identified by conventional sequence alignment using the publicly available BlastP software from the National Center for Biotechnology Information (NCBI).

Accordingly, the adenovirus E2a gene which comprises a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in a sequence homologous to the sequence of SEQ ID NO:1 with a degree of identity of at least 63% may be encoded by a wide range of different nucleotide sequences representing the sequence of the E2a gene with regulatory elements of an adenovirus serotype other than the serotype with SEQ ID NO:1. Such nucleotide sequence variants are within the scope of the present invention.

In some embodiments of the invention, a variant of the sequence of SEQ ID NO: 1, which is a naturally-occurring sequence of the E2a gene with regulatory elements of an adenovirus serotype other than the serotype with SEQ ID NO: 1 is the sequence of SEQ ID NO: 19 and corresponds to the sequence of the genome of adenovirus serotype 54 (NCBI Reference Sequence: NC\_012959.1) from 21166 bp to 25369 bp. The naturally-occurring variant of the sequence of the E2a gene of an adenovirus serotype, other than serotype with SEQ ID NO: 1, with the sequence of SEQ ID NO: 19 is provided for illustrative purposes only and should not be construed as limiting naturally-occurring variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype of SEQ ID NO: 1. Further, the designation of

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in the sequence of SEQ ID NO: 1 corresponds to the designation of

- i. deletions of sequence regions corresponding to 1534-2036 bp, 2115-2340 bp, 2545-3198 bp; or
- ii. deletion of a sequence region corresponding to 1534-3198 bp; or
- iii. deletion of a sequence region corresponding to 1534-2036 bp; or
- iv. deletion of a sequence region corresponding to 2115-2340 bp; or
- v. deletion of a sequence region corresponding to 2545-3198 bp; or

- vi. deletions of sequence regions corresponding to 1534-2036 bp, 2115-2340 bp; or
  - vii. deletions of sequence regions corresponding to 1534-2036 bp, 2545-3198 bp; or
  - viii. deletions of sequence regions corresponding to 2115-2340 bp, 2545-3198 bp
- in the sequence of SEQ ID NO: 19 (see Figure 2).

The adenovirus E2a gene which comprises a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in the codon-optimized variant of the sequence of SEQ ID NO:1 may also be encoded by a wide range of distinct nucleotide sequences due to redundancy of the genetic code. It is well within the skill of those trained in the art to create these alternative sequences encoding one and the same amino acid sequences. Such nucleotide sequence variants are within the scope of the present invention.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 1 is the sequence of SEQ ID NO: 10. The sequence with SEQ ID NO: 1 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 10 have 92% identity. The codon-optimized variant of the sequence of SEQ ID NO: 1 with the sequence of SEQ ID NO: 10 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 1.

Thus, the nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in a sequence homologous to the sequence of SEQ ID NO:1 with a degree of identity of at least 63% is also part of the present invention.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous

to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, includes the sequence of SEQ ID NO: 2 or a sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62%.

According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 2 with a degree of identity of at least 62% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 2, or a codon-optimized variant of the sequence of SEQ ID NO: 2.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene other than the serotype with SEQ ID NO: 2, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp in the sequence of SEQ ID NO: 1, is the sequence of SEQ ID NO: 20. The sequence of SEQ ID NO: 20 is a nucleic acid comprising a modified sequence of the E2a gene of adenovirus serotype 54, which comprises deletions of sequence regions corresponding to the regions of 1534-2036 bp, 2115-2340 bp, 2545-3198 bp in the sequence of SEQ ID NO: 19. The sequences of the E2a gene of the adenovirus serotype with SEQ ID NO: 2 and serotype with SEQ ID NO: 20 have 62% identity. A variant of the modified sequence of the E2a gene of an adenovirus serotype, other than serotype with SEQ ID NO: 2, with the sequence of SEQ ID NO: 20 is provided for illustrative purposes only and should not be construed as limiting the variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 2.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 2 is the sequence of SEQ ID NO: 11. The sequence with SEQ ID NO: 2 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 11 have 88% identity. The codon-optimized variant of the sequence of SEQ ID NO: 2 with the sequence of SEQ ID NO: 11 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 2.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, includes the sequence of SEQ ID NO: 3 or a sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61%.

According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 3 with a degree of identity of at least 61% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 3, or a codon-optimized variant of the sequence of SEQ ID NO: 3.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene other than the serotype with SEQ ID NO: 3, which comprises a deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1, is the sequence of SEQ ID NO: 21. The sequence of SEQ ID NO: 21 is a nucleic acid comprising a modified sequence of the E2a gene of adenovirus serotype 54, which comprises a deletion of a sequence region corresponding to the region of 1534-3198 bp in the sequence of SEQ ID NO: 19. The sequences of the E2a gene of the adenovirus serotype with SEQ ID NO: 3 and serotype with SEQ ID NO: 21 have 61% identity. A variant of the modified sequence of the E2a gene of an adenovirus serotype, other than serotype with SEQ ID NO: 3, with the

sequence of SEQ ID NO: 21 is provided for illustrative purposes only and should not be construed as limiting the variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 3.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 3 is the sequence of SEQ ID NO: 12. The sequence with SEQ ID NO: 3 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 12 have 87% identity. The codon-optimized variant of the sequence of SEQ ID NO: 3 with the sequence of SEQ ID NO: 12 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 3.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 1696-2321 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, includes the sequence of SEQ ID NO: 4 or a sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64%.

