FACTOR IX ENCODING NUCLEOTIDES

Field of the Invention

The present invention relates to polynucleotides comprising a nucleotide sequence encoding Factor IX, viral particles comprising the polynucleotides and treatments utilising the polynucleotides.

Background to the Invention

Haemophilia B, an X-linked life threatening bleeding disorder affects 1:30,000 males. Current treatment involves frequent intravenous injections (2-3 times per week) of Factor IX (FIX) protein. This treatment is highly effective at arresting bleeding but it is not curative and is extremely expensive (£150,000/patient/year), thus making it unaffordable by the majority of haemophilia B patients in the world. Gene therapy for haemophilia B offers the potential for a cure through persistent, endogenous production of Factor IX following the transfer of a functioning copy of the Factor IX gene to an affected patient.

The present application relates to a gene therapy approach for treating haemophilia B, involving administering a vector comprising a polynucleotide encoding Factor IX. Such a gene therapy approach would avoid the need for frequent intravenous injections of Factor IX. However, it is difficult to provide an effective gene therapy vector, *i.e.* one that allows for a high level of Factor IX expression and of the expression of Factor IX which is highly active.

Summary of the Invention

The present application demonstrates that various modifications to a polynucleotide comprising a Factor IX nucleotide sequence can help to improve the expression level and the activity of the expressed Factor IX polypeptide. For example, the present application demonstrates that the following can improve the efficacy of a polynucleotide comprising a Factor IX nucleotide sequence for treatment of haemophilia B:

- using a codon optimised sequence;
- maintaining a portion of the Factor IX polypeptide that is not codon optimised;
- including an intron or a fragment of an intron;
- providing sequences flanking the intron or fragment of an intron that are not codon optimised;
- using a gain of function mutation;
- using a specific promoter; and/or
- maintaining an AAV genome, comprising the nucleotide, in single stranded form.

These modifications provide a Factor IX sequence which is expressed highly, and which encodes a highly active Factor IX polypeptide or fragment thereof. As demonstrated in the Examples, the polynucleotide of the invention expresses, and provides overall Factor IX activity, at higher levels than other Factor IX encoding polynucleotides, for example those disclosed in WO16/075473.

Accordingly, in a first aspect of the invention, there is provided a polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence comprises a coding sequence that encodes a Factor IX protein or fragment thereof and wherein a portion of the coding sequence is not wild type.

In a second aspect of the invention, there is provided a polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence comprises a coding sequence that encodes a Factor IX protein or a fragment thereof and the coding sequence comprises: (i) a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1; and (ii) a sequence that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 15.

In a third aspect of the invention, there is provided a polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence encodes a Factor IX protein or fragment thereof and has at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identity to SEQ ID NO. 5.

In a fourth aspect of the invention, there is provided a viral particle comprising a recombinant genome comprising the polynucleotide of the invention.

In a fifth aspect of the invention, there is provided a composition comprising the polynucleotide or viral particle of the invention and a pharmaceutically acceptable excipient.

In a sixth aspect of the invention, there is provided a method of treatment comprising administering an effective amount of the polynucleotide or viral particle of the invention to a patient.

In a seventh aspect of the invention, there is provided a use of the polynucleotide, viral particle or composition of the invention in the manufacture of a medicament for use in a method of treatment.

In an eighth aspect, the polynucleotide, viral particle or composition of the invention is for use in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7, thereby treating a disease.

In a ninth aspect, the polynucleotide, viral particle or composition of the invention is for use in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7.

In a tenth aspect, there is provided a method of expressing a polypeptide encoded by a Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7 comprising administering the polynucleotide, viral particle or composition of the invention to a patient.

In an eleventh aspect, there is provided a method of treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7 comprising administering the polynucleotide, viral particle or composition of the invention to a patient.

In a twelfth aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7, thereby treating a disease.

In a thirteenth aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7.

In a fourteenth aspect, the polynucleotide, viral particle or composition of the invention is for use in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, thereby treating a disease.

In a fifteenth aspect, the polynucleotide, viral particle or composition of the invention is for use in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

In a sixteenth aspect, there is provided a method of expressing a polypeptide encoded by a Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6 comprising administering the polynucleotide, viral particle or composition of the invention to a patient.

In a seventeenth aspect, there is provided a method of treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6 comprising administering the polynucleotide, viral particle or composition of the invention to a patient.

In an eighteenth aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, thereby treating a disease.

In a nineteenth aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

In a twentieth aspect, the polynucleotide, viral particle or composition of the invention is for use in treating a disease by expressing Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

In a twenty-first aspect, there is provided a method of treating a disease by expressing Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, comprising administering an effective amount of the polynucleotide, viral particle or composition of the invention to a patient.

In a twenty-second aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in treating a disease by expressing Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

In a twenty-third aspect, the polynucleotide, viral particle or composition of the invention is for use in achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B.

In a twenty-fourth aspect, there is provided a method of achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B, comprising administering an effective amount of the polynucleotide, viral particle or composition of the invention to a patient.

In a twenty-fifth aspect there is provided a use of the polynucleotide, viral particle or composition of the invention in achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B.

In a twenty-sixth aspect, there is provided the polynucleotide, viral particle or composition of the invention for use in treating haemophilia B comprising administering the polynucleotide, viral particle or composition to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient.

In a twenty-seventh aspect, there is provided a method of treating haemophilia B comprising administering the polynucleotide, viral particle or composition of the invention to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient.

In a twenty-eighth aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in treating haemophilia B comprising administering the polynucleotide, viral particle or composition to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient.

Description of the Figures

Figure 1 – Schematic of FIX transgene cassettes ssLP1.FIXco, ssHLP2.TI-codop-FIX-GoF (HTFG) and ssHLP2.TI-ACNP-FIX-GoF (HTAG). ITR = Inverted Terminal Repeat; HLP2 and LP1 are transcription regulatory elements of SEQ ID NOs: 6 and 7 respectively; E = Exon; T1 = Truncated Intron 1A; WT = Wild Type; CO = Codon Optimised; ACNP = a codon optimised sequence of the invention.

Figure 2 – Results from HUH7 transduction with AAV2/Mut C vectors – Experiment 1; (A) shows the level of FIX antigen in supernatant; (B) shows the level of FIX antigen after normalisation using the number of vector genomes present in cell lysate; Error bars represent mean ± SD of n=2. 1e3 = 1 x 10^3; 5e3 = 5 x 10^3; 1e4 = 1 x 10^4; MOI = multiplicity of infection.

Figure 3 – Results from HUH7 transduction with AAV2/Mut C vectors – Experiment 2, showing the level of FIX antigen in supernatant. Error bars represent mean ± SD of n=3. 1e3 = 1 x 10^3; 5e3 = 5 x 10^3; 1e4 = 1 x 10^4, MOI = multiplicity of infection.

Figure 4 – Results from HUH7 transduction with AAV2/Mut C vectors – Experiment 2; A) shows the level of FIX antigen in supernatant; (B) shows the level of FIX antigen after normalisation using the number of vector genomes present in cell lysate. Error bars represent mean ± SD of n=3. 1e3 = 1 x 10^3; 5e3 = 5 x 10^3; 1e4 = 1 x 10^4; MOI = multiplicity of infection.

Figure 5 – Combined data (from Experiments 1 and 2) for AAV2/Mut C transduction of HUH7 cells at MOI 5 x 10^3. Error bars represent mean ± SD of n = 12. P= 0.001 by Student's T-test. 5e3 = 5 x 10^3; MOI = multiplicity of infection.

Figure 6 – The activity of FIX for MOI 1x 10^3, 5 x 10^3 and 1 x 10^4 is shown after HUH7 transduction with AAV2/Mut C vectors (Experiment 1). Error bars represent mean ± SD of n=2 duplicate wells. 1e3 = 1 x 10^3; 5e3 = 5 x 10^3; 1e4 = 1 x 10^4; MOI = multiplicity of infection.

Figure 7 – The activity of FIX is shown after HUH7 transduction with AAV2/Mut C vectors (Experiment 2). Error bars represent mean ± SD of n=3. 1e3 = 1 x 10^3; 5e3 = 5 x 10^3; 1e4 = 1 x 10^4; MOI = multiplicity of infection.

Figure 8 – The activity of FIX is shown after HUH7 transduction with AAV2/Mut C vectors (Experiment 3). Error bars represent mean ± SD of n=3. 1e3 = 1 x 10^3; 5e3 = 5 x 10^3; 1e4 = 1 x 10^4; MOI = multiplicity of infection.

Figure 9 - Combined data (from Experiments 1 and 2; Figures 6 and 7) showing activity of FIX for MOI 5 x 10^3 shown after HUH7 transduction with AAV2/Mut C vectors. Error bars represent mean ± SD of n = 12. 5e3 = 5 x 10^3; MOI = multiplicity of infection. Statistical significance determined using a Student's T-test (p=0.0195).

Figure 10 – Normalised level of human FIX in murine plasma after administration of AAV2/8.LP1.FIXco and AAV2/8.HLP2.HTFG (Experiment 3). FIX:Ag levels were normalised to vector copies/cell. Error bars represent mean ± SD of n=4 mice. P-value <0.05 (Student's T-test)

Figure 11 – Comparison of alternate codon optimisation of FIX in C57BL/6 mice (Experiment 4). Mice were injected with AAV2/8 vectors containing ssHLP2.HTFG, ssHLP2.HTAG or scLP1.FIXco (control). The level of FIX antigen (A) was assessed 3 weeks post-injection (P=0.0007 between ssHLP2.HTAG and sc.LP1.FIXco and p=0.0198 between ssHLP2.HTAG and ssHLP2.HTFG). Antigen levels were normalised to vector genome (B) (p=0.0009 between ssHLP2.HTAG and sc.LP1.FIXco and p=0.0039 between ssHLP2.HTAG and ssHLP2.HTFG). n=4 mice. P-values were determined using one-way ANOVA (multiple comparison).

Figure 12 – Comparison of alternate codon optimisation of FIX in C57BL/6 mice (Experiment 4). Mice were injected with AAV2/8 vectors containing ssHLP2.HTFG, ssHLP2.HTAG or scLP1.FIXco (control). The level of FIX activity was assessed 3 weeks post-injection. n=4 mice. P=0.0008 between ssHLP2.HTAG and scLP1.FIXco and p=0.01 between ssHLP2.HTAG and ssHLP2.HTFG; p values were determined using one-way ANOVA (multiple comparison).

Figure 13 – Schematic of Factor IX structure. The numbers above the schematic represent amino acid positions in the complete Factor IX polypeptide including the signal peptide and the pro-peptide region (encoded by SEQ ID NO. 9). The numbers below the schematic represent equivalent amino acid positions in mature Factor IX (which corresponds to the portion of coding sequence in SEQ ID NO.19).

Figure 14 - Sequence listing.

Figure 15 – Mean Factor IX activity in two haemophilia B patients that had received AAV2/Mut C vectors containing ssHLP2.HTAG. Two haemophilia patients were administered 4.5 x 10^{11} vg/kg of AAV2/Mut C vectors containing ssHLP2.HTAG. The Factor IX activity levels were assessed regularly during 0-55 weeks post-administration. The patients achieved stable average Factor IX activity levels of 40.5 +/- 4.5% at 52 weeks as assessed by one-stage clotting assay using the SynthASil™ kit. The patients were also administered prophylactic prednisolone at weeks 6-12 following vector administration

(infusion). "*FIX:C (%)*" describes the FIX activity level; see left y axis. "*ALT (IU/L)*" describes the alanine aminotransferase level; see right y axis.

Detailed Description

General definitions

Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by a person skilled in the art to which this invention belongs.

In general, the term "*comprising*" is intended to mean including but not limited to. For example, the phrase "*a polynucleotide comprising a Factor IX nucleotide sequence*" should be interpreted to mean that the polynucleotide has a Factor IX nucleotide sequence, but the polynucleotide may contain additional nucleotides.

In some embodiments of the invention, the word "*comprising*" is replaced with the phrase "*consisting of*". The term "*consisting of*" is intended to be limiting. For example, the phrase "*a polynucleotide consisting of a Factor IX nucleotide sequence*" should be understood to mean that the polynucleotide has a Factor IX nucleotide sequence and no additional nucleotides.

The terms "*protein*" and "*polypeptide*" are used interchangeably herein, and are intended to refer to a polymeric chain of amino acids of any length.

For the purpose of this invention, in order to determine the percent identity of two sequences (such as two polynucleotide or two polypeptide sequences), the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in a first sequence for optimal alignment with a second sequence). The nucleotide residues at nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide residue as the corresponding position in the second sequence, then the nucleotides are identical at that position. The percent identity between the two sequences

is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = number of identical positions /total number of positions in the reference sequence x 100).

Typically the sequence comparison is carried out over the length of the reference sequence. For example, if the user wished to determine whether a given ("test") sequence is 95% identical to SEQ ID NO. 5, SEQ ID NO. 5 would be the reference sequence. For example, to assess whether a sequence is at least 80% identical to SEQ ID NO. 5 (an example of a reference sequence), the skilled person would carry out an alignment over the length of SEQ ID NO. 5, and identify how many positions in the test sequence were identical to those of SEQ ID NO. 5. If at least 80% of the positions are identical, the test sequence is at least 80% identical to SEQ ID NO. 5. If the sequence is shorter than SEQ ID NO. 5, the gaps or missing positions should be considered to be non-identical positions.

The skilled person is aware of different computer programs that are available to determine the homology or identity between two sequences. For instance, a comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In an embodiment, the percent identity between two amino acid or nucleic acid sequences is determined using the Needleman and Wunsch (1970) algorithm which has been incorporated into the GAP program in the Accelrys GCG software package (available at http://www.accelrys.com/products/gcg/), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6.

For the purposes of the present invention, the term *"fragment"* refers to a contiguous portion of a sequence. For example, a fragment of SEQ ID NO. 5 of 50 nucleotides refers to 50 contiguous nucleotides of SEQ ID NO. 5.

A polynucleotide

In one aspect, the present invention provides a polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence comprises a coding

sequence that encodes a Factor IX protein or fragment thereof and wherein a portion of the Factor IX nucleotide sequence is not wild type.

The polynucleotide may further comprise one or more of the following features. The polynucleotide may comprise a portion that is not codon optimised. The polynucleotide may comprise an intron or a fragment of an intron. The polynucleotide may comprise a mutation in a codon corresponding to codon 384 of wild type Factor IX.

The term "*polynucleotide*" refers to a polymeric form of nucleotides of any length, deoxyribonucleotides, ribonucleotides, or analogs thereof. For example, the polynucleotide may comprise DNA (deoxyribonucleotides) or RNA (ribonucleotides). The polynucleotide may consist of DNA. The polynucleotide may be mRNA. Since the polynucleotide may comprise RNA or DNA, all references to T (thymine) nucleotides may be replaced with U (uracil).

A Factor IX nucleotide sequence

The polynucleotide comprises a Factor IX nucleotide sequence. The Factor IX nucleotide sequence comprises a coding sequence that encodes the Factor IX protein or fragment thereof.

A "*coding sequence*" is a sequence that encodes a polynucleotide, and excludes non coding regions such as introns. A coding sequence may be interrupted by non-coding nucleotides (*e.g.* an intron), but only nucleotides that encode the polypeptide should be considered to be part of the coding sequence. For example, a coding sequence that encodes a Factor IX protein will comprise any codons that encode an amino acid forming part of the Factor IX protein that is expressed from that coding sequence, irrespective of whether those codons are contiguous in sequence or separated by one or more non-coding nucleotides. In other words, a polynucleotide which contains stretches of coding nucleotides interrupted by a stretch of non-coding nucleotides will be considered to comprise a "*coding sequence*" consisting of the non-contiguous coding stretches immediately juxtaposed (*i.e.* minus the non-coding stretch). However, herein, the stop codon will be considered to be part of the full length coding sequence.

The term *"sequence that encodes"* refers to a nucleotide sequence comprising codons that encode the encoded polypeptide. For example, a nucleotide sequence that encodes a Factor IX protein or fragment thereof comprises codons that encode the amino acid sequence of the Factor IX protein or fragment thereof. A suitable nucleotide sequence is provided in SEQ ID NO. 5.

The following Table describes codons that encode each amino acid:

Amino Acid	**Codon**	**Amino Acid**	**Codon**	**Amino Acid**	**Codon**
Phenylalanine	TTC TTT	Proline	CCT CCC CCA CCG	Asparagine	AAT AAC
Leucine	TTA TTG CTT CTC CTA CTG	Threonine	ACT ACC ACA ACG	Lysine	AAA AAG
Isoleucine	ATT ATC ATA	Alanine	GCT GCC GCA GCG	Aspartic Acid	GAT GAC
Methionine	ATG	Tyrosine	TAT TAC	Glutamic Acid	GAA GAG
Valine	GTT GTC GTA GTG	Histidine	CAT CAC	Cysteine	TGT TGC
Serine	TCT TCC TCA TCG AGT AGC	Glutamine	CAA CAG	Tryptophan	TGG
Arginine	CGT CGC CGA CGG AGA AGG	Glycine	GGT GGC GGA GGG		

The corresponding RNA codons will contain Us in place of the Ts in the Table above.

One aspect of the present invention provides a polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence encodes a Factor IX protein or fragment thereof and has at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identity to SEQ ID NO. 5. Optionally, the Factor IX nucleotide sequence comprises a coding sequence and a portion of the coding sequence is codon optimised.

In general, the Factor IX nucleotide sequence may be at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 1200, at least 1350, or at least 1650 nucleotides of SEQ ID NO. 5. The Factor IX nucleotide sequence may be at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a contiguous fragment of at least 1200, at least 1350, or at least 1650 nucleotides of SEQ ID NO. 5. The Factor IX nucleotide sequence may be at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.5. For example, the Factor IX nucleotide sequence may be at least 98% identical to SEQ ID NO.5.

Factor IX protein or fragment thereof

The polynucleotide comprises a Factor IX nucleotide sequence comprising a coding sequence that encodes a Factor IX protein or fragment thereof.

Wild type Factor IX is a serine protease, which forms part of the coagulation cascade. Lack of or mutated Factor IX can lead to reduced blood clotting and the disease haemophilia B. A typical wild type Factor IX polypeptide is encoded by SEQ ID NO. 9 (sometimes referred to as Factor IX Mälmo B) or SEQ ID NO. 19. An alternative wild type Factor IX polypeptide differs from that encoded by SEQ ID NO. 9 at codon 194, for example codon 194 may encode threonine ("Malmo A") instead of alanine.

Factor IX (e.g. a Factor IX of SEQ ID NO. 16) as initially expressed as a precursor "immature" form, comprising a hydrophobic signal peptide (amino acids 1-28 of SEQ ID NO. 16), a pro-peptide region (amino acids 29-46 of SEQ ID NO. 16) and a mature polypeptide region, as set out in Figure 13. The mature (zymogen) form of Factor IX lacks the hydrophobic signal peptide and the pro-peptide region. The term *"mature Factor IX"* refers to a Factor IX polypeptide that does not comprise the hydrophobic signal peptide or the pro-peptide region, such as SEQ ID NO. 8.

During clotting the single-chain zymogen form is cleaved by Factor XIa or Factor VIIa to produce an active two-chain form (Factor IXa), with the two chains linked by a disulphide bridge. The activated form can catalyse the hydrolysis of an arginine-isoleucine bond in Factor X to form Factor Xa. Wild type Factor IX is inhibited by thrombin. The wild type Factor IX protein has four protein domains, a Gla domain, two tandem copies of the EGF domain and a C-terminal trypsin-like peptidase domain which is responsible for catalytic cleavage.

The term "*Factor IX protein*" refers to the single-chain zymogen form of Factor IX, the activated two-chain form and variants thereof, and may refer to the mature Factor IX polypeptide or a Factor IX polypeptide comprising the pro-peptide region and/or the signal peptide region.

Preferably the Factor IX fragment is at least 200, at least 250, at least 300, between 200 and 461, between 250 and 461, or between 300 and 461 amino acids in length. In an embodiment, the Factor IX protein or fragment thereof comprises a sequence:

a) at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 8; or

b) at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of SEQ ID NO. 8 at least 200, at least 250, at least 300, between 200 and 415, between 250 and 415, or between 300 and 415 amino acids in length.

In an embodiment, the Factor IX protein or fragment thereof is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.16; or at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of SEQ ID NO.16 at least 200, at least 250, at least 300, between 200 and 461, between 250 and 461, or between 300 and 461 amino acids in length. In an embodiment, the Factor IX protein or fragment thereof is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.16; or at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of SEQ ID NO.16 at least 300, or between 300 and 461 amino acids in length. The Factor IX protein or fragment thereof may have a sequence of SEQ ID NO: 16 or SEQ ID NO: 8.

Preferably the Factor IX protein or fragment thereof is functional. A functional Factor IX protein or fragment is one which carries out hydrolysis of an arginine-isoleucine bond in Factor X to form Factor Xa.

It is within the abilities of the person skilled in the art to determine whether a Factor IX protein or fragment encoded by a Factor IX nucleotide sequence is functional. The skilled person merely needs to express the Factor IX nucleotide sequence, and test whether the expressed protein is active. For example, the skilled person could prepare a viral particle of the invention comprising the Factor IX nucleotide sequence linked to an operable promoter, and transduce cells with the viral particle under conditions suitable for expression of the Factor IX protein or fragment thereof. The activity of the expressed Factor IX protein or fragment thereof can be analysed using a chromogenic assay, such as the activity assay described in Example 3.

For example, a suitable chromogenic assay is as follows. Factor IX is mixed with thrombin, phospholipids, calcium, thrombin activated Factor VIII and Factor XIa. Under these conditions, the Factor XIa activates the Factor IX to form Factor IXa, and the activity of the Factor IXa can catalyse cleavage of a chromogenic substrate (SXa-11) to produce pNA. The level of pNA generated can be measured by determining absorbance at 405 nm, and this is proportional to the activity of the Factor IX in the sample.