According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 4, or a codon-optimized variant of the sequence of SEQ ID NO: 4.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene other than the serotype with SEQ ID NO: 4, which comprises a deletion of a sequence region corresponding to the region of 1696-2321 bp in the sequence of SEQ ID NO: 1, is the sequence of SEQ ID NO: 22. The sequence of SEQ ID NO: 22 is a nucleic acid comprising a modified sequence of the E2a gene of adenovirus serotype 54, which comprises a deletion of a sequence region corresponding to the region of 1534-2036 bp in the sequence of SEQ ID NO: 19. The sequences of the E2a gene of the adenovirus serotype with SEQ ID NO: 4 and serotype with SEQ ID NO: 22 have 64% identity. A variant of the modified sequence of the E2a gene of an adenovirus serotype, other than serotype with SEQ ID NO: 4, with the sequence of SEQ ID NO: 22 is provided for illustrative purposes only and should not be construed as limiting the variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 4.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 4 is the sequence of SEQ ID NO: 13. The sequence with SEQ ID NO: 4 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 13 have 91% identity. The codon-optimized variant of the sequence of SEQ ID NO: 4 with the sequence of SEQ ID NO: 13 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 4.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 2400-2634 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1

with a degree of identity of at least 63%, includes the sequence of SEQ ID NO: 5 or a sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64%.

According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 5, or a codon-optimized variant of the sequence of SEQ ID NO: 5.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene other than the serotype with SEQ ID NO: 5, which comprises a deletion of a sequence region corresponding to the region of 2400-2634 bp in the sequence of SEQ ID NO: 1, is the sequence of SEQ ID NO: 23. The sequence of SEQ ID NO: 23 is a nucleic acid comprising a modified sequence of the E2a gene of adenovirus serotype 54, which comprises a deletion of a sequence region corresponding to the region of 2115-2340 bp in the sequence of SEQ ID NO: 19. The sequences of the E2a gene of the adenovirus serotype with SEQ ID NO: 5 and serotype with SEQ ID NO: 23 have 64% identity. A variant of the modified sequence of the E2a gene of an adenovirus serotype, other than serotype with SEQ ID NO: 5, with the sequence of SEQ ID NO: 23 is provided for illustrative purposes only and should not be construed as limiting the variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 5.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 5 is the sequence of SEQ ID NO: 14. The sequence with SEQ ID NO: 5 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 14 have 91.5% identity. The codon-optimized variant of the sequence of SEQ ID NO: 5 with the sequence of SEQ ID NO: 14 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 5.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, includes the sequence of SEQ ID NO: 6 or a sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61%.

According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 6, or a codon-optimized variant of the sequence of SEQ ID NO: 6.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene other than the serotype with SEQ ID NO: 6, which comprises a deletion of a sequence region corresponding to the region of 2839-3492 bp in the sequence of SEQ ID NO: 1, is the sequence of SEQ ID NO: 24. The sequence of SEQ ID NO: 24 is a nucleic acid comprising a modified sequence of the E2a gene of adenovirus serotype 54, which comprises a deletion of a sequence region corresponding to the region of 2545-3198 bp in the sequence of SEQ ID NO: 19. The sequences of the E2a gene of the adenovirus serotype with SEQ ID NO: 6 and serotype with SEQ ID NO: 24 have 61% identity. A variant of the modified sequence of the E2a gene of an adenovirus serotype, other than serotype with SEQ ID NO: 6, with the

sequence of SEQ ID NO: 24 is provided for illustrative purposes only and should not be construed as limiting the variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 6.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 6 is the sequence of SEQ ID NO: 15. The sequence with SEQ ID NO: 6 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 15 have 91% identity. The codon-optimized variant of the sequence of SEQ ID NO: 6 with the sequence of SEQ ID NO: 15 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 6.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp and 2400-2634 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, includes the sequence of SEQ ID NO: 7 or a sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63%.

According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 7, or a codon-optimized variant of the sequence of SEQ ID NO: 7.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene other than the serotype with SEQ ID NO: 7, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp and 2400-2634 bp in the sequence of SEQ ID NO: 1, is the sequence of SEQ ID NO: 25. The sequence of SEQ ID NO: 25 is a nucleic acid comprising a modified sequence of the E2a gene of adenovirus serotype 54, which comprises deletions of sequence regions corresponding to the regions of 1534-2036 bp, 2115-2340 bp in the sequence of SEQ ID NO: 19. The sequences of the E2a gene of the adenovirus serotype with SEQ ID NO: 7 and serotype with SEQ ID NO: 25 have 63% identity. A variant of the modified sequence of the E2a gene of an adenovirus serotype, other than serotype with SEQ ID NO: 7, with the sequence of SEQ ID NO: 25 is provided for illustrative purposes only and should not be construed as limiting the variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 7.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 7 is the sequence of SEQ ID NO: 16. The sequence with SEQ ID NO: 7 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 16 have 90% identity. The codon-optimized variant of the sequence of SEQ ID NO: 7 with the sequence of SEQ ID NO: 16 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 7.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp and 2839-3492 in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of

SEQ ID NO: 1 with a degree of identity of at least 63%, includes the sequence of SEQ ID NO: 8 or a sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62%.

According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 8, or a codon-optimized variant of the sequence of SEQ ID NO: 8.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene other than the serotype with SEQ ID NO: 8, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1, is the sequence of SEQ ID NO: 26. The sequence of SEQ ID NO: 26 is a nucleic acid comprising a modified sequence of the E2a gene of adenovirus serotype 54, which comprises deletions of sequence regions corresponding to the regions of 1534-2036 bp and 2545-3198 bp in the sequence of SEQ ID NO: 19. The sequences of the E2a gene of the adenovirus serotype with SEQ ID NO: 8 and serotype with SEQ ID NO: 26 have 62% identity. A variant of the modified sequence of the E2a gene of an adenovirus serotype, other than serotype with SEQ ID NO: 8, with the sequence of SEQ ID NO: 26 is provided for illustrative purposes only and should not be construed as limiting the variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 8.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 8 is the sequence of SEQ ID NO: 17. The sequence with SEQ ID NO: 8 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 17 have 89% identity. The codon-optimized variant of the sequence of SEQ ID NO: 8 with the sequence of SEQ ID NO: 17 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 8.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 2400-2634 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, includes the sequence of SEQ ID NO: 9 or a sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61%.

According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 9, or a codon-optimized variant of the sequence of SEQ ID NO: 9.

In some embodiments of the invention, a nucleic acid comprising a modified sequence of the adenovirus E2a gene other than the serotype with SEQ ID NO: 9, which comprises deletions of sequence regions corresponding to the regions of 2400-2634 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1, is the sequence of SEQ ID NO: 27. The sequence of SEQ ID NO: 27 is a nucleic acid comprising a modified sequence of the E2a gene of adenovirus serotype 54, which comprises deletions of sequence regions corresponding to the regions of 2115-2340 bp, 2545-3198 bp in the sequence of SEQ ID NO: 19. The sequences of the E2a gene of the adenovirus serotype with SEQ ID NO: 9 and serotype with SEQ ID NO: 27 have 61% identity. A variant of the modified sequence of the E2a gene of an adenovirus serotype,

other than serotype with SEQ ID NO: 9, with the sequence of SEQ ID NO: 27 is provided for illustrative purposes only and should not be construed as limiting the variants of the sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 9.

In some embodiments of the invention, the codon-optimized variant of the sequence of SEQ ID NO: 9 is the sequence of SEQ ID NO: 18. The sequence with SEQ ID NO: 9 and the codon-optimized variant thereof with the sequence of SEQ ID NO: 18 have 90% identity. The codon-optimized variant of the sequence of SEQ ID NO: 9 with the sequence of SEQ ID NO: 18 is provided for illustrative purposes only and should not be construed as limiting the codon-optimized variants of the sequence of SEQ ID NO: 9.

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 2 or a sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62%. According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 2, or a codon-optimized variant of the sequence of SEQ ID NO: 2.

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 3 or a sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61%. According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 3, or a codon-optimized variant of the sequence of SEQ ID NO: 3.

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 4 or a sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64%. According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 4, or a codon-optimized variant of the sequence of SEQ ID NO: 4.

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 5 or a sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64%. According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 5, or a codon-optimized variant of the sequence of SEQ ID NO: 5.

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 6 or a sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61%. According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61% is a respective sequence of the

E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 6, or a codon-optimized variant of the sequence of SEQ ID NO: 6.

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 7 or a sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63%. According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 7, or a codon-optimized variant of the sequence of SEQ ID NO: 7.

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 8 or a sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62%. According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 8, or a codon-optimized variant of the sequence of SEQ ID NO: 8.

In one aspect, the present invention relates to a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 9 or a sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61%. According to the present invention, the sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 9, or a codon-optimized variant of the sequence of SEQ ID NO: 9.

In some embodiments, the present invention is an isolated nucleic acid.

An "isolated" nucleic acid molecule is one which is identified and separated from at least one nucleic acid molecule-impurity. An isolated nucleic acid molecule is different from the form or set in which it is found under natural conditions. Thus, an isolated nucleic acid molecule is different from a nucleic acid molecule that exists in cells under natural conditions.

The sequences of SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, representing modified sequences of the adenovirus E2a gene with deletions of sequence regions corresponding to the regions of

- i. 1696-2321 bp, 2400-2634 bp, 2839-3492 bp (SEQ ID NO: 2),
- ii. 1696-3492 bp (SEQ ID NO: 3),
- iii. 1696-2321 bp (SEQ ID NO: 4),
- iv. 2400-2634 bp (SEQ ID NO: 5),
- v. 2839-3492 bp (SEQ ID NO: 6),
- vi. 1696-2321 bp, 2400-2634 bp (SEQ ID NO: 7),
- vii. 1696-2321 bp, 2839-3492 bp (SEQ ID NO: 8),
- viii. 2400-2634 bp, 2839-3492 bp (SEQ ID NO: 9).

in the sequence of SEQ ID NO: 1, respectively, are provided here and below for illustrative purposes only. As described earlier, those skilled in the art will appreciate that by sequences with a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in the sequence of SEQ ID NO: 1 may be meant any sequences close to SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9.

Codon-optimized sequences of SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, representing modified sequences of the adenovirus E2a gene with deletions of sequence regions corresponding to the regions of

- i. 1696-2321 bp, 2400-2634 bp, 2839-3492 bp (SEQ ID NO: 11),
- ii. 1696-3492 bp (SEQ ID NO: 12),
- iii. 1696-2321 bp (SEQ ID NO: 13),
- iv. 2400-2634 bp (SEQ ID NO: 14),
- v. 2839-3492 bp (SEQ ID NO: 15),
- vi. 1696-2321 bp, 2400-2634 bp (SEQ ID NO: 16),
- vii. 1696-2321 bp, 2839-3492 bp (SEQ ID NO: 17),
- viii. 2400-2634 bp, 2839-3492 bp (SEQ ID NO: 18).

in the codon-optimized sequence of SEQ ID NO: 10, respectively, are provided here and below for illustrative purposes only. As described earlier, those skilled in the art will appreciate that by sequences with a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in the sequence of SEQ ID NO: 10 may be meant any sequences close to SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18.

The sequences of SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, representing modified sequences of the adenovirus E2a serotype 54 gene with deletions of sequence regions corresponding to the regions of

- i. 1534-2036 bp, 2115-2340 bp, 2545-3198 bp (SEQ ID NO: 20),
- ii. 1534-3198 bp (SEQ ID NO: 21),
- iii. 1534-2036 bp (SEQ ID NO: 22),
- iv. 2115-2340 bp (SEQ ID NO: 23),
- v. 2545-3198 bp (SEQ ID NO: 24),
- vi. 1534-2036 bp, 2115-2340 bp (SEQ ID NO: 25),
- vii. 1534-2036 bp, 2545-3198 bp (SEQ ID NO: 26),
- viii. 2115-2340 bp, 2545-3198 bp (SEQ ID NO: 27).

in the sequence of SEQ ID NO: 19, respectively, are provided here and below for illustrative purposes only. As described earlier, those skilled in the art will appreciate that by sequences with a modification selected from the group:

- i. deletions of sequence regions corresponding to 1534-2036 bp, 2115-2340 bp, 2545-3198 bp; or
- ii. deletion of a sequence region corresponding to 1534-3198 bp; or
- iii. deletion of a sequence region corresponding to 1534-2036 bp; or
- iv. deletion of a sequence region corresponding to 2115-2340 bp; or
- v. deletion of a sequence region corresponding to 2545-3198 bp; or
- vi. deletions of sequence regions corresponding to 1534-2036 bp, 2115-2340 bp; or
- vii. deletions of sequence regions corresponding to 1534-2036 bp, 2545-3198 bp; or
- viii. deletions of sequence regions corresponding to 2115-2340 bp, 2545-3198 bp

in the sequence of SEQ ID NO: 19 may be meant any sequences close to SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27.