The activity can be normalised to compensate for different concentrations of Factor IX in the sample, by measuring the concentration of Factor IX in the sample using a standard ELISA assay, such as the assay described in Example 4, and dividing the activity by the Factor IX concentration. For example, an antibody that binds to Factor IX could be bound to a plate. The sample, comprising the Factor IX at unknown concentration, could be passed over the plate. A second detection antibody that binds to Factor IX could be applied to the plate, and any excess washed off. The detection antibody that remains (*i.e.* is not washed off) will be bound to Factor IX. The detection antibody could be linked to an enzyme such as horse radish peroxidase. The level of detection antibody that binds to the Factor IX on the plate could be measured by measuring the amount of the detection antibody. For example, if the detection antibody is linked to horse radish peroxidase, the horse radish peroxidase can catalyse the production of a blue reaction product from a substrate such as TMB (3,3',5,5'-tetramethylbenzidine), and the level of the blue product can be detected by absorbance at 450 nm. The level of the blue product is proportional to the amount of detection antibody that remained after the washing step, which is proportional to the amount of Factor IX in the sample.

Optionally, the Factor IX protein or fragment thereof has an activity greater than that of the Factor IX polypeptide encoded by SEQ ID NO. 9, SEQ ID NO. 19, or SEQ ID NO. 12. Optionally, the activity is measured using a chromogenic substrate which is specific for Factor IX, i.e. a substrate which may be altered by Factor IXa to provide a chromogenic signal. A suitable chromogenic substrate is SXa-11.

In an embodiment, the Factor IX protein or fragment thereof comprises a mutation at a position corresponding to position 384 of wild type Factor IX. For example, position 384 (numbering from the start of the signal peptide, *i.e.* a position corresponding to amino acid 384 of SEQ ID NO. 16) of wild type Factor IX is an arginine residue (R384), but this can be replaced by a different residue. In an embodiment, R384 is replaced with a small, hydrophobic amino acid. For example, the small, hydrophobic amino acid could be alanine, isoleucine, leucine, valine or glycine. Preferably, the Factor IX protein or fragment thereof comprises a leucine at a position corresponding to position 384 in wild type Factor IX, as shown in SEQ ID NO. 16.

A mutation at a position corresponding to position 384 of the wild type sequence may cause a gain-of-function (GoF) mutation, resulting in Factor IX that is hyperfunctional. The advantage of expressing a Factor IX protein containing a mutation at position 384 is that a relatively small increase in protein amount produces a larger increase in overall protein activity.

It is within the abilities of the person skilled in the art to determine whether a given polypeptide has a mutation at a position corresponding to position 384. The person skilled in the art merely needs to align the sequence of the polypeptide sequence with that of a wild type (precursor, immature) Factor IX polypeptide, and determine whether the residue of the former that aligns with the 384th residue of the latter is an arginine. If not, the polypeptide has a mutation at a position corresponding to position 384 in wild type Factor IX. The alignment may be performed using any suitable algorithm such as that of Needleman and Wunsch described above.

A portion of the coding sequence is not wild type

A portion of the coding sequence may not be wild type. The wild type Factor IX-encoding nucleotide sequence is represented by SEQ ID NO. 9, and a coding sequence that comprises a portion differing from that of SEQ ID NO. 9 comprises a portion that is not wild type (providing such portion also differs from other Factor IX coding sequences which are regarded also as wild type, for example the Malmo A variant mentioned previously).

In an embodiment, the portion of the coding sequence that is not wild type is codon optimised. To identify whether a coding sequence comprises a portion that is codon optimised, one can align the coding sequence with SEQ ID NO. 9. If any portions of the sequence are not identical to SEQ ID NO. 9, the user should then determine whether they are codon optimised, *i.e.*, whether they comprise at least one codon that has been replaced with a favoured codon, *i.e.*, one of TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG. If the portion

that is not wild type comprises at least one codon that has been replaced with a favoured codon, then it is codon optimised. Preferably, a contiguous portion of the coding sequence is codon optimised. However, in some embodiments, the portion of the coding sequence which is codon optimised could be split over 2, 3, 4 or 5 regions of the coding sequence. Optionally, the portion of the coding sequence which is not codon optimised is split over less than 3 or less than 2 regions of the coding sequence. A nucleotide sequence can be codon optimised by replacing codons with other codons that are favoured (*i.e.* reflective of codon bias) in a particular organ or a particular organism (so-called favoured codons). Such a codon optimisation improves expression of the nucleotide sequence in the particular organ or particular organism. For example, if a nucleotide sequence is codon optimised for the human liver, the nucleotide sequence is modified to increase the number of codons that are favoured in the human liver. The skilled person would appreciate that codon-optimising a sequence may not entail changing every codon as at some positions a "*favoured codon*" may already be present.

Such codon optimisation may be subject to other factors. For example, it can be seen that the presence of CpGs has an adverse effect on expression and so the user may decide not to use favoured codons if their use at certain positions introduces CpGs into the sequence; this will still be considered to be codon optimisation. In an embodiment, a favoured codon that ends with a C nucleotide will not be included in the portion of the coding sequence that is codon optimised, where the next codon in the sequence begins with a G. For example, codon CTC encodes leucine. CTC should not be used for encoding leucine where the next codon in the sequence begins with a G, such as codon GTT.

The present application discloses that certain codons are favoured for expression in the human liver and that reducing the CpG content of a coding sequence, whilst maintaining a high proportion of those favoured codons, improves expression of the coding sequence. The favoured codons are TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG.

In one embodiment, the portion of the coding sequence that is codon optimised is codon optimised for expression in the liver, optionally the human liver. A portion of the coding

sequence that is codon optimised for expression in the liver may comprise a higher proportion of codons that are favoured in the liver, such as favoured codons TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG.

In an embodiment, the following codons are collectively overrepresented in the portion of the coding sequence that is not wild type or is codon optimised: TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG. By *"collectively overrepresented"*, is meant that the total number of favoured codons in the portion of the coding sequence which is codon optimised or not wild type is higher than the total number of the favoured codons in the corresponding portion of a wild type Factor IX nucleotide sequence (such as that as SEQ ID NO. 9 or SEQ ID NO. 19).

In a preferred embodiment, in the portion of the coding sequence that is codon optimised there is a greater frequency of the following codons compared to the corresponding portion of a wild type Factor IX nucleotide sequence (such as that of SEQ ID NO.9): TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG. Optionally, the following codons are collectively overrepresented in the portion of the coding sequence that is codon optimised, except where their presence results in a CpG: TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG. Optionally, the portion of the coding sequence that is codon optimised comprises at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, or at least 65%, at least 70% or at least 73% of codons selected from the group consisting of: TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG.

The codon usage in a codon optimised portion of a polynucleotide of the invention (HLP2.T1-ACNP-FIX-GoF) is compared with the codon usage in a corresponding stretch of wild type Factor IX nucleotide sequence (SEQ ID NO.9) in the following table.

Table 1

Amino Acid	Codon	HTAG	%age of codons	Wild type
Phe	TTT	5	26	12
	TTC	14	74	9
Leu	CTT	0		9
	CTC	1	5	5
	CTA	0		2
	CTG	19	95	3
	TTA	0		6
	TTG	0		3
Ile	ATT	5	24	17
	ATC	16	76	7
	ATA	0		1
Met	TTG	0		0
	ATG	2	100	6
Val	GTT	0		22
	GTC	1	3	3
	GTA	0		5
	GTG	33	97	7
Ser	TCT	6	25	6
	TCC	1	4	5
	TCA	0		6
	TCG	0		0
	AGT	0		7
	AGC	17	71	3
Pro	CCT	3	23	4
	CCC	8	62	3
	CCA	2	15	8
	CCG	0		0
Thr	ACT	11	39	13
	ACC	17	61	7
	ACA	0		10

	ACG	0		1
Ala	GCT	11	55	9
	GCC	9	45	5
	GCA	0		8
	GCG	0		0
Tyr	TAT	7	50	11
	TAC	7	50	5
His	CAT	3	33	6
	CAC	6	67	4
Gln	CAA	1	8	7
	CAG	11	92	7
Asn	AAT	7	27	15
	AAC	19	73	17
Lys	AAA	1	4	12
	AAG	24	96	16
Asp	GAT	7	39	12
	GAC	11	61	7
Glu	GAA	1	3	33
	GAG	35	97	10
Cys	TGT	9	43	19
	TGC	12	57	5
Trp	TGG	7	88	7
	TGA	1	13	0
Arg	CGT	0		1
	CGC	0		1
	CGA	0		6
	CGG	0		3
	AGA	3	20	8
	AGG	12	80	1
Gly	GGT	0		8
	GGC	21	66	9
	GGA	0		15
	GGG	11	34	4

The total number of favoured codons in SEQ ID NO. 9 in this region is 120 (30% of the sequence). On the other hand, the total number of favoured codons in the codon optimised portion of HTAG is 293 (73% of the codons).

It is straightforward to determine whether a given portion of a polynucleotide comprises favoured codons. In order to determine the frequency of each codon used in a portion of a nucleotide sequence, the skilled person merely needs to enter the sequence of that portion into one of the readily available algorithms that looks at codon usage and review the results. Alternatively, the user could simply count them.

The codons that are replaced in the codon optimised portion of HTAG compared to the corresponding region of SEQ ID NO.9 are set out in the following table.

Table 2

Amino Acid	Codon replacements	Frequency
Pro	CCA to CCC	2
	CCA to CCT	2
	CCT to CCC	3
Leu	TTA to CTG	5
	CTC to CTG	3
	CTT to CTG	6
	TTG to CTG	2
	CTA to CTG	1
	TTA to TTG	0
Gly	GGC to GGG	1
	GGA to GGC	9
	GGT to GGG	4
	GGT to GGC	3
	GGG to GGC	2
	GGA to GGG	5
Ile	ATT to ATC	12
	ATC to ATT	1
	ATA to ATC	1
Val	GTA to GTG	4
	GTC to GTG	3
	GTG to GTA	

	GTT to GTG	20
	GTA to GTC	1
Lys	AAA to AAG	10
	AAC to AAG	
	AAG to AAA	1
Tyr	TAT to TAC	3
	TAC to TAT	1
Gln	CAA to CAG	6
	CAG to CAA	1
His	CAT to CAC	2
Glu	GAA to GAG	27
Cys	TGT to TGC	9
	TGC to TGT	1
Ser	AGT to AGC	3
	TCC to AGC	4
	AGT to TCT	3
	TCA to AGC	3
	TCT to AGC	4
	TCA to TCT	1
Ala	GCA to GCC	3
	GCA to GCT	4
	GCT to GCC	2
Arg	CGA to AGG	5
	AGA to AGG	5
	CGT to AGG	1
	CGG to AGG	1
	CGG to AGA	1
Thr	ACA to ACC	6
	ACT to ACC	4
	ACA to ACT	3
	ACG to ACC	1
Phe	TTT to TTC	5
Asp	GAT to GAC	5
	GAC to GAT	1
Asn	AAT to AAC	6
stop	TAA to TGA	1
GoF mutation		1

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 4, or at least 5 codons that encode phenylalanine is/are replaced with TTC compared to a reference wild type Factor IX sequence;

b) at least 60%, at least 65%, or at least 70% of the codons that encode phenylalanine are TTC;

c) at least 60%, at least 65%, or at least 70% of the codons that encode phenylalanine are TTC and the remainder are TTT; and/or

d) the codons that encode phenylalanine are TTC, except where the following codon starts with a G.

For example, when we say at least 1 of codon A is replaced with at least 1 of codon B, this refers to replacement of codon A with codon B in at least 1 position compared to a wild type sequence, such as SEQ ID NO. 9. To determine whether such a replacement has taken place, one merely needs to align the test sequence to a wild type Factor IX sequence and see which codons are different. If at least 1 codon in the test sequence corresponding to codon A of wild type Factor IX is codon B in the test sequence, then at least 1 of codon A has been replaced by codon B. For example, if the first codon is TTT in the test sequence and TTC in the wild type Factor IX sequence, the test sequence comprises at least 1 of codon TTC replaced with TTT.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 10, at least 15, or at least 16 codons that encode leucine is/are replaced with CTG compared to a reference wild type Factor IX sequence;

b) at least 90%, or at least 94% of the codons that encode leucine are CTG; and/or

c) at least 90%, or at least 95% of the codons that encode leucine are CTG and the remainder are CTC.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 10, at least 11, or at least 12 codons that encode isoleucine is/are replaced with ATC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon ATC is/are replaced with ATT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 60%, at least 70%, or at least 75% of the codons that encode isoleucine are ATC;

d) at least 60%, at least 70%, or at least 75% of the codons that encode isoleucine are ATC and the remainder are ATT; and/or

e) the codons that encode isoleucine are ATC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 10, at least 15, at least 20, or at least 25 codons that encode valine is/are replaced with GTG compared to a reference wild type Factor IX sequence;

b) at least 1 codon that encodes valine is/are replaced with GTC compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 90%, or at least 95% of the codons that encode valine are GTG; and/or

d) at least 80%, at least 90%, or at least 95% of the codons that encode valine are GTG and the remainder are GTC.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 10, at least 12, or at least 13 codons that encode serine is/are replaced with AGC compared to a reference wild type Factor IX sequence;

b) at least 1, at least 2, or at least 4 codons that encode serine is/are replaced with TCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 60%, at least 65%, or at least 70% of the codons that encode serine are AGC; and/or

d) at least 60%, at least 65%, or at least 70% of the codons that encode serine are AGC and the remainder are TCT or TCC.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, or at least 5 codons that encode proline is/are replaced with CCC compared to a reference wild type Factor IX sequence;

b) at least 1 codons that encode proline is/are replaced with CCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 50%, at least 55%, or at least 60% of the codons that encode proline are CCC;

d) at 50%, at least 55%, or at least 60% of the codons that encode proline are CCC and the remainder are CCA or CCT; and/or

e) the codons that encode proline are CCC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 6, at least 7, at least 8, or at least 10 codons that encode threonine is/are replaced with ACC compared to a reference wild type Factor IX sequence;

b) at least 1, or at least 2, codons that encode threonine is/are replaced with ACT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 45%, at least 50%, or at least 55% of the codons that encode threonine are ACC;

d) at least 45%, at least 50%, or at least 55% of the codons that encode threonine are ACC and the remainder are ACT; and/or

e) the codons that encode threonine are ACC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 3, or at least 4 codons that encode alanine is/are replaced with GCC compared to a reference wild type Factor IX sequence;

b) at least 1, at least 2, or at least 3 codons that encode alanine is/are replaced with GCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 35%, at least 40%, or at least 43% of the codons that encode alanine are GCC;

d) at least 35%, at least 40%, or at least 45% of the codons that encode alanine are GCC and the remainder are GCT; and/or

e) the codons that encode alanine are GCC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 1, or at least 2 codons that encode tyrosine is/are replaced with TAC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon TAC is/are replaced with TAT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 40%, at least 45%, or at least 48% of the codons that encode tyrosine are TAC;

d) at least 40%, at least 45%, or at least 48% of the codons that encode tyrosine are TAC and the remainder are TAT; and/or

e) the codons that encode tyrosine are TAC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 1 codons that encode histidine is/are replaced with CAC compared to a reference wild type Factor IX sequence;

b) at least 50%, at least 60%, or at least 65% of the codons that encode histidine are CAC;

c) at least 50%, at least 60%, or at least 65% of the codons that encode histidine are CAC and the remainder are CAT; and/or

d) the codons that encode histidine are CAC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 4, or at least 5 codons that encode glutamine is/are replaced with CAG compared to a reference wild type Factor IX sequence;

b) at least 1 of codon CAG is/are replaced with CAA compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 85%, or at least 90% of the codons that encode glutamine are CAG; and/or

d) at least 80%, at least 85%, or at least 90% of the codons that encode glutamine are CAG and the remainder are CAA.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 4, or at least 5 codons that encode asparagine is/are replaced with AAC compared to a reference wild type Factor IX sequence;

b) at least 60%, at least 65%, or at least 70% of the codons that encode asparagine are AAC;

c) at least 60%, at least 65%, or at least 70% of the codons that encode asparagine are AAC and the remainder are AAT; and/or

d) the codons that encode asparagine are AAC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 7, at least 8, or at least 9 codons that encode lysine is/are replaced with AAG compared to a reference wild type Factor IX sequence;

b) at least 1 of codon AAG is/are replaced with AAA compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 90%, or at least 95% of the codons that encode lysine are AAG; and/or

d) at least 80%, at least 90%, or at least 95% of the codons that encode lysine are AAG and the remainder are AAA.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 3, or at least 4 codons that encode aspartate is/are replaced with GAC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon GAC is/are replaced with GAT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 45%, at least 50%, or at least 60% of the codons that encode aspartate are GAC;

d) at least 45%, at least 50%, or at least 60% of the codons that encode aspartate are GAC and the remainder are GAT; and/or

e) the codons that encode aspartate are GAC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 15, at least 20, at least 25, or at least 26 codons that encode glutamate is/are replaced with GAG compared to a reference wild type Factor IX sequence;

b) at least 80%, at least 90%, or at least 95% of the codons that encode glutamate are GAG; and/or

c) at least 80%, at least 90%, or at least 95% of the codons that encode glutamate are GAG and the remainder are GAA.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 6, at least 7, or at least 8 codons that encode cysteine is/are replaced with TGC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon TGC is/are replaced with TGT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 40%, at least 50%, or at least 55% of the codons that encode cysteine are TGC;

d) at least 40%, at least 50%, or at least 55% of the codons that encode cysteine are TGC and the remainder are TGT; and/or

e) the codons that encode cysteine are TGC, except where the following codon starts with a G.

In an embodiment, in the portion of the coding sequence that is codon optimised in the portion of the coding sequence that is codon optimised the codons that encode tryptophan are TGG.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 8, at least 10, or at least 11 codons that encode arginine is/are replaced with AGG compared to a reference wild type Factor IX sequence;

b) at least 1 codon that encodes arginine is/are replaced with AGA compared to a reference wild type Factor IX sequence;

c) at least 60%, at least 70%, or at least 75% of the codons that encode arginine are AGG; and/or

d) at least 60%, at least 70%, or at least 75% of the codons that encode arginine are AGG and the remainder are AGA.

Preferably at least 60%, at least 70%, or at least 75% of the codons that encode arginine are AGG.

In an embodiment, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 10, at least 12, or at least 13 codons that encode glycine is/are replaced with GGC compared to a reference wild type Factor IX sequence;

b) at least 5, at least 6, at least 7, or at least 8 codons that encode glycine is/are replaced with GGG compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 50%, at least 55%, or at least 60% of the codons that encode glycine are GGC;

d) at least 50%, at least 55%, or at least 60% of the codons that encode glycine are GGC and the remainder are GGG; and/or

e) the codons that encode glycine are GGC, except where the following codon starts with a G.

In an embodiment, the portion of the coding sequence that is codon optimised comprises codons that encode phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine, and glycine.

In an embodiment, the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine and glycine, and in the codon optimised portion:

a) at least 5 codons that encode phenylalanine is/are replaced with TTC compared to a reference wild type Factor IX sequence;

b) at least 16 codons that encode leucine is/are replaced with CTG compared to a reference wild type Factor IX sequence;

c) at least 12 codons that encode isoleucine is/are replaced with ATC compared to a reference wild type Factor IX sequence;

d) at least 25 codons that encode valine is/are replaced with GTG compared to a reference wild type Factor IX sequence;

e) at least 13 codons that encode serine is/are replaced with AGC compared to a reference wild type Factor IX sequence;

f) at least 5 codons that encode proline is/are replaced with CCC compared to a reference wild type Factor IX sequence;

g) at least 10 codons that encode threonine is/are replaced with ACC compared to a reference wild type Factor IX sequence;

h) at least 4 codons that encode alanine is/are replaced with GCC compared to a reference wild type Factor IX sequence;

i) at least 2 codons that encode tyrosine is/are replaced with TAC compared to a reference wild type Factor IX sequence;

j) at least 1 codons that encode histidine is/are replaced with CAC compared to a reference wild type Factor IX sequence;

k) at least 5 codons that encode glutamine is/are replaced with CAG compared to a reference wild type Factor IX sequence;

l) at least 5 codons that encode asparagine is/are replaced with AAC compared to a reference wild type Factor IX sequence;

m) at least 9 codons that encode lysine is/are replaced with AAG compared to a reference wild type Factor IX sequence;

n) at least 4 codons that encode aspartate is/are replaced with GAC compared to a reference wild type Factor IX sequence;

o) at least 26 codons that encode glutamate is/are replaced with GAG compared to a reference wild type Factor IX sequence;

p) at least 8 codons that encode cysteine is/are replaced with TGC compared to a reference wild type Factor IX sequence;

q) the codons that encode tryptophan are TGG;

r) at least 11 codons that encode arginine is/are replaced with AGG compared to a reference wild type Factor IX sequence; and

s) at least 13 codons that encode glycine is/are replaced with GGC compared to a reference wild type Factor IX sequence.

In an embodiment, the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine and glycine, and in the codon optimised portion:

a) at least 70% of the codons that encode phenylalanine are TTC;
b) at least 94% of the codons that encode leucine are CTG;
c) at least 75% of the codons that encode isoleucine are ATC;
d) at least 95% of the codons that encode valine are GTG;
e) at least 70% of the codons that encode serine are AGC;
f) at least 60% of the codons that encode proline are CCC;
g) at least 55% of the codons that encode threonine are ACC;
h) at least 43% of the codons that encode alanine are GCC;
i) at least 48% of the codons that encode tyrosine are TAC;
j) at least 65% of the codons that encode histidine are CAC;
k) at least 90% of the codons that encode glutamine are CAG;
l) at least 70% of the codons that encode asparagine are AAC;
m) at least 95% of the codons that encode lysine are AAG;
n) at least 60% of the codons that encode aspartate are GAC;
o) at least 95% of the codons that encode glutamate are GAG;
p) at least 55% of the codons that encode cysteine are TGC;
q) the codons that encode tryptophan are TGG;
r) at least 75% of the codons that encode arginine are AGG; and
s) at least 60% of the codons that encode glycine are GGC.