All nucleic acids, comprising a modified sequence of the adenovirus E2a gene, according to the invention encode the adenovirus E2a protein.

In some embodiments, nucleic acids, comprising a modified sequence of the adenovirus E2a gene, according to the invention are used to produce vectors based on adeno-associated viruses.

#### *Helper plasmid*

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or

- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in the sequence of SEQ ID NO:1 or in a sequence homologous to the sequence of SEQ ID NO:1 with a degree of identity of at least 63%.

The terms "helper plasmid" and "pHelper" are used interchangeably herein. In the present invention, "helper plasmid", "pHelper" refer to a plasmid having functions necessary for producing vectors based on AAV.

The term "adeno-associated virus" refers to a virus of the family Parvoviridae. The genomic organization of all known AAV serotypes is very similar. The genome of AAV is a linear, single-stranded DNA molecule that is less than about 5000 nucleotides (nt) in length. Inverted terminal repeats (ITRs) flank the unique coding nucleotide sequences of replication of non-structural proteins (Rep) and structural proteins (Cap). The Cap gene encodes the VP proteins (VP1, VP2, and VP3) which form the capsid. The terminal 145 nucleotides are self-complementary and are organized such that an energetically stable intramolecular duplex forming a T-shaped hairpin may be formed. Such hairpin structures function as an origin for virus DNA replication, serving as primers for the cellular DNA polymerase complex. Following wild-type AAV (wtAAV) infection in mammalian cells, the Rep genes (e.g. Rep78 and Rep52) are expressed using the P5 promoter and the P19 promoter, respectively, and the both Rep proteins have a certain function in the replication of the viral genome. A splicing event in the Rep open reading frame (Rep ORF) results in the expression of actually four Rep proteins (e.g. Rep78, Rep68, Rep52, and Rep40). However, it has been shown that the unspliced mRNA encoding Rep78 and Rep52 proteins is sufficient for AAV vector production in mammalian cells (see US20160032254).

The terms "adeno-associated virus", "AAV" are used interchangeably herein.

The terms "vector based on adeno-associated virus", "vector based on AAV", "rAAV" are used interchangeably herein.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and comprises the sequence of SEQ ID NO: 2 or a sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62%, respectively. As described above, the sequence homologous to the sequence of SEQ ID NO: 2 with a degree of identity of at least 62% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 2, or a codon-optimized variant of the sequence of SEQ ID NO: 2.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree

of identity of at least 63%, and comprises the sequence of SEQ ID NO: 3 or a sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61%, respectively. As described above, the sequence homologous to the sequence of SEQ ID NO: 3 with a degree of identity of at least 61% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 3, or a codon-optimized variant of the sequence of SEQ ID NO: 3.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 1696-2321 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and comprises the sequence of SEQ ID NO: 4 or a sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64%, respectively. As described above, the sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 4, or a codon-optimized variant of the sequence of SEQ ID NO: 4.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 2400-2634 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and comprises the sequence of SEQ ID NO: 5 or a sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64%, respectively. As described above, the sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 5, or a codon-optimized variant of the sequence of SEQ ID NO: 5.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and comprises the sequence of SEQ ID NO: 6 or a sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61%, respectively. As described above, the sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 6, or a codon-optimized variant of the sequence of SEQ ID NO: 6.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp and 2400-2634 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and comprises the sequence of SEQ ID NO: 7 or a sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63%, respectively. As described

above, the sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 7, or a codon-optimized variant of the sequence of SEQ ID NO: 7.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and comprises the sequence of SEQ ID NO: 8 or a sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62%, respectively. As described above, the sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 8, or a codon-optimized variant of the sequence of SEQ ID NO: 8.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 2400-2634 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and comprises the sequence of SEQ ID NO: 9 or a sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61%, respectively. As described above, the sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 9, or a codon-optimized variant of the sequence of SEQ ID NO: 9.

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 2 or a sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62%.

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 3 or a sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61%.

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 4 or a sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64%.

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 5 or a sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64%.

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 6 or a sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61%.

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 7 or a sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63%.

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 8 or a sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62%.

In one aspect, the present invention relates to a helper plasmid for producing vectors based on adeno-associated viruses comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises the sequence of SEQ ID NO: 9 or a sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61%.

In some embodiments of the invention, the helper plasmid for producing vectors based on adeno-associated viruses comprises a nucleic acid comprising a modified sequence of the adenovirus E2a gene (any one of the above variants), nucleic acid comprising a sequence of the adenovirus VA RNA gene, and nucleic acid comprising a sequence of the adenovirus E4 gene.

In some embodiments of the invention, the sequence of the adenovirus VA RNA gene comprises SEQ ID NO: 28.

In some embodiments of the invention, the sequence of the adenovirus E4 gene is a full-length naturally-occurring sequence of the E4 gene, sequence of E4 orf6/7, or sequence of E4 orf6.

In some embodiments of the invention, the full-length naturally-occurring sequence of the adenovirus E4 gene comprises SEQ ID NO: 29.

In some embodiments of the invention, the sequence of adenovirus E4 orf6/7 is a sequence encoding SEQ ID NO: 30.

In some embodiments of the invention, the sequence of adenovirus E4 orf6 is a sequence encoding SEQ ID NO: 31.

#### *Use*

In one aspect, the present invention relates to the use of a nucleic acid comprising any one of the variants of the modified sequence of the adenovirus E2a gene described herein, or of a helper plasmid comprising a nucleic acid comprising any one of the variants of the modified sequence of the adenovirus E2a gene described herein for producing vectors based on adeno-associated viruses.