In an embodiment, the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine and glycine, and in the codon optimised portion:

a) at least 70% of the codons that encode phenylalanine are TTC and the remainder are TTT;

b) at least 94% of the codons that encode leucine are CTG and the remainder are CTC;

c) at least 75% of the codons that encode isoleucine are ATC and the remainder are ATT;

d) at least 95% of the codons that encode valine are GTG;

e) at least 70% of the codons that encode serine are AGC;

f) at least 60% of the codons that encode proline are CCC and the remainder are CCA or CCT;

g) at least 55% of the codons that encode threonine are ACC and the remainder are ACT;

h) at least 43% of the codons that encode alanine are GCC and the remainder are GCT;

i) at least 48% of the codons that encode tyrosine are TAC and the remainder are TAT;

j) at least 65% of the codons that encode histidine are CAC and the remainder are CAT;

k) at least 90% of the codons that encode glutamine are CAG and the remainder are CAA;

l) at least 70% of the codons that encode asparagine are AAC and the remainder are AAT;

m) at least 95% of the codons that encode lysine are AAG and the remainder are AAA;

n) at least 60% of the codons that encode aspartate are GAC and the remainder are GAT;

o) at least 95% of the codons that encode glutamate are GAG and the remainder are GAA;

p) at least 55% of the codons that encode cysteine are TGC and the remainder are TGT;

q) the codons that encode tryptophan are TGG;

r) at least 75% of the codons that encode arginine are AGG and the remainder are AGA; and

s) at least 60% of the codons that encode glycine are GGC and the remainder are GGG.

The reference wild type Factor IX sequence may be SEQ ID NO. 9 or SEQ ID NO. 19.

The portion that is codon optimised can correspond to a sequence encoding part of or an entire Factor IX protein. For example, the Factor IX protein could be a full length coding sequence (such as a sequence encoding SEQ ID NO. 8 or SEQ ID NO. 16) or a variant thereof, and the entire coding sequence could be codon optimised. Hence, reference herein to *"a portion of the coding sequence is codon optimised"* should be understood to mean *"at least a portion of the coding sequence is codon optimised"*. In some embodiments, however, a portion of the coding sequence is not codon optimised, for example a portion of the coding sequence is not codon optimised for expression in the liver. In some embodiments, the portion of the coding sequence that is codon optimised is at least 800, at least 900, at least 1100, less than 1500, less than 1300, less than 1200, between 800 and 1500, between 900 and 1300, between 1100 and 1200, or around 1191 nucleotides in length.

In an embodiment, the portion of the coding sequence that is codon optimised comprises exon 3 or a portion of at least 10, at least 15, at least 20, less than 25, between 10 and 25, between 15 and 25, or between 20 and 25 nucleotides of exon 3. In a further embodiment, the portion of the coding sequence that is codon optimised comprises exon 4 or a portion of at least 80, at least 90, at least 100, less than 114, between 80 and 114, between 90 and 114, or between 100 and 114 nucleotides of exon 4. In a further embodiment, the portion of the coding sequence that is codon optimised comprises exon 5 or a portion of at least 90, at least 100, at least 110, less than 129, between 90 and 129, between 100 and 129, or between 110 and 129 nucleotides of exon 5. In a further embodiment, the portion of the coding sequence that is codon optimised comprises exon 6 or a portion of at least 150, at least 180, at least 200, less than 203, between 150 and 203, between 180 and 203, or between 200 and 203 nucleotides of exon 6. In a further embodiment, the portion of the coding sequence that is codon optimised comprises exon 7 or a portion of at least 70, at least 80, at least 90, at least 100, less than 115, between 70 and 115, between 80 and 115,

between 90 and 115, or between 100 and 115 nucleotides of exon 7. In a further embodiment, the portion of the coding sequence that is codon optimised comprises exon 8 or a portion of at least 400, at least 450, at least 500, less than 548, between 400 and 548, between 450 and 548, or between 500 and 548 nucleotides of exon 8.

Exon 3 comprises nucleotides 253-277 of wild type Factor IX (such as a Factor IX of SEQ ID NO. 9), or a corresponding sequence in a non-wild-type Factor IX nucleotide sequence. Exon 4 comprises nucleotides 278-391 of wild type Factor IX (such as a Factor IX of SEQ ID NO: 9), or a corresponding sequence in a non-wild-type Factor IX nucleotide sequence. Exon 5 comprises nucleotides 392-520 of wild type Factor IX (such as a Factor IX of SEQ ID NO: 9), or a corresponding sequence in a non-wild-type Factor IX nucleotide sequence. Exon 6 comprises nucleotides 521-723 of wild type Factor IX (such as a Factor IX of SEQ ID NO: 9), or a corresponding sequence in a non-wild-type Factor IX nucleotide sequence. Exon 7 comprises nucleotides 724-838 of wild type Factor IX (such as a Factor IX of SEQ ID NO: 9), or a corresponding sequence in a non-wild-type Factor IX nucleotide sequence. Exon 8 comprises nucleotides 839-1386 of wild type Factor IX (such as a Factor IX of SEQ ID NO: 9), or a corresponding sequence in a non-wild-type Factor IX nucleotide sequence.

Preferably a portion of at least 20 nucleotides of exon 3, a portion of at least 100 nucleotides of exon 4, a portion of at least 110 nucleotides of exon 5, a portion of at least 180 nucleotides of exon 6, a portion of at least 100 nucleotides of exon 7, and a portion of at least 500 nucleotides of exon 8 are codon optimised. The portion of the coding sequence that is codon optimised may comprise exon 3, exon 4, exon 5, exon 6, exon 7 and exon 8. In an embodiment, the portion of the coding sequence that is codon optimised comprises exon 3, exon 4, exon 5, exon 6, exon 7 and exon 8.

In an embodiment, the portion of the coding sequence that is codon optimised comprises a portion of exon 2, and the portion of exon 2 is less than 160, less than 150, less than 100, less than 75, less than 60, at least 20, at least 30, at least 40, at least 50, between 20 and 160, between 30 and 150, between 30 and 100, between 40 and 75, or around 56 nucleotides in length. Exon 2 comprises nucleotides 89-252 of wild type Factor IX (such

as a Factor IX of SEQ ID NO: 9), or a corresponding sequence in a non-wild-type Factor IX nucleotide sequence. In a preferred embodiment, the portion of the coding sequence that is codon optimised comprises a portion of exon 2 that is between 30 and 100 nucleotides in length.

It is within the capabilities of the person skilled in the art to determine whether a portion of a sequence encoding a Factor IX protein or fragment thereof corresponds, for example, to exon 8 of wild type Factor IX. The person skilled in the art merely needs to perform a sequence alignment of the sequence encoding the Factor IX protein or fragment thereof with exon 8 using a suitable alignment algorithm such as that of Needleman and Wunsch described above, and determine whether at least part of the nucleotide sequence has greater than 90%, greater than 95%, or greater than 98% identity to exon 8 of SEQ ID NO. 9 (as described above, exon 8 of SEQ ID NO. 9 consists of nucleotides 839-1386 of SEQ ID NO.9).

As discussed above, providing a polynucleotide sequence comprising a coding sequence that is partially or wholly codon optimised can ensure that the encoded polypeptide is expressed at a high level. In one embodiment, a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at higher levels compared to the reference wild type Factor IX sequence. The reference wild type Factor IX sequence may be SEQ ID NO: 9. In an embodiment, a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at higher levels compared to a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO: 7. In an embodiment, a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at higher levels compared to a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO: 18 and a transcription regulatory element of SEQ ID NO: 6.

In an embodiment, a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at a level at least 1.5, at least 2, at least 2.5, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX

nucleotide sequence of SEQ ID NO. 12 or SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 7 or SEQ ID NO. 6. Optionally, a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at a level at least 1.5, at least 2, at least 2.5, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7. Optionally, a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at a level at least 1.5, at least 2, at least 2.5, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

In an aspect, the polynucleotide, viral particle or composition of the invention is for use in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7, thereby treating a disease. In an aspect, the polynucleotide, viral particle or composition of the invention is for use in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7. In an aspect, there is provided a method of expressing a polypeptide encoded by a Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7 comprising administering the polynucleotide, viral particle or composition of the invention to a patient. In an aspect, there is provided a method of treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7 comprising administering the polynucleotide, viral particle or composition of the invention to a patient. In an aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in expressing a polypeptide encoded by the Factor

IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7, thereby treating a disease. In an aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7. In an embodiment, expressing a polypeptide encoded by the Factor IX nucleotide sequence comprises administering the polynucleotide, viral particle or composition of the invention to a patient.

In an aspect, the polynucleotide, viral particle or composition of the invention is for use in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, thereby treating a disease. In an aspect, the polynucleotide, viral particle or composition of the invention is for use in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6. In an aspect, there is provided a method of expressing a polypeptide encoded by a Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6 comprising administering the polynucleotide, viral particle or composition of the invention to a patient. In an aspect, there is provided a method of treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6 comprising administering the polynucleotide, viral particle or composition of the invention to a patient. In an aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in expressing a polypeptide encoded by the Factor

IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, thereby treating a disease. In an aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6. In an embodiment, expressing a polypeptide encoded by the Factor IX nucleotide sequence comprises administering the polynucleotide, viral particle or composition of the invention to a patient. Optionally, the disease is haemophilia, for example haemophilia B.

The skilled person may determine whether the Factor IX nucleotide sequence is expressed at higher levels compared to a reference sequence by transducing host cells with a viral particle comprising the Factor IX nucleotide sequence, and some cells with a vector comprising the reference sequence. The cells may be cultured under conditions suitable for expressing the Factor IX protein or fragment thereof encoded by the Factor IX nucleotide sequence, and the level of expressed Factor IX protein can be compared. The level of expressed Factor IX protein can be assessed using an ELISA such as that described in the section entitled "*Factor IX protein or fragment thereof*". Suitable host cells include cultured human liver cells, such as Huh 7 cells.

As discussed above, the presence of CpGs (*i.e.* CG dinucleotides) may reduce expression efficiency. This is because CpGs may be methylated, and their methylation may lead to gene silencing thereby reducing expression. For this reason, it is preferred that the portion of the coding sequence that is codon optimised comprises a reduced number of CpGs compared to a corresponding portion of a reference wild type Factor IX sequence. In a preferred embodiment, the portion of the coding sequence that is codon optimised comprises than 40, less than 20, less than 10, less than 5, or less than 1 CpG. Preferably, the portion of the coding sequence that is codon optimised is CpG free, *i.e.* contains no (0) CG dinucleotides.

In an embodiment, the portion of the coding sequence that is codon optimised is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 800, at least 900, at least 1100, less than 1191, less than 1100, less than 1000, between 800 and 1191, between 900 and 1191, or around 1191 nucleotides of SEQ ID NO. 1. In an embodiment, the portion of the coding sequence that is codon optimised is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 1. In an embodiment, the portion of the coding sequence that is codon optimised is at least 95% identical to a fragment of between 900 and 1191 nucleotides of SEQ ID NO. 1. In an embodiment, the portion of the coding sequence that is codon optimised is at least 95%, or at least 98% identical to SEQ ID NO. 1.

The present invention provides a polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence comprises a coding sequence that encodes a Factor IX protein or a fragment thereof and the coding sequence comprises a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1 and a sequence that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 15. Optionally, the sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, or at least 99.8% identical to SEQ ID NO. 1 is codon optimised.

Portion of the coding sequence that is not codon optimised

In an embodiment, the Factor IX nucleotide sequence comprises a portion that is not codon optimised. The portion that is not codon optimised may be a contiguous portion. Including a portion that is not codon optimised may improve expression of the coding sequence, as the portion that is not codon optimised may interact beneficially with other portions of the coding sequence such as an intron or a fragment of an intron. For example, the Factor IX nucleotide sequence may comprise an intron, or a fragment of an intron, and in such cases flanking the intron or the fragment of an intron with wild type Factor IX sequence may help to ensure correct splicing.

The portion that is not codon optimised is not modified to include a greater number of favoured codons compared to the wild type sequence. For example, the portion that is not codon optimised may comprise a similar number of favoured codons to a wild type sequence. The portion that is not codon optimised may comprise less than 50% of codons TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG. Optionally, the portion that is not codon optimised comprises less than 50%, less than 45%, or less than 40% codons selected from the group consisting of: TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC, and GAG.

Optionally, the portion that is not codon optimised is not codon optimised for expression in human liver cells. In an embodiment, the portion that is not codon optimised comprises substantially the same number of favoured codons as a corresponding portion of SEQ ID NO. 9. For example, the portion that is not codon optimised may comprise at least 90% of the number of favoured codons as a corresponding portion of SEQ ID NO. 9.

Optionally, the portion that is not codon optimised is at least 100, at least 150, at least 170, at least 190, less than 250, less than 225, less than 200, or around 195 nucleotides in length.

As discussed in more detail below, the Factor IX nucleotide sequence may comprise an intron or a fragment of an intron. In such cases, the intron or the fragment of an intron may be flanked by the portion that is not codon optimised, *i.e.* some of the portion that is not codon optimised may be adjacent to the 3' end of the intron or the fragment of an intron and some of the portion that is not codon optimised may be adjacent to the 5' end of the intron or the fragment of the intron. The intron or the fragment of an intron may be between exon 1 and exon 2. In such cases, it is advantageous to include a portion that is not codon optimised which portion comprises a portion of exon 1 and a portion of exon 2.

Optionally, the portion that is not codon optimised comprises exon 1 or a portion of at least 60, at least 70, at least 80, between 60 and 88, between 70 and 88, or between 80 and 88 contiguous nucleotides of exon 1. Exon 1 comprises nucleotides 1-88 of wild type Factor

IX (such as a Factor IX of SEQ ID NO: 9), or a corresponding sequence in a non-wild-type Factor IX nucleotide sequence. Part of exon 1 may encode the signal peptide region and the pro-peptide region. Optionally, the portion that is not codon optimised comprises or does not comprise the signal peptide and/or pro-peptide regions. Exon 1 may also comprise an additional non-coding stretch of 29 nucleotide at the 5' end. If the Factor IX nucleotide sequence comprises an intron or a fragment of an intron, it is preferable that the portion that is not codon optimised comprises a portion of exon 1 that is adjacent to the intron or the fragment of an intron. For example, if the intron or the fragment of an intron is between exon 1 and exon 2, it is preferable that the portion that is not codon optimised comprises a portion of exon 1 that corresponds to nucleotides 80-88, 70-88, 60-88, 40-88, or 20-88 of SEQ ID NO.9.

Optionally, the portion that is not codon optimised comprises a portion of at least 50, at least 75, at least 80, at least 90, at least 100, less than 140, less than 120, between 50 and 140, between 75 and 120, or around 107 nucleotides of exon 2. For example, if the intron or the fragment of an intron is between exon 1 and exon 2, it is preferable that the portion that is not codon optimised comprises a portion of exon 2 that corresponds to nucleotides 89-100, 89-120, 89-140, 89-160, 89-180, or 89-196 of SEQ ID NO.9.

The portion that is not codon optimised may comprise CpGs, For example, the portion that is not codon optimised may comprise the same number of CpGs as a corresponding portion of SEQ ID NO. 9. The portion that is not codon optimised may comprise at least 1, at least 1.5, or at least 2 CpGs per 100 nucleotides. The portion that is not codon optimised may comprise at least 1, at least 2, at least 3, between 1 and 5, between 2 and 5, or around 5 CpGs.

The portion that is not codon optimised may be at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 100, at least 150, at least 175, less than 195, less than 190, or less than 180 of SEQ ID NO. 15 or SEQ ID NO: 2. The portion that is not codon optimised may be at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 15 or SEQ ID NO: 2. For

example, the portion that is not codon optimised may be at least 98% identical to SEQ ID NO. 15 or SEQ ID NO. 2.

The portion that is not codon optimised may be wild type. SEQ ID NO. 9 is an example of a wild type Factor IX nucleotide coding sequence. Thus, the portion that is not codon optimised may be at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a corresponding portion of SEQ ID NO: 9.

The Factor IX nucleotide sequence may comprise an intron or a fragment of an intron

The Factor IX nucleotide sequence may comprise an intron or a fragment of an intron that interrupts the coding sequence. An intron is a sequence of nucleotides that is excised during the process of expression, and does not form part of the coding sequence.

A genomic wild type Factor IX nucleotide sequence comprises introns, that interrupt the Factor IX coding sequence. The presence of an intron may assist in maintaining a high level of expression of wild type Factor IX. Thus, it may be advantageous to include an intron, or at least a fragment of an intron, in a Factor IX nucleotide sequence of the invention. For example, the Factor IX nucleotide sequence may comprise an intron or a fragment of an intron that corresponds to intron 1 in wild type Factor IX. Suitably, the intron is a fragment of intron 1A of wild type Factor IX, such as SEQ ID NO: 3. It has been found that truncating the sequence of intron 1 causes expression of the Factor IX nucleotide sequence to be increased. It is thought that the truncation of intron 1 to form intron 1A may delete a repressor element in the intron. Truncation of the intron 1 sequence also results in the Factor IX nucleotide sequence being shorter which allows more efficient packaging of the Factor IX nucleotide sequence into a viral delivery system in gene therapy embodiments.

The fragment of an intron may be less than 500, less than 400, less than 350, less than 300, at least 100, at least 200, at least 250, at least 290, between 100 and 500, between 200 and 400, between 250 and 350, or around 299 nucleotides. The fragment of an intron may be at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least

99.5%, at least 99.8%, or 100% identical to a fragment of at least 100, at least 200, at least 250, or at least 290 nucleotides of SEQ ID NO. 3. The intron or fragment of an intron may be at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.3. For example, the intron or the fragment of an intron may be at least 95%, or at least 98% identical to SEQ ID NO.3.

Preferably, the intron or the fragment of an intron interrupts the portion that is not codon optimised *i.e.* the intron is 5' to a portion that is not codon optimised and 3' to a portion that is not codon optimised in the Factor IX nucleotide sequence. An intron is "*flanked by*" a sequence that is not codon optimised if the nucleotides immediately 3' and 5' of the intron or close to the 3' and 5' sections of the intron are not codon optimised. *"Close to the intron"* refers to within 1, within 2, within 3, within 4, within 5, within 6, within 7, within 8, within 8 or within 10 nucleotides of the intron. As discussed above, flanking the intron or the fragment of an intron with a nucleotide sequence that is not codon optimised may help to ensure correct splicing. Optionally, the intron or the fragment of an intron is flanked by at least 60, at least 70, at least 80, at least 90, or at least 100 nucleotides that are not codon optimised. For example, an intron is flanked by 60 nucleotides that are not codon optimised if 40 nucleotides that are immediately 3' of the intron and 20 nucleotides that are immediately 5' of the intron are not codon optimised, or if 30 nucleotides that are immediately 3' of the intron and 30 nucleotides that are immediately 5' of the intron are not codon optimised. Optionally, the intron or the fragment of an intron is flanked by between 110 and 120 nucleotides that are not codon optimised at the 5' end (*e.g.* immediately 5' of the intron) and between 100 and 110 nucleotides that are not codon optimised at the 3' end (*e.g.* immediately 3' of the intron).

The intron or the fragment of an intron may be positioned between portions of the coding sequence corresponding to exon 1 and exon 2 of a Factor IX nucleotide sequence. If the intron or the fragment of an intron corresponds to a fragment of intron 1 in wild type Factor IX, it is preferable that the intron or the fragment of an intron is between portions of the coding sequence corresponding to exon 1 and exon 2 of a Factor IX nucleotide sequence.

The polynucleotide may further comprise a transcription regulatory element

The polynucleotide may comprise a transcription regulatory element.

In one embodiment, the transcription regulatory element is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 6. In an embodiment, the polynucleotide comprises a transcription regulatory element that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO: 6. Optionally, the polynucleotide comprises a transcription regulatory element of SEQ ID NO: 6.

Any appropriate transcription regulatory element may be used, such as HLP2, HLP1, LP1, HCR-hAAT, ApoE-hAAT, and LSP, which are all liver specific transcription regulatory elements. These transcription regulatory elements are described in more detail in the following references: HLP1: McIntosh J. *et al.*, Blood 2013 Apr 25, 121(17):3335-44; LP1: Nathwani *et al.*, Blood. 2006 April 1, 107(7): 2653–2661; HCR-hAAT: Miao *et al.,* Mol Ther. 2000;1: 522-532; ApoE-hAAT: Okuyama *et al.,* Human Gene Therapy, 7, 637-645 (1996); and LSP: Wang *et al.,* Proc Natl Acad Sci U S A. 1999 March 30, 96(7): 3906–3910. The HLP2 transcription regulatory element has a sequence of SEQ ID NO: 6.

The transcription regulatory element may comprise a promoter and/or an enhancer, such as the promoter element and/or enhancer element from HLP2, HLP1, LP1, HCR-hAAT, ApoE-hAAT, and LSP. Each of these transcription regulatory elements comprises a promoter, an enhancer, and optionally other nucleotides.

In an embodiment, the transcription regulatory element comprises an enhancer which is the human apolipoprotein E (ApoE) hepatic locus control region (HCR; Miao et al (2000), Molecular Therapy 1(6):522), or a fragment thereof. In an embodiment, the transcription regulatory element comprises a fragment of the HCR enhancer which is a fragment of at least 80, at least 90, at least 100, less than 192, between 80 and 192, between 90 and 192, between 100 and 250, or between 117 and 192 nucleotides in length. Optionally, the fragment of the HCR enhancer is between 100 and 250 nucleotides in length.

A suitable HCR enhancer element fragment is described in SEQ ID NO. 13. Optionally, the transcription regulatory element comprises an enhancer that is at least 80, at least 90, at least 100, less than 192, between 80 and 192, between 90 and 192, between 100 and 250, or between 117 and 192 nucleotides in length and the enhancer comprises a polynucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical SEQ ID NO. 13. Optionally, the transcription regulatory element comprises an enhancer that is between 117 and 192 nucleotides in length and the enhancer comprises a polynucleotide sequence that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical SEQ ID NO. 13.

Optionally, the transcription regulatory element comprises an enhancer that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 90, at least 100, or at least 110 nucleotides of SEQ ID NO. 13. Optionally, the polynucleotide comprises an enhancer that is at least 80%, at least 85%, at least 90%, at least 95% at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 13. Optionally, the polynucleotide comprises an enhancer that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 13. Optionally, the polynucleotide comprises an enhancer of SEQ ID NO. 13.

In an embodiment, the transcription regulatory element comprises a promoter which is a human alpha-1 anti-trypsin promoter (A1AT; Miao et al (2000), Molecular Therapy 1(6):522), or a fragment thereof. Optionally, a fragment of an A1AT promoter which is at least 100, at least 120, at least 150, at least 180, less than 255, between 100 and 255, between 150 and 225, between 150 and 300, or between 180 and 255 nucleotides in length. Optionally, the fragment of an A1AT promoter is between 150 and 300 nucleotides in length.

A suitable A1AT promoter fragment is described in SEQ ID NO. 14. Optionally, the transcription regulatory element comprises a promoter that is at least 100, at least 120, at least 150, at least 180, less than 255, between 100 and 255, between 150 and 300, or between 180 and 255 nucleotides in length and the promoter comprises a polynucleotide

sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14. Optionally, the transcription regulatory element comprises a promoter that is between 180 and 255 nucleotides in length and the promoter comprises a polynucleotide sequence that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14. Optionally, the polynucleotide comprises a promoter that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 100, at least 120, or at least 150 nucleotides of SEQ ID NO. 14. Optionally, the polynucleotide comprises a promoter that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14. Optionally, the polynucleotide comprises a promoter that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14. Optionally, the polynucleotide comprises a promoter of SEQ ID NO. 14.

If the polynucleotide is intended for expression in the liver, the promoter may be a liver-specific promoter. Optionally, the promoter is a human liver-specific promoter. A "*liver-specific promoter*" is a promoter that provides a higher level of expression in liver cells compared to other cells in general. For example, the skilled person can determine whether a promoter is a liver-specific promoter by comparing expression of the polynucleotide in liver cells (such as Huh 7 cells) with expression of the polynucleotide in cells from other tissues. If the level of expression is higher in the liver cells, compared to the cells from other tissues, the promoter is a liver-specific promoter.

Gain of Function Mutation

The Factor IX protein or fragment thereof may comprise a gain of function mutation. A gain of function mutation is a mutation that increases the activity of the Factor IX protein or fragment thereof. For example, the gain of function mutation may result in a Factor IX protein or fragment thereof that has an activity at least 1.5-fold, at least 2-fold, at least 2.5-fold, at least 3-fold, at least 4-fold, at least 5-fold, at least 6-fold, at least 6.5-fold, at least

7-fold, at least 7.5-fold, or at least 8-fold or more greater than wild type Factor IX (such as the Factor IX encoded by SEQ ID NO. 9 or SEQ ID NO. 19).

The Factor IX protein or fragment thereof may comprise a mutation at a position corresponding to position 384 of wild type Factor IX (corresponding to codon 384 of SEQ ID NO. 9 or amino acid 384 of the immature polypeptide encoded by SEQ ID NO. 9). A mutation at a position corresponding to position 384 of wild type Factor IX may be a gain of function mutation. For example, replacement of arginine 384 with leucine can lead to a substantial increase in activity.

Whether or not a Factor IX protein comprises a mutation at a position corresponding to position 384 in Factor IX can be determined by aligning the Factor IX protein with SEQ ID NO. 16 using a suitable algorithm such as that of Needleman and Wunsch described above, and determining whether the amino acid that aligns to amino acid 384 (which is leucine in SEQ ID NO. 16) is an arginine residue. If the amino acid that aligns to amino acid 384 of SEQ ID NO. 16 is not an arginine residue then the Factor IX protein has a mutation at a position corresponding to position 384 of wild type Factor IX.

Whether or not a mutation is a gain of function mutation can be determined by comparing the activity of a Factor IX protein comprising the mutation with the activity of a reference Factor IX protein that is identical except for the putative gain of function mutation. The relative activities of these two proteins can be determined using a chromogenic assay such as that discussed under the heading "*Factor IX protein or fragment thereof*". If the activity of the Factor IX protein comprising the mutation is higher than the activity of the reference protein, the mutation is a gain of function mutation.

Accordingly, the Factor IX nucleotide sequence may comprise a codon that encodes a mutation at a position corresponding to position 384 in Factor IX. For example, the Factor IX nucleotide sequence may comprise a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX that is a small, hydrophobic amino acid. The small, hydrophobic amino acid may be alanine, leucine, isoleucine, glycine, or

valine. For example, the small, hydrophobic amino acid may be alanine or leucine. Preferably the small, hydrophobic amino acid is leucine.

The codon that encodes a mutation at a position corresponding to position 384 in wild type Factor IX can be a codon that encodes leucine such as CTX, where X is any nucleotide. Preferably, X is C or G. The codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX may be CTC, such as in SEQ ID NO. 4. In alternative embodiments, the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX is TTG or CTG, such as in SEQ ID NO. 11 or SEQ ID NO. 26. For example, reference to SEQ ID NO: 1 herein may be replaced by reference to the corresponding portions of SEQ ID NOs: 26 or 11. In other words, SEQ ID NO:1 may be substituted at nucleotide 957 (C) with G, or at nucleotides 955 (C) and 957 (C) with T and G respectively.

The polynucleotide may comprise a Factor IX sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 1200, at least 1350, or at least 1650 nucleotides of SEQ ID NO. 5. For example, the Factor IX nucleotide sequence may be at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 5.

Suitably,

(i) the Factor IX nucleotide sequence comprises a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1; and

(ii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine.

Suitably,

(i) the Factor IX nucleotide sequence comprises a coding sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1;

(ii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine; and

(iii) the polynucleotide comprises an enhancer element that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 13.

Suitably,

(i) the Factor IX nucleotide sequence comprises a coding sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1;

(ii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine; and

(iii) the polynucleotide comprises a promoter element that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14.

Suitably,

(i) the Factor IX nucleotide sequence comprises a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1;

(ii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine; and

(iii) the polynucleotide comprises a transcription regulatory element that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 6.

Suitably,

(i) the Factor IX nucleotide sequence comprises a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1;

(ii) the Factor IX nucleotide sequence comprises a sequence that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO: 2; and

(iii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine.

Suitably, the Factor IX nucleotide sequence comprises an intron or a fragment of an intron, and the fragment of an intron is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 3.

A viral particle comprising the polynucleotide

The invention further provides a viral particle comprising a recombinant genome comprising polynucleotides of the invention. For the purposes of the present invention, the term "*viral particle*" refers to all or part of a virion. For example, the viral particle comprises a recombinant genome and may further comprise a capsid. The viral particle may be a gene therapy vector. Herein, the terms "*viral particle*" and "*vector*" are used interchangeably. For the purpose of the present application, a "*gene therapy*" vector is a viral particle that can be used in gene therapy, *i.e.* a viral particle that comprises all the required functional elements to express a transgene, such as a Factor IX nucleotide sequence, in a host cell after administration.

Suitable viral particles include a parvovirus, a retrovirus, a lentivirus or a herpes simplex virus. The parvovirus may be an adeno-associated virus (AAV). The viral particle is preferably a recombinant adeno-associated viral (AAV) vector or a lentiviral vector. More preferably, the viral particle is an AAV viral particle. The terms AAV and rAAV are used interchangeably herein.

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 5,000 nucleotides in length. Inverted terminal repeats (ITRs) flank the unique coding nucleotide sequences for the non-structural replication (Rep) proteins and the structural (VP) proteins. The VP proteins (VP1, -2 and -3) form the capsid. The terminal 145 nt are self-complementary and are organized so that an energetically stable intramolecular duplex forming a T-shaped hairpin may be formed. These hairpin structures function as an origin for viral DNA replication, serving as primers for the cellular DNA polymerase complex. Following wild type (wt) AAV infection in mammalian cells the Rep genes (*i.e.* encoding Rep78 and Rep52 proteins) are expressed from the P5 promoter and the P19 promoter, respectively, and both Rep proteins have a function in the replication of the viral genome. A splicing event in the Rep ORF results in the expression of actually four Rep proteins (i.e. Rep78, Rep68, Rep52 and Rep40). However, it has been shown that the unspliced mRNA, encoding Rep78 and Rep52 proteins, in mammalian cells are sufficient for AAV vector production. Also in insect cells the Rep78 and Rep52 proteins suffice for AAV vector production.

The recombinant viral genome of the invention may comprise ITRs. It is possible for an AAV vector of the invention to function with only one ITR. Thus, the viral genome comprises at least one ITR, but, more typically, two ITRs (generally with one either end of the viral genome, i.e. one at the 5' end and one at the 3' end). There may be intervening sequences between the polynucleotide and one or more of the ITRs. The polynucleotide of the invention may be incorporated into a viral particle located between two regular ITRs or located on either side of an ITR engineered with two D regions.

AAV sequences that may be used in the present invention for the production of AAV vectors can be derived from the genome of any AAV serotype. Generally, the AAV

serotypes have genomic sequences of significant homology at the amino acid and the nucleic acid levels, provide an identical set of genetic functions, produce virions which are essentially physically and functionally equivalent, and replicate and assemble by practically identical mechanisms. For the genomic sequence of the various AAV serotypes and an overview of the genomic similarities see e.g. GenBank Accession number U89790; GenBank Accession number J01901; GenBank Accession number AF043303; GenBank Accession number AF085716; Chiorini *et al*, 1997; Srivastava *et al*, 1983; Chiorini *et al*, 1999; Rutledge *et al*, 1998; and Wu *et al*, 2000. AAV serotype 1, 2, 3, 3B, 4, 5, 6, 7, 8, 9, 10, 11 or 12 may be used in the present invention. The sequences from the AAV serotypes may be mutated or engineered when being used in the production of gene therapy vectors.

Optionally, an AAV vector comprises ITR sequences which are derived from AAV1, AAV2, AAV4 and/or AAV6. Preferably the ITR sequences are AAV2 ITR sequences. Herein, the term AAVx/y refers to a viral particle that comprises some components from AAVx (wherein x is a AAV serotype number) and some components from AAVy (wherein y is the number of the same or different serotype). For example, an AAV2/8 vector may comprise a portion of a viral genome, including the ITRs, from an AAV2 strain, and a capsid derived from an AAV8 strain.

In an embodiment, the viral particle is an AAV viral particle comprising a capsid. AAV capsids are generally formed from three proteins, VP1, VP2 and VP3. The amino acid sequence of VP1 comprises the sequence of VP2. The portion of VP1 which does not form part of VP2 is referred to as VP1unique or VP1U. The amino acid sequence of VP2 comprises the sequence of VP3. The portion of VP2 which does not form part of VP3 is referred to as VP2unique or VP2U. Preferably the capsid is an AAV5 capsid or a Mut C capsid. The Mut C capsid may have at least 96%, at least 98%, at least 99%, at least 99.5%, at least 99.8% identity or 100% identity to SEQ ID NO.10. The AAV capsid may have at least 96%, at least 98%, at least 99%, at 99.5%, at least 99.8%, or 100% identity to SEQ ID NO. 17. In an alternative embodiment, the capsid has a VP2U and/or VP3 of SEQ ID NO. 17 and a VP1U sequence having at least 96%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identity to SEQ ID NO: 20, SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24 or SEQ ID NO: 25.

A viral particle of the invention may be a "*hybrid*" particle in which the viral ITRs and viral capsid are from different parvoviruses, such as different AAV serotypes. Preferably, the viral ITRs and capsid are from different serotypes of AAV, in which case such viral particles are known as transcapsidated or pseudotyped. Likewise, the parvovirus may have a "*chimeric*" capsid (*e. g.*, containing sequences from different parvoviruses, preferably different AAV serotypes) or a "*targeted*" capsid (e. g., a directed tropism).

In some embodiments, the recombinant AAV genome comprises intact ITRs, comprising functional terminal resolution sites (TRS). Such an AAV genome may contain one or two resolvable ITRs, i.e. ITRs containing a functional TRS at which site-specific nicking can take place to create a free 3' hydroxyl group which can serve as a substrate for DNA polymerase to unwind and copy the ITR. Preferably, the recombinant genome is single-stranded (*i.e.*, it is packaged into the viral particle in a single-stranded form). Optionally, the recombinant genome is not packaged in self-complementary configuration, *i.e.* the genome does not comprise a single covalently-linked polynucleotide strand with substantial self-complementary portions that anneal in the viral particle. Alternatively, the recombinant genome may be packaged in "*monomeric duplex*" form. "*Monomeric duplexes*" are described in WO 2011/122950. The genome may be packaged as two substantially complementary but non-covalently linked polynucleotides which anneal in the viral particle.

The viral particle may further comprise a poly A sequence. The poly A sequence may be positioned downstream of the nucleotide sequence encoding a functional Factor IX protein. The poly A sequence may be a bovine growth hormone poly A sequence (bGHpA). The poly A sequence may be between 250 and 270 nucleotides in length.

The viral particle of the invention optionally expresses highly in host cells. For example, on transduction in Huh7 cells, the viral particle expresses Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX protein expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a

Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

Optionally, after transduction into a population of Huh7 cells, the viral particle expresses Factor IX protein, or a fragment thereof, having a Factor IX activity greater than the activity of Factor IX expressed from a comparable viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 transduced into a comparable population of Huh7 cells. Optionally, after transduction into a population of Huh7 cells, the viral particle expresses Factor IX protein, or a fragment thereof, having a Factor IX activity greater than the activity of Factor IX expressed from a comparable viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 18 and a transcription regulatory element of SEQ ID NO. 6 transduced into a comparable population of Huh7 cells.

In an aspect, the polynucleotide, viral particle or composition of the invention is for use in treating a disease by expressing Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6. In aspect, there is provided a method of treating a disease by expressing Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, comprising administering an effective amount of the polynucleotide, viral particle or composition of the invention to a patient. In an aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in treating a disease by expressing Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a

transcription regulatory element of SEQ ID NO. 6. Optionally, the viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, is a comparable viral particle. Optionally, expressing the Factor IX protein or a fragment thereof comprises administering the polynucleotide, viral particle or composition of the invention to a patient. Optionally, the disease is haemophilia, for example haemophilia B.

In such embodiments, the term *"comparable viral particle"* refers to a viral particle that is the same as an AAV viral particle of the invention, except the comparable viral particle comprises a different Factor IX nucleotide sequence and a different transcription regulatory element (those of SEQ ID NO: 12 and SEQ ID NO: 7 or SEQ ID NO: 18 and SEQ ID NO: 6). Optionally, the activity is assessed using a chromogenic assay such as the chromogenic assay discussed above. In this case, however, the activity is not normalised for the Factor IX concentration, so the activity is a function of the level of expression as well as the inherent activity of the Factor IX protein.

Compositions, methods and uses

In a further aspect of the invention, there is provided a composition comprising the polynucleotide or vector/viral particle of the invention and a pharmaceutically acceptable excipient.

The pharmaceutically acceptable excipients may comprise carriers, diluents and/or other medicinal agents, pharmaceutical agents or adjuvants, etc. Optionally, the pharmaceutically acceptable excipients comprise saline solution. Optionally, the pharmaceutically acceptable excipients comprise human serum albumin.

The invention further provides a polynucleotide, vector/viral particle or composition of the invention for use in a method of treatment. Optionally the method of treatment comprises

administering an effective amount of the polynucleotide or vector/viral particle of the invention to a patient.

The invention further provides a method of treatment comprising administering an effective amount of the polynucleotide or vector/viral particle of the invention to a patient.

The invention further provides use of the polynucleotide, vector/viral particle or composition of the invention in the manufacture of a medicament for use in a method of treatment. Optionally the method of treatment comprises administering an effective amount of the polynucleotide or vector/viral particle of the invention to a patient. Optionally the method of treatment is a gene therapy. A "*gene therapy*" involves administering a vector/viral particle of the invention that is capable of expressing a transgene (such as a Factor IX nucleotide sequence) in the host to which it is administered.

Optionally, the method of treatment is a method of treating a coagulopathy such as haemophilia (for example haemophilia A or B) or Van Willebrands' disease. Preferably, the coagulopathy is characterised by increased bleeding and/or reduced clotting. Optionally, the method of treatment is a method of treating haemophilia, for example haemophilia B. In some embodiments, the patient is a patient suffering from haemophilia B. Optionally the patient has antibodies or inhibitors to Factor IX. Optionally, the polynucleotide and/or vector/viral particle is administered intravenously. Optionally, the polynucleotide and/or vector/viral particle is for administration only once (i.e. a single dose) to a patient.

When haemophilia B is "*treated*" in the above method, this means that one or more symptoms of haemophilia are ameliorated. It does not mean that the symptoms of haemophilia are completely remedied so that they are no longer present in the patient, although in some methods, this may be the case. The method of treatment may result in one or more of the symptoms of haemophilia B being less severe than before treatment. Optionally, relative to the situation pre-administration, the method of treatment results in an increase in the amount/concentration of circulating Factor IX in the blood of the patient, and/or the overall level of Factor IX activity detectable within a given volume of blood of

the patient, and/or the specific activity (activity per amount of Factor IX protein) of the Factor IX in the blood of the patient.

A "*therapeutically effective amount*" refers to an amount effective, at dosages and for periods of time necessary, to achieve the desired therapeutic result, such as raising the level of functional factor IX in a subject (so as to lead to functional factor IX production at a level sufficient to ameliorate the symptoms of haemophilia B).

Optionally, the vector/viral particle is administered at a dose of less than 1×10^{11}, less than 1×10^{12}, less than 5×10^{12}, less than 2×10^{12}, less than 1.5×10^{12}, less than 3×10^{12}, less than 1×10^{13}, less than 2×10^{13}, or less than 3×10^{13} vector genomes per kg of weight of patient. Optionally, the polynucleotide, viral particle or composition is administered in a dose of at least 4.5×10^{11}, or between 4.5×10^{11} and 1×10^{12} vector genomes per kg of weight of patient (vg/kg). Optionally, the polynucleotide, viral particle or composition is administered in a dose of less than 5×10^{11} vg/kg. Optionally, the polynucleotide, viral particle or composition is administered in a dose of between 4.5×10^{11} vg/kg and 5×10^{11} vg/kg, or between 4.5×10^{11} vg/kg and 4.9×10^{11} vg/kg. Optionally, the polynucleotide or viral particle that is administered is produced from a mammalian cell, and/or possesses characteristics which result from use of mammalian viral vector production cells and distinguish from vectors produced in insect viral vector production cells (e.g. baculovirus system).

Optionally, administration of a given dose of vectors/viral particles – quantified in terms of the number of vector genomes - is achieved using qPCR to titrate vector genomes. In principle, the qPCR primers can be designed to bind any part of the recombinant vector genome which is not common to wild type genomes, but is it recommended against using primer template sequences very close to the ITRs as doing so can lead to an exaggerated vector genome titre measurement. Optionally, vector genomes are quantified by quantitative polymerase chain reaction (qPCR) utilising primers directed towards a promoter region, for example the promoter region of the transgene cassette (also sometimes known as a transgene expression cassette).

Methods of performing qPCR are known to those of skill in the art. Using a real-time PCR cycler and DNA-binding dye such as SYBR Green (ThermoFisher Scientific) the amplification of nascent double-stranded amplicons can be detected in real time. Known quantities of the qPCR template genetic material (*e.g.* the promoter region) may then be serially diluted to create a standard curve, and sample vector genome titre interpolated from the standard curve.

Optionally, the dose of vector/viral particle that is administered is selected such that the subject expresses Factor IX at an activity of 10%-90%, 20%-80%, 30%-70%, 25%-50%, 20%-150%, 30%-140%, 40%-130%, 50%-120%, 60%-110% or 70%-100% of the Factor IX activity of a non-haemophilic healthy subject. Optionally, the polynucleotide, viral particle or composition is administered at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient having haemophilia B. Optionally, the polynucleotide, viral particle or composition is administered at a dose effective to achieve a stable Factor IX activity level of at least 25% in a patient having haemophilia B. Optionally, the polynucleotide, viral particle or composition is administered at a dose effective to achieve a stable Factor IX activity level of at least 30% in a patient having haemophilia B. Optionally, the polynucleotide, viral particle or composition is administered at a dose effective to achieve a stable Factor IX activity level of at least 35% in a patient having haemophilia B. Optionally, the polynucleotide, viral particle or composition is administered at a dose effective to achieve a stable Factor IX activity level of at least or around 40% in a patient having haemophilia B.

In an aspect, the polynucleotide, viral particle or composition of the invention is for use in achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B. In an aspect, there is provided a method of achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B, comprising administering an effective amount of the polynucleotide, viral particle or composition of the invention to a patient. In a further aspect there is provided a use of the polynucleotide, viral particle or composition of the invention in achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B. In a further aspect, there is provided the polynucleotide, viral particle or composition of the invention for use in treating haemophilia B comprising

administering the polynucleotide, viral particle or composition to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient having haemophilia B. In a further aspect, there is provided a method of treating haemophilia B comprising administering the polynucleotide, viral particle or composition of the invention to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient having haemophilia B. In a further aspect, there is provided a use of the polynucleotide, viral particle or composition of the invention in treating haemophilia B comprising administering the polynucleotide, viral particle or composition to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient having haemophilia B. Optionally, treating haemophilia B or achieving a stable Factor IX activity level comprises administering the polynucleotide, viral particle or composition of the invention to a patient. Optionally, to achieve a stable Factor IX activity level, such as a stable Factor IX activity level of at least 20%, at least 30% or at least 40%, the polynucleotide, viral particle or composition is administered in a dose of less than 5×10^{11} vg/kg or between 4.5×10^{11} vg/kg and 4.9×10^{11} vg/kg. Optionally, the polynucleotide or viral particle that is administered is produced from a mammalian cell and/or possesses characteristics which result from use of mammalian viral vector production cells and distinguish from vectors produced in insect viral vector production cells (*e.g.* baculovirus system).

A Factor IX activity level of at least X% (*e.g.* at least 20%) refers to a Factor IX activity level that is at least X% (e.g. 20%) of the normal Factor IX level range. The person skilled in the art would readily understand what is meant by reference to a %-of-normal Factor IX activity level, which is determined in routine clinical practice by reference to international standards. Pooled normal (healthy) human plasma has a Factor IX concentration of approximately 5µg/ml.