In one aspect, the present invention relates to the use of a nucleic acid comprising any one of the variants of the modified sequence of the adenovirus E2a gene described herein for producing vectors based on adeno-associated viruses.

In one aspect, the present invention relates to the use of a helper plasmid comprising a nucleic acid comprising any one of the variants of the modified sequence of the adenovirus E2a gene described herein for producing vectors based on adeno-associated viruses.

In some variants of the use, the helper plasmid comprising a nucleic acid comprising any one of the variants of the modified sequence of the adenovirus E2a gene described herein is used to produce vectors based on adeno-associated viruses by means of transfecting a host cell with this helper plasmid, and also with plasmids comprising sequences of the gene of interest and genes encoding AAV proteins.

The terms "production" and "obtaining" are used interchangeably herein.

In some embodiments of the invention, variants of the modified sequence of the adenovirus E2a gene according to the invention may be integrated into the genome of a host cell to produce a packaging cell line or a producer cell line for viral vectors. Such approaches to producing viral vectors are well known to those skilled in the art and have been described in the literature (Imre Kovesdi et al., Adenoviral producer cells, *Viruses*. 2010 Aug;2(8):1681-1703).

In one aspect, the present invention relates to the use of variants of the modified sequence of the adenovirus E2a gene according to the invention to produce recombinant viral vectors based on adenoviruses by means of integrating a portion of the adenovirus genome, comprising any one of the variants of the modified sequence of the adenovirus E2a gene described herein into the genome of a host cell, and also by integration, into the genome of a replication-defective recombinant adenovirus, of the gene of interest.

In some aspects, the present invention relates to the use of variants of the modified sequence of the adenovirus E2a gene according to the invention to produce recombinant viral vectors based on adenoviruses by means of integrating, into the genome of the host cell, a portion of the adenovirus genome, which does not comprise the sequence of the E2a gene, and also by integration, into the genome of the replication-defective recombinant adenovirus, of the gene of interest and a nucleic acid comprising any one of the variants the modified sequence of the adenovirus E2a gene according to the invention. Such approaches are well known to those skilled in the art and have been described in the literature (M I Gorziglia et al., Generation of an adenovirus vector lacking E1, e2a, E3, and all of E4 except open reading frame 3, *J Virol*. 1999 Jul;73(7):6048-55; H Zhou et al., A new vector system with inducible E2a cell line for production of higher titer and safer adenoviral vectors, *Virology*. 2000 Sep 30;275(2):348-57).

In one aspect, the present invention relates to the use of variants of the modified sequence of the adenovirus E2a gene according to the invention for producing recombinant viral vectors based on adeno-associated viruses by means of integrating, into the genome of the host cell, a portion of the adenovirus genome comprising a nucleic acid comprising any one of the variants of the modified sequence of the adenovirus E2a gene according to the invention, and/or a portion of the adeno-associated virus genome and/or of an expression cassette with the gene of interest and ITR. Such approaches are well known to those skilled in the art and have been described in the literature (Satoki Nakamura et al., Development of packaging cell lines for generation of adeno-associated virus vectors by lentiviral gene transfer of trans-complementary components, 2004 Oct;73(4): 285-94, see Abstract; Nagarathinam Selvaraj et al., Detailed

**WO 2025/038011**

**PCT/RU2024/050178**

Protocol for the Novel and Scalable Viral Vector Upstream Process for AAV Gene Therapy Manufacturing,  
Hum Gene Ther. August 2021; 32(15-16): 850–861).

**Brief description of drawings**

**Figure 1** is a graph which shows the productivity of generation of rAAV serotypes 2, 5, 6 and 9 in HEK293 cells following transfection of all necessary sequences for rAAV generation, including sequences encoding the E2a gene and regulatory elements corresponding to the following nucleic acids: to a nucleic acid comprising the natural sequence of the adenovirus E2a gene (SEQ ID NO: 1), a nucleic acid comprising deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp in the sequence of SEQ ID NO: 1 (SEQ ID NO: 2), a nucleic acid comprising a deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1 (SEQ ID NO: 3).

Productivity of rAAV generation following three-plasmid transfection:

1 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA and E2a genes, where the naturally-occurring sequence of the adenovirus E2a gene comprises SEQ ID NO: 1, with regulatory elements; pRepCap plasmid encoding the Rep and Cap genes of serotype 2; plasmid encoding a transgene based on an antibody to VEGF/C5,

2 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA genes and a modified sequence of the E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp. n. in the sequence of SEQ ID NO: 1 and includes the sequence of SEQ ID NO: 2, with regulatory elements; plasmid pRepCap encoding the Rep and Cap genes of serotype 2; plasmid encoding a transgene based on an antibody to VEGF/C5,

3 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA genes and a modified sequence of the E2a gene, which comprises a deletion of a sequence region corresponding to the region of 1696-3492 bp. n. in the sequence of SEQ ID NO: 1 and includes the sequence of SEQ ID NO: 3, with regulatory elements; plasmid pRepCap encoding the Rep and Cap genes of serotype 2; plasmid encoding a transgene based on an antibody to VEGF/C5,

4 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA and E2a genes, where the naturally-occurring sequence of the adenovirus E2a gene comprises SEQ ID NO: 1, with regulatory elements; pRepCap plasmid encoding the serotype 2 Rep and serotype 5 Cap genes; plasmid encoding a transgene based on coagulation factor IX protein,

5 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA genes and a modified sequence of the E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp. n. in the sequence of SEQ ID NO: 1 and includes the sequence of SEQ ID NO: 2, with regulatory elements; plasmid pRepCap encoding the serotype 2 Rep and serotype 5 Cap genes; plasmid encoding a transgene based on coagulation factor IX protein,