The term "*stable Factor IX activity level*" refers to a Factor IX activity level that maintains at or above a certain level for a continuous period of at least 5 weeks. In some embodiments, the Factor IX activity level maintains at or above a certain level for a continuous period of at least 10, at least 15, at least 20, at least 30, at least 40, or at least 50 weeks. For example, a patient has a stable Factor IX activity level of at least 20% if the

activity level maintains at at least 20% for a continuous period of at least 5 weeks. In such an example, the Factor IX activity level may continue to be at at least 20% following the at least 5 weeks and thus maintains at at least 20% for a cumulative continuous period of at least 10, at least 15, at least 20, at least 30 or at least 40, or at least 50 weeks. Optionally, a patient has a stable Factor IX activity level if the Factor IX activity level maintains at or above a certain level for a continuous period of at least 20 weeks. Optionally, a patient administered with the polynucleotide, viral particle or composition may have a stable Factor IX activity level of at least 20%, at least 25%, at least 30% or at least 35%. Optionally, a patient administered with the polynucleotide, viral particle or composition may have a stable Factor IX activity level of at least or around 40%.

Optionally, the Factor IX activity level is stable after at least 10 weeks, at least 15 weeks, at least 20 weeks, at least 30 weeks, at least 40 weeks, or at least 50 weeks from administration of the polynucleotide, viral particle or composition. For example, where a patient has a stable Factor IX activity level of at least 20% after at least 10 weeks from when the patient is administered with the polynucleotide, viral particle or composition, there is a Factor IX activity level of at least 20% that maintains at at least 20% for a continuous period of at least 5, at least 10, at least 15, at least 20, at least 30 or at least 40, or at least 50 weeks following the initial at least 10 weeks from administration. Optionally, a patient has a stable Factor IX activity level after at least 10 weeks from administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at or above a certain level for a continuous period of at least 20 weeks or at least 40 weeks (following the initial at least 10 weeks from administration). Optionally, the stable Factor IX activity level after at least 10 weeks, at least 15 weeks, at least 20 weeks, at least 30 weeks, or at least 40 weeks from administration may be at least 20%, at least 25%, at least 30% or at least 35%. Optionally, the stable Factor IX activity level after at least 10 weeks, at least 15 weeks, at least 20 weeks, at least 30 weeks, or at least 40 weeks from administration may be at least or around 40%. In some embodiments, a patient has a stable Factor IX activity level of at least 20% after at least 10 weeks from administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at least 20% for a continuous period of at least at least 20 weeks or at least 40 weeks (following the initial at least 10 weeks from administration). In some

embodiments, a patient has a stable Factor IX activity level of at least 40% after at least 10 weeks from administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at least 40% for a continuous period of at least 20 weeks or at least 40 weeks (following the initial at least 10 weeks from administration).

Optionally, the Factor IX activity level is at or above a certain level (*e.g.* 20%, 25%, 30%, 35%, or 40%) at a time point at least 10, at least 20, at least 30, at least 40 or at least 50 weeks after administration of the polynucleotide, viral particle or composition. For example, the Factor IX activity level is at or above a certain level (*e.g.* 20%, 25%, 30%, 35%, or 40%) at a time point of around 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51 or 52 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least 20% at a time point at around 10 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least 25% at a time point at around 10 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least 40% at a time point at around 10 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least 20% at a time point at around 20 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least 25% at a time point at around 20 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least or around 40% at a time point at around 20 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least 20% at a time point at around 52 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least 25% at a time point at around 52 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least or around 40% at a time point at around 52 weeks after administration of the polynucleotide, viral particle or composition.

When the Factor IX activity level is at or above a certain level (*e.g.* 20%, 25%, 30%, 35%, or 40%) at a time point at least 10, at least 20, at least 30, at least 40 or at least 50 weeks

after administration of the polynucleotide, viral particle or composition, the Factor IX activity level may also be stable. In such embodiments where the Factor IX activity level is also stable, the continuous period of at least 5 weeks (or at least 10, at least 15, at least 20, at least 30 or at least 40, or at least 50 weeks) may immediately precede, comprise, or immediately follow the time point at least 10, at least 20, at least 30, at least 40 or at least 50 weeks after administration of the polynucleotide, viral particle or composition. Optionally, the Factor IX activity level is at least 20% at a time point at around 10 weeks after administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at least 20% for a continuous period of at least 40 weeks immediately after the time point. Optionally, the Factor IX activity level is at least 25% at a time point at around 10 weeks after administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at least 25% for a continuous period of at least 40 weeks immediately after the time point. Optionally, the Factor IX activity level is at least 40% at a time point at around 10 weeks after administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at least 40% for a continuous period of at least 40 weeks immediately after the time point. Optionally, the Factor IX activity level is at least 20% at a time point at around 52 weeks after administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at least 20% for a continuous period of at least 40 weeks immediately preceding the time point. Optionally, the Factor IX activity level is at least 25% at a time point at around 52 weeks after administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at least 25% for a continuous period of at least 40 weeks immediately preceding the time point. Optionally, the Factor IX activity level is at least 40% at a time point at around 52 weeks after administration of the polynucleotide, viral particle or composition, and the Factor IX activity level maintains at at least 40% for a continuous period of at least 40 weeks immediately preceding the time point.

In patients, such as human patients, Factor IX activity may be measured by taking a blood sample from the patient, or by performing an assay on a blood sample that has been taken from the patient.

Optionally, Factor IX activity level can be measured using a clotting assay. Optionally, the clotting assay is a one-stage clotting assay. Optionally, the clotting assay, such as a one-stage clotting assay, uses a SynthASil™ kit. In one example, a suitable clotting assay (a one-stage clotting assay) is as follows.

Since Factor IX is part of the clotting cascade, a Factor IX polypeptide that has increased activity will catalyse blood clotting more quickly than a Factor IX polypeptide that has a lower activity. The Factor IX polypeptide is mixed with platelet poor plasma, and incubated at 37°C. Then phospholipid and a contact activation pathway activator such as SynthASil™ APTT (Activated Partial Thromboplastin Time) reagent are added. Calcium is then added, and the user measures the time taken for clotting to occur. The clot formation can be assessed using spectrophotometry or by a magnetic steel ball method. For example, clot formation may be deemed to have occurred with the optical density of the mixture exceeds a threshold or when the mixture has congealed sufficiently to alter the movement of the magnetic ball, which is detected by a sensor.

Optionally, the Factor IX activity is measured using a SynthASil™ one stage clotting assay kit to determine the Factor IX activity in the blood sample.

Optionally, the patient is a patient who is refractory to treatment with Factor IX having an activity lower than the activity of Factor IX expressed from a viral particle comprising (i) a Factor IX nucleotide sequence of SEQ ID NO. 4 or 5 and (ii) a transcription regulatory element of SEQ ID NO. 6. Optionally, the patient is a patient who is a patient refractory to treatment with a Factor IX polypeptide that is expressed at a level lower than that encoded by a nucleotide sequence comprising (i) a Factor IX nucleotide sequence of SEQ ID NO. 4 or 5 and (ii) a transcription regulatory element of SEQ ID NO.6. As demonstrated in the examples, a viral particle comprising (i) a Factor IX nucleotide sequence of SEQ ID NO. 4 or 5 and (ii) a transcription regulatory sequence of SEQ ID NO.6 expresses a Factor IX polypeptide at higher levels and said polypeptide has higher activity compared to a polypeptide expressed from other Factor IX nucleotide sequences. Thus, a viral particle comprising (i) a Factor IX nucleotide sequence of SEQ ID NO. 4 or 5 and (ii) a transcription regulatory sequence of SEQ ID NO.6 can be used to treat patients that need

higher levels of Factor IX, lower doses of viral particle or who have trouble expressing Factor IX from a viral particle comprising alternative Factor IX nucleotide sequences, for example patients who have lost part of their liver and so are less able to express high levels of Factor IX nucleotide sequence from liver cells.

Optionally, the method or use comprises prophylactic co-administration of a steroid, *i.e.* a suitable dose of a steroid is administered to the patient. The co-administration may be concomitant, *i.e.* the patient receives the dose of steroid at the same time as receiving a dose of the polynucleotide, vector or composition of the invention. Alternatively, the steroid may be administered before or after the polynucleotide, vector or composition of the invention is administered to the patient. Optionally, the steroid is administered at 6 to 12 weeks after administration of the polynucleotide, vector or composition of the invention to a patient.

Optionally, the steroid is selected from the group consisting of prednisone, bethamethasone, prednisolone, triamcinolone, methylprednisolone and dexamethasone. Optionally the steroid is prednisolone.

Sequence Listing

Table 3

Sequence identity number	Sequence description
1	Codon optimised portion of TI-ACNP-FIX-GoF coding sequence
2	Wild type portion of TI-ACNP-FIX-GoF coding sequence, including intron
3	Truncated FIX intron 1A
4	Coding sequence of TI-ACNP-FIX-GoF
5	Coding sequence of TI-ACNP-FIX-GoF Factor IX sequence, including intron
6	HLP2 transcription regulatory element sequence
7	LP1 transcription regulatory element sequence
8	"Mature" Factor IX amino acid sequence encoded by SEQ ID NO. 4
9	Wild type Factor IX (Malmo B variant) coding sequence
10	Mut C capsid polypeptide sequence

11	FIXco coding sequence with TTG GoF codon
12	FIXco coding sequence
13	Enhancer element from HLP2
14	Promoter element from HLP2
15	Wild type portion of TI-ACNP-FIX-GoF coding sequence, excluding the intron
16	"Immature" Factor IX amino acid sequence encoded by SEQ ID NO. 4
17	AAV5 capsid polypeptide sequence
18	Coding sequence of TI-codop-FIX-GoF Factor IX sequence, including intron
19	Wild type Factor IX (Malmo B variant) coding sequence corresponding to mature FIX polypeptide
20	AAV2-5 hybrid VP1u variant 1
21	AAV2-5 hybrid VP1u variant 2
22	AAV2-5 hybrid VP1u variant 3
23	AAV2-5 hybrid VP1u variant 4
24	AAV2-5 hybrid VP1u variant 5
25	AAV2-5 hybrid VP1u variant 6
26	FIXco coding sequence with CTG GoF codon

Examples

In the following examples, experiments were performed with recombinant AAV carrying the FIX transgene cassettes ssLP1.FIXco (FIXco herein), ssHLP2.TI-codop-FIX-GoF (HTFG herein) and ssHLP2.TI-ACNP-FIX-GoF (HTAG herein). SsHLP2.TI-codop-FIX-GoF is a version of the ssHLP2.TI-codop-FIX construct disclosed in WO2016/075473 modified to encode leucine (L) instead of arginine (R) at position 384 of the encoded FIX polypeptide. These cassettes are shown in Figure 1. ssLP1.FIXco contains a fully codon-optimised FIX coding sequence (SEQ ID NO: 12) preceded 5' by an SV40 intron, with expression driven by the LP1 promoter (SEQ ID NO: 7). ssHLP2.TI-codop-FIX-GoF and ssHLP2.TI-ACNP-FIX-GoF share the structure of having the shorter HLP2 transcription regulatory element (SEQ ID NO. 6) 5' to a FIX coding sequence which is interrupted by a truncated version of the native intron 1A and in which the exon 1 and part of the exon 2 nucleotide sequence is wild type (non-codon-optimised), with the remainder of the coding sequence codon-optimised (SEQ ID NO. 18 for HTFG and SEQ ID NO. 5 for HTAG). The nucleotide sequence of the codon-optimised portions is what differs between the respective two constructs. Unlike the wild type FIX protein encoded by ssLP1.FIXco, ssHLP2.TI-

codop-FIX-GoF and ssHLP2.TI-ACNP-FIX-GoF encode a hyper-active FIX having an arginine (R) to leucine (L) substitution at position 384 of the FIX polypeptide.

Example 1 - Methods

AAV vector production and quantification

1. AAV vector stocks were prepared by standard triple plasmid transfection of human embryonic kidney (HEK293) cells with a combination of plasmids consisting of a vector genome plasmid, an adenoviral helper plasmid, and a packaging plasmid containing AAV Rep and Cap (AAV8 or AAVMut C) functions. As the recombinant AAV particles contained a genome based on AAV2, and a capsid from serotype 8 or a synthetic capsid comprising portions from two serotypes ('Mut C'; SEQ ID NO: 10), they are referred to as 'pseudotyped'.

2. Vectors were purified by density gradient centrifugation with iodixanol.

3. Vector genomes were titred by qPCR with primers directed to the promoter region.

In vitro transduction and detection of FIX expression

1. HUH7 cells were plated at $5x10^5$ cells per well in 12-well plates.

2. Cells were then stimulated with mitomycin C for 1 hour before transduction with AAV particles carrying a FIX-encoding transgene cassette.

3. Five days after transduction, supernatant was collected and analysed for the level of FIX using a commercially available ELISA kit (Stago Asserachrom IX:Ag kit Ref #00943). Activity of FIX was analysed using the commercially available chromogenic kit from Quadratech (Biophen FIX (6) kit Ref #221806).

4. Vector genome DNA was extracted using the Qiagen DNeasy Blood and Tissue kit (Ref# 69506) according to the manufacturer's instructions, and quantified by qPCR.

Detection of FIX in vivo

1. Adult C57BL/6 mice were injected with $5x10^{10}$ vector genomes (vg) of AAV particles carrying a FIX-encoding transgene cassette via the tail vein (n=4 per group).

2. Two weeks after injection mice were anaesthetised and blood collected via cardiac puncture, added to sodium citrate (1/10 dilution), and centrifuged at 3000xg for 15 minutes at 4°C to collect the plasma, which was frozen at -80°C for analysis.

3. Liver was harvested and snap frozen in liquid nitrogen before storage at -80°C for DNA extraction (Qiagen DNeasy Blood and Tissue kit, Ref# 69506) for vector genome analysis.

4. The level of FIX present in murine plasma was determined using a FIX ELISA kit (Stago Asserachrom IX:Ag kit REF 00943). The activity of human FIX was determined using the commercially available chromogenic kit from Quadratech (Biophen FIX (6) kit Ref #221806).

Example 2 – Analysis of *in vitro* FIX transgene expression using ELISA

HUH7 (human hepatocyte) cells were cultured in standard cell culture conditions. FIX transgenes were expressed by treating HUH7 cells with mitomycin C for 1 hour then subsequently transducing with pseudotyped ssAAV2/Mut C. AAV particles were first generated by transfection of HEK293 cells with recombinant vector genome plasmid, in addition to AAV helper and packaging plasmids, and culturing for a further 48 hours. ssAAV2/Mut C vectors were purified from the HEK293 cells by density gradient centrifugation and iodixanol. Vector genomes were titred by qPCR utilizing primers directed towards the promoter region of the transgene expression cassette. FIX expression cassettes were compared to determine their relative ability to express a FIX transgene *in*

vitro by measuring FIX levels 5 days post-transduction. Vectors being evaluated were ssAAV2/Mut C.HLP2.TI-codop-FIX-GoF and ssAAV2/Mut C.HLP2.TI-ACNP-FIX-GoF. FIX levels in the culture supernatant were analysed through the use of a commercially available ELISA kit (Stago Asserachrom IX:Ag kit Ref #00943). In two separate experiments (see Figure 2B and 4B) FIX expression levels – as derived from ELISA assays utilizing the supernatant of HUH7 cultured cells – were normalised against copies of the vector genome per cell following the harvesting of HUH7 cell DNA using the Qiagen DNeasy Blood and Tissue kit (Ref #69506).

5 days post-transduction with ssAAV2/Mut C.HLP2.TI-codop-FIX-GoF and ssAAV2/Mut C.HLP2.TI-ACNP-FIX-GoF at a MOI of $1x10^3$ vector genomes (vg), FIX levels were greater in HUH7 cells transduced with ssAAV2/Mut C.HLP2.TI-ACNP-FIX-GoF than in HUH7 cells transduced with ssAAV2/Mut C.HLP2.TI-codop-FIX-GoF (n=2; Figure 2A, 3 and 4A). Similarly, when identical FIX expression assays were performed in HUH7 cells with increased MOI ($5x10^3$ and $1x10^4$), FIX levels were greater in cells transduced with ssAAV2/Mut C.HLP2.TI-ACNP-FIX-GoF than in cells transduced with ssAAV2/Mut C.HLP2.TI-codop-FIX-GoF (n=2; Figure 2A, 3 and 4A). When FIX expression levels were normalised against viral vector genome copies per cell, at each of the three MOIs tested ($1x10^3$, $5x10^3$ and $1x10^4$) FIX levels were consistently higher in HUH7 cells transduced with ssAAV2/Mut C.HLP2.TI-ACNP-FIX-GoF relative to ssAAV2/MutC.HLP2.TI-codop-FIX-GoF (n=2; Figure 2B and 4B). When the $5x10^3$ transduction data from the experiments of Figures 2, 3 and 4 is combined, it shows significantly superior expression from ssAAV2/Mut C.HLP2.TI-ACNP-FIX-GoF relative to ssAAV2/Mut C.HLP2.TI-codop-FIX-GoF (Figure 5).

Example 3 – Analysis of *in vitro* FIX transgene activity

FIX activity was assessed by harvesting HUH7 cell supernatant and using the BIOPHEN Factor IX kit (Quadratech #221806, #222101, #223201). Partially codon-optimised FIX transgenes (HLP2.TI-codop-FIX-GoF and HLP2.TI-ACNP-FIX-GoF) were compared *in vitro* to determine relative FIX activity following ssAAV2/Mut C transduction (at a MOI of $1x10^3$, $5x10^3$ and $1x10^4$) of HUH7 cells. Supernatant was isolated from the HUH7 cells

5 days post-transduction, and FIX activity was determined using the BIOPHEN Factor IX kit. Regardless of the MOI, greater mean FIX activity was observed in supernatant derived from cells transduced with ssAAV2/Mut C.HLP2.TI-ACNP-FIX-GoF (Figure 6-8). When the $5x10^3$ transduction data from the experiments of Figures 6 and 7 is combined, it shows significantly superior expression from ssAAV2/Mut C.HLP2.TI-ACNP-FIX-GoF relative to ssAAV2/Mut C.HLP2.TI-codop-FIX-GoF (Figure 9).

Example 4 – Analysis of *in vivo* FIX transgene expression using ELISA

FIX transgenes were expressed in C57Bl/6 mice following transduction by tail-vein injection with $5x10^{10}$ vector genomes (vg) of pseudotyped ssAAV2/8 vectors. AAV particles were first generated by transfection of HEK293 cells with recombinant vector genome plasmid, in addition to AAV helper and packaging plasmids, and culturing for a further 48 hours. ssAAV2/8 vectors were purified from the HEK293 cells by density gradient centrifugation and iodixanol. Vector genomes were titred by qPCR utilizing primers directed towards the promoter region of the transgene expression cassette. FIX expression cassettes LP1.FIXco, HLP2.TI-codop-FIX-GoF and HLP2.TI-ACNP-FIX-GoF were compared to determine their relative ability to express a FIX transgene.

In a first experiment involving ssAAV2/8.HLP2.TI-codop-FIX-GoF and ssAAV2/8.LP1.FIXco, 2 weeks post-dosing blood was collected from anaesthetised mice via cardiac puncture. Subsequently, plasma was isolated via the addition of sodium citrate (1/10 dilution) and centrifugation at 3000x*g* for 15 minutes at 4ºC. Circulating levels of FIX were determined using a FIX ELISA kit (Stago Asserachrom IX:Ag kit Ref #00943).

FIX expression levels were normalised against copies of vector genome per cell following the harvesting of mouse liver. Normalised FIX expression levels were determined as being significantly higher ($p < 0.05$) after transduction with ssAAV2/8.HLP2.TI-codop-FIX-GoF relative to ssAAV2/8.LP1.FIXco (n=4 mice; Figure 10).

In a further experiment the partially codon-optimised FIX transgenes (HLP2.TI-codop-FIX-GoF and HLP2.TI-ACNP-FIX-GoF) were compared *in vivo* to determine relative FIX expression following C57Bl/6 mouse transduction with ssAAV2/8. Concurrently, FIX expression was determined following transduction of C57Bl/6 mice with the scAAV2/8.LP1.FIXco vector. Plasma was isolated from the mice 3 weeks post-dosing, and FIX antigen levels were determined using the ELISA assay. Mean FIX expression levels were lowest in mice transduced with scAAV2/8.LP1.FIXco, whilst levels were greater in mice transduced with ssAAV2/8.HLP2.TI-codop-FIX-GoF (n=4 mice; Figure 11A). Mice transduced with ssAAV2/8.HLP2.TI-ACNP-FIX-GoF had significantly greater FIX expression than mice transduced with ssAAV2/8.HLP2.TI-codop-FIX-GoF (n=4 mice; Figure 11A). When FIX expression levels were normalised against viral vector genome copies per cell the trend in FIX expression was maintained, whereby ssAAV2/8.HLP2.TI-ACNP-FIX-GoF produces significantly more FIX than both ssAAV2/8.HLP2.TI-codop-FIX-GoF and scAAV2/8.LP1.FIXco (n=4; Figure 11B). Furthermore, ssAAV2/8.HLP2.TI-codop-FIX-GoF exhibited greater FIX expression than scAAV2/8.LP1.FIXco (n=4 mice; Figure 11B).

Example 5 – Analysis of *in vivo* FIX transgene activity

BIOPHEN Factor IX kit (Quadratech #221806, #222101, #223201) is a chromogenic assay for measuring Factor IX activity in human citrated plasma or in Factor IX concentrates, using a manual chromogenic method.

In the presence of thrombin, phospholipids and calcium, first Factor XIa, supplied in the assay at a constant concentration and in excess, activates FIX, present in the tested sample, into FIXa, which forms an enzymatic complex with thrombin activated factor VIII:C, also supplied in the assay at a constant concentration and in excess, phospholipids (PLPs) and Calcium, that activates Factor X, present in the assay system, into Factor Xa. This activity is directly related to the amount of Factor IX, which is the limiting factor. Generated Factor Xa is then exactly measured by its specific activity on Factor Xa chromogenic substrate (SXa-11). Factor Xa cleaves the substrate and releases pNA. The amount of pNA generated is directly proportional to the Factor IXa activity. Finally, there is a direct

relationship between the amount of Factor IX in the assayed sample and the Factor Xa activity generated, measured by the amount of pNA released, determined by colour development at 405nm.

The partially codon-optimised FIX transgenes (HLP2.TI-codop-FIX-GoF and HLP2.TI-ACNP-FIX-GoF) were compared *in vivo* to determine relative FIX activity following C57Bl/6 mouse transduction with ssAAV2/8. Concurrently, FIX activity was determined following transduction of C57Bl/6 mice with scAAV2/8.LP1.FIXco. Plasma was isolated from the mice 3 weeks post-dosing, and FIX activity was determined using the BIOPHEN Factor IX kit. Mean FIX activity was lowest in mice transduced with scAAV2/8.LP1.FIXco, whilst activity was greater in mice transduced with ssAAV2/8.HLP2.TI-codop-FIX-GoF (n=4 mice; Figure 12). Mice transduced with ssAAV2/8.HLP2.TI-ACNP-FIX-GoF had significantly greater FIX activity than mice transduced with ssAAV2/8.HLP2.TI-codop-FIX-GoF (n=4 mice; Figure 12).

Example 6

Two haemophilia B patients were treated with $4.5x10^{11}$ vg/ kg of AAV2/Mut C vectors containing ssHLP2.HTAG. Both patients were known to have severe hemophilia B with endogenous FIX activity <1% at screening. Patient 1 was receiving prophylaxis at baseline and continued to do so for the first four days of the study. Within 4 weeks of receiving a single infusion of AAV2/Mut C vectors containing ssHLP2.HTAG, both patients achieved FIX activity levels of >30%. The patients achieved stable Factor IX activity levels of 40.5 +/- 4.5% at 52 weeks as assessed by one-stage clotting assay using SynthASil™ kit. Prophylactic steroids (prednisolone) were administered during weeks 6 to 12.

Neither patient has experienced spontaneous or prolonged bleeding events. Patient 1 has had a traumatic bleed due to a cut to the right finger at Week 10 post treatment. Significantly, the patient did not require any exogenous coagulation factor and the event resolved on the same day. The patient's FIX activity at the time of the event was measured as 38%.

The invention described herein also relates to the following aspects:

1. A polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence comprises a coding sequence that encodes a Factor IX protein or fragment thereof and wherein a portion of the coding sequence is not wild type.

2. The polynucleotide of aspect 1, wherein the portion of the coding sequence that is not wild type is codon optimised.

3. A polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence comprises a coding sequence that encodes a Factor IX protein or a fragment thereof and the coding sequence comprises:
   (i) a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1; and
   (ii) a sequence that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 15.

4. The polynucleotide of aspect 3, wherein the sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, or at least 99.8% identical to SEQ ID NO. 1 is codon optimised.

5. A polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence encodes a Factor IX protein or fragment thereof and has at least 97%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identity to SEQ ID NO. 5.

6. The polynucleotide of aspect 5, wherein the Factor IX nucleotide sequence comprises a coding sequence and a portion of the coding sequence is codon optimised.

7. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises DNA or RNA.

8. The polynucleotide of any one of aspects 2, 4, 6 or 7, wherein the portion of the coding sequence that is codon optimised is a contiguous portion.

9. The polynucleotide of aspect 2, 4, 6, 7 or 8, wherein the portion of the coding sequence that is codon optimised is codon optimised for expression in the human liver.

10. The polynucleotide of any one of the preceding aspects, wherein a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at higher levels compared to a reference wild type Factor IX nucleotide sequence.

11. The polynucleotide of any one of aspects 2, 4, 6 or 7, wherein the portion of the coding sequence that is codon optimised is at least 800, at least 900, at least 1100, less than 1500, less than 1300, less than 1200, between 800 and 1500, between 900 and 1300, between 1100 and 1200, or around 1191 nucleotides in length.

12. The polynucleotide of any one of aspects 2, 4 or 6-11, wherein the portion of the coding sequence that is codon optimised comprises 1, 2, 3, 4, 5 or all of:

a) exon 3 or a portion of at least 10, at least 15, at least 20, less than 25, between 10 and 25, between 15 and 25, or between 20 and 25 nucleotides of exon 3;

b) exon 4 or a portion of at least 80, at least 90, at least 100, less than 114, between 80 and 114, between 90 and 114, or between 100 and 114 nucleotides of exon 4;

c) exon 5 or a portion of at least 90, at least 100, at least 110, less than 129, between 90 and 129, between 100 and 129, or between 110 and 129 nucleotides of exon 5;

d) exon 6 or a portion of at least 150, at least 180, at least 200, less than 203, between 150 and 203, between 180 and 203, or between 200 and 203 nucleotides of exon 6;

e) exon 7 or a portion of at least 70, at least 80, at least 90, at least 100, less than 115, between 70 and 115, between 80 and 115, between 90 and 115, or between 100 and 115 nucleotides of exon 7; and/or

f) exon 8 or a portion of at least 400, at least 450, at least 500, less than 548, between 400 and 548, between 450 and 548, or between 500 and 548 nucleotides of exon 8.

13. The polynucleotide of aspect 12, wherein the portion of the coding sequence that is codon optimised comprises a), b), c), d), e) and f).

14. The polynucleotide of aspect 12 or aspect 13, wherein the portion of the coding sequence that is codon optimised comprises a portion of at least 20 nucleotides of exon 3, a portion of at least 100 nucleotides of exon 4, a portion of at least 110 nucleotides of exon 5, a portion of at least 180 nucleotides of exon 6, a portion of at least 100 nucleotides of exon 7, and a portion of at least 500 nucleotides of exon 8.

15. The polynucleotide of any one of aspects 12-14, wherein the portion of the coding sequence that is codon optimised comprises exon 3, exon 4, exon 5, exon 6, exon 7, and exon 8.

16. The polynucleotide of any one of aspects 2, 4 or 6-15, wherein the portion of the coding sequence that is codon optimised comprises a portion of exon 2, and the portion of exon 2 is less than 160, less than 150, less than 100, less than 75, less than 60, at least 20, at least 30, at least 40, at least 50, between 20 and 160, between 30 and 150, between 30 and 100, between 40 and 75, or around 56 nucleotides in length.

17. The polynucleotide of any one of aspects 2, 4 or 6-16, wherein the portion of the coding sequence that is codon optimised comprises a portion of exon 2 that is between 30 and 100 nucleotides in length.

18. The polynucleotide of any one of aspects 2, 4 or 6-17, wherein the portion of the coding sequence that is codon optimised comprises a reduced number of CpGs compared to a corresponding portion of a reference wild type Factor IX sequence.

19. The polynucleotide of aspect 18, wherein the portion of the coding sequence that is codon optimised comprises less than less than 40, less than 20, less than 18, less than 10, less than 5, or less than 1 CpG.

20. The polynucleotide of aspect 18 or 19, wherein the portion of the coding sequence that is codon optimised is CpG free.

21. The polynucleotide of any one of aspects 2, 4 or 6-20, wherein, in the portion of the coding sequence that is codon optimised, at least 40%, at least 45%, at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, or at least 73% of the codons are selected from the group consisting of:

a) TTC;
b) CTG;
c) ATC;
d) GTG;
e) GTC;
f) AGC;
g) CCC;
h) ACC;
i) GCC;
j) TAC;
k) CAC;
l) CAG;
m) AAC;
n) AAA;
o) AAG;
p) GAC;
q) TGC;
r) AGG;
s) GGC; and
t) GAG.

22. The polynucleotide of any one of aspects 2, 4 or 6-21, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 4, or at least 5 codons that encode phenylalanine is/are replaced with TTC compared to a reference wild type Factor IX sequence;

b) at least 60%, at least 65%, or at least 70% of the codons that encode phenylalanine are TTC;

c) at least 60%, at least 65%, or at least 70% of the codons that encode phenylalanine are TTC and the remainder are TTT; and/or

d) the codons that encode phenylalanine are TTC, except where the following codon starts with a G.

23. The polynucleotide of any one of aspects 2, 4 or 6-22, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 10, at least 15, or at least 16 codons that encode leucine is/are replaced with CTG compared to a reference wild type Factor IX sequence;

b) at least 90%, or at least 94% of the codons that encode leucine are CTG; and/or

c) at least 90%, or at least 94% of the codons that encode leucine are CTG and the remainder are CTC.

24. The polynucleotide of any one of aspects 2, 4, 6-23, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 10, at least 11, or at least 12 codons that encode isoleucine is/are replaced with ATC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon ATC is/are replaced with ATT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 60%, at least 70%, or at least 75% of the codons that encode isoleucine are ATC;

d) at least 60%, at least 70%, or at least 75% of the codons that encode isoleucine are ATC and the remainder are ATT; and/or

e) the codons that encode isoleucine are ATC, except where the following codon starts with a G.

25. The polynucleotide of any one of aspects 2, 4 or 6-24, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 10, at least 15, at least 20, or at least 25 codons that encode valine is/are replaced with GTG compared to a reference wild type Factor IX sequence;

b) at least 1 codon that encodes valine is/are replaced with GTC compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 90%, or at least 95% of the codons that encode valine are GTG; and/or

d) at least 80%, at least 90%, or at least 95% of the codons that encode valine are GTG and the remainder are GTC.

26. The polynucleotide of any one of aspects 2, 4 or 6-25, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 10, at least 12, or at least 13 codons that encode serine is/are replaced with AGC compared to a reference wild type Factor IX sequence;

b) at least 1, at least 2, or at least 4 codons that encode serine is/are replaced with TCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 60%, at least 65%, or at least 70% of the codons that encode serine are AGC; and/or

d) at least 60%, at least 65%, or at least 70% of the codons that encode serine are AGC and the remainder are TCT or TCC.

27. The polynucleotide of any one of aspects 2, 4 or 6-26, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, or at least 5 codons that encode proline is/are replaced with CCC compared to a reference wild type Factor IX sequence;

b) at least 1 codons that encode proline is/are replaced with CCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 50%, at least 55%, or at least 60% of the codons that encode proline are CCC;

d) at 50%, at least 55%, or at least 60% of the codons that encode proline are CCC and the remainder are CCA or CCT; and/or

e) the codons that encode proline are CCC, except where the following codon starts with a G.

28. The polynucleotide of any one of aspects 2, 4 or 6-27, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 6, at least 7, at least 8, or at least 10 codons that encode threonine is/are replaced with ACC compared to a reference wild type Factor IX sequence;

b) at least 1, or at least 2, codons that encode threonine is/are replaced with ACT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 45%, at least 50%, or at least 55% of the codons that encode threonine are ACC;

d) at least 45%, at least 50%, or at least 55% of the codons that encode threonine are ACC and the remainder are ACT; and/or

e) the codons that encode threonine are ACC, except where the following codon starts with a G.

29. The polynucleotide of any one of aspects 2, 4 or 6-28, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 3, or at least 4 codons that encode alanine is/are replaced with GCC compared to a reference wild type Factor IX sequence;

b) at least 1, at least 2, or at least 3 codons that encode alanine is/are replaced with GCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 35%, at least 40%, or at least 43% of the codons that encode alanine are GCC;

d) at least 35%, at least 40%, or at least 45% of the codons that encode alanine are GCC and the remainder are GCT; and/or

e) the codons that encode alanine are GCC, except where the following codon starts with a G.

30. The polynucleotide of any one of aspects 2, 4 or 6-29, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 1, or at least 2 codons that encode tyrosine is/are replaced with TAC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon TAC is/are replaced with TAT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 40%, at least 45%, or at least 48% of the codons that encode tyrosine are TAC;

d) at least 40%, at least 45%, or at least 48% of the codons that encode tyrosine are TAC and the remainder are TAT; and/or

e) the codons that encode tyrosine are TAC, except where the following codon starts with a G.

31. The polynucleotide of any one of aspects 2, 4 or 6-30, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 1 codons that encode histidine is/are replaced with CAC compared to a reference wild type Factor IX sequence;

b) at least 50%, at least 60%, or at least 65% of the codons that encode histidine are CAC;

c) at least 50%, at least 60%, or at least 65% of the codons that encode histidine are CAC and the remainder are CAT; and/or

d) the codons that encode histidine are CAC, except where the following codon starts with a G.

32. The polynucleotide of any one of aspects 2, 4 or 6-31, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 4, or at least 5 codons that encode glutamine is/are replaced with CAG compared to a reference wild type Factor IX sequence;

b) at least 1 of codon CAG is/are replaced with CAA compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 85%, or at least 90% of the codons that encode glutamine are CAG; and/or

d) at least 80%, at least 85%, or at least 90% of the codons that encode glutamine are CAG and the remainder are CAA.

33. The polynucleotide of any one of aspects 2, 4 or 6-32, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 4, or at least 5 codons that encode asparagine is/are replaced with AAC compared to a reference wild type Factor IX sequence;

b) at least 60%, at least 65%, or at least 70% of the codons that encode asparagine are AAC;

c) at least 60%, at least 65%, or at least 70% of the codons that encode asparagine are AAC and the remainder are AAT; and/or

d) the codons that encode asparagine are AAC, except where the following codon starts with a G.

34. The polynucleotide of any one of aspects 2, 4 or 6-33, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 7, at least 8, or at least 9 codons that encode lysine is/are replaced with AAG compared to a reference wild type Factor IX sequence;

b) at least 1 of codon AAG is/are replaced with AAA compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 90%, or at least 95% of the codons that encode lysine are AAG; and/or

d) at least 80%, at least 90%, or at least 95% of the codons that encode lysine are AAG and the remainder are AAA.

35. The polynucleotide of any one of aspects 2, 4 or 6-34, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 1, at least 2, at least 3, or at least 4 codons that encode aspartate is/are replaced with GAC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon GAC is/are replaced with GAT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 45%, at least 50%, or at least 60% of the codons that encode aspartate are GAC;

d) at least 45%, at least 50%, or at least 60% of the codons that encode aspartate are GAC and the remainder are GAT; and/or

e) the codons that encode aspartate are GAC, except where the following codon starts with a G.

36. The polynucleotide of any one of aspects 2, 4 or 6-35, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 15, at least 20, at least 25, or at least 26 codons that encode glutamate is/are replaced with GAG compared to a reference wild type Factor IX sequence;

b) at least 80%, at least 90%, or at least 95% of the codons that encode glutamate are GAG; and/or

c) at least 80%, at least 90%, or at least 95% of the codons that encode glutamate are GAG and the remainder are GAA.

37. The polynucleotide of any one of aspects 2, 4, or 6-36, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 6, at least 7, or at least 8 codons that encode cysteine is/are replaced with TGC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon TGC is/are replaced with TGT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 40%, at least 50%, or at least 55% of the codons that encode cysteine are TGC;

d) at least 40%, at least 50%, or at least 55% of the codons that encode cysteine are TGC and the remainder are TGT; and/or

e) the codons that encode cysteine are TGC, except where the following codon starts with a G.

38. The polynucleotide of any one of aspects 2, 4, or 6-37, wherein, in the portion of the coding sequence that is codon optimised the codons that encode tryptophan are TGG.

39. The polynucleotide of any one of aspects 2, 4, or 6-38, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 8, at least 10, or at least 11 codons that encode arginine is/are replaced with AGG compared to a reference wild type Factor IX sequence;

b) at least 1 codon that encodes arginine is/are replaced with AGA compared to a reference wild type Factor IX sequence;

c) at least 60%, at least 70%, or at least 75% of the codons that encode arginine are AGG; and/or

d) at least 60%, at least 70%, or at least 75% of the codons that encode arginine are AGG and the remainder are AGA.

40. The polynucleotide of any one of aspects 2, 4, or 6-39, wherein, in the portion of the coding sequence that is codon optimised:

a) at least 5, at least 10, at least 12, or at least 13 codons that encode glycine is/are replaced with GGC compared to a reference wild type Factor IX sequence;

b) at least 5, at least 6, at least 7, or at least 8 codons that encode glycine is/are replaced with GGG compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 50%, at least 55%, or at least 60% of the codons that encode glycine are GGC;

d) at least 50%, at least 55%, or at least 60% of the codons that encode glycine are GGC and the remainder are GGG; and/or

e) the codons that encode glycine are GGC, except where the following codon starts with a G.

41. The polynucleotide of any one of aspects 2, 4, or 6-40, wherein the portion of the coding sequence that is codon optimised comprises codons that encode phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine, and glycine.

42. The polynucleotide of any one of aspects 2, 4, or 6-41, wherein the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine, and glycine, and in the codon optimised portion:

a) at least 5 codons that encode phenylalanine is/are replaced with TTC compared to a reference wild type Factor IX sequence;

b) at least 16 codons that encode leucine is/are replaced with CTG compared to a reference wild type Factor IX sequence;

c) at least 12 codons that encode isoleucine is/are replaced with ATC compared to a reference wild type Factor IX sequence;

d) at least 25 codons that encode valine is/are replaced with GTG compared to a reference wild type Factor IX sequence;

e) at least 13 codons that encode serine is/are replaced with AGC compared to a reference wild type Factor IX sequence;

f) at least 5 codons that encode proline is/are replaced with CCC compared to a reference wild type Factor IX sequence;

g) at least 10 codons that encode threonine is/are replaced with ACC compared to a reference wild type Factor IX sequence;

h) at least 4 codons that encode alanine is/are replaced with GCC compared to a reference wild type Factor IX sequence;

i) at least 2 codons that encode tyrosine is/are replaced with TAC compared to a reference wild type Factor IX sequence;

j) at least 1 codons that encode histidine is/are replaced with CAC compared to a reference wild type Factor IX sequence;

k) at least 5 codons that encode glutamine is/are replaced with CAG compared to a reference wild type Factor IX sequence;

l) at least 5 codons that encode asparagine is/are replaced with AAC compared to a reference wild type Factor IX sequence;

m) at least 9 codons that encode lysine is/are replaced with AAG compared to a reference wild type Factor IX sequence;

n) at least 4 codons that encode aspartate is/are replaced with GAC compared to a reference wild type Factor IX sequence;

o) at least 26 codons that encode glutamate is/are replaced with GAG compared to a reference wild type Factor IX sequence;

p) at least 8 codons that encode cysteine is/are replaced with TGC compared to a reference wild type Factor IX sequence;

q) the codons that encode tryptophan are TGG;

r) at least 11 codons that encode arginine is/are replaced with AGG compared to a reference wild type Factor IX sequence; and

s) at least 13 codons that encode glycine is/are replaced with GGC compared to a reference wild type Factor IX sequence.

43. The polynucleotide of any one of aspects 2, 4, or 6-42, wherein the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine, and glycine, and in the codon optimised portion:

a) at least 70% of the codons that encode phenylalanine are TTC;

b) at least 94% of the codons that encode leucine are CTG;

c) at least 75% of the codons that encode isoleucine are ATC;

d) at least 95% of the codons that encode valine are GTG;

e) at least 70% of the codons that encode serine are AGC;

f) at least 60% of the codons that encode proline are CCC;

g) at least 55% of the codons that encode threonine are ACC;

h) at least 43% of the codons that encode alanine are GCC;

i) at least 48% of the codons that encode tyrosine are TAC;

j) at least 65% of the codons that encode histidine are CAC;

k) at least 90% of the codons that encode glutamine are CAG;

l) at least 70% of the codons that encode asparagine are AAC;

m) at least 95% of the codons that encode lysine are AAG;

n) at least 60% of the codons that encode aspartate are GAC;

o) at least 95% of the codons that encode glutamate are GAG;

p) at least 55% of the codons that encode cysteine are TGC;

q) the codons that encode tryptophan are TGG;

r) at least 75% of the codons that encode arginine are AGG; and

s) at least 60% of the codons that encode glycine are GGC.

44. The polynucleotide of any one of aspects 2, 4, or 6-43, wherein the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine, and glycine, and in the codon optimised portion:

a) at least 70% of the codons that encode phenylalanine are TTC and the remainder are TTT;

b) at least 94% of the codons that encode leucine are CTG and the remainder are CTC;

c) at least 75% of the codons that encode isoleucine are ATC and the remainder are ATT;

d) at least 95% of the codons that encode valine are GTG;

e) at least 70% of the codons that encode serine are AGC;

f) at least 60% of the codons that encode proline are CCC and the remainder are CCA or CCT;

g) at least 55% of the codons that encode threonine are ACC and the remainder are ACT;

h) at least 43% of the codons that encode alanine are GCC and the remainder are GCT;

i) at least 48% of the codons that encode tyrosine are TAC and the remainder are TAT;

j) at least 65% of the codons that encode histidine are CAC and the remainder are CAT;

k) at least 90% of the codons that encode glutamine are CAG and the remainder are CAA;

l) at least 70% of the codons that encode asparagine are AAC and the remainder are AAT;

m) at least 95% of the codons that encode lysine are AAG and the remainder are AAA;

n) at least 60% of the codons that encode aspartate are GAC and the remainder are GAT;

o) at least 95% of the codons that encode glutamate are GAG and the remainder are GAA;

p) at least 55% of the codons that encode cysteine are TGC and the remainder are TGT;

q) the codons that encode tryptophan are TGG;

r) at least 75% of the codons that encode arginine are AGG and the remainder are AGA; and

s) at least 60% of the codons that encode glycine are GGC and the remainder are GGG.

45. The polynucleotide of any one of aspects 10-44, wherein the reference wild type Factor IX sequence is SEQ ID NO. 9 or SEQ ID NO. 19.

46. The polynucleotide of any one of aspects 2, 4 or 6-45, wherein the portion of the coding sequence that is codon optimised is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 800, at least 900, at least 1100, less than 1191, less than 1100, less than 1000, between 800 and 1191, between 900 and 1191, or around 1191 nucleotides of SEQ ID NO. 1.

47. The polynucleotide of aspect 46, wherein the portion of the coding sequence that is codon optimised is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 1.

48. The polynucleotide of aspect 46 or 47, wherein the portion of the coding sequence that is codon optimised is at least 95% identical to a fragment of between 900 and 1191 nucleotides of SEQ ID NO. 1.

49. The polynucleotide of any one of aspects 46-48, wherein the portion of the coding sequence that is codon optimised is at least 95%, or at least 98% identical to SEQ ID NO. 1.