6 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA genes and a modified sequence of the E2a gene, which comprises a deletion of a sequence region corresponding to the regions of 1696-3492 bp. n. in the sequence of SEQ ID NO: 1 and includes the sequence of SEQ ID NO: 3, with regulatory elements; plasmid pRepCap encoding the serotype 2 Rep and serotype 5 Cap genes; plasmid encoding a transgene based on coagulation factor IX protein,

7 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA and E2a genes, where the naturally-occurring sequence of the adenovirus E2a gene comprises SEQ ID NO: 1, with regulatory elements; pRepCap plasmid encoding the serotype 2 Rep and serotype 6 Cap genes; plasmid encoding a transgene based on coagulation factor VIII protein,

8 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA genes and a modified sequence of the E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp. n. in the sequence of SEQ ID NO: 1 and includes the sequence of SEQ ID NO: 2, with regulatory elements; plasmid pRepCap encoding the serotype 2 Rep and serotype 6 Cap genes; plasmid encoding a transgene based on coagulation factor VIII protein,

9 – when using the pHelper plasmid comprising a naturally-occurring sequence of the E4, VA RNA genes and a modified sequence of the E2a gene, which comprises a deletion of a sequence region corresponding to the regions of 1696-3492 bp. n. in the sequence of SEQ ID NO: 1 and includes the sequence of SEQ ID NO: 3, with regulatory elements; plasmid pRepCap encoding the serotype 2 Rep and serotype 6 Cap genes; plasmid encoding a transgene based on coagulation factor VIII protein,

10 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA and E2a genes, where the naturally-occurring sequence of the adenovirus E2a gene comprises SEQ ID NO: 1, with regulatory elements; pRepCap plasmid encoding the serotype 2 Rep and serotype 9 Cap genes; plasmid encoding a transgene based on survival motor neurone SMN1 protein,

11 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA genes and a modified sequence of the E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp. n. in the sequence of SEQ ID NO: 1 and includes the sequence of SEQ ID NO: 2, with regulatory elements; plasmid pRepCap encoding the serotype 2 Rep and serotype 9 Cap genes; plasmid encoding a transgene based on survival motor neurone SMN1 protein,

12 – when using the pHelper plasmid comprising naturally-occurring sequences of the E4, VA RNA genes and a modified sequence of the E2a gene, which comprises a deletion of a sequence region corresponding to the regions of 1696-3492 bp. n. in the sequence of SEQ ID NO: 1 and includes the sequence of SEQ ID NO: 3, with regulatory elements; plasmid pRepCap encoding the serotype 2 Rep and serotype 9 Cap genes; plasmid encoding a transgene based on survival motor neurone SMN1 protein,

**Figure 2** is a schematic view of a region of the genome of adenovirus serotype 2 comprising the sequence of the E2a gene with regulatory elements (A; corresponds to NCBI Ref. No. AC\_000007.1 from 22393 bp to 27214 bp) and a region of the genome of adenovirus serotype 54 comprising the sequence of the E2a gene with regulatory elements (B; corresponds to NCBI Reference Sequence: NC\_012959.1 from 21166 bp to 25369 bp) indicating deletions of the sequence regions described herein and the correspondence thereof to one another. The sequence deletion regions are numbered relative to SEQ ID NO: 1 (for A) and SEQ ID NO: 19 (for B).

#### **Examples (implementation of the invention)**

The following examples are provided for better understanding of the invention. These examples are for purposes of illustration only and are not to be construed as limiting the scope of the invention in any manner.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended embodiments.

#### **Materials and general methods**

##### **Recombinant DNA techniques**

Standard methods were used to manipulate DNA as described in Sambrook, J. et al, Molecular cloning: A laboratory manual; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2012. The molecular biological reagents were used according to the manufacturer protocols. Briefly, plasmid DNA was produced for further manipulation in E. coli cells grown under selective antibiotic pressure so that the plasmids were not lost in the cell population. We isolated the plasmid DNA from cells using commercial kits, measured the concentration, and used it for cloning by restriction endonuclease treatment or PCR amplification. The DNA fragments were ligated to each other using ligases and transformed into bacterial cells for the selection of clones and further production. All resulting genetic constructs were confirmed by restriction patterns and complete Sanger sequencing.

##### **DNA sequence determination**

DNA sequences were determined by Sanger sequencing. DNA and protein sequences were analyzed and sequence data was processed in SnapGene Viewer 4.2 or higher for sequence creation, mapping, analysis, annotation and illustration.

##### **Culturing cell cultures**

The experiments used HEK293 (Human Embryonic Kidney clone 293) cell lines. The suspended HEK293 cells used to produce AAV were cultured under standard conditions at 37°C and 5% CO<sub>2</sub> on a complete culture medium without FBS and antibiotic. Cell viability was assessed using Trypan Blue stain and disposable cell counting chambers using an automatic Countess II counter.

##### **Generation of viral particles of AAV recombinant vectors**

To produce recombinant AAV viral particles comprising a transgene based on various proteins, we used HEK293 producer cells which were transfected with 3 plasmids:

- A plasmid comprising an AAV expression cassette for expression of the transgene;
- A plasmid comprising serotype 2, 5, 6 or 9 Cap genes and AAV2 serotype Rep genes. Each gene, using alternative reading frames, encodes several protein products;
- A plasmid comprising adenovirus Ad2 genes necessary for the assembly and packaging of AAV capsids: E2a, E4 and VA RNA.

The titer of the viral particles was determined by quantitative PCR with primers and a sample that were specific for the region of the recombinant viral genome and expressed as the copy number of viral genomes per 1 ml.

**Statistical data analysis**

The results indicate an average value  $\pm$  standard deviation (SD), one-way analysis of variance (ANOVA) followed by Dunnett's multiple pairwise comparisons was employed to compare the experiment results, and they were determined to be statistically significant.

**Example 1. Evaluation of productivity of generation of recombinant viral vectors based on AAV while using modified adenovirus E2a gene sequence and helper plasmid based thereon**

To the naturally-occurring sequence of the adenovirus E2a gene with regulatory elements comprising SEQ ID NO:1 or a sequence homologous to the sequence of SEQ ID NO:1 with a degree of identity of at least 63% introduced were deletions

- i. 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. 1696-3492 bp

Produced were variants of modified sequences of the adenovirus E2a gene with regulatory elements, which were further tested for the level of productivity while generating vectors based on adeno-associated virus serotypes rAAV2, rAAV5, rAAV6 and rAAV9.