50. The polynucleotide of any one of the preceding aspects, wherein the coding sequence comprises a portion that is not codon optimised.

51. The polynucleotide of aspect 50, wherein the portion that is not codon optimised is at least 100, at least 150, at least 170, at least 190, less than 250, less than 225, less than 200, or around 195 nucleotides.

52. The polynucleotide of any one of aspects 50 or 51, wherein the portion that is not codon optimised comprises exon 1 or a portion of at least 60, at least 70, at least 80, between 60 and 88, between 70 and 88, or between 80 and 88 nucleotides of exon 1.

53. The polynucleotide of any one of aspects 50-52, wherein the portion that is not codon optimised comprises a portion of at least 50, at least 75, at least 80, at least 90, at least 100, less than 140, less than 120, between 50 and 140, between 75 and 120, or around 107 nucleotides of exon 2.

54. The polynucleotide of any one of aspects 50-53, wherein the portion that is not codon optimised comprises CpGs.

55. The polynucleotide of aspect 54, wherein the portion that is not codon optimised comprises at least 1 or at least 2 CpGs per 100 nucleotides.

56. The polynucleotide of any one of aspects 50-55, wherein the portion that is not codon optimised comprises less than 50%, less than 45%, less than 40%, or less than 35% codons selected from the group consisting of:

a) TTC;

b) CTG;

c) ATC;

d) GTG;
e) GTC;
f) AGC;
g) CCC;
h) ACC;
i) GCC;
j) TAC;
k) CAC;
l) CAG;
m) AAC;
n) AAA;
o) AAG;
p) GAC;
q) TGC;
r) AGG;
s) GGC; and
t) GAG.

57. The polynucleotide of any one of aspects 50-56, wherein the portion that is not codon optimised is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 100, at least 150, at least 175, less than 195, less than 190, or less than 180 nucleotides of SEQ ID NO. 15.

58. The polynucleotide of aspect 57, wherein the portion that is not codon optimised is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 15.

59. The polynucleotide of any one of aspects 50-58, wherein the portion that is not codon optimised is wild type.

60. The polynucleotide of any one of aspects 50-59, wherein the portion that is not codon optimised is at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO: 15.

61. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide further comprises an intron or a fragment of an intron that interrupts the coding sequence.

62. The polynucleotide of aspect 61, wherein the intron or the fragment of an intron is a portion of a wild type Factor IX intron.

63. The polynucleotide of aspect 61 or 62, wherein the fragment of an intron is less than 500, less than 400, less than 350, less than 300, at least 100, at least 200, at least 250, at least 290, between 100 and 500, between 200 and 400, between 250 and 350, or around 299 nucleotides.

64. The polynucleotide of any one of aspects 61-63, wherein the fragment of an intron is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of at least 100, at least 200, at least 250, or at least 290 nucleotides of SEQ ID NO. 3.

65. The polynucleotide of any one of aspects 61-64, wherein the intron or the fragment of an intron is at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.3.

66. The polynucleotide of aspect 65, wherein the intron or the fragment of an intron is at least 95%, or at least 98% identical to SEQ ID NO.3.

67. The polynucleotide of any one of aspects 61-66, wherein the intron or the fragment of an intron interrupts the portion that is not codon optimised.

68. The polynucleotide of aspect 67, wherein the intron or the fragment of an intron is flanked by at least 60, at least 70, at least 80, at least 90, or at least 100 nucleotides that are not codon optimised.

69. The polynucleotide of aspect 68, wherein the intron or the fragment of an intron is flanked by between 110 and 120 nucleotides that are not codon optimised at the 5' end and between 100 and 110 nucleotides that are not codon optimised at the 3' end.

70. The polynucleotide of any one of aspects 61-69, wherein the intron or the fragment of an intron is positioned between exon 1 and exon 2.

71. The polynucleotide of any one of aspects 61-70, wherein the intron or the fragment of the intron is a fragment of native intron 1 (intron 1a).

72. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide further comprises a transcription regulatory element.

73. The polynucleotide of aspect 72, wherein the transcription regulatory element comprises a liver-specific promoter.

74. The polynucleotide of aspect 72 or aspect 73, wherein the transcription regulatory element comprises an A1AT promoter or a fragment of an A1AT promoter.

75. The polynucleotide of aspect 74, wherein the fragment of an A1AT promoter is at least 100, at least 120, at least 150, at least 180, less than 255, between 100 and 255, between 150 and 225, between 150 and 300, or between 180 and 255 nucleotides in length.

76. The polynucleotides of aspect 75, wherein the fragment of an A1AT promoter is between 150 and 300 nucleotides in length.

77. The polynucleotides, of any one of aspects 72-76, wherein the transcription regulatory element comprises an enhancer.

78. The polynucleotide of aspect 77, wherein the enhancer is an HCR enhancer or a fragment of an HCR enhancer.

79. The polynucleotide of aspect 78, wherein the fragment of an HCR enhancer is a fragment of at least 80, at least 90, at least 100, less than 192, between 80 and 192, between 90 and 192, between 100 and 250, or between 117 and 192 nucleotides in length.

80. The polynucleotide of aspect 79, wherein the fragment of an HCR enhancer is between 100 and 250 nucleotides in length.

81. The polynucleotide of any one of aspects 72-80, wherein the transcription regulatory element is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 6.

82. The polynucleotide of aspect 81, wherein the transcription regulatory element has a sequence of SEQ ID NO. 6.

83. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises an enhancer that is at least 80%, at least 85%, at least 90%, at least 95% at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 13.

84. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises an enhancer that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 13.

85. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises an enhancer of SEQ ID NO. 13.

86. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises a promoter that is at least 80%, at least 85%, at least 90%, at least 95% at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14.

87. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises a promoter that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14.

88. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises a promoter of SEQ ID NO. 14.

89. The polynucleotide of any one of the preceding aspects, wherein the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to codon 384 of wild type factor IX, and wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes alanine or leucine.

90. The polynucleotide of aspect 89, wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX is CTX, wherein X is any nucleotide.

91. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises a Factor IX nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment at least 1200, at least 1350, or at least 1650 nucleotides of SEQ ID NO. 5.

92. The polynucleotide of any one of the preceding aspects, wherein the polynucleotide comprises a Factor IX nucleotide sequence that is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.5.

93. The polynucleotide of any one of the preceding aspects, wherein:

(i) the Factor IX nucleotide sequence comprises a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1; and

(ii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine.

94. The polynucleotide of any one of the preceding aspects, wherein:

(i) the Factor IX nucleotide sequence comprises a coding sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1;

(ii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine; and

(iii) the polynucleotide comprises a promoter element that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14 and/or an enhancer element that is at least 98%, at least 99%, at least 99.5%, at least 99.8% or 100% identical to SEQ ID NO. 13.

95. The polynucleotide of any one of the preceding aspects, wherein:

(i) the Factor IX nucleotide sequence comprises a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1;

(ii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine; and

(iii) the polynucleotide comprise a transcription regulatory element that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 6.

96. The polynucleotide of any one of the preceding aspects, wherein:

(i) the Factor IX nucleotide sequence comprises a sequence that is at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO.1;

(ii) the Factor IX nucleotide sequence comprises a sequence that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a corresponding portion of SEQ ID NO: 2; and

(iii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine.

97. The polynucleotide of any one of aspects 95 or 96, wherein the Factor IX nucleotide sequence comprises an intron or a fragment of an intron, and the fragment of an intron is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 3.

98. The polynucleotide of any one of the preceding aspects, wherein:

(i) the Factor IX nucleotide sequence comprises a coding sequence and a portion of the coding sequence is not codon optimised; and

(ii) the Factor IX nucleotide sequence comprises a codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX encodes leucine.

99. The polynucleotide of any one of the preceding aspects, wherein a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at higher levels compared to a polypeptide encoded by a nucleotide sequence

comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7.

100. The polynucleotide of any one of the preceding aspects, wherein a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at higher levels compared to a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

101. The polynucleotide of any one of the preceding aspects, wherein a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 or SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 7 or SEQ ID NO. 6.

102. A viral particle comprising a recombinant genome comprising the polynucleotide of any one of the preceding aspects.

103. The viral particle of aspect 102, which is an AAV, adenoviral, or lentiviral viral particle.

104. The viral particle of aspect 103, which is an AAV viral particle.

105. The viral particle of any one of aspects 102-104, wherein the recombinant genome further comprises:

a) AAV2 ITRs;

b) a poly A sequence;

c) an origin of replication; and/or

d) two resolvable ITRs.

106. The viral particle of aspect 105, wherein the recombinant genome is single-stranded and/or comprises two resolvable ITRs.

107. The viral particle of any one of aspects 102-106, wherein the viral particle comprises a capsid selected from the group consisting of:

(i) a capsid having at least 96%, at least 98%, at least 99%, at least 99.5%, at least 99.8% identity or 100% identity to SEQ ID NO.10;

(ii) a capsid having at least 96%, at least 98%, at least 99%, at 99.5%, at least 99.8%, or 100% identity to SEQ ID NO. 17;

(iii) AAVMutC; and

(iv) AAV5.

108. The viral particle of any one of aspects 102-107, wherein on transduction into Huh7 cells, the viral particle expresses Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

109. The viral particle of aspect 108, wherein the activity is measured using a chromogenic substrate which is specific for Factor Xa.

110. The polynucleotide or viral particle of any one of the preceding aspects, wherein the Factor IX protein fragment is at least 200, at least 250, at least 300, between 200 and 415, between 250 and 415, or between 300 and 415 amino acids in length.

111. The polynucleotide or viral particle of any one of the preceding aspects, wherein the Factor IX protein or fragment thereof comprises a sequence:

a) at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 8; or

b) at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to a fragment of SEQ ID NO. 8 at least 200, at least 250, at least 300, between 200 and 415, between 250 and 415, or between 300 and 415 amino acids in length.

112. A composition comprising the polynucleotide or viral particle of any one of the preceding aspects and a pharmaceutically acceptable excipient.

113. The polynucleotide, viral particle or composition of any one of the preceding aspects for use in a method of treatment.

114. The polynucleotide, viral particle or composition for use of aspect 113, wherein the method of treatment comprises administering an effective amount of the polynucleotide, composition or viral particle of any one of aspects 1-112 to a patient.

115. A method of treatment comprising administering an effective amount of the polynucleotide, composition or viral particle of any one of aspects 1-112 to a patient.

116. Use of the polynucleotide, viral particle or composition of any one of aspects 1-112 in the manufacture of a medicament for use in a method of treatment.

117. The use of aspect 116, wherein the method of treatment comprises administering an effective amount of the polynucleotide, composition or viral particle of any one of aspects 1-112 to a patient.

118. The polynucleotide, viral particle, composition, use or method of any one of aspects 113-117, wherein the method of treatment is a method of treating haemophilia.

119. The polynucleotide, viral particle, composition, use or method of aspect 118, wherein the haemophilia is haemophilia B.

120. The polynucleotide, viral particle, composition, use or method of aspect 119, wherein the patient has antibodies or inhibitors to Factor IX.

121. The polynucleotide, viral particle or composition of any one of aspects 1-112 for use in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide

sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7, thereby treating a disease.

122. The polynucleotide, viral particle or composition of any one of aspects 1-112 for use in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7.

123. A method of expressing a polypeptide encoded by a Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7 comprising administering the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

124. A method of treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7 comprising administering the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

125. Use of the polynucleotide, viral particle or composition of any one of aspects 1-112 in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7, thereby treating a disease.

126. Use of the polynucleotide, viral particle or composition of any one of aspects 1-112 in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded

by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 12 and a transcription regulatory element of SEQ ID NO. 7.

127. The polynucleotide, viral particle or composition of any one of aspects 1-112 for use in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, thereby treating a disease.

128. The polynucleotide, viral particle or composition of any one of aspects 1-112 for use in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

129. A method of expressing a polypeptide encoded by a Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6 comprising the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

130. A method of treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6 comprising administering the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

131. Use of the polynucleotide, viral particle or composition of any one of aspects 1-112 in expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide

sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, thereby treating a disease.

132. Use of the polynucleotide, viral particle or composition of any one of aspects 1-112 in treating a disease by expressing a polypeptide encoded by the Factor IX nucleotide sequence at a level at least 2, or at least 3 times greater than a polypeptide encoded by a nucleotide sequence comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

133. The polynucleotide, viral particle or composition for use, or use of any one of aspects 121-122, 125-128, or 131-132, wherein expressing a polypeptide encoded by the Factor IX nucleotide sequence comprises administering the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

134. The polynucleotide, viral particle or composition of any one of aspects 1-112 for use in treating a disease by expressing Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

135. A method of treating a disease by expressing Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6, comprising administering an effective amount of the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

136. Use of the polynucleotide, viral particle or composition of any one of aspects 1-112 in treating a disease by expressing Factor IX protein or a fragment thereof having a

Factor IX activity greater than the activity of Factor IX expressed from a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7 and/or a viral particle comprising a Factor IX nucleotide sequence of SEQ ID NO. 18 and a transcription regulatory element of SEQ ID NO. 6.

137. The polynucleotide, viral particle or composition for use, or use of any one of aspects 134 or 136, wherein expressing the Factor IX protein or a fragment thereof comprises administering the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

138. The polynucleotide, viral particle or composition for use, method of treatment or use of any one of aspects 121-122, 124-128, 130-137, wherein the disease is haemophilia.

139. The polynucleotide, viral particle or composition for use, method of treatment or use of aspect 138, wherein the disease is haemophilia b.

140. The polynucleotide, viral particle or composition of any one of aspects 1-112 for use in achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B.

141. A method of achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B, comprising administering an effective amount of the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

142. Use of the polynucleotide, viral particle or composition of any one of aspects 1-112 in achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B.

143. The polynucleotide, viral particle or composition of any one of aspects 1-112 for use in treating haemophilia B comprising administering the polynucleotide, viral particle

or composition to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient.

144. A method of treating haemophilia B comprising administering the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient.

145. Use of the polynucleotide, viral particle or composition of any one of aspects 1-112 in treating haemophilia B comprising administering the polynucleotide, viral particle or composition to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient.

146. The polynucleotide, viral particle, or composition for use or use of aspect 140, 142, 143 or 145, wherein treating haemophilia B or achieving a stable Factor IX activity level comprises administering the polynucleotide, viral particle or composition of any one of aspects 1-112 to a patient.

147. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 141, 144, 146-148, wherein the Factor IX activity level is stable after at least 10 weeks, at least 15 weeks, at least 20 weeks, at least 30 weeks, at least 40 weeks, or at least 50 weeks from administration of the polynucleotide, viral particle or composition.

148. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 123, 124, 129, 130, 133, 135, 137-147, wherein the patient is a patient who is refractory to treatment with Factor IX having an activity lower than the activity of Factor IX expressed from a viral particle comprising (i) a Factor IX nucleotide sequence of SEQ ID NO. 4 or 5 and (ii) a transcription regulatory element of SEQ ID NO. 6.

149. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 123, 124, 129, 130, 133, 135, 137-148, wherein the patient is a patient

who is refractory to treatment with a Factor IX polypeptide that is expressed at a level lower than that encoded by a nucleotide sequence comprising (i) a Factor IX nucleotide sequence of SEQ ID NO. 4 or 5 and (ii) a transcription regulatory sequence of SEQ ID NO. 6.

150. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 121-149, wherein the polynucleotide, viral particle or composition is administered at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient having haemophilia B.

151. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 121-150, wherein the polynucleotide, viral particle or composition is administered at a dose of at least 4.5×10^{11} vg/kg, or at a dose of less than 5×10^{11} vg/kg, or at a dose of between 4.5×10^{11} and 1×10^{12} vg/kg, or at a dose of between 4.5×10^{11} vg/kg and 4.9×10^{11} vg/kg.

152. The polynucleotide, viral particle or composition for use, method, or use of aspect 151, wherein the dose is determined by quantifying the number of vector genomes using qPCR, optionally wherein the qPCR uses primers which bind to a promoter region, optionally wherein the promoter region is the promoter region of a transgene cassette.

153. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 121-152, wherein the polynucleotide or viral particle that is administered is produced from a mammalian cell and/or possesses characteristics which result from use of mammalian viral vector production cells and distinguish from vectors produced in insect viral vector production cells (e.g. baculovirus system).

154. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 140-153 wherein the stable Factor IX activity level is a stable Factor IX activity level of at least 25% or of at least 30%.

155. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 140-154 wherein the stable Factor IX activity level is a stable Factor IX activity level of at least 35%.

156. The polynucleotide, viral particle, or composition for use, method or use of any one of aspects 140-155, wherein the stable Factor IX activity level is of at least or around 40%.

157. The polynucleotide, viral particle, or composition for use, method, or use of any one of aspects 140-156, wherein the Factor IX activity level is at least 20% at a time point at least 50 weeks or at around 52 weeks after administration of the polynucleotide, viral particle or composition.

158. The polynucleotide, viral particle, or composition for use, method, or use of aspect 157, wherein the Factor IX activity level maintains at at least 20% for a continuous period of at least 40 weeks immediately preceding the time point.

159. The polynucleotide, viral particle, or composition for use, method, or use of any one of aspects 140-156, wherein the Factor IX activity level is at least 25% at a time point at least 50 weeks or at around 52 weeks after administration of the polynucleotide, viral particle or composition.

160. The polynucleotide, viral particle, or composition for use, method, or use of aspect 159, wherein the Factor IX activity level maintains at at least 25% for a continuous period of at least 40 weeks immediately preceding the time point.

161. The polynucleotide, viral particle, or composition for use, method, or use of any one of aspects 140-156, wherein the Factor IX activity is at least or around 40% at a time point at least 50 weeks or at around 52 weeks after administration of the polynucleotide, viral particle or composition.

162. The polynucleotide, viral particle, or composition for use, method, or use of aspect 161, wherein the Factor IX activity level maintains at at least 40% for a continuous period of at least 40 weeks immediately preceding the time point.

163. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 140-162, wherein the Factor IX activity is measured using a clotting assay to determine the Factor IX activity in a blood sample from the patient, optionally wherein the clotting assay uses a SynthASil™ kit.

164. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 140-162, wherein the Factor IX activity is measured by taking a blood sample from the patient and using a clotting assay to determine the Factor IX activity in the blood sample, optionally wherein the clotting assay uses a SynthASil™ kit.

165. The polynucleotide, viral particle or composition for use, method, or use of aspect 163 or 164, wherein the clotting assay is a one-stage clotting assay, optionally wherein the one-stage clotting assay uses a SynthASil™ kit.

166. The polynucleotide, viral particle or composition for use, method, or use of any one of aspects 113-165, wherein the method or use comprises co-administration of a steroid.

167. The polynucleotide, viral particle or composition for use, method or use of aspect 166, wherein the steroid is prednisolone.

CLAIMS

1. A polynucleotide comprising a Factor IX nucleotide sequence, wherein:
(i) the Factor IX nucleotide sequence comprises a coding sequence that encodes a Factor IX protein or fragment thereof;
(ii) a portion of the coding sequence is codon optimised;
(iii) the portion of the coding sequence that is codon optimised is at least 1100 nucleotides in length;
(iv) in the portion that is codon optimised at least 60% of the codons are selected from the group consisting of TTC, CTG, ATC, GTG, GTC, AGC, CCC, ACC, GCC, TAC, CAC, CAG, AAC, AAA, AAG, GAC, TGC, AGG, GGC and GAG;
(v) the portion that is codon optimised comprises a reduced number of CpGs compared to a corresponding portion of SEQ ID NO. 9;
(vi) the polynucleotide further comprises a transcription regulatory element comprising:
    (a) an A1AT promoter or a fragment of an A1AT promoter; and/or
    (b) an HCR enhancer or a fragment of an HCR enhancer; and
(vii) the Factor IX nucleotide sequence comprises a codon that encodes leucine at a position corresponding to position 384 of wild type Factor IX.

2. The polynucleotide of the preceding claim, wherein the portion of the coding sequence that is codon optimised is codon optimised for expression in the human liver.

3. The polynucleotide of any one of the preceding claims, wherein:
(i) a polypeptide encoded by the Factor IX nucleotide sequence is expressed in human liver cells at higher levels compared to a reference wild type Factor IX nucleotide sequence; and/or
(ii) the portion of the coding sequence that is codon optimised is at least 800, at least 900, at least 1100, less than 1500, less than 1300, less than 1200, between 800 and 1500, between 900 and 1300, between 1100 and 1200, or around 1191 nucleotides in length; and/or

(iii) the portion of the coding sequence that is codon optimised comprises 1, 2, 3, 4, 5 or all of:

a) exon 3 or a portion of at least 10, at least 15, at least 20, less than 25, between 10 and 25, between 15 and 25, or between 20 and 25 nucleotides of exon 3;

b) exon 4 or a portion of at least 80, at least 90, at least 100, less than 114, between 80 and 114, between 90 and 114, or between 100 and 114 nucleotides of exon 4;

c) exon 5 or a portion of at least 90, at least 100, at least 110, less than 129, between 90 and 129, between 100 and 129, or between 110 and 129 nucleotides of exon 5;

d) exon 6 or a portion of at least 150, at least 180, at least 200, less than 203, between 150 and 203, between 180 and 203, or between 200 and 203 nucleotides of exon 6;

e) exon 7 or a portion of at least 70, at least 80, at least 90, at least 100, less than 115, between 70 and 115, between 80 and 115, between 90 and 115, or between 100 and 115 nucleotides of exon 7; and/or

f) exon 8 or a portion of at least 400, at least 450, at least 500, less than 548, between 400 and 548, between 450 and 548, or between 500 and 548 nucleotides of exon 8; and/or

(iv) the portion of the coding sequence that is codon optimised comprises a portion of at least 20 nucleotides of exon 3, a portion of at least 100 nucleotides of exon 4, a portion of at least 110 nucleotides of exon 5, a portion of at least 180 nucleotides of exon 6, a portion of at least 100 nucleotides of exon 7, and a portion of at least 500 nucleotides of exon 8; and/or

(v) the portion of the coding sequence that is codon optimised comprises exon 3, exon 4, exon 5, exon 6, exon 7, and exon 8; and/or

(vi) the portion of the coding sequence that is codon optimised comprises a portion of exon 2, and the portion of exon 2 is less than 160, less than 150, less than 100, less than 75, less than 60, at least 20, at least 30, at least 40, at least 50, between 20 and 160, between 30 and 150, between 30 and 100, between 40 and 75, or around 56 nucleotides in length; and/or

(vii) the portion of the coding sequence that is codon optimised comprises a portion of exon 2 that is between 30 and 100 nucleotides in length.