To show increased level of productivity while generating rAAV, produced were plasmids encoding target genetic constructs, in particular comprising

- a. an E2a gene sequence with regulatory elements which includes SEQ ID NO:1 or a sequence homologous to the sequence of SEQ ID NO:1 with a degree of identity of at least 63%,
- b. a modified sequence of the E2a gene with regulatory elements, which comprises deletions of sequence regions corresponding to regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 2 or a sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62%, and
- c. a sequence of the E2a gene with regulatory elements, which comprises a deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 3 or a sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61%.

Also produced were the remaining plasmids necessary for generating viral vectors

- i. based on adeno-associated virus serotype 2 carrying a nucleotide sequence encoding an antibody to VEGF/C5,
- ii. based on adeno-associated virus serotype 5 carrying a nucleotide sequence encoding the coagulation factor IX(FIX) protein,
- iii. based on adeno-associated virus serotype 6 carrying a nucleotide sequence encoding the coagulation factor VIII(FVIII) protein, and
- iv. based on adeno-associated virus serotype 9 carrying a nucleotide sequence encoding the survival motor neurone (SMN1) protein.

It has been shown that the developed modified sequences encoding the E2a gene and its regulatory elements led to increased productivity of generation of rAAV as compared to an analogous initial sequence encoding the E2a gene and regulatory elements with SEQ ID NO:1 (Figure 1):

- while generating rAAV serotype 2 with a transgene based on an antibody to VEGF/C5, by 2.41 times in the case of using the sequence of the adenovirus E2a gene with deletions of sequence regions corresponding to regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp in the sequence of SEQ ID NO: 1, which includes SEQ ID NO: 2, and by 3.02 times in the case of using the sequence of the adenovirus E2a gene with deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1, which includes SEQ ID NO: 3;

- while generating rAAV serotype 5 with a transgene based on coagulation factor IX protein, by 1.45 times in the case of using the sequence of the adenovirus E2a gene with deletions of sequence regions corresponding to regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp in the sequence of SEQ ID NO: 1, which includes SEQ ID NO: 2, and by 1.69 times in the case of using the sequence of the adenovirus E2a gene with deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1, which includes SEQ ID NO: 3;

- while generating rAAV serotype 6 with a transgene based on coagulation factor VIII protein, by 1.81 times in the case of using the sequence of the adenovirus E2a gene with deletions of sequence regions corresponding to regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp in the sequence of SEQ ID NO: 1, which includes SEQ ID NO: 2, and by 1.83 times in the case of using the sequence of the adenovirus E2a gene with deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1, which includes SEQ ID NO: 3;

- while generating rAAV serotype 9 with a transgene based on survival motor neurone (SMN1) protein, by 1.8 times in the case of using the sequence of the adenovirus E2a gene with deletions of sequence regions corresponding to regions of 1696-2321 bp, 2400-2634 bp, 2839-3492 bp in the sequence of SEQ ID NO: 1, which includes SEQ ID NO: 2, and by 1.97 times in the case of using the sequence of the adenovirus E2a gene with deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1, which includes SEQ ID NO: 3.

Similar data were obtained after comparing the codon-optimized variant of the adenovirus E2a gene with SEQ ID NO: 10 to its modified variants, which comprise deletions of i. 1696-2321 bp, 2400-2634 bp, 2839-3492 bp (SEQ ID NO: 11) or ii. 1696-3492 bp (SEQ ID NO: 12), and after comparing the sequence of the adenovirus E2a gene of a serotype other than SEQ ID NO: 1 with SEQ ID NO: 19 with its modified variants, which comprise deletions of i. 1534-2036 bp, 2115-2340 bp, 2545-3198 bp (SEQ ID NO: 20) or ii. 1534-3198 bp (SEQ ID NO: 21).

Thus, the produced nucleic acids which encode the E2a gene with its regulatory elements and comprise deletions of sequence regions corresponding to the regions of

- i. 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. 1696-3492 bp

in the sequence of SEQ ID NO:1 or in the sequence homologous to the sequence of SEQ ID NO:1 with a degree of identity of at least 63%, provided is increased productivity of recombinant viral vectors based on AAV of various serotypes as compared with the natural sequence of the adenovirus E2a gene comprising SEQ ID NO:1 or with a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%.

**CLAIMS**

1. A nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a modification selected from the group:

- i. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp, 2839-3492 bp; or
- ii. deletion of a sequence region corresponding to 1696-3492 bp; or
- iii. deletion of a sequence region corresponding to 1696-2321 bp; or
- iv. deletion of a sequence region corresponding to 2400-2634 bp; or
- v. deletion of a sequence region corresponding to 2839-3492 bp; or
- vi. deletions of sequence regions corresponding to 1696-2321 bp, 2400-2634 bp; or
- vii. deletions of sequence regions corresponding to 1696-2321 bp, 2839-3492 bp; or
- viii. deletions of sequence regions corresponding to 2400-2634 bp, 2839-3492 bp

in the sequence of SEQ ID NO:1 or in a sequence homologous to the sequence of SEQ ID NO:1 with a degree of identity of at least 63%.

2. The nucleic acid according to claim 1, wherein the sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 1, or a codon-optimized variant of the sequence of SEQ ID NO: 1.

3. The nucleic acid according to any one of claims 1-2, comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp, 2400-2634 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 2 or a sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62%, respectively.

4. The nucleic acid according to claim 3, wherein the sequence homologous to the sequence of SEQ ID NO: 2 with a degree of identity of at least 62% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 2, or a codon-optimized variant of the sequence of SEQ ID NO: 2.

5. The nucleic acid according to any one of claims 1-2, comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 1696-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 3 or a sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61%, respectively.

6. The nucleic acid according to claim 5, wherein the sequence homologous to the sequence of SEQ ID NO: 3 with a degree of identity of at least 61% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 3, or a codon-optimized variant of the sequence of SEQ ID NO: 3.

7. The nucleic acid according to any one of claims 1-2, comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 1696-

2321 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 4 or a sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64%, respectively.

8. The nucleic acid according to claim 7, wherein the sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 4, or a codon-optimized variant of the sequence of SEQ ID NO: 4.

9. The nucleic acid according to any one of claims 1-2, comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 2400-2634 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 5 or a sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64%, respectively.

10. The nucleic acid according to claim 9, wherein the sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 5, or a codon-optimized variant of the sequence of SEQ ID NO: 5.

11. The nucleic acid according to any one of claims 1-2, comprising a modified sequence of the adenovirus E2a gene, which comprises a deletion of a sequence region corresponding to the region of 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 6 or a sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61%, respectively.

12. The nucleic acid according to claim 11, wherein the sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 6, or a codon-optimized variant of the sequence of SEQ ID NO: 6.

13. The nucleic acid according to any one of claims 1-2, comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp and 2400-2634 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 7 or a sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63%, respectively.

14. The nucleic acid according to claim 13, wherein the sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 7, or a codon-optimized variant of the sequence of SEQ ID NO: 7.

15. The nucleic acid according to any one of claims 1-2, comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 1696-2321 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 8 or a sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62%, respectively.

16. The nucleic acid according to claim 15, wherein the sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62% is a sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 8, or a codon-optimized variant of the sequence of SEQ ID NO: 8.

17. The nucleic acid according to any one of claims 1-2, comprising a modified sequence of the adenovirus E2a gene, which comprises deletions of sequence regions corresponding to the regions of 2400-2634 bp and 2839-3492 bp in the sequence of SEQ ID NO: 1 or in a sequence homologous to the sequence of SEQ ID NO: 1 with a degree of identity of at least 63%, and includes the sequence of SEQ ID NO: 9 or a sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61%, respectively.

18. The nucleic acid according to claim 17, wherein the sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 9, or a codon-optimized variant of the sequence of SEQ ID NO: 9.

19. A nucleic acid comprising a modified sequence of the adenovirus E2a gene, which comprises a sequence selected from the group:

i. the sequence of SEQ ID NO: 2 or a sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62%; or

ii. the sequence of SEQ ID NO: 3 or a sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61%; or

iii. the sequence of SEQ ID NO: 4 or a sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64%; or

iv. the sequence of SEQ ID NO: 5 or a sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64%; or

v. the sequence of SEQ ID NO: 6 or a sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61%; or

vi. the sequence of SEQ ID NO: 7 or a sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63%; or

vii. the sequence of SEQ ID NO: 8 or a sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62%; or

viii. the sequence of SEQ ID NO: 9 or a sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61%.

20. The nucleic acid according to claim 19, wherein

i. the sequence homologous to the sequence of SEQ ID NO: 2 with identity of at least 62% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 2, or a codon-optimized variant of the sequence of SEQ ID NO: 2; or

ii. the sequence homologous to the sequence of SEQ ID NO: 3 with identity of at least 61% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 3, or a codon-optimized variant of the sequence of SEQ ID NO: 3; or

iii. the sequence homologous to the sequence of SEQ ID NO: 4 with identity of at least 64% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 4, or a codon-optimized variant of the sequence of SEQ ID NO: 4; or

iv. the sequence homologous to the sequence of SEQ ID NO: 5 with identity of at least 64% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 5, or a codon-optimized variant of the sequence of SEQ ID NO: 5; or

v. the sequence homologous to the sequence of SEQ ID NO: 6 with identity of at least 61% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 6, or a codon-optimized variant of the sequence of SEQ ID NO: 6; or

vi. the sequence homologous to the sequence of SEQ ID NO: 7 with identity of at least 63% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 7, or a codon-optimized variant of the sequence of SEQ ID NO: 7; or

vii. the sequence homologous to the sequence of SEQ ID NO: 8 with identity of at least 62% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 8, or a codon-optimized variant of the sequence of SEQ ID NO: 8; or

viii. the sequence homologous to the sequence of SEQ ID NO: 9 with identity of at least 61% is a respective sequence of the E2a gene of an adenovirus serotype other than the serotype with SEQ ID NO: 9, or a codon-optimized variant of the sequence of SEQ ID NO: 9.

21. A helper plasmid for producing vectors based on adeno-associated viruses, comprising a nucleic acid comprising a modified sequence of the adenovirus E2a gene according to any one of claims 1-20.

22. The helper plasmid for producing vectors based on adeno-associated viruses according to claim 21, comprising a nucleic acid comprising a sequence of the adenovirus VA RNA gene and a nucleic acid comprising a sequence of the adenovirus E4 gene.

23. The helper plasmid for producing vectors based on adeno-associated viruses according to claim 22, wherein the sequence of the adenovirus VA RNA gene comprises SEQ ID NO: 28.

24. The helper plasmid for producing vectors based on adeno-associated viruses according to claim 22, wherein the sequence of the adenovirus E4 gene is a full-length naturally-occurring sequence of the E4 gene, sequence of E4 orf6/7, or sequence of E4 orf6.

25. The helper plasmid for producing vectors based on adeno-associated viruses according to claim 24, wherein the full-length naturally-occurring sequence of the adenovirus E4 gene comprises SEQ ID NO: 29.

26. The helper plasmid for producing vectors based on adeno-associated viruses according to claim 24, wherein the sequence of adenovirus E4 orf6/7 is a sequence encoding SEQ ID NO: 30.

27. The helper plasmid for producing vectors based on adeno-associated viruses according to claim 24, wherein the sequence of adenovirus E4 orf6 is a sequence encoding SEQ ID NO: 31.

28. Use of the nucleic acid comprising a modified sequence of the adenovirus E2a gene according to any one of claims 1-20 or the helper plasmid according to any one of claims 21-27 for producing vectors based on adeno-associated viruses.