4. The polynucleotide of any one of the preceding claims, wherein:
(i) the portion of the coding sequence that is codon optimised comprises less than less than 40, less than 20, less than 18, less than 10, less than 5, or less than 1 CpG; and/or
(ii) the portion of the coding sequence that is codon optimised is CpG free.

5. The polynucleotide of any one of the preceding claims wherein, in the portion of the coding sequence that is codon optimised:

(i) a) at least 1, at least 2, at least 4, or at least 5 codons that encode phenylalanine is/are replaced with TTC compared to a reference wild type Factor IX sequence;
b) at least 60%, at least 65%, or at least 70% of the codons that encode phenylalanine are TTC;
c) at least 60%, at least 65%, or at least 70% of the codons that encode phenylalanine are TTC and the remainder are TTT; and/or
d) the codons that encode phenylalanine are TTC, except where the following codon starts with a G; and/or

(ii) a) at least 5, at least 10, at least 15, or at least 16 codons that encode leucine is/are replaced with CTG compared to a reference wild type Factor IX sequence;
b) at least 90%, or at least 94% of the codons that encode leucine are CTG; and/or
c) at least 90%, or at least 94% of the codons that encode leucine are CTG and the remainder are CTC; and/or

(iii) a) at least 5, at least 10, at least 11, or at least 12 codons that encode isoleucine is/are replaced with ATC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon ATC is/are replaced with ATT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 60%, at least 70%, or at least 75% of the codons that encode isoleucine are ATC;

d) at least 60%, at least 70%, or at least 75% of the codons that encode isoleucine are ATC and the remainder are ATT; and/or

e) the codons that encode isoleucine are ATC, except where the following codon starts with a G; and/or

(iv) a) at least 10, at least 15, at least 20, or at least 25 codons that encode valine is/are replaced with GTG compared to a reference wild type Factor IX sequence;

b) at least 1 codon that encodes valine is/are replaced with GTC compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 90%, or at least 95% of the codons that encode valine are GTG; and/or

d) at least 80%, at least 90%, or at least 95% of the codons that encode valine are GTG and the remainder are GTC; and/or

(v) a) at least 5, at least 10, at least 12, or at least 13 codons that encode serine is/are replaced with AGC compared to a reference wild type Factor IX sequence;

b) at least 1, at least 2, or at least 4 codons that encode serine is/are replaced with TCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 60%, at least 65%, or at least 70% of the codons that encode serine are AGC; and/or

d) at least 60%, at least 65%, or at least 70% of the codons that encode serine are AGC and the remainder are TCT or TCC; and/or

(vi) a) at least 1, at least 2, or at least 5 codons that encode proline is/are replaced with CCC compared to a reference wild type Factor IX sequence;

b) at least 1 codons that encode proline is/are replaced with CCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 50%, at least 55%, or at least 60% of the codons that encode proline are CCC;

d) at 50%, at least 55%, or at least 60% of the codons that encode proline are CCC and the remainder are CCA or CCT; and/or

e) the codons that encode proline are CCC, except where the following codon starts with a G; and/or

(vii) a) at least 6, at least 7, at least 8, or at least 10 codons that encode threonine is/are replaced with ACC compared to a reference wild type Factor IX sequence;

b) at least 1, or at least 2, codons that encode threonine is/are replaced with ACT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 45%, at least 50%, or at least 55% of the codons that encode threonine are ACC;

d) at least 45%, at least 50%, or at least 55% of the codons that encode threonine are ACC and the remainder are ACT; and/or

e) the codons that encode threonine are ACC, except where the following codon starts with a G; and/or

(viii) a) at least 1, at least 2, at least 3, or at least 4 codons that encode alanine is/are replaced with GCC compared to a reference wild type Factor IX sequence;

b) at least 1, at least 2, or at least 3 codons that encode alanine is/are replaced with GCT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 35%, at least 40%, or at least 43% of the codons that encode alanine are GCC;

d) at least 35%, at least 40%, or at least 45% of the codons that encode alanine are GCC and the remainder are GCT; and/or

e) the codons that encode alanine are GCC, except where the following codon starts with a G; and/or

(ix) a) at least 1, or at least 2 codons that encode tyrosine is/are replaced with TAC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon TAC is/are replaced with TAT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 40%, at least 45%, or at least 48% of the codons that encode tyrosine are TAC;

d) at least 40%, at least 45%, or at least 48% of the codons that encode tyrosine are TAC and the remainder are TAT; and/or

e) the codons that encode tyrosine are TAC, except where the following codon starts with a G; and/or

(x) a) at least 1 codons that encode histidine is/are replaced with CAC compared to a reference wild type Factor IX sequence;

b) at least 50%, at least 60%, or at least 65% of the codons that encode histidine are CAC;

c) at least 50%, at least 60%, or at least 65% of the codons that encode histidine are CAC and the remainder are CAT; and/or

d) the codons that encode histidine are CAC, except where the following codon starts with a G; and/or

(xi) a) at least 1, at least 2, at least 4, or at least 5 codons that encode glutamine is/are replaced with CAG compared to a reference wild type Factor IX sequence;

b) at least 1 of codon CAG is/are replaced with CAA compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 85%, or at least 90% of the codons that encode glutamine are CAG; and/or

d) at least 80%, at least 85%, or at least 90% of the codons that encode glutamine are CAG and the remainder are CAA; and/or

(xii) a) at least 1, at least 2, at least 4, or at least 5 codons that encode asparagine is/are replaced with AAC compared to a reference wild type Factor IX sequence;

b) at least 60%, at least 65%, or at least 70% of the codons that encode asparagine are AAC;

c) at least 60%, at least 65%, or at least 70% of the codons that encode asparagine are AAC and the remainder are AAT; and/or

e) the codons that encode asparagine are AAC, except where the following codon starts with a G; and/or

(xiii) a) at least 5, at least 7, at least 8, or at least 9 codons that encode lysine is/are replaced with AAG compared to a reference wild type Factor IX sequence;

b) at least 1 of codon AAG is/are replaced with AAA compared to a reference wild type Factor IX sequence;

c) at least 80%, at least 90%, or at least 95% of the codons that encode lysine are AAG; and/or

d) at least 80%, at least 90%, or at least 95% of the codons that encode lysine are AAG and the remainder are AAA; and/or

(xiv) a) at least 1, at least 2, at least 3, or at least 4 codons that encode aspartate is/are replaced with GAC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon GAC is/are replaced with GAT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 45%, at least 50%, or at least 60% of the codons that encode aspartate are GAC;

d) at least 45%, at least 50%, or at least 60% of the codons that encode aspartate are GAC and the remainder are GAT; and/or

e) the codons that encode aspartate are GAC, except where the following codon starts with a G; and/or

(xv) a) at least 15, at least 20, at least 25, or at least 26 codons that encode glutamate is/are replaced with GAG compared to a reference wild type Factor IX sequence;

b) at least 80%, at least 90%, or at least 95% of the codons that encode glutamate are GAG; and/or

c) at least 80%, at least 90%, or at least 95% of the codons that encode glutamate are GAG and the remainder are GAA; and/or

(xvi) a) at least 5, at least 6, at least 7, or at least 8 codons that encode cysteine is/are replaced with TGC compared to a reference wild type Factor IX sequence;

b) at least 1 of codon TGC is/are replaced with TGT compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 40%, at least 50%, or at least 55% of the codons that encode cysteine are TGC;

d) at least 40%, at least 50%, or at least 55% of the codons that encode cysteine are TGC and the remainder are TGT; and/or

e) the codons that encode cysteine are TGC, except where the following codon starts with a G; and/or

(xvii) the codons that encode tryptophan are TGG; and/or

(xviii) a) at least 5, at least 8, at least 10, or at least 11 codons that encode arginine is/are replaced with AGG compared to a reference wild type Factor IX sequence;

b) at least 1 codon that encodes arginine is/are replaced with AGA compared to a reference wild type Factor IX sequence;

c) at least 60%, at least 70%, or at least 75% of the codons that encode arginine are AGG; and/or

d) at least 60%, at least 70%, or at least 75% of the codons that encode arginine are AGG and the remainder are AGA; and/or

(xix) a) at least 5, at least 10, at least 12, or at least 13 codons that encode glycine is/are replaced with GGC compared to a reference wild type Factor IX sequence;

b) at least 5, at least 6, at least 7, or at least 8 codons that encode glycine is/are replaced with GGG compared to a reference wild type Factor IX sequence, where the following codon starts with a G;

c) at least 50%, at least 55%, or at least 60% of the codons that encode glycine are GGC;

d) at least 50%, at least 55%, or at least 60% of the codons that encode glycine are GGC and the remainder are GGG; and/or

e) the codons that encode glycine are GGC, except where the following codon starts with a G.

6. The polynucleotide of any one of the preceding claims wherein the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine, and glycine, and in the codon optimised portion:

a) at least 5 codons that encode phenylalanine is/are replaced with TTC compared to a reference wild type Factor IX sequence;

b) at least 16 codons that encode leucine is/are replaced with CTG compared to a reference wild type Factor IX sequence;

c) at least 12 codons that encode isoleucine is/are replaced with ATC compared to a reference wild type Factor IX sequence;

d) at least 25 codons that encode valine is/are replaced with GTG compared to a reference wild type Factor IX sequence;

e) at least 13 codons that encode serine is/are replaced with AGC compared to a reference wild type Factor IX sequence;

f) at least 5 codons that encode proline is/are replaced with CCC compared to a reference wild type Factor IX sequence;

g) at least 10 codons that encode threonine is/are replaced with ACC compared to a reference wild type Factor IX sequence;

h) at least 4 codons that encode alanine is/are replaced with GCC compared to a reference wild type Factor IX sequence;

i) at least 2 codons that encode tyrosine is/are replaced with TAC compared to a reference wild type Factor IX sequence;

j) at least 1 codons that encode histidine is/are replaced with CAC compared to a reference wild type Factor IX sequence;

k) at least 5 codons that encode glutamine is/are replaced with CAG compared to a reference wild type Factor IX sequence;

l) at least 5 codons that encode asparagine is/are replaced with AAC compared to a reference wild type Factor IX sequence;

m) at least 9 codons that encode lysine is/are replaced with AAG compared to a reference wild type Factor IX sequence;

n) at least 4 codons that encode aspartate is/are replaced with GAC compared to a reference wild type Factor IX sequence;

o) at least 26 codons that encode glutamate is/are replaced with GAG compared to a reference wild type Factor IX sequence;

p) at least 8 codons that encode cysteine is/are replaced with TGC compared to a reference wild type Factor IX sequence;

q) the codons that encode tryptophan are TGG;

r) at least 11 codons that encode arginine is/are replaced with AGG compared to a reference wild type Factor IX sequence; and

s) at least 13 codons that encode glycine is/are replaced with GGC compared to a reference wild type Factor IX sequence.

7. The polynucleotide of any one of the preceding claims, wherein the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine,

glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine, and glycine, and in the codon optimised portion:

a) at least 70% of the codons that encode phenylalanine are TTC;

b) at least 94% of the codons that encode leucine are CTG;

c) at least 75% of the codons that encode isoleucine are ATC;

d) at least 95% of the codons that encode valine are GTG;

e) at least 70% of the codons that encode serine are AGC;

f) at least 60% of the codons that encode proline are CCC;

g) at least 55% of the codons that encode threonine are ACC;

h) at least 43% of the codons that encode alanine are GCC;

i) at least 48% of the codons that encode tyrosine are TAC;

j) at least 65% of the codons that encode histidine are CAC;

k) at least 90% of the codons that encode glutamine are CAG;

l) at least 70% of the codons that encode asparagine are AAC;

m) at least 95% of the codons that encode lysine are AAG;

n) at least 60% of the codons that encode aspartate are GAC;

o) at least 95% of the codons that encode glutamate are GAG;

p) at least 55% of the codons that encode cysteine are TGC;

q) the codons that encode tryptophan are TGG;

r) at least 75% of the codons that encode arginine are AGG; and

s) at least 60% of the codons that encode glycine are GGC.

8. The polynucleotide of any one of the preceding claims, wherein the portion of the coding sequence that is codon optimised comprises codons encoding phenylalanine, leucine, isoleucine, valine, serine, proline, threonine, alanine, tyrosine, histidine, glutamine, asparagine, lysine, aspartate, glutamate, cysteine, tryptophan, arginine, and glycine, and in the codon optimised portion:

a) at least 70% of the codons that encode phenylalanine are TTC and the remainder are TTT;

b) at least 94% of the codons that encode leucine are CTG and the remainder are CTC;

c) at least 75% of the codons that encode isoleucine are ATC and the remainder are ATT;

d) at least 95% of the codons that encode valine are GTG;

e) at least 70% of the codons that encode serine are AGC;

f) at least 60% of the codons that encode proline are CCC and the remainder are CCA or CCT;

g) at least 55% of the codons that encode threonine are ACC and the remainder are ACT;

h) at least 43% of the codons that encode alanine are GCC and the remainder are GCT;

i) at least 48% of the codons that encode tyrosine are TAC and the remainder are TAT;

j) at least 65% of the codons that encode histidine are CAC and the remainder are CAT;

k) at least 90% of the codons that encode glutamine are CAG and the remainder are CAA;

l) at least 70% of the codons that encode asparagine are AAC and the remainder are AAT;

m) at least 95% of the codons that encode lysine are AAG and the remainder are AAA;

n) at least 60% of the codons that encode aspartate are GAC and the remainder are GAT;

o) at least 95% of the codons that encode glutamate are GAG and the remainder are GAA;

p) at least 55% of the codons that encode cysteine are TGC and the remainder are TGT;

q) the codons that encode tryptophan are TGG;

r) at least 75% of the codons that encode arginine are AGG and the remainder are AGA; and

s) at least 60% of the codons that encode glycine are GGC and the remainder are GGG.

9. The polynucleotide of any one of claims 4-8, wherein the reference wild type Factor IX sequence is SEQ ID NO. 9 or SEQ ID NO. 19.

10. The polynucleotide of any one of the preceding claims, wherein the transcription regulatory element comprises a liver-specific promoter.

11. The polynucleotide of any one of the preceding claims, wherein the fragment of an A1AT promoter is:

(i) at least 100, at least 120, at least 150, at least 180, less than 255, between 100 and 255, between 150 and 225, between 150 and 300, or between 180 and 255 nucleotides in length; and/or

(ii) between 150 and 300 nucleotides in length.

12. The polynucleotide of any one of the preceding claims, wherein the fragment of an HCR enhancer is:

(i) a fragment of at least 80, at least 90, at least 100, less than 192, between 80 and 192, between 90 and 192, between 100 and 250, or between 117 and 192 nucleotides in length; and/or

(ii) between 100 and 250 nucleotides in length.

13. The polynucleotide of any one of the preceding claims wherein:

(i) the transcription regulatory element is at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 6; and/or

(ii) the transcription regulatory element has a sequence of SEQ ID NO. 6.

14. The polynucleotide of any one of the preceding claims, wherein:

(i) the polynucleotide comprises an enhancer that is at least 80%, at least 85%, at least 90%, at least 95% at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 13; and/or

(ii) the polynucleotide comprises an enhancer that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 13; and/or

(iii) the polynucleotide comprises an enhancer of SEQ ID NO. 13; and/or
(iv) the polynucleotide comprises a promoter that is at least 80%, at least 85%, at least 90%, at least 95% at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14; and/or
(v) the polynucleotide comprises a promoter that is at least 98%, at least 99%, at least 99.5%, at least 99.8%, or 100% identical to SEQ ID NO. 14; and/or
(vi) the polynucleotide comprises a promoter of SEQ ID NO. 14.

15. The polynucleotide of any one of the preceding claims, wherein the codon that encodes an amino acid at a position corresponding to position 384 of wild type Factor IX is CTX, wherein X is any nucleotide, optionally wherein X is C or G.

16. A polynucleotide comprising a Factor IX nucleotide sequence, wherein the Factor IX nucleotide sequence comprises a coding sequence that encodes a Factor IX protein or fragment thereof and wherein a portion of the coding sequence is not wild type.

17. A viral particle comprising a recombinant genome comprising the polynucleotide of any one of the preceding claims.

18. The viral particle of claim 17:
(i) which is an AAV, adenoviral, or lentiviral viral particle; and/or
(ii) which is an AAV viral particle; and/or
(iii) wherein the recombinant genome further comprises:
a) AAV2 ITRs;
b) a poly A sequence;
c) an origin of replication; and/or
d) two resolvable ITRs; and/or
(iv) wherein the recombinant genome is single-stranded.

19. The viral particle of claim 17 or claim 18, wherein on transduction into Huh7 cells the viral particle expresses Factor IX protein or a fragment thereof having a Factor IX activity greater than the activity of Factor IX expressed from a viral particle

comprising a Factor IX nucleotide sequence of SEQ ID NO: 12 and a transcription regulatory element of SEQ ID NO. 7.

20. The polynucleotide or viral particle of any one of the preceding claims, wherein the Factor IX fragment is at least 200, at least 250, at least 300, between 200 and 415, between 250 and 415, or between 300 and 415 amino acids in length.

21. A composition comprising the polynucleotide or viral particle of any one of the preceding claims and a pharmaceutically acceptable excipient.

22. The polynucleotide, viral particle or composition of any one of the preceding claims for use in a method of treatment.

23. The polynucleotide, viral particle or composition for use of claim 22, wherein the method of treatment comprises administering an effective amount of the polynucleotide, viral particle or composition of any one of claims 1-21 to a patient.

24. The polynucleotide, viral particle, or composition, for use of claim 22 or claim 23, wherein:
(i) the method of treatment is a method of treating haemophilia; and/or
(ii) the method of treatment is a method of treating haemophilia B; and/or
(iii) the patient has antibodies or inhibitors to Factor IX.

25. The polynucleotide, viral particle or composition of any one of claims 1-21 for use in achieving a stable Factor IX activity level of at least 20% in a patient having haemophilia B, optionally wherein achieving a stable Factor IX activity level comprises administering the polynucleotide, viral particle or composition of any one of claims 1-21 to a patient.

26. The polynucleotide, viral particle or composition of any one of claims 1-21 for use in treating haemophilia B comprising administering the polynucleotide, viral particle or composition to a patient at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient.

27. The polynucleotide, viral particle, or composition for use of claim 25 or 26, wherein:

(a) the Factor IX activity level is stable after at least 10 weeks, at least 15 weeks, at least 20 weeks, at least 30 weeks, at least 40 weeks, or at least 50 weeks from administration of the polynucleotide, viral particle or composition;

(b) the patient is a patient who is refractory to treatment with Factor IX having an activity lower than the activity of Factor IX expressed from a viral particle comprising (i) a Factor IX nucleotide sequence of SEQ ID NO. 4 or 5 and (ii) a transcription regulatory element of SEQ ID NO. 6;

(c) the patient is a patient who is refractory to treatment with a Factor IX polypeptide that is expressed at a level lower than that encoded by a nucleotide sequence comprising (i) a Factor IX nucleotide sequence of SEQ ID NO. 4 or 5 and (ii) a transcription regulatory sequence of SEQ ID NO. 6;

(d) the polynucleotide, viral particle or composition is administered at a dose effective to achieve a stable Factor IX activity level of at least 20% in a patient having haemophilia B;

(e) the polynucleotide, viral particle or composition is administered at a dose of at least 4.5×10^{11} vg/kg, or at a dose of less than 5×10^{11} vg/kg, or at a dose of between 4.5×10^{11} and 1×10^{12} vg/kg, or at a dose of between 4.5×10^{11} vg/kg and 4.9×10^{11} vg/kg, optionally wherein the dose is determined by quantifying the number of vector genomes using qPCR, optionally wherein the qPCR uses primers which bind to a promoter region, optionally wherein the promoter region is the promoter region of a transgene cassette;

(f) the polynucleotide or viral particle that is administered is produced from a mammalian cell and/or possesses characteristics which result from use of mammalian viral vector production cells and distinguish from vectors produced in insect viral vector production cells (e.g. baculovirus system);

(g) the stable Factor IX activity level is a stable Factor IX activity level of at least 25% or of at least 30%, optionally wherein the stable Factor IX activity level is a stable Factor IX activity level of at least 35%, optionally wherein the stable Factor IX activity level is of at least or around 40%;

(h) the Factor IX activity level is at least 20% at a time point at least 50 weeks or at around 52 weeks after administration of the polynucleotide, viral particle or

composition, optionally wherein the Factor IX activity level maintains at at least 20% for a continuous period of at least 40 weeks immediately preceding the time point;

(i) the Factor IX activity level is at least 25% at a time point at least 50 weeks or at around 52 weeks after administration of the polynucleotide, viral particle or composition, optionally wherein the Factor IX activity level maintains at at least 25% for a continuous period of at least 40 weeks immediately preceding the time point;

(j) the Factor IX activity is at least or around 40% at a time point at least 50 weeks or at around 52 weeks after administration of the polynucleotide, viral particle or composition, optionally wherein the Factor IX activity level maintains at at least 40% for a continuous period of at least 40 weeks immediately preceding the time point;

(k) the Factor IX activity is measured using a clotting assay to determine the Factor IX activity in a blood sample from the patient, optionally wherein the clotting assay uses a SynthASil™ kit;

(l) the Factor IX activity is measured by taking a blood sample from the patient and using a clotting assay to determine the Factor IX activity in the blood sample, optionally wherein the clotting assay uses a SynthASil™ kit;

(m) the clotting assay is a one-stage clotting assay, optionally wherein the one-stage clotting assay uses a SynthASil™ kit; and/or

(n) the use comprises co-administration of a steroid, optionally wherein the steroid is prednisolone